BIOPHYSICAL CHARACTERIZATION OF THE PLAKIN FAMILY

by

CAEZAR AL-JASSAR

A thesis submitted to the University of Birmingham

For the degree of

DOCTOR OF PHILOSOPHY

September 2012

Supervisors:

Dr Martyn Chidgey and Professor Michael Overduin

Abstract

The desmosome is an inter-cellular complex required for strong cell-cell adhesion in skin and heart tissue. The plakin family of proteins are important to the desmosome as their main function is to act as linkers between the cell surface and cytoskeleton. Plakin proteins are defined by the presence of a plakin domain which is composed of spectin repeats, a central coiled coil rod domain and a C-terminal tail containing plakin repeat domains and a linker domain. The structures, dynamics and interactions of these component domains have remained largely unknown. Moreover, many mutations in the desmosomal genes that cause arrhythmogenic right ventricular cardiomyopathy (ARVC) and various skin blistering diseases are known but their consequences remain undefined. Consequently this thesis used biophysical techniques to better understand plakin family function but also to shed light into the specific effects of pathological mutations.

The primary aims of the thesis were to investigate the plakin domain structure in solution, whether the linker domain constituted a structured domain and whether there were structural differences between plakin repeat domains. In order to achieve this complementary biophysical techniques including nuclear magnetic resonance (NMR) spectroscopy, small angle X-ray scattering (SAXS) and X-ray crystallography were used. Specifically, the N and C termini from plakin family members desmoplakin, envoplakin and periplakin were explored.

The study generated novel insights into a variety of areas. These included the modular organization of plakin domains in solution. The non-linearity and flexibility of serial spectrin repeat constructs in the N-terminal plakin domain was established, contrary to previous predictions. The envoplakin C-terminal plakin repeat domain (PRD) was successfully crystallized, with the structure revealing novel features including a putative alternative binding groove. The discovery of a novel structural domain at the C-terminal tail of desmoplakin,

envoplakin and periplakin was made, revealing a possible mechanism in cytoskeletal preference. Finally the molecular consequences of a series of pathogenic point mutations known to cause ARVC were defined, including dramatic destabilisation and subtle conformational effects.

Together this provides a better understanding of the specific roles of the plakin family's N- and C- termini in linking to the desmosomal machinery on the cell membrane and in intermediate filaments, respectively. The conservation and divergences of the key features across the plakin family illustrates how desmosomal domain structure, dynamics, stability and interactions play essential yet specialised roles in preserving the integrity and flexibility of strong cell-cell junctions within distinct tissues. The results set out a foundation of work to better understand how the functional N and C termini "arms" of plakin proteins allow desmosomal assembly and intermediate filament binding.

Quotes

Dr Sinan Al-Jassar:

"In science, enough, is never enough"

from my father

Professor Robert Huber:

"To think big, you must start small"

Personally communicated to myself and the Lai lab at Cardiff University

Chinese proverb:

"A person who asks a question is a fool for five minutes. He who does not ask a question remains a fool forever"

A quote that spurred me on at Bristol Cathedral School

Professor Linus Pauling:

"The best way to have an idea, is to have lots of ideas"

Jehst – UK hip hop rapper:

"The news is confusin', it's all manufactured,

Pop-propaganda, political actors.

Political factions and critical factors,

Get kidnapped and sympathise with your captors"

Never believe things at face value

Acknowledgements

I have many people I wish to acknowledge for their incredible support they have given me during this PhD. Firstly I would like to thank both Dr Martyn Chidgey and Professor Michael Overduin. I want to thank them both for their excellent tutorage, support, encouragement, wisdom and for giving me an immense amount of freedom which made me become an independent scientist.

I must also acknowledge the excellent support I was given from the entire Overduin lab including Dr Tim Knowles for his encouragement, tutorage and "character building" (!) and Dr Mark Jeeves for his patience, helping me with NMR experiments and numerous attempts to try and teach me NMR theory. I must also thank Dr Claudia Fogl and Miss Penelope Rodriguez-Zamora for conducting some experiments with me to validate some of my thoughts. I hope we can finish off our fruitful ideas together! I would also like to thank and apologise to the excellent technicians: Pooja, Sandya, Jas and Eleni. Also, I couldn't have done this PhD without the fun times I had with fellow PhD students Piv (aka Professor YORKE), Mike (aka GIGGSY) Riyaz, Nazia and Vicky.

I would also like to acknowledge my collaborators. Dr Fiyaz Mohammed for being an excellent champion of X-ray crystallography and helping me hugely with my X-ray crystallography work. Dr Pau Bernadó for your enthusiasm and tutorage in SAXS analysis. I also wish to thank those at the Birmingham NMR facility for access to the NMR spectrometers. I also thank Dr Hennie Bikker and Dr Elijah Behr for their data which helped in the publication of some of my results. I must also thank the numerous other scientists I harassed by email for answers to my questions.

A heartfelt thank you must go to Teddy who without your constant and unwavering support, my PhD would have been unbearable. I also wish to thank my sister Gemma for always being there!

I think it is also important to acknowledge the excellent teaching I had at Bristol Cathedral School, with whom my science career started. I want to thank especially the inspirational Mr John Pearson who brought out the ability to question and probe everything about life. My science career in structural biology would not have existed without the amazing enthusiasm and encouragement from Dr Leon D'Cruz at Cardiff University. Your future patients are lucky to have you as their doctor!

Finally, this thesis is dedicated to my father and mother: Dr Sinan Al-Jassar and Dr Jayran Dabbaghian. Without your risks, leaving your families behind and enduring hardship upon coming to the UK my life would not have been the happy or successful one I have today.

Publications and presentations

Conferences attended

BIONMR "exchange" IRB, Barcelona – gave a lecture to department **(Oct 2011)**

EMB SAXS Conference, ESRF – Workshop on SAXS analysis **(May 2011)**

ICSG,University of Toronto - Presented a poster **(May 2011)**

4th Annual West Midlands Biophysics forum, University of Warwick – Gave a lecture **(Apr 2011)**

Research and enterprise gala, University of Birmingham – Presented a poster **(Apr 2011)**

Symposium and workshop on Cell free sciences, Matsuyama, Japan **(Sep 2010)**

ISMB Symposium, ISMB **(Jun 2010)**

3rd Annual West Midlands Biophysics forum, University of Leicester – Gave a lecture **(Apr 2010)**

International course on biological magnetic resonance – Gave a lecture and poster **(Jun 2009)**

Details of posters and talks given externally

"*The desmosomal complex – A story solved by crystallography, SAXS and NMR?"* Barcelona **(Oct 2011)**

"*Plakin domain architecture in solution and its relation to cardiomyopathies*" - Toronto **(May 2011)**

"*Molecular insights into plakin domain architectecture"* – Warwick **(Apr 2011)**

"*How can solving a protein structure help those afflicted with heart disease?*" – Birmingham **(Apr 2011)**

"*Biophysical characterisation of desmosomal proteins*" Leicester **(Apr 2011)**

"*Biophysical characterisation of desmosomal proteins*" Erice **(Jun 2009)**

Awards and grants

Based primarily on my work from my PhD, we were awarded a postdoctoral Wellcome Trust grant **(2011)**

Chosen to represent school of cancer sciences in the University of Birmingham postgraduate prospectus **(2011)**

Nominated for the University of Birmingham "Michael O'Rourke PhD Publication Award" **(2011)**

Travel award for the International course on biological magnetic resonance **(2009)**

Relevant publications

1. Al-Jassar C, Knowles T, Jeeves M, Kami K, Bikker H, Overduin M, Chidgey M (2011) The nonlinear structure of the desmoplakin plakin domain and the effects of cardiomyopathy-linked mutations, J Mol Biol, 411:1049-61. (In appendix A1).

2. Tong, M, Rajesh S, Jeeves M, Al-Jassar C, Overduin M et al. Structural insight into the autoinhibition and activation of human calcium calmodulin dependent protein kinase 1 delta by Ca^{2+}/calmodulin, Nature structural and molecular biology (Submitted).

3. Al-Jassar C, Bernado P, Overduin M, Chidgey M – "Spectrin repeats are separated by flexible hinges in plakin domains" – In preparation for PLoS One.

4. Al-Jassar C*, Fogl C*, Mohammed F^, Jeeves M^, White S, Knowles T, Wilcox B, Overduin M, Chidgey M – "Molecular mechanisms of plakin repeat domain interactions to the cytoskeleton" – In preparation for the Journal of Biological Chemistry.

5. Al-Jassar C*, Rodriguez-Zamora P*, Fogl C, Jeeves M, Overduin M, Chidgey M – "The NMR structure of the desmoplakin linker domain reveals insights into plakin protein cytoskeletal binding preferences" – In preparation for Nature Structural and Molecular Biology.

* and ^ denote these authors contributed equally.

N.B Relevant publication 3 is likely to be submitted in 2012

N.B Relevant publications 4 and 5 are likely to be submitted in 2013

Contents

Abbreviations

AUC	**Analytical ultracentrifugation**
ARVC	**Arrhythmogenic Right Ventricular Cardiomyopathy**
CD	**Circular Dichroism**
CT	**C-terminal region**
DNA	**Deoxyribonucleic Acid**
DP	**Desmoplakin**
DSC	**Desmocollin**
DSG	**Desmoglein**
E. coli	***Escherischia coli***
EOM	**Ensemble optimization method**
FID	**Free Induction Decay**
GAJOE	**Genetic Algorithm Judging Optimization of Ensembles**
h	**Hour**
HEPES	**(4-(2-hydroxythyl)-1-peperazineethanesulphonic acid)**
HSQC	**Heteronuclear single quantum coherence**
IF	**Intermediate Filaments**
IPTG	**Isopropyl-β-D-1-thiogalactopyranoside**
kDa	**Kilodalton**
LB broth	**Luria Bertani broth**
Min	**minute**
Mw	**Molecular Weight**
OD	**Optical Density**
NaCl	**Sodium Chloride**
NMR	**Nuclear Magnetic Resonance**
NSD	**Normalized Spatial Discrepancy**
PAGE	**Polyacrylamide Gel Electrophoresis**
PBS	**Phosphate Buffered Saline**
PG	**Plakoglobin**
PKP	**Plakophilin**
PPK	**Palmoplantar Keratoderma**
PR	**Plakin Repeat**

PRD	**Plakin Repeat Domain**
PPK	**Palmoplantar Keratoderma**
RANCH	**RANdom CHain**
RMSD	**Root Mean Square Deviation**
SDS	**Sodium Dodecyl Sulfate**
SH3	**SRC Homology 3**
SNP	**Single Nucleotide Polymorphism**
SPPK	**Striate Palmoplantar Keratoderma**
SR	**Spectrin Repeat**
SAXS	**Small Angle X-ray scattering**
SUMO	**Small Ubiquitin-like Modifier**
TEV	**Tobacco Etch Virus**
T_m	**Melting temperature**
TRIS	**Tris(hydroxymethyl)aminomethane**
WT	**Wild Type**

Chapter 1 - Introduction

The following introduction will give a broad overview of the desmosomal field. The introductions of the following results chapters will then expand on the information presented in this chapter.

1.1 Cell adhesion

In order for a multicellular organism to develop fully functioning organs, tissues are required and these in turn are made up of specialized differentiated cells. For cells to become part of a specific tissue they must adhere to one another. Adhesion amongst cells is therefore critical in maintaining tissue integrity. How the cells adhere to one another is based on the tissue and organ and is maintained by adhesion molecules in cell junctions. There are two types of cell junctions; cell-cell (or intercellular) and cell-matrix junctions (Wegener, 2001). Cell junctions are found in abundance in epithelial tissues (such as the skin, intestine, stomach, bladder, uterus and pancreas) and the myocardium. They maintain the strength and flexibility of epithelial and myocardial tissues. Intercellular junctions, as their name implies, connect one cell to another. Cell-matrix junctions on the other hand adhere cells to the extracellular matrix. Cell junctions are further divided into three groups: occluding junctions, anchoring junctions and communicating junctions (Wegener, 2001). Occluding junctions which include tight junctions are specialized to create a diffusion barrier by sealing cells together. Communicating junctions such as gap junctions allow free transfer of small molecules and solutes between adjacent cells. Anchoring junctions are specialized to confer strong cell to cell and cell to extracellular matrix adhesion and are summarized in figure 1.1.

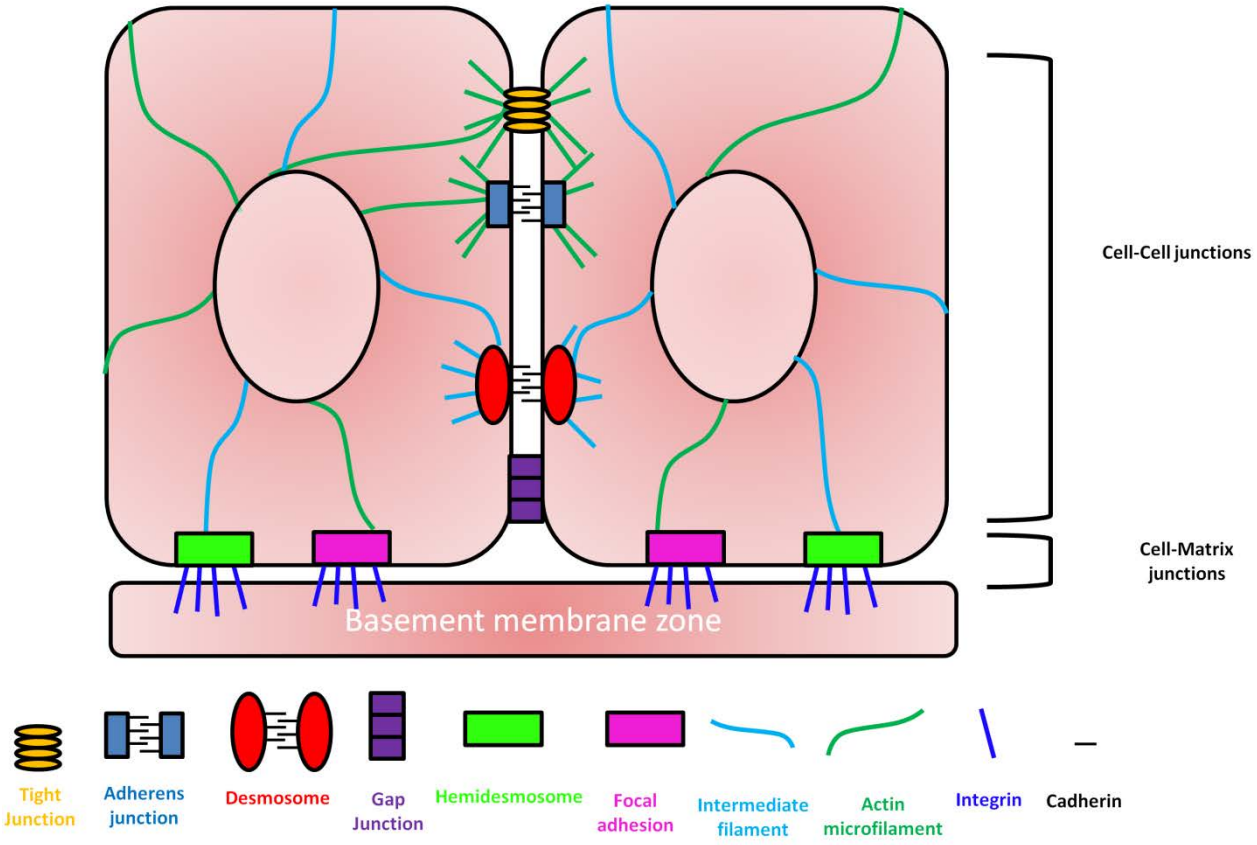


Figure 1.1 Schematic diagram of cell-cell and cell-matrix adhesion junctions

Tight junctions (yellow), adherens junctions (blue), desmosomes (red) and gap junctions (purple) are all cell-cell junctions which connect two basal epithelial cells. Hemidesmosomes (green) and focal adhesions (purple) represent cell-matrix junctions and connect epithelial cells to the basement membrane zone. Also shown are elements of the cell cytoskeleton including intermediate filaments (cyan) and actin microfilaments (dark green). Actin microfilaments are connected to adherens junctions and focal adhesions. Desmosomes and hemidesmosomes are connected to intermediate filaments. Integrins (dark blue) anchor hemidesmosomes and focal adhesions to the basement membrane zone via the plasma membrane. Adhesion between desmosomes and adherens junctions are mediated by cadherins (black) in the extracellular space, which attach two adjacent cells together (adapted from Green and Gaudry, 2000).

1.11 Anchoring junctions

Anchoring junctions have an adhesive interface in the intercellular space between adjacent cells that is mediated by transmembrane molecules (discussed below) which interact to seal the adjacent cells together. These transmembrane molecules are often connected to the cytoskeleton or to signalling molecules by “adaptor proteins”. Anchoring junctions are further subdivided into four classes, adherens junctions, desmosomes, hemidesmosomes and focal adhesions (Alberts et al., 2002). Adherens junctions and desmosomes engage in intercellular adhesion and contain transmembrane proteins from the cadherin family. Hemidesmosomes and focal adhesions engage in cell-matrix adhesion with their transmembrane proteins from the integrin family. The cytoskeletal network binds to both desmosomes and hemidesmosomes while adherens junctions and focal adhesions are connected to the actin microfilament cytoskeleton (discussed further in 1.2) (Alberts et al., 2002). Although architecturally similar, these junctions are composed of varying and distinct protein components.

1.2 The cell cytoskeleton

The cytoskeleton is a eukaryotic cell structure that provides the cell with mechanical strength and maintains its shape and integrity (Insall and Machesky, 2001). It is made up of three major components, actin microfilaments, microtubules and intermediate filaments (IFs). These components are assembled from soluble precursors to form larger complex networks. Assembly is carefully controlled by a variety of different cellular processes. Actin microfilaments have important roles in cell migration, contraction and polarity (Insall and Machesky, 2001). Microtubules are essential for processes such as intracellular vesicle, organelle and protein delivery, and are the largest filaments with a diameter of 25 nm. IFs play important roles in intracellular mechanical strength and are therefore present in tissues that encounter frequent mechanical stress such as the epidermis and muscles (Foisner, 2001). Plakin family members are known to act as cytolinkers, connecting elements of the cytoskeleton to each other and to junctional complexes at the membrane, and will be discussed further in section 1.4.

1.21 Intermediate filaments

Intermediate filaments are unique relative to other cytoskeletal family members (Herrmann and Aebi, 2004). They are made up of elongated α-helical protein building blocks. The α-helical "building blocks" are unstable in solution and multimerize to create a hydrophobic core in order to stabilize the IF whilst exposing their hydrophilic amino acids on the surface. This in turn makes IFs fibrous and rope-like, which makes them flexible and more resistant to mechanical stress than the other cytoskeletal elements (Foisner, 2001).

Intermediate filament proteins are expressed in a tissue-specific manner and are encoded by a total of 70 genes (Hesse et al., 2001). They are divided into six sequence homology classes (SHC) based on their primary structure, gene structure, assembly properties and tissue expression patterns. These are acidic keratins (SHC I), basic keratins (SHC II), desmin-vimentin-

type proteins (SHC III), neurofilaments (SHC IV), nuclear lamins (SHC V) and eye-lens filaments (SHC VI). Plakin family members have been shown to bind to keratins, vimentin and desmin (Favre et al., 2011, Fontao et al., 2003, Kazerounian et al., 2002, Meng et al., 1997b, Stappenbeck et al., 1993, Lapouge et al., 2006). This study is particularly interested in the SHCIII desmin-vimentin type proteins given their relevance to plakin family binding.

There are several features of IFs that are revealed by primary sequence analysis. Firstly, the heptad repeat $(a\text{-}b\text{-}c\text{-}d\text{-}e\text{-}f\text{-}g)_n$ is observed which are typical of α-helical two stranded coiled coils like that observed for IFs. The *a* and *d* positions are typically occupied by hydrophobic residues such as leucine, isoleucine or valine. There are 3.5 residues per turn and is left-handed to compensate for a small structural difference of the parallel α-helix. IFs are made up of a tripartite structure with a central ~310 amino acid α-helical rod domain which is flanked by N (head) and C (tail)-termini of variable length. It is thought that while the C-terminus is dispensable (Hatzfeld and Weber, 1990), the N-terminus is essential for filament assembly (Herrmann and Aebi, 2004).

1.22 Intermediate filament assembly

Keratins (SHC I and II) form heterodimers while the desmin/vimentin type (SHC III) and nuerofilaments (SHC IV) form homopolymers *in vitro*. Denatured vimentin filaments form dimers at a concentration of 6 M urea. Vimentin assembles into a mix of dimers and tetratmers under further dilution to 4.5 M urea. In buffers of low ionic strength (e.g. 5mM Tris-HCl) and in the absence of urea, vimentin forms long and uniform filaments (Herrmann and Aebi, 2004). Increasing the ionic strength (≥130 mM) by the addition of salts initiates multimeric filament formation. *In vivo*, vimentin is generally observed in the polymerized state although there are reports of an unpolymerized pool of unknown significance (Blikstad, 1983). Furthermore, it is thought vimentin interconverts between the two states, however, the mechanism underlying this is unclear.

1.3 Desmosomes

Desmosomes were first discovered by Giulio Bizzozero in 1864 (Bizzozero, 1864) and were named desmosomes by Josef Schaffer in 1920 (Schaffer, 1920). In 1974 the molecular components of the desmosome were characterized by Skerrow and Matoltsy (Skerrow and Matoltsy, 1974) who identified 8 proteins by SDS-PAGE analysis. They also correctly assumed that the cadherins were responsible for intercellular adhesion. Desmosomes are known for their important role in strong intercellular adhesion and the maintenance of tissue integrity. Since desmosomes are connected to the IF cytoskeletal network, they are thought to have a role in resisting mechanical stress imparted on the cell from intracellular and extracellular sources (Green and Gaudry, 2000). This is inferred from the plethora of studies which show that breakage of the links of the desmosome to the cytoskeletal network cause a variety of different diseases (Russell et al., 2004, Vasioukhin et al., 2001, Bierkamp et al., 1996, Thomason et al., 2010).

The desmosome is typically described as containing three ordered regions: 1) the extracellular core domain (ECD), 2) the outer dense plaque (ODP) and 3) the inner dense plaque (IDP). The cytoskeletal network is connected to the desmosome via the IDP (Garrod and Chidgey, 2008). While the thickness of both the ODP and IDP combined is approximately 15-20 nm, the diameter of an entire desmosome is less than 1 µm. However, the actual size and morphology may also vary within and between different tissues.

1.31 Desmosome composition

The desmosome is comprised of proteins from three distinct families, the cadherin family, the armadillo family and the plakin family (figure 1.2) (Garrod and Chidgey, 2008). Depending on tissue location, expression patterns from within the families vary. These include the cadherin proteins desmoglein (DSG1-4), desmocollin (DSC1-3) and the armadillo proteins: plakophilin

(PKP1-3) and plakoglobin. The armadillo protein plakoglobin only has one isoform while desmoplakin has two isoforms DSPI and DSPII. Although these proteins form the minimal requirement for desmosomal formation, additional proteins such as periplakin and envoplakin (see 1.42) may be required depending on cell specific requirements.

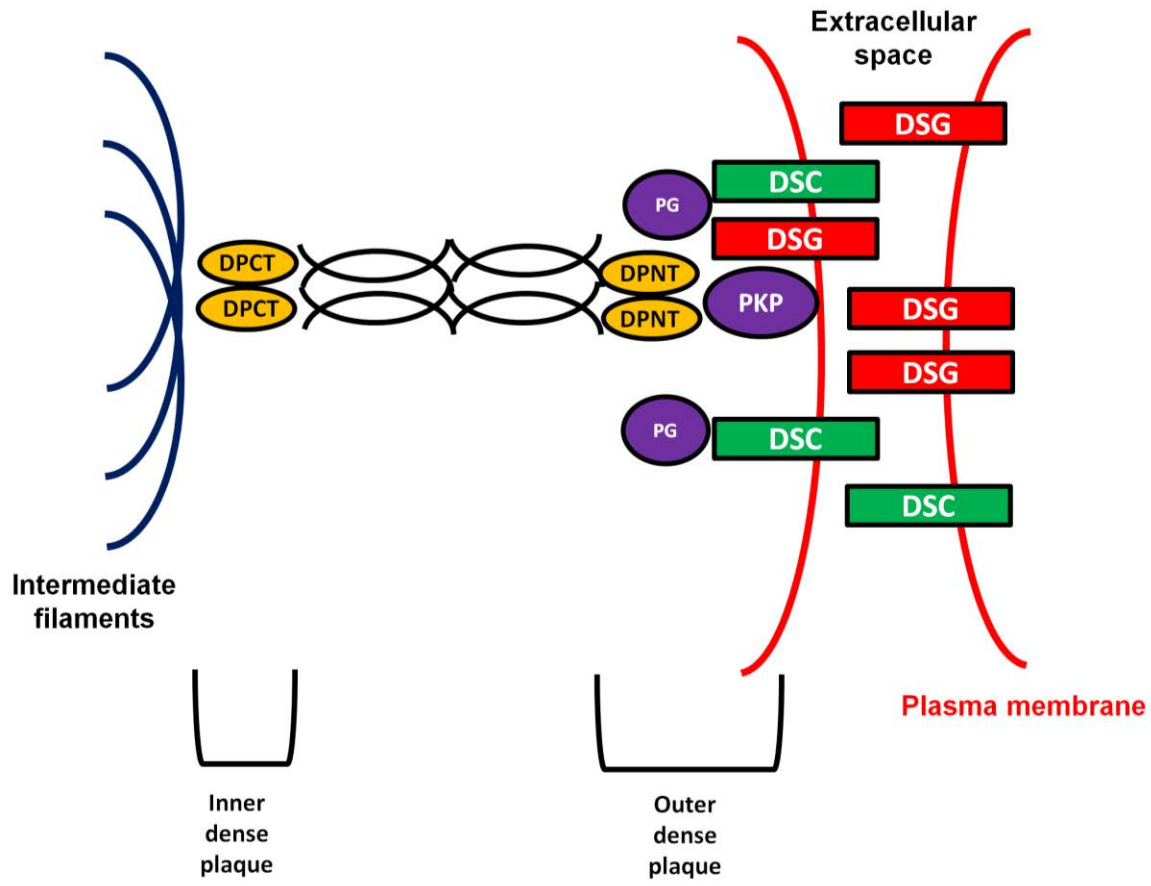


Figure 1.2 Molecular composition of the desmosome

A schematic diagram of the desmosome joining two adjacent cells (only one shown in detail) showing the relative positions of the armadillo proteins plakophilin (PKP) and plakoglobin (PG) (both purple), the cadherin proteins desmocollin (DSC) (green) and desmoglein (DSG) (red) and the obligatory plakin protein desmoplakin (N and C-terminal domains shown in yellow separated by a coiled coil domain). Also indicated is the location of what is termed the “outer dense plaque” and the “inner dense plaque”. Plakin proteins bind to intermediate filaments (dark blue) at their C-termini. Adapted from (South, 2004).

1.32 Cadherin family

In humans the desmosomal cadherin family are made up of seven proteins DSC1-3 and DSG1-4. The cadherin family form the “molecular glue” of two inter-cellular desmosomes. They are single-pass membrane spanning proteins. Their N-terminal extracellular domain made up of 5 homologous cadherin domains. The cadherin domains are separated by putative flexible linkers which accommodate 3 calcium ions (Garrod et al., 2005). Studies have shown interactions in a homophilic (i.e. between desmoglein and desmoglein) and heterophilic (i.e. between desmoglein and desmocollin) fashion both in *cis* (between cadherins of the same cell) and *trans* (between cadherins of adjacent cells) (Nie et al., 2011). Since the N-terminal domains contain multiple cadherin modules and are highly homologous, it is likely this portion of the protein is predominantly folded.

Their second domain is their transmembrane domain which anchors the proteins in the membrane. Finally, the cadherin family have a cytoplasmic tail domain which interacts with the intra-cellular components of the desmosome (Garrod and Chidgey, 2008). The C-terminal tails (which contain the least homology amongst one another) are intrinsically disordered and form a “hub” for interacting with a variety of different desmosomal proteins (Kami et al., 2009), a typical feature of intrinsically disordered regions (Haynes et al., 2006). It is still unclear how this desmosomal “hub” correctly assembles with multiple protein partners to form a characteristic desmosome. However, multiple interactions have been characterized between different desmosomal proteins and these are currently under further investigation by Martyn Chidgey’s lab.

1.33 Armadillo family

Armadillo family proteins are characterized by their highly conserved and alpha helical central ARM repeat domains. The arm repeats (of which there are a variable number depending on the

armadillo family member) typically consist of 42 amino acids resulting in one alpha helix (Choi and Weis, 2005). The central Arm repeat domain is flanked by unstructured N and C-termini (JK3rd et al., 2000). The most characterized of all the armadillo family members is β-catenin (Choi and Weis, 2005, Haegel et al., 1995, Zhurinsky et al., 2000). The armadillo desmosomal components homologous to β-catenin are plakoglobin and plakophilin. β-catenin localizes to adherens junctions only and is unable to interact with the desmosome (Troyanovsky et al., 1996). Crystal structures are available for β-catenin (Huber et al., 1997), PKP1 (Choi and Weis, 2005), PKP2a (Kirchner et al., 2012) and plakoglobin (Choi et al., 2009). They reveal multiple motifs consisting of 3 α-helices organised into the shape of a right handed helix superhelix. In turn, this creates a positively charged groove on its surface which interacts with ligands.

Plakoglobin interactions occur via its conserved central ARM repeat domain for which repeats 1-3 are required for interaction with the C-terminal tail of DSG1 and DSC2a (Chitaev et al., 1998). Other protein interactions include desmoplakin (Kowalczyk et al., 1997), PKP2 and PKP3 (Bonné et al., 2003, Chen et al., 2002), however, the binding sites have not been identified. A study investigating the deletion of the plakoglobin C-terminus observed larger desmosomes which suggested that plakoglobin regulates the control of desmosome size by its interactions with other desmosomal proteins (Palka and Green, 1997). Plakoglobin is also implicated in signal transduction (Garrod and Chidgey, 2008). Plakoglobin is highly homologous to β-catenin (a modulator of the Wnt/ β-catenin pathway) and the two molecules are thought to compete for proteins involved in this pathway. Whether plakoglobin activates or represses this pathway is unclear as is whether it plays a distinct role to that of β-catenin (Zhurinsky et al., 2000). β-catenin knock out studies show that plakoglobin cannot compensate for the lack of β-catenin (Bierkamp et al., 1996, Haegel et al., 1995, Ruiz, 1996). Suppression of desmoplakin in cardiac myocytes results in translocation of plakoglobin to the nucleus. This in turn leads to suppression

of the Wnt/ β-catenin pathway (Garcia-Gras et al., 2006). Interestingly plakoglobin had been suggested to bind to the N-terminal of desmoplakin (Bornslaeger et al., 2001, Kowalczyk et al., 1997) however unpublished results from our lab have shown that a direct interaction between the two is unlikely to occur *in vitro*. Therefore it appears that plakoglobin interaction with the desmosome is likely to be more complex than originally thought.

There are three plakophilins (PKP): PKP1, PKP2 and PKP3. They each contain a central 9 ARM repeat domain, a large unstructured N-terminal domain and a short unstructured C-terminal domain. Unlike plakoglobin they contain a large and putative flexible region between repeats 5 and 6 (Choi and Weis, 2005). PKP members are localized to the desmosome and nucleus and bind a wide variety of proteins including desmoplakin, cadherins, β-catenin, plakoglobin and intermediate filaments (Zhurinsky et al., 2000). Recently it was also discovered that PKP can bind to single stranded DNA (Sobolik-Delmaire et al., 2010) as well as other DNA associated proteins; RNA polymerase III and transcription factor TFIIIB (Mertens et al., 2001). Hence it is possible that they play both a structural and signalling role in association with desmosomes. Interestingly, all interactions of PKP so far documented are mediated via its unstructured head domain (Choi and Weis, 2005).

1.4 Plakin family

The plakin family typically provide a link between cytoskeletal proteins and junctional complexes. So far seven members have been identified including desmoplakin, periplakin, envoplakin, plectin, bullous pemphigoid antigen-1 (BPAG1), epiplakin and microtubule-actin crosslinking factor (MACF) (Sonnenberg and Liem, 2007a). Four of these proteins (desmoplakin, periplakin, envoplakin and plectin) are known to localize to the desmosome but of the four only desmoplakin is essential for desmosomal adhesion. Typically, plakin proteins consist of a N-terminal head domain including a globular plakin domain (which itself is made up of spectrin

repeat (SR) modules), a central coiled coil rod domain and a C-terminal tail with a “linker” domain and variable numbers of plakin repeat domains (PRDs) (discussed in further detail in 1.62). The focus of the thesis was to investigate the biophysical properties of the proteins desmoplakin, envoplakin and periplakin and will therefore be discussed in further detail. A schematic diagram of the structure of desmosomally localized plakin proteins is given in figure 1.3.

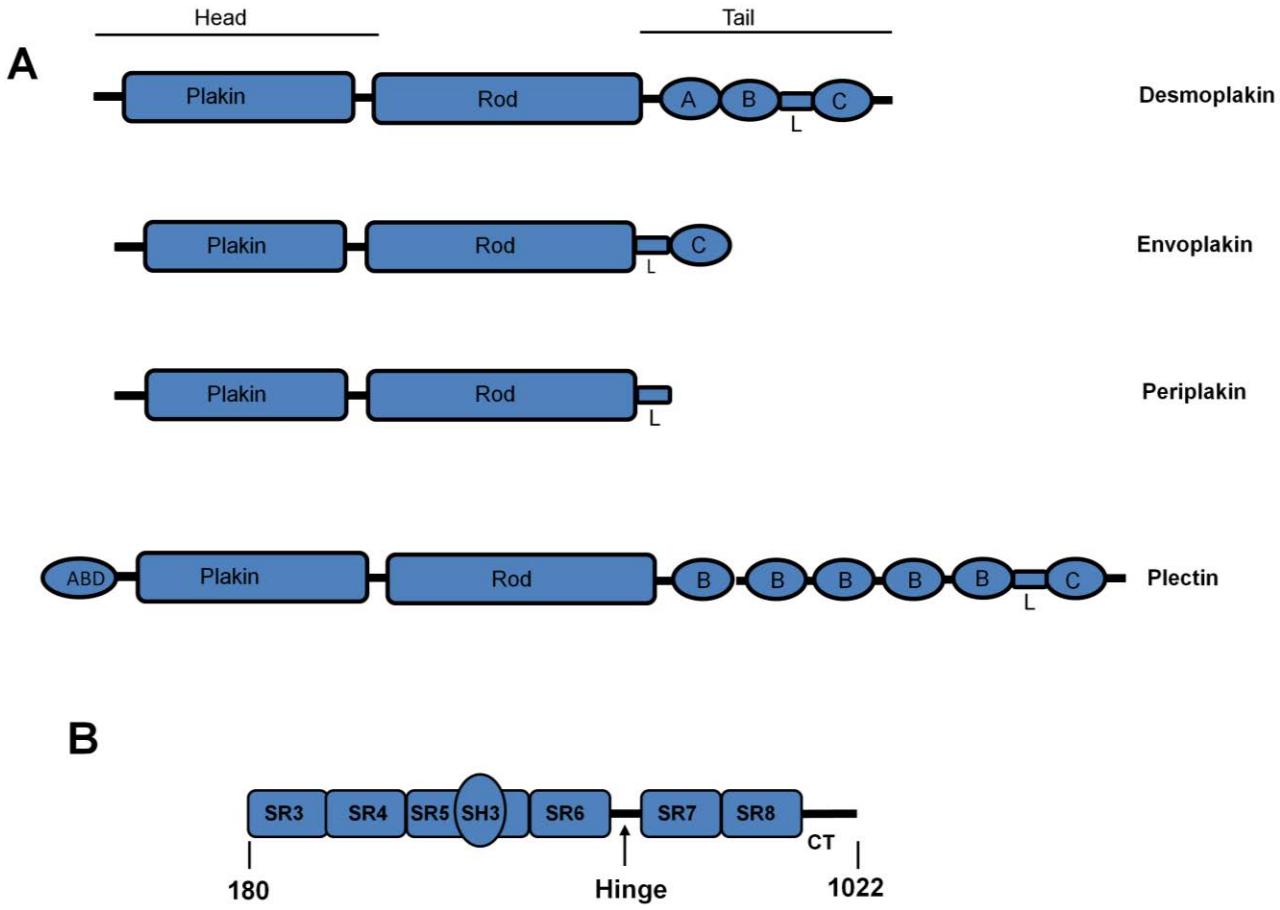


Figure 1.3 Schematic diagram of plakin family members that are associated with desmosomes

A: The tripartite structures of desmoplakin, envoplakin, periplakin and plectin are shown. At the N-terminal (head) region a plakin domain (Plakin) is present in all plakin family members with the exception of epiplakin. The middle rod domain (Rod) consists of a coiled coil. Finally the C-terminal (tail) region consists of a variable number of PRDs (termed A, B and C) and a linker domain (L). Additionally, plectin contains a unique N-terminal actin binding domain (ABD). B: The plakin domain of desmoplakin consists of 6 SRs and an SH3 domain. This represents a typical plakin domain of a plakin family member although the number of SRs can vary between family members. Also indicated is a hinge located between SR6 and SR7 and a flexible "CT" region (based on Sonnenberg et al, 2008; Choi and Weis 2011 and Al-Jassar et al, 2011).

1.41 Desmoplakin

Desmoplakin is an obligatory component of the desmosomal complex. Its N-terminal head region interacts with plakoglobin and plakophilin (Bornslaeger et al., 2001, Kowalczyk et al., 1997). Its coiled coil rod domain region facilitates homo-dimerization (Green et al., 1990), whereas the C-terminal tail interacts with intermediate filament proteins and contains 3 PRDs (designated A, B and C). Two isoforms created by alternative splicing are known; DPI (~332 kDa) and DPII (~260 kDa). They are essentially identical except that DPII has a shorter rod domain. The reason why two splice variants are required is currently unknown. Both are expressed widely and abundantly with the exception of the heart where DPII is only expressed at very low levels (Garrod and Chidgey, 2008). Using antibodies against the N and C-termini of desmoplakin the orientation of molecule was determined which suggested the N-terminal end orientates towards the membrane region while the C-terminal end orientates towards the intracellular region (North et al., 1999). This corroborated previous findings which suggested the N-terminal end of the protein is required for desmosomal assembly whereas the C-terminal end is required for binding to cytoskeletal proteins. Desmoplakin's role is thus primarily to mediate interactions between a variety of intermediate filaments (keratin, desmin and vimentin) and the desmosome. A putative phosphorylation site at the extreme C-terminal tail (S2849) has also been shown to weaken interaction with keratin when phosphorylated as indicated by a point mutation (S2849G) disrupting assembly into desmosomes (Stappenbeck et al., 1994).

1.42 Envoplakin and periplakin

Both envoplakin (~210 kDa) and periplakin (~195 kDa) were identified as protein constituents of the epidermal cornified envelope (Ruhrberg et al., 1997). This is an area that is found in the outermost layers of the skin that contributes to its barrier function. Both proteins are found localized to desmosomes and the interdesmosomal membrane where they contribute to a

scaffold upon which the cornified envelope assembles (Sonnenberg and Liem, 2007a). Periplakin and envoplakin form a functional dimeric unit through heterodimerization facilitated by their rod domain (Ruhrberg et al., 1997). Periplakin is also thought to form homodimers via its rod domain based on a scoring method described by (Parry et al., 1977) which correctly deduced that keratin would form heterodimers. In addition, heterodimerization is thought to be facilitated by the linker domains at the C-terminal tails of both envoplakin and periplakin (Karashima and Watt, 2002). Surprisingly when the periplakin or envoplakin genes were disrupted in mice, cornified envelope assembly remained intact with little observable differences (Aho et al., 2004, Määttä et al., 2001). A triple gene disruption of envoplakin, periplakin and involucrin revealed impairment to the epidermal barrier including cornified envelopes that were ultrastructurally abnormal with impaired desquamation (Sevilla et al., 2007). Therefore it appears that the three proteins work together for effective cornified envelope formation. Autoantibodies have also been found targeting the linker domains of envoplakin and periplakin in the skin disease paraneoplastic pemphigus (Kiyokawa et al., 1998, Mahoney et al., 1998); whether the antibodies are important in the aetiology of the disease is unknown.

Binding partners of envoplakin and periplakin have been identified. Kazrin (a desmosomally localized protein of unknown significance and structure) binds to the first 131 amino acids of periplakin, a likely unstructured region (Groot et al., 2004). Plectin has also been shown to bind to periplakin at the extreme N-terminal (Boczonadi et al., 2007). However, no binding partners have been identified for the N-terminal region of envoplakin. The linker domain is the sole domain of the periplakin C-terminal tail and is known to bind keratin 8 and vimentin (Kazerounian et al., 2002). The binding propensity of the envoplakin C-terminal tail (which consists of a linker domain and one PRD) is less clear as to whether it can bind vimentin with or without the presence of periplakin (Karashima and Watt, 2002).

1.5 Biophysical properties of plakin proteins

1.51 The N-terminal head domain of plakin proteins

The N-terminal head region of plakin proteins contain a N-terminal tail of variable length followed by a globular region called the plakin domain. Unpublished results from our lab showed that the extreme N-terminal region of desmoplakin contained a predominantly α-helical secondary structure however, was unable to form a folded tertiary structure as shown by 1D-NMR (Unpublished results Dr Keichiro Kami, 2008). Conversely, the preceeding plakin domain is relatively well studied. It was originally proposed to contain 6 α-helical fragments (NN, Z, Y, X, W, V) which formed α-helical bundles (Virata, 1992). However, in 2007 the α-helical bundles were identified as being members of the spectrin superfamily (Jefferson et al., 2007). Also present are a SH3 domain and CT region (Sonnenberg et al., 2007, Al-Jassar et al., 2011). Unfortunately, the previously outdated nomenclature is still being used in very recent publications despite the proven presence of spectrin repeats in multiple plakin family members.

1.52 The plakin domain is a member of the spectrin superfamily

1.53 The spectrin repeat

To date, spectrin, dystrophin, utrophin, α-actinin and plakin proteins are known members of the spectrin superfamily (Broderick and Winder, 2005, Jefferson et al., 2007). Spectrin repeats (SRs) form the structural platform of the spectrin superfamily (Djinovic-Carugo et al., 2002). Interestingly, the SR is unique to the animal kingdom (Broderick and Winder, 2005). A SR is a left-handed anti-parallel triple helical coiled often found in proteins with multiple copies (4-20 plus) (Yan et al., 1993). They are characterized by a typical heptad repeat structure termed *a,b,c,d,e,f,g* where *a* and *d* are usually occupied by apolar residues involved in hydrophobic packing of the SR similarly to those observed in IFs. In addition to these interactions, charged

side chains at positions *g* and *e* from adjacent coils stabilize the fold of the domain. SRs are often found in tandems connected through a linker region (Law et al., 2003). The linker region is often an α-helical structure which contributes to their linear and non-globular structure (Law et al., 2003). Serial chains of SRs are therefore thought to form linear structures, which is corroborated by the multitude of crystal and NMR structures of SRs (Ylänne et al., 2001). However, a paper published with my colleagues based on data presented in this thesis (chapter 4) showed that the desmoplakin plakin domain is non-linear (appendix A1). The study was the first of its kind to investigate the solution structure of a protein containing over 3 spectrin repeats, and contrasts with the exclusively linear conformation of previous structures solved by X-ray crystallography (Al-Jassar et al., 2011).

The main function of SRs is to withstand mechanical stress (Law et al., 2003). This is unsurprising given that SRs are located in proteins expressed in tissues known to encounter mechanical and/or abrasive forces. For instance spectrin is known to withstand mechanical stress which the red blood cell encounters and plakin family members are tethered to cytoskeletal proteins which function to resist mechanical stress. Furthermore they are known to bind a diverse range of ligands including (and not limited to) ankyrin, F-actin, α-catenin and titin (Broderick and Winder, 2005).

1.54 The SH3 domain

The SH3 domain is a well characterized protein module known to bind proline rich motifs (Kaneko et al., 2008). First characterized from the protein spectrin, it was reported as being an independent unit with no preference of location on its host protein (Musacchio et al., 1992). This has a bearing on the plakin domain SH3 domains which is discussed further in 1.61. They are small (55-70 amino acids) β barrel domains made up of 5 anti parallel β-sheets that are tightly compacted to form a spherical structure. SH3 domains contain a highly variable RT (arginine-

threonine) loop. This is located between the first and second β-strands and contains an irregular anti parallel structure separated by acidic residues. The RT loop is often the site of ligand interaction which binds to motifs typically with PxxP (proline rich) motifs (Musacchio et al., 1994, Yu et al., 1994).

1.6 Biophysical data on plakin proteins

1.61 X-ray crystallography data of constructs from plakin domains

To date four crystal structures have been elucidated for the plakin domains from BPAG1 (Jefferson et al., 2007), plectin (Ortega et al., 2011, Sonnenberg et al., 2007) and desmoplakin (Choi and Weis, 2011) some of which were published during the collection of data on similar constructs for this study. Published findings including the solution structure of the entire plakin domain and its modular organization is discussed in detail in chapters 3 and 4 and will therefore not be discussed in this chapter.

The structures elucidated from various segments of the plakin domain from BPAG1, plectin and desmoplakin reveal a high level of structural homology to typical SR structures. Early studies also revealed the presence of a SH3 domain within the plakin domain located within SR5 (Sonnenberg et al., 2007). This was not surprising as the protein α-spectrin also contains a SH3 domain (Musacchio et al., 1992). In 2011, two crystal structures revealed that the SH3 domain of plectin and desmoplakin is located within loop B of SR5 and interacts with SR4 in a fashion that rigidifies the plakin domain (Choi and Weis, 2011, Ortega et al., 2011). The result was a non-canonical SH3 domain incapable of binding ligands through the conventional method of the RT loop. It is unlikely the SH3 domain is capable of binding ligands as its primary function appears to be rigidification of the SR3-6 region of the plakin domain. Although crystal structures of the more homologous N-terminal region from several plakin domains have been published the more

divergent C-terminal region remained uncharacterized. Furthermore SR6 from desmoplakin, the most C-terminal SR characterized to date at atomic resolution was revealed to be only “spectrin repeat like” according to the authors (Choi and Weis, 2011).

1.61 The central coiled coil rod domain

The coiled coil central domain has yet to be structurally elucidated in any plakin protein, despite its importance in protein multimerization and extension (Green et al., 1990).

1.62 X-ray crystallographic data on plakin C-terminal tails

The C-terminal tail is comprised of a highly homologous linker domain and variable numbers of PRDs. To date no biophysical data is available on the linker domains however are implicated in binding to cytoskeletal proteins (discussed previously in 1.2). The PRDs of desmoplakin however have been characterized by Bill Weis and his group (2002) . Through sequence homology they were originally predicted to resemble ankyrin repeats which are 30-34 amino acid motifs in a helix-turn-helix conformation which are packed in a linear array by intra and inter-domain interactions (Li et al., 2006). Bill Weis’ group elucidated the crystal structures of desmoplakin PRDs B and C, revealing a novel structure containing 4.5 repeats of a 38 amino acid motif and termed it the plakin repeat (PR) motif. The motif itself consists of a β-hairpin followed by two anti parallel α-helices. Additionally a negatively charged residue at position 4 on the first β-strand of the β-hairpin interacts with a positively charged residue at position 19 of helix 1 to “fix” the β-hairpin into place. Furthermore, multiple hydrophobic contacts are used to stabilize the two anti parallel α-helices. The entire PRD forms a globular and compact structure by further inter domain hydrophobic contacts. A positively charged groove was located in both molecular surfaces of desmoplakin PRDs B and C (Choi et al., 2002). This was proposed to be the putative binding site for vimentin. Manual docking revealed that the crystal structure of vimentin (PDB code: 1GK4) was able to fit into this groove (Choi et al., 2002). However,

although polymerized vimentin bound to variable constructs from the desmoplakin C-terminal tail including PRDs B and C, the presence of high salt (>1 M) was unable to disrupt interaction between the two proteins implying other interactions may play a more significant role in complex formation (Choi et al., 2002).

1.7 Inherited diseases that affect desmosomes

Mutations across all protein families of the desmosome have been identified as causative agents of inherited human skin and heart diseases (with the exception of envoplakin, periplakin, DSG3, DSC1 and PKP3). Disease causing mutations in many of the proteins that make up the desmosome are sufficient to impair desmosomal function (summarized in table 1.1). In turn this would weaken adhesive strength and tissue integrity. Since mutations often cause any combination of skin, hair and heart defects they are referred to as cardio-cutaneous syndromes (Bolling and Jonkman, 2009).

Gene	**Mode of inheritance**	**Phenotypes**		
		Skin	**Hair**	**Heart**
DSG1	Dominant	SPPK		
DSG2	Dominant			ARVC
	Recessive			ARVC
DSG4	Recessive		Hypotrichosis	
DSC2	Dominant			ARVC
DSC2	Recessive	PPK, woolly hair and ARVC		
DSC3	Recessive	Hypotrichosis		
Desmoplakin	Dominant	SPPK		ARVC
	Dominant	SPPK, woolly hair and ARVC		
	Recessive	Carvajal syndrome		
	Recessive	Bullous dermatosis, PPK, total alopecia and ARVC		
	Recessive	Focal and diffuse PPK and woolly hair		
Plakoglobin	Dominant			ARVC
	Recessive	Naxos disease		
PKP1	Recessive	Ectodermal-dysplasia skin fragility syndrome		

PKP2	Dominant			ARVC
	Recessive			ARVC

Table 1.1 Dominant and recessive mutations in desmosomal genes and their disease outcomes

Desmosomal genes which give rise to a variety of different skin and/or heart diseases. SPPK: Striate Palmoplantar Keratoderma. ARVC: Arrhythmogenic Right Ventricular Cardiomyopathy. PPK: Palmoplantar Keratoderma.

1.71 Skin diseases and the desmosome

Striate palmoplantar keratoderma (SPPK) and pemphigus vulgaris are two of the skin diseases known to affect the desmosome. Hereditary PPKs are a clinically heterogenous group which cause thickening of the epidermis in the palms and soles of the feet. They are usually caused by autosomal dominant mutations from multiple desmosomal genes. They are further subdivided into 3 groups – focal, diffuse and punctate PPK. Focal PPK is restricted to areas which encounter mechanical and frictional stress. In diffuse PPK a more uniform thickening of the epidermis is observed. Finally punctate PPKs are characterized by numerous pinhead-sized hyperkeratotic papules that are irregularly distributed and localized to the palms and soles. Pemphigus vulgaris is a potentially fatal autoimmune disease whereby autoantibodies are generated against proteins of the desmosome, typically DSG3 and DSG1. The underlying cause of autoantibody generation is unknown although genetic or environmental factors may contribute (Amagai et al., 1991). The disease is characterized by skin blistering deep in the epidermis which can lead to fluid loss and infection. The link between autoantibody generation and skin blistering is unclear. It has been postulated that autoantibody binding to extracellular domains of the cadherin families may initiate downstream effects leading to desmosome disassembly (Chidgey, 2011).

1.72 Cardio-cutaneous diseases and the desmosome

Naxos disease and Carvajal syndrome are both genetic diseases of the skin and heart that are caused by autosomal recessive mutations in the desmoplakin and plakoglobin genes respectively. They are characterized by PPK, woolly hair and heart disease. Whereas some suggest both diseases are a single entity caused by mutations in two different proteins, others suggest they are slightly different with Naxos disease causing right ventricular effects and Carvajal syndrome causing left-dilated cardiomyopathy. A similar syndrome caused by mutations in desmoplakin whereby the patients exhibited PPK and woolly hair but no cardiac defects was called “skin fragility woolly hair syndrome” described by Whittock et al (2002).

1.73 Arrhythmogenic Right Ventricular Cardiomyopathy (ARVC)

ARVC is a genetic heart disease resulting in fibro-fatty tissue replacement in the right ventricule which leads to ventricular arrhythmias. Although its prevalence is reportedly low (estimates range from 1/1000-5000) it is one of the most common causes of unexpected sudden cardiac death in the young, particularly in young fit athletes (Awad et al., 2008). The prevalence may even be higher based on a recent study from Finland which suggested that as many as 1/200 may be carriers of a mutation predisposing them to ARVC (Lahtinen et al., 2011). Furthermore, low reported prevalence in the general population may be due to inadequate autopsies in which pathologists are not required to delve further into the underlying genetic cause (Papadakis et al., 2009). A genetic cause is thought to be the main cause in 50% of the cases and to date 7 genes have been implicated of which 5 encode cardiac desmosomal proteins (Awad et al., 2008). Studies have revealed that mutations in the cardiac desmosomal genes account for most of the reported cases of ARVC in the order of: (from high to low) PKP2>DP>DSG2>DSC2>PG (van Tintelen et al., 2006).

Historically ARVC was not officially described clinically until 1982 when Marcus et al reported 24 cases in which the electrophysiological effects (i.e. on electrocardiograms) were first described (Marcus et al., 1982). It was not until 1994 when McKenna et al set out a number of clinical criteria known as the “Task force criteria” designed to aid diagnosis of ARVC (McKenna et al., 1994). Criteria were separated as major and minor criterions in which, two major, four minor, or one major plus two minor criteria constituted diagnosis of ARVC in a patient. Major criteria included: reduced right ventricular ejection, fibrofatty replacement in the myocardium by biopsy, and “familial disease confirmed at necropsy or surgery”. Minor criteria included: altered electrocardiogram waves and familial history of sudden cardiac death. Despite these efforts to improve ARVC diagnosis it remains notoriously difficult to diagnose given the similarities to other cardiomyopathies (such as long QT syndrome). Differentiation between other cardiomyopathies often requires specialist examination and may still be ambiguous (Personal communication: Professor William McKenna, 2011). McKenna and colleagues recently updated the taskforce criteria further to facilitate diagnosis in which, amongst other suggestions, desmosomal mutations were upgraded to a major criterion (Marcus et al., 2010). A further study by an independent group found that these recent modifications significantly improved diagnosis of ARVC in patients when compared to the old taskforce criteria (Protonotarios et al., 2011).

Despite improvements to ARVC diagnosis, the molecular underpinning of the disease remains poorly understood. Two main theories have been proposed to explain the link between desmosomal gene mutations and the formation of fibrofatty tissue and subsequent cardiomyopathy. The first and simplest theory is an entirely structural model which proposes that a defective myocyte adhesion leads to cell death. It has been suggested that right ventricular scar formation can lead to arrhythmias such as those experienced during ARVC. The model also proposes that environmental factors such as exercise or virally induced inflammation could

exacerbate the already impaired myocyte adhesion. The second theory is more complex, and involves the Wnt/β-catenin signalling pathway. Disruption of desmosomal protein binding to one another is thought to be a major contributor to ARVC (Yang et al., 2006, Awad et al., 2008)). Plakoglobin, a desmosomal armadillo family member was shown to translocate to the nucleus upon disruption of binding to desmoplakin and suppress canonical Wnt/β-catenin signalling (Garcia-Gras et al., 2006). As a result of nuclear localization of plakoglobin Wnt signalling mediated inhibition of adipogenesis (the differentiation of cells into fatty tissue) was prevented.

Genetic screening of ARVC patients is becoming more important. The main benefit is that relatives of those affected, who have a higher risk of developing ARVC are forewarned. Genetic screening is an additional aid for diagnosis with mutations resulting in insertions, deletions, frameshift or premature termination which result in significantly altered protein products, frequently observed in those affected by ARVC. However, those with missense mutations are harder to interpret given proteins adopt tertiary structures and potentially altered binding sites to desmosomal ligands have not been localized comprehensively. Figure 1.4 summarizes all the missense mutations classified as pathogenic on desmoplakin from the ARVC database.

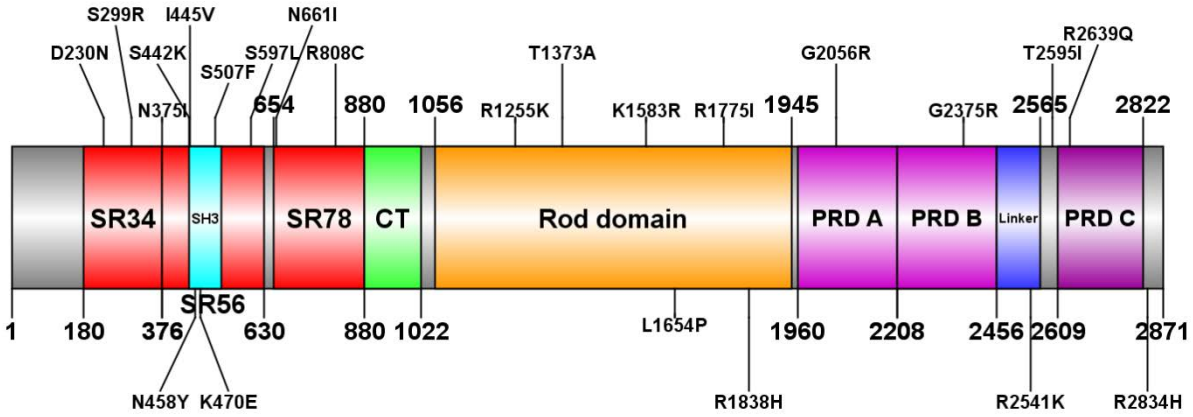


Figure 1.4 Pathogenic missense mutations in desmoplakin that have been reported in the ARVC database

Highlighted are the SRs of the plakin domain (red), the SH3 domain (cyan), the CT region (green), the rod domain (orange), the PRDs (purple) and the linker domain (blue). Most (~50%) of the reported mutations are located in the plakin domain. Image created by Domain Graph version 2.0 (Ren et al, 2009).

1.8 Hypotheses

A number of hypotheses prompted the work in this study. They are based on previous studies of plakin family members.

The first assumed that by ascertaining the structure of the plakin domain it would help explain its essential role in tethering the desmosome to the cytoskeletal network. The desmoplakin N-terminal region which includes the plakin domain is known to bind both plakophilin and plakoglobin (Bornslaeger et al., 2001). Through this mechanism it is thought that the desmosome assembles correctly, since it was previously shown that disruption of these interactions led to improper desmosomal development (Russell et al., 2004). Prior to this study being conducted the structures of only the first 2 SRs of BPAG1 (Jefferson et al., 2007) and plectin (Sonnenberg et al., 2007) were known (these are homologous to the first 2 SRs of desmoplakin's plakin domain). However, 4 more C-terminal spectrin repeats (and a SH3 domain) of unknown significance are present in the desmoplakin plakin domain. How the entire plakin domain folded in light of these additional domains was unknown.

The second hypothesis assumed that there would be subtle structural differences between plakin family members which gave them their specialized roles in the cell. Desmoplakin is essential for murine embryonic development (Gallicano et al., 1998), Conversely, single or double gene disruptions of envoplakin and periplakin, plakin proteins also localized to the desmosome, produced little observable effects (Aho et al., 2004, Määttä et al., 2001). By comparing domains from throughout the full length sequences (particularly from the plakin domains and C-terminal tails) of the plakin family members desmoplakin, periplakin and envoplakin it was hoped that it would clarify why certain proteins have specialized roles in the cell.

The third hypothesis presumed that specialized domains at the C-terminal tail give rise to specific cytoskeletal binding for different plakin family members. Both PRDs and linker domains are present in desmoplakin and envoplakin. However, only a linker domain is present at the C-terminal tail of the periplakin protein. Desmoplakin has a preference of binding keratin over desmin/vimentin (Meng et al., 1997b). whereas desmin/vimentin is preferred over cytokeratins for envoplakin and periplakin (Kazerounian et al., 2002, Karashima and Watt, 2002). The structure of PRDs are only known from desmoplakin and it is presumed that subtle differences between PRDs of different plakin family members may give rise to these cytoskeletal preferences. In addition to this the linker domain which has been shown to be critical for binding cytoskeletal proteins (Karashima and Watt, 2002, Kazerounian et al., 2002) has not been structurally characterized. By structurally characterizing and comparing a variety of constructs from the C-terminal tails of desmoplakin, envoplakin and periplakin it was hoped it would explain cytoskeletal preferences and possible binding mechanisms.

The final hypothesis presumed that by characterizing the structural implications of a number of different missense mutations thought to cause ARVC, it would help explain the ambiguity and complexity of the genotype-phenotype link. Prior to this study being conducted nothing was known of the biophysical impact of an ARVC mutation on any desmosomal protein. By gaining this information it was hoped it would ultimately aid cardiologists make more informed decisions when presented with mutation data from a patient thought to suffer from ARVC.

1.9 Aims

To address the above mentioned hypotheses the following aims were created:

1) To express and purify a number of constructs from desmoplakin, envoplakin and periplakin

2) To purify plakin domains from desmoplakin, envoplakin and periplakin and assess their domain architecture by Circular Dichroism (CD), Analytical Ultracentrifugation (AUC), Nuclear Magnetic Resonance (NMR) and Small Angle X-ray Scattering (SAXS).
3) To compare the binding of desmoplakin and envoplakin PRDs to the cytoskeletal protein vimentin and ascertain whether structural differences between PRDs lead to differential binding to vimentin.
4) To crystallize the envoplakin PRD C and compare its structure to that of its homologue desmoplakin PRD C.
5) To compare the structure of the linker domains of desmoplakin, envoplakin and periplakin and ascertain their overall fold using NMR and SAXS.
6) To ascertain whether a variety of ARVC mutants caused biophysical differences in desmoplakin protein domains when compared to their wild type (WT) counterparts.

Overall it was hoped that the information gained would give a clearer picture into the mechanisms which allow the desmosome to be tethered to the IF cytoskeletal network via plakin proteins.

Chapter 2 - Materials and methods

2.1 Molecular biology

Three plakin proteins were investigated in this project. Constructs were designed using GenBank human cDNA sequences for desmoplakin (accession number: NM_004415), envoplakin (accession number: BC126103) and periplakin (accession number: BC114620)

2.11 Construct Design

Constructs were either provided kindly by Dr Martyn Chidgey or purchased from either ShineGene (Shanghai, China) or GenScript (New Jersey, USA). Depending on the predicted solubility of the construct, different vectors were used. For example, in low solubility constructs, the presence of glutathione-S-transferase (GST), a known solubility enhancing tag was present at the N-terminal of proteins when expressed from pGex6p-1 vectors. All constructs used are shown in tables 2.1-2.4. Some constructs in table 2.1 (flagged with an asterisk) were chemically synthesized so that cysteine residues were replaced by serine residues in the final protein product in order to enhance expression and purification. Constructs inserted into pET21b+ vectors contained an N terminal methionine residue and a C-terminal 6x His tag of ***LGHHHHHH***. Constructs inserted into pET45b+ vectors contained N terminal ***MAHHHHHHVENLYFQG*** residues which include a Tobacco Etch Virus (TEV) protease site to facilitate proteolytic removal of the 6x His tag. Constructs inserted into pProEX HTc vectors contained a vector encoded N terminal 6x his tag, followed by a spacer, followed by a TEV protease site (see figure 2.12 for more details). Constructs were designed using information from the literature (Sonnenberg et al., 2007, Karashima and Watt, 2002) and results from the Globprot server (Linding et al., 2003) and the recombinant protein solubility prediction tool at http://www.biotech.ou.edu/ (Wilkinson and Harrison, 1991)

Name	Residues	Codon Optimi-zed?	Vector	Restriction Sites (5' and 3')	Tag
Spectrin Repeat 3-4 (SR34*)	180-375	Yes	pET21b+	*Nde*I and *Bam*H1	6x His
Spectrin Repeat 3-4 (SR34*)	180-375	Yes	pET-SUMO	N/A	SUMO
Spectrin Repeat 5-6 (SR56*)	373-630	Yes	pET21b+	*Nde*I and *Bam*H1	6x His
Spectrin Repeat 5-6 K470E (SR56-K470E*)	373-630	Yes	pET21b+	*Nde*I and *Bam*H1	6x His
Desmoplakin SH3*	440-518	Yes	pET21b+	*Nde*I and *Bam*H1	6x His
Spectrin Repeat 7-8 (SR78*)	654-883	Yes	pET45b+	*Pml*I and *Bam*H1	6x His
Spectrin Repeat 7-8 R808C (SR78-R808C*)	654-883	Yes	pET45b+	*Pml*I and *Bam*H1	6x His
Spectrin Repeat 8-C-terminal region (SR8-CT*)	771-1022	Yes	pET45b+	*Pml*I and *Bam*H1	6x His
Desmoplakin plakin domain	180-1022	No	pGex6p-1	*Bam*HI and *Xho*I	GST and 6x His
Desmoplakin plakin domain_K470E	180-1022	No	pGex6P-1	*Bam*HI and *Xho*I	GST and 6x His
Desmoplakin plakin domain_R808C	180-1022	No	pGex6P-1	*Bam*HI and *Xho*I	GST and 6x His
Desmoplakin PRD B	2209-2456	No	pProEX HTc	*Eco*RI and *Xho*I	6x His
Desmoplakin PRD B_G2375R	2209-2456	No	pProEX HTc	*Eco*RI and *Xho*I	6x His
Desmoplakin Linker	2454-2565	No	pGex6p-1	*Bam*HI and *Xho*I	GST
Desmoplakin Linker_R2541K	2454-2565	No	pGex6P-1	*Bam*HI and *Xho*I	GST
Desmoplakin PRD C	2609-2822	No	pProEX HTc	*Eco*RI and *Xho*I	6x His
Desmoplakin PRD C_R2639Q	2609-2822	No	pProEX HTc	*Eco*RI and *Xho*I	6x His

Table 2.1 Desmoplakin constructs

Name	Residues	Codon Optimized?	Vector	Restriction Sites (5' and 3')	Tag
Periplakin SH3 domain (Peri-SH3)	381-464	No	pET21b+	*Nde*I and *Xho*I	6x His
Periplakin plakin domain	131-871	No	pGex6p-1	*Bam*HI and *Xho*I	GST and 6x His
Periplakin Linker domain	1646-1756	No	pGex6p-1	*Bam*HI and *Xho*I	GST

Table 2.2 Periplakin constructs

Name	Residues	Codon Optimized?	Vector	Restriction Sites (5' and 3')	Tag
Envoplakin SH3	395-478	No	pET21b+	*Nde*I and *Xho*I	6x His
Envoplakin plakin domain	144-891	No	pGex6p-1	*Bam*HI and *Xho*I	GST and 6x His
Envoplakin_Linker	1674-1784	No	pGex6p-1	*Bam*HI and *Xho*I	GST and 6x His
Envoplakin linker_WA	1674-1784	No	pGex6p-1	*Bam*HI and *Xho*I	GST and 6x His
Envoplakin PRD Small	1829-2000	No	pProEX HTc	*Eco*RI and *Xho*I	6x His
Envoplakin PRD Medium	1822-2014	No	pProEX HTc	*Eco*RI and *Xho*I	6x His
Envoplakin PRD Large	1812-2030	No	pProEX HTc	*Eco*RI and *Xho*I	6x His

Table 2.3 Envoplakin constructs

Name	Residues	Codon Optimized?	Vector	Restriction Sites (5' and 3')	Tag
Vimentin full length	1-466	No	pET21a+	*Nde*I and *Xho*I	None

Table 2.4 Vimentin constructs

2.12 Vectors used

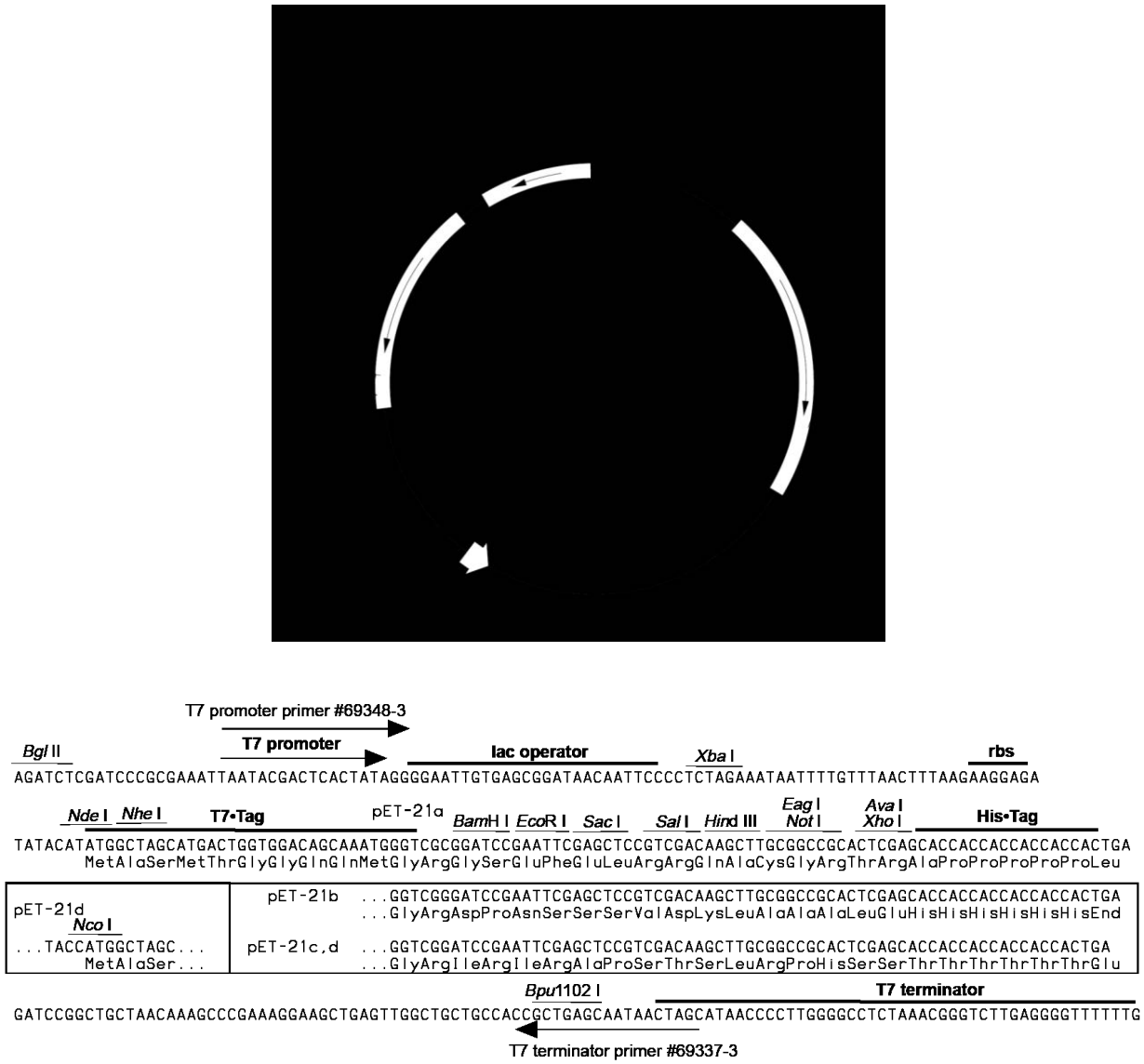
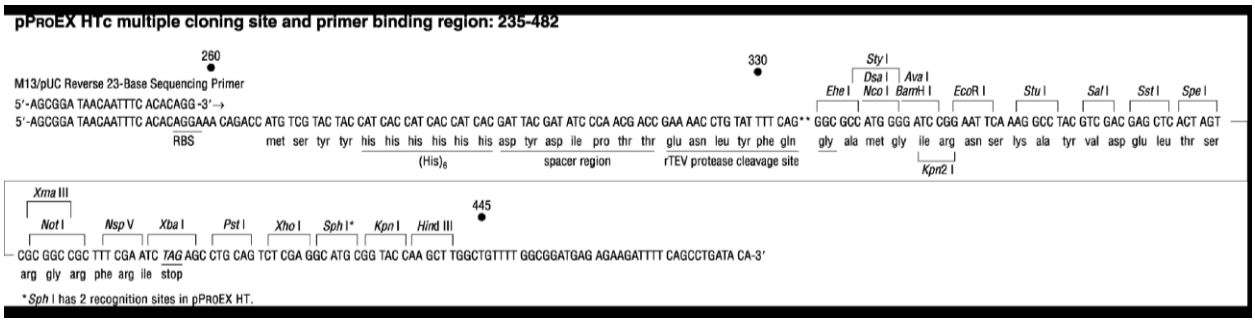


Figure 2.11 pET21a-d

pET21a+ and pET21b+ were used for vimentin full length and plakin domain subdomain constructs, respectively. Key features include a T7 promoter for expression in *E.coli* and ampicillin resistance. Since all constructs contained 5' *Nde*I and 3' *Xho*I sites no T7-tag was included in the expressed proteins. Vimentin in pET21a+ has no tag while constructs in pET21b+ had a non-cleavable C-terminal his-tag. (Image from Novagen manual).



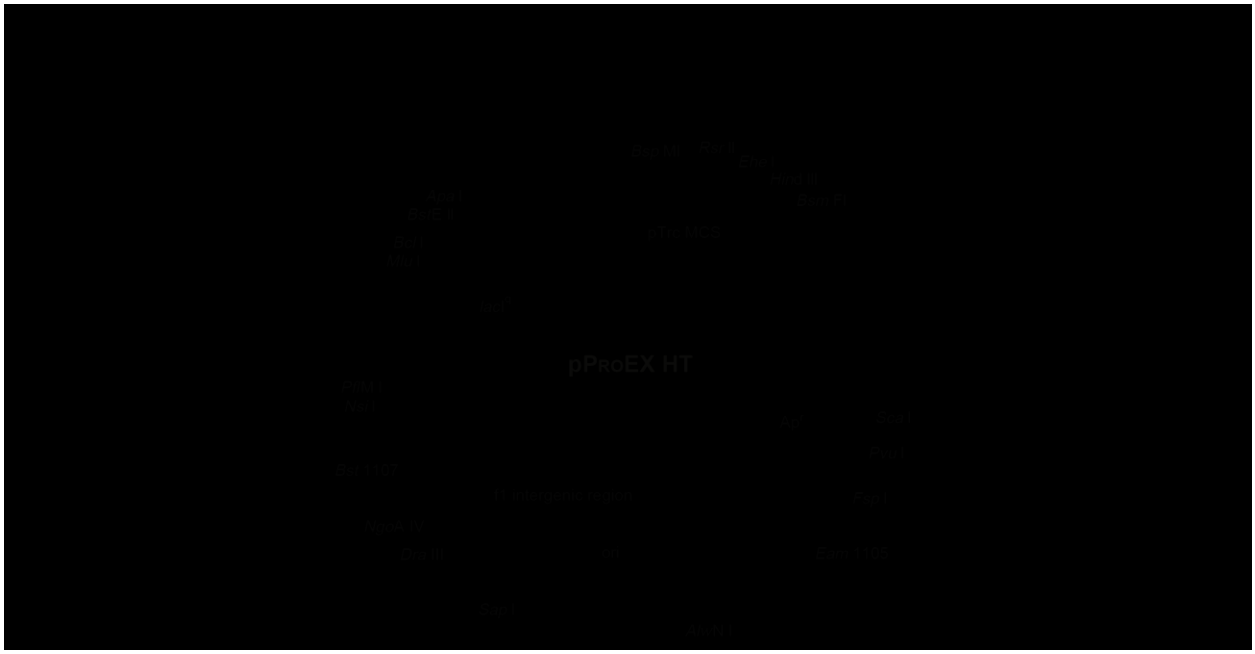


Figure 2.12 - pProEX HTc

The pProEX HTc vector was used primarily for C-terminal tail constructs (see Tables 2.1-2.3). Key features include a TRc promoter that can be induced using Isopropyl-β-D-Thio-Galactoside (IPTG), ampicillin resistance and a TEV protease recognition site that allows cleavage of the 6x His tag from the recombinant protein product. An artificial STOP codon is also encoded by the vector DNA. (Image from Invitrogen manual).

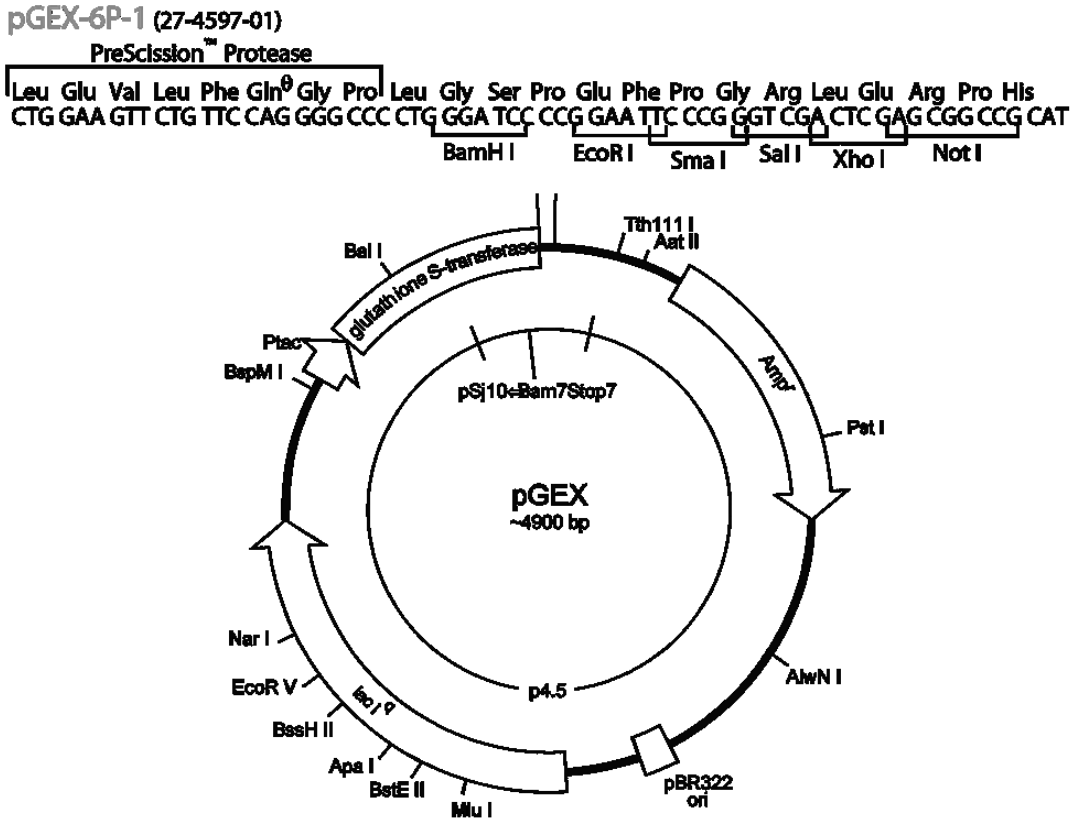
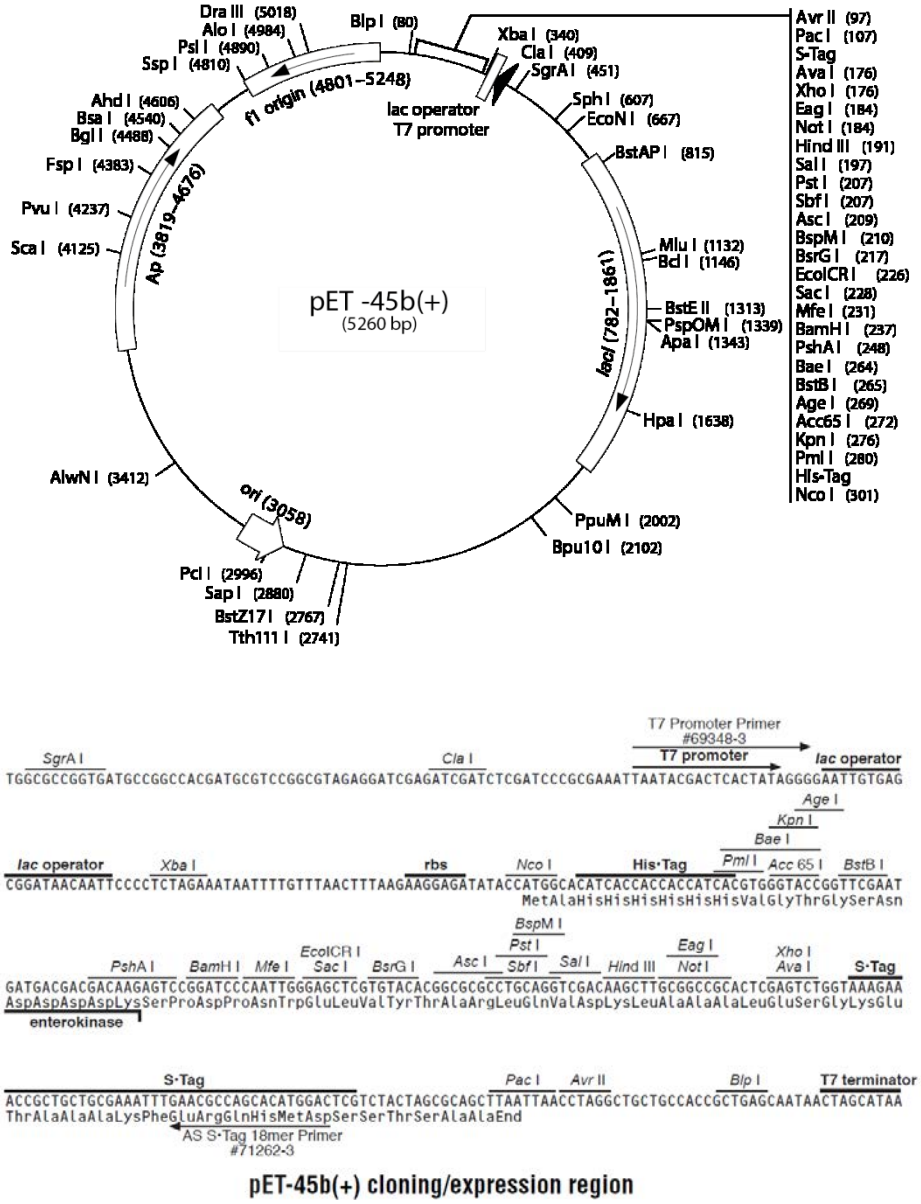


Figure 2.13 - pGex6p-1

pGex6p1 was used for all GST fusion proteins. Key features include T7 promoter for expression in *E.coli*, ampicillin resistance and a PreScission protease cleavable GST tag. (Image from GE Healthcare manual).



pET-45b(+) cloning/expression region

Figure 2.14 - pET 45b+

pET45b+ was used for the constructs SR78 and SR8-CT only, with an added TEV protease site. Key features include a T7 promoter, ampicillin resistance and an N-terminal 6x His tag. An artificial STOP codon was also encoded in the vector DNA. (Image from Novagen manual).

2.2 Protein expression

2.21 Transformation into DH5α cells

1 µl of plasmid DNA at 10-500 ng/µl was used to transform 20 µl of competent DH5α cells (Invitrogen). After a 30 min equilibration at 4 °C, a heat shock was applied at 42 °C for 30 sec. Super Optimal Broth – with catabolite repression (SOC) media (80 µl) (Invitrogen) was added and the cells were then grown at 37 °C, shaking at 220 rpm for 1 h. The transformed cells were then plated onto Luria Broth (LB) agar plates containing the appropriate antibiotic and grown overnight at 37 °C.

2.22 Isolation of plasmid DNA and DNA sequencing

A single colony was picked and used to inoculate a 10 ml LB culture containing the appropriate antibiotic and grown overnight at 37 °C and shaking at 220 rpm. Cells were harvested at 4000 g and plasmid DNA was extracted and purified using the QIAPREP Spin Miniprep Kit (Qiagen) according to the manufacturer's instructions. DNA concentration was measured using a Nanodrop system (Thermo Scientific). Plasmid DNA was sequenced at the Functional Genomics Laboratory (University of Birmingham, UK) using 10 µl of plasmid DNA at a concentration of 10-100 ng/µl and 25 ng of sequencing primer. Once sequence verified, the construct was transformed into *E.coli* strain BL21 (DE3) (Novagen) as described in 2.21.

2.23 Site Directed Mutagenesis

Site directed mutagenesis reactions were conducted using the QuickChange Lightning site directed mutagenesis kit (Agilent Technologies) according to the manufacturer's instructions. All PCR reactions were conducted on a Eppendorf Thermal Cycler (Eppendorf). Mutagenic primers were designed using the online primer design tool (www.agilent.com/genomics/qcpd) and synthesised by Invitrogen.

2.24 Small scale protein expression/solubility trials

For constructs that needed optimization for expression and/or solubility, small scale expression trials were conducted. A 10-200 ml culture containing the appropriate antibiotic was inoculated with a single colony. The culture was incubated at 37 °C with shaking at 220rpm until an O.D of 0.4-0.6 had been reached. Prior to induction cultures were incubated for 15-30 min at 18 °C, 20 °C, 25 °C, 28 °C or 37 °C , depending on size of the culture. The culture was divided into two so that a non IPTG induced and IPTG induced fraction could be compared by sodium dodecylsulphate polyacrylamide gel electrophoresis (SDS-Page) analysis. IPTG was added to a final concentration of 1 mM. The cultures were left to grow for either 4 h or overnight. Samples (10µl) were taken for analysis by SDS-Page from the non-IPTG and IPTG induced cultures.

2.25 Overnight growth of starter cultures and large scale expression in LB media

Starter cultures were prepared by inoculating LB media (10 ml) with single bacterial colonies. Starter cultures were grown overnight at 37 °C (shaking at 220 rpm) and expanded by adding to LB media (10ml per l) with appropriate antibiotic selection. Cultures were incubated at 37 °C until an O.D of 0.4-0.6 had been reached. Flasks were left to equilibrate at their optimal expression temperature, which typically was 18 °C for 45-60 mins. Cells were then induced with IPTG at a final concentration of 1 mM. Cells were then grown either for a further 4 h or overnight except in the case of those expressing vimentin which were induced with IPTG (final concentration of 1 mM) and then grown at 37 °C for a further 3-5 hours to form inclusion bodies.

2.26 Large scale expression of ^{15}N labelled proteins for NMR analysis in minimal media

Proteins selected for ^{15}N (and those including ^{13}C) labelling were grown in M9 minimal media (appendix A2). Proteins were expressed as described in 2.25 except the LB media was replaced

with M9 minimal media. Nutrient mix (appendix A2) was added to the minimal media after filter sterilization (0.2 µm) and before inoculation with the starter culture. After induction with IPTG (final concentration 1 mM), cells were grown overnight.

2.3 Protein Purification

2.31 Harvesting cells

E.coli cells were harvested by centrifugation at 4000 rcf for 20 min at 4 °C (Avanti J-20XP centrifuge with JLA 8.1 rotor). Cells were resuspended with either PBS or His-equilibration buffer (Appendix A2) and one complete protease inhibitor cocktail tablet (Roche Applied sciences). Cells were either stored at -80 °C or used immediately.

2.32 Mechanical cell lysis and separation of insoluble fraction

Typically, cell pellets (from cultures of 200ml or more) were mechanically homogenized and lysed using 3 passes through an Emulsiflex C-3 (Avestin, Canada). Throughout lysis the resuspended cell pellet was kept at 4 °C. For smaller volumes a sonicator was used (Soniprep 150 from MSE). The homogenized cell pellet was then spun at 25000 rcf for 45-60 min (Avanti J-25, JA25.50 rotor) at 4 °C. The soluble fraction was then decanted and filtered (0.2 µm filter).

Since a wide variety of solubility tags were used for protein purification an overview of how each tag was purified is given below.

2.33 Protein purification of GST-tagged proteins using GSTrap columns

Bacterial lysates were loaded onto a 5 ml GSTrap HP column (GE Healthcare) pre-equilibrated with at least 5 column volumes of GSTrap equilibration buffer (appendix A2). The column was then washed with at least 5 column volumes of GSTrap wash buffer (appendix A2). Finally the protein was eluted using GSTrap elution buffer (appendix A2). Samples were taken at each

purification stage for analysis by SDS-Page. Eluted protein was transferred to SnakeSkin dialysis tubing (Thermo Scientific) and dialyzed in 2 l of either His equilibration buffer (if His-tag was present on the protein to be purified) or PBS together with 25 units of PreScission protease (GE Healthcare). Pre and post cleavage samples were taken for SDS-Page analysis. The purification was continued only if over 90% cleavage had been achieved.

2.34 Protein purification of His-tagged proteins using HisTrap™ columns

His tagged proteins were purified with HisTrap columns (GE Healthcare). His-tagged proteins were loaded onto pre-equilibrated 5 ml HisTrap HP columns (GE Healthcare) with His equilibration buffer. These were washed with at least 5 column volumes of His wash buffer (appendix A2). Proteins were either eluted using a single step gradient or by a linear gradient using an Äkta prime system (Äkta prime Plus). If using a single step gradient proteins were eluted using 15-25 ml of HisTrap elution buffer in 1-2ml fractions. If using linear gradient purification, buffer A would be the His wash buffer while buffer B would be the His elution buffer eluted in 2ml fractions. Samples were analysed by SDS-Page to gauge purity. Fractions sufficiently pure were pooled. If further purification was required the pooled eluates were purified using size exclusion chromatography (see 2.36). In some cases proteins tagged with both GST and His were purified using GSTrap columns (as above) and then further purified (post cleavage of GST) by Nickel chelation chromatography.

2.35 Cleavage and removal of His-tag using Tobacco Etch Virus (TEV) protease

Proteins expressed in the pET45b+ and pProEX HTc vectors contained a TEV protease recognition site that allows the His tag to be cleaved from the purified protein. Pooled HisTag eluates were transferred to SnakeSkin dialysis tubing (Thermo Scientific) and dialyzed in His equilibration buffer in the presence of 10-50 µl 5mg/ml TEV protease (kindly provided by either Mrs Pooja Sridhar or Mrs Sandya Rajesh, University of Birmingham) at 4 °C and left overnight.

Pre and post cleavage samples were taken for SDS-Page analysis. If >90% cleavage had occurred the sample was purified as stated in 2.33. Those fractions containing the desired protein were then pooled.

2.36 Purification of proteins by size exclusion chromatography

Proteins were purified by size exclusion chromatography using either Superdex 75 or Superdex 200 column systems (GE Healthcare) depending on the molecular mass of the protein to be purified. Purification and analysis was implemented using the Äkta purifier FPLC system with Unicorn software. Depending on volume, proteins were either loaded onto 10/300 GL column or 26/60 PG columns. A flow rate of 0.5 ml/min was used for the 10/300 GL column with 0.5 ml fraction sizes, while 2.5 ml/min was used for the 26/60 PG column with 4ml fraction sizes. Samples of fractions containing the highest peaks at 280 nm were analysed by SDS-Page and those with sufficient quantity and purity of the protein were pooled.

2.37 Vimentin purification

Large scale expression of vimentin was performed as described in section 2.25. Essentially a modified version of the Garboczi method (Garboczi et al., 1992) was utilized in which the final 0 M urea buffer had been previously shown to contain folded vimentin (Quinlan et al., 1986). Vimentin was expressed in inclusion bodies which were retained in the insoluble fraction. Hence in the case of vimentin the soluble fraction was discarded and the insoluble fraction re-suspended and washed with Triton wash buffer (appendix A2) and centrifuged at 13500 g for 15 min. This was repeated 3 times. A final wash step was carried out using triton wash buffer without triton X-100. The final insoluble pellet was re-suspended in solubilisation buffer (appendix A2) and left on a rocker overnight at 4 °C. Once the pellet had solubilized, the sample was centrifuged at 15000 g for 30 min. Vimentin was retained in the soluble fraction, with the insoluble fraction discarded. Folded vimentin was obtained following the step wise removal of

urea in the solubilisation buffer by dialysis. The soluble fraction was dialyzed in 10 mM Tris-HCl (pH 8), 5 mM DTT, 4 M urea overnight, then 10mM Tris-HCl pH 8, 5mM DTT, 2M urea overnight and finally 10mM Tris-HCl pH 7, 5mM DTT overnight at room temperature. In preparation for NMR titrations the sample was dialyzed in the same buffer as the labelled protein.

2.4 Biochemical analysis of proteins

2.41Sodium dodecylsulphate polyacrylamide gel electrophoresis gels

Samples (1-15 µl) were added to an equal volume of 2x Laemelli buffer (Sigma-Aldrich) and loaded on to sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-Page) 4-12% pre-cast bis-tris gels (Bio-Rad). Precision Plus molecular weight markers (Bio-Rad) (10 µl) were also loaded on to all gels. Gels were run at 180 V for 40min in 2-(N-morpholino)ethanesulfonic acid MES SDS running buffer (Bio-Rad). Gels were stained with Instant blue™ (BioRad) and left on a shaker for at least 1 h. Gels were washed thoroughly at least three times with water and an image taken for future reference.

2.42 Concentrating proteins and calculation of concentration

Proteins were concentrated using Amicon Ultra™ (Amicon) or Vivaspin (Vivaspin) concentrators using appropriate molecular weight cut offs. These were centrifuged at 4000 g at 4 °C using a centrifuge (Eppendorf Centrifuge 5810R). Final protein concentration was measured using a UV spectrophotometer (Shimadzu UV-1700 Pharmaspec) with an absorption set to 280nm. The optical density reading and ProtParam (used at http://web.expasy.org/protparam/) parameters for each individual protein were input into the following equation:

Concentration (M) = (A_{280nm} x dilution factor) / (extinction co-efficient ($M^{-1}cm^{-1}$) x path length (cm))

2.5 Biophysical analysis

2.51 Analytical Ultracentrifugation (AUC)

Analytical Ultracentrifugation experiments were carried out in collaboration with Mrs Rosemary Parslow at the Birmingham Biophysical Characterization facility (University of Birmingham). Proteins were supplied at three concentrations between 0.05-5 mg/ml in PBS. Samples were centrifuged at 25000 and 40000 rpm for 20h at 20 °C on a Beckman XL-1 analytical ultracentrifuge (Beckman Coulter) and detected at 280nm. Sedimentation profiles were analysed using the SEDFIT software (Brown and Schuck, 2006).

2.52 Circular Dichroism

Circular dichroism (CD) is a spectroscopic technique in which secondary structure can be ascertained. It uses circular polarized light which has a property of moving in a circular motion along a vector. The motion is both left and right handed and if absorbed equally, no difference is observed. However, if absorbed by optically active chiral molecules such as proteins, a difference in speed and extent of both the left and right handed motion causes a difference in the polarization of light that is recordable. As a result CD values are measured as a change in absorbance or more accurately:

$$\Delta\varepsilon \equiv \varepsilon_L - \varepsilon_R.$$

Where $\Delta\varepsilon$ is the observed difference in the polarization of light, ε_L is the left handed component and ε_R is the right handed component.

All proteins subjected to CD analysis were dialyzed into 20 mM sodium phosphate pH 7.2-7.5. A reduction in Cl^- ions is critical to maximizing signal to noise at lower wavelengths of ~180-195nm. Proteins at a concentration of 0.1 mg/ml were transferred to a 1 mm cuvette with a capacity of ~300 µl. Measurements were conducted on a JASCO-J810 spectrometer, at 25 °C

with a spectrum collected between 240nm-180nm, the range in which peptide bonds absorb polarized light. Data were deconvoluted using the DICHROWEB server (Lobley et al., 2002) which is a database of CD spectra from proteins of known secondary structure. This gives an estimated percentage per secondary structure element.

2.53 Thermal stability measured by Thermofluor™

Thermofluor™ is an assay used to measure protein thermal stability. It relies on a fluorescent dye (SYPRO Orange) which binds to hydrophobic residues in the core of the protein. An increase in fluorescent signal is observed when the protein begins to unfold as a result of temperature increase. For folded proteins a sigmoidal curve is often observed which often tails off. The “tailing off” is attributed to degradation of the dye at high temperatures according to the developers of the assay (Ericsson et al., 2006). Since disordered proteins do not have hydrophobic cores, this assay is less sensitive to measuring their thermal stability, however, can give a qualitative suggestion that they are unfolded. This however, is not definitive, given that some folded proteins may have hydrophobic patches on the surface which skews the fluorescent signal. The assay can also be used to identify optimal buffer conditions. Optimized buffer conditions can be observed by higher melting temperatures since this correlates with the thermal stability of the protein. Screening for optimized buffer conditions involved using a 96 well plate with each well containing a different buffer of varying buffer conditions developed by the Overduin lab.

Fluorescent dye (2 µl of a 10-20x stock) was added to the protein (20-100 µM) (final volume 200 µl). 2 µl was of the protein/dye mix was added to 18 µl of the buffer of interest. The sample was loaded into the wells of a 96 well Peqlab plate (Low profile reaction plate) and this was placed in an RT- Mx3005P PCR machine (Stratagene). The experiment was carried out at 25 °C -90 °C and fluorescence was filtered through custom interference excitation (492nm) and emission

(568nm) filters. The data was either plotted from raw data or normalized using Microsoft Excel and T_m values calculated using the XLFit 5 program (IDBS solutions) to fit the data to a Boltzmann model represented by the equation:

$$intensity = \left(A + \left(\frac{B - A}{1 + e\left(\frac{C - temperature}{D}\right)} \right) \right)$$

Where *C* is the T_m; *A* is pretransitional and *B* is posttransitional intensities and *D* is a slope factor. A sigmoidal curve fit was applied which scales the data as suggested by previous protocols (Ericsson et al., 2006).

2.54 Limited Proteolysis and N-terminal sequencing

In an effort to identify protease resistant fragments (separated by flexible and exposed linkers) by limited proteolysis two alternative experiments were carried out. Proteins at 10 mg/ml were subjected to limited proteolysis using a variety of proteases from the Proti-Ace™ Kit (Hampton Research) according to the manufacturer's instructions. Alternatively, proteins at 10 mg/ml were subjected to 10-100 ng (final concentration) of trypsin digestion. Trypsin typically cleaves at the carboxyl side of the amino acids lysine or arginine. Typically limited proteolysis reactions were carried out at 37 °C for 1 h or overnight, or at room temperature for 1 h-9 days. Samples were collected at different time points for SDS-Page analysis. Protease resistant fragments were excised from the SDS-Page gel and subjected to N-terminal sequencing by Alta Bioscience (University of Birmingham).

2.6 NMR experiments

2.61 NMR Introduction

NMR is a powerful technique in which a variety of parameters can be acquired, these include (but are not limited to): determination of the solution structure of a protein at atomic resolution, gauging overall fold, backbone assignment (the sequential assignment of a peak in the HSQC to a residue on the protein chain), side chain assignment and protein-protein interactions. It has an advantage over X-ray crystallography (discussed in 2.7) by lacking the rate limiting step of requiring proteins to crystallize. It also has the added advantage of being able to be performed in near-physiological solutions in which proteins are more likely to assume native conformations. Unlike X-ray crystallography NMR has a size limitation whereby proteins greater than 30 kDa are often too challenging for 3D structural determination due to significant line broadening and peak overlap. Although full NMR structural determination was not carried out on any protein, a brief overview of the theory will be given.

2.62 NMR theory

NMR uses the properties of ^{1}H, ^{13}C, and ^{15}N to acquire information on proteins. All atomic nuclei possess two properties: one called spin and the other magnetism. Spin values are determined by the proton and neutron content of the nucleus. Only spin values that are integral or half integral are amenable for NMR as these are associated with a magnetic field and as such these types of nuclei are called spin active. ^{1}H, ^{13}C, ^{15}N are all spin half nuclei. These nuclei have spins that can be in one of two orientations e.g. a lower energy state and a higher energy state. They can thus align with the magnetic field in one of the two orientations.

Magnetism is a property in which nuclei with non zero spin (in this case ^{1}H, ^{13}C and ^{15}N) also possess a magnetic moment. When the external magnetic field is applied from the spectrometer

the spin of each nuclei precess about the axis of the magnetic field. The nuclear spins then either align with or against the field. A key feature of NMR is that nuclei align slightly more with than against the field. If the nuclei aligned equally with and against the field, no signal would be observed. The frequency at which the spins precess is termed the Larmor frequency. The introduction of a radiofrequency (RF) pulse (a second oscillating magnetic field perpendicular to the static magnetic field) causes fluctuations to the magnetic field when the RF pulse is removed. This causes the spins to precess back to their original state, (called the ground state) in a process termed relaxation. An electrical signal is then generated as a result of relaxation in the metal coil surrounding the NMR tube which is recorded as the Free Induction Decay (FID). This is then processed by Fourier transform (a mathematical function that deconvolutes a signal made up of a multitude of different signals which is plot as a frequency spectrum) and is used to plot a 1-dimensional proton spectrum. The Fourier transform in effect converts a function of time to a function of frequency.

The Larmor frequency for a given nucleus is proportional to the applied magnetic field. However, for nuclei within proteins surrounded by moving electrons, the situation is more complicated. For each nucleus the local magnetic field around the nucleus is altered caused by an opposing magnetic field produced by the surrounding electrons. This in turn reduces the external magnetic field the nucleus encounters and alters the Larmor frequency. This “shielding” effect is termed the chemical shift phenomenon. This alters the FID which is composed of a multitude of different frequencies, which when Fourier transformed reveals a spectrum of chemical shifts with each peak corresponding to an individual chemical environment. Since the Larmor frequency is proportional to the applied magnetic field the scale is described in parts per million (ppm). As a result this unit is also independent of the magnet size.

2.63 Preparation of the sample for NMR analysis

Proteins of > 90 % purity as determined by SDS-PAGE analysis, and of a concentration of at least 100 µM were analyzed by NMR in the presence of 10 % D_2O. Samples were analysed using Agilent 600 MHz, 800 MHz or 900 MHz spectrometers (Biomolecular NMR facility, University of Birmingham). Each experiment used a triple-resonance cryogenic probe and z-axis pulsed-field gradients. Parallel to the data acquisition experiment another experiment was conducted in which the D_2O signal was measured. This is the lock signal. Locking the D_2O signal prevents “magnetic drift” which can lead to errors in the magnetic pulses of the NMR experiment. This corrects the magnet so that the signal returns to its original value.

Firstly, the spectrometer is tuned and shimmed using automated gradient shimming on the z dimension. The transmitter offset was also optimized using automated methods. Since the molarity of water is at 55 M, in vast excess to the typical protein concentration, this can have an adverse effect on data acquisition and must therefore be suppressed. This can be conducted using the software package of the spectrometer automatically. The signal of the protein can now be acquired. The simplest of all experiments is a 90 ° pulse (as in 1D-NMR) and this needs optimization. The 90 ° pulse is the length of time (in µs) the RF pulse is switched on for to allow the magnetization of nuclei to rotate onto the y axis from the z axis. By the magnetization precessing back to the z axis from the xy plane it gives the maximum signal possible from the protein. In addition the in-built pulse programs within the Agilent software package must also be manually checked. These include: length and power of the water suppression pulse, the recovery time and the phase correction pulse.

Since one of the main aims of this project was to elucidate the structure of a novel domain, many constructs were gauged by 1-D and 2D-Heteronuclear single quantum coherence (HSQC) spectra, to assess overall fold (how these were assessed is discussed in section 2.65). Since

many domains were of sufficient quality for NMR assignment high impact constructs were taken further to 3D experiments which were carried out in collaboration with Dr Mark Jeeves. These included: SR78 and the envoplakin SH3 domain.

2.64 1D NMR

One-dimensional NMR can give limited information on the folded state of the protein. These experiments were often used to gauge the extent of folding of unlabelled proteins or as a test to ascertain fold before HSQC spectra of labelled proteins. These experiments consist of a simple 90 ° pulse on proton followed by signal acquisition. Typically they were carried out on an Agilent 600 MHz Direct Drive NMR spectrometer equipped with a HCN 5 mm cryogenic probe at 25 °C using standard 1D Biopack experiments (Agilent).

2.65 2D-HSQC NMR

Heteronuclear Single Quantum Coherence (HSQC) is a 2-D NMR method which correlates attached protons and ^{15}N labelled amide groups, and was used primarily in this study to assess the foldedness of proteins. Typically a 2D NMR method records the NMR signal as a function of two time variables, t_1 and t_2. The resulting free induction decay is then fourier transformed twice which is a function of two frequency variables. It can thus be broken down into four stages. Firstly, the preparation stage where the sample is excited by one or more pulses. Secondly, the resulting magnetization of the first stage is allowed to evolve as the time period t_1. The third stage, the mixing period contains further pulse(s). The fourth stage records the signal as a function of the second time variable t_2. Once processed, the result is a 2D spectrum with a ^{15}N y axis and ^{1}H x axis in ppm.

HSQC spectra are predominantly used for backbone and secondary structure assignment and binding studies which in combination with 3D experiments can be used to assess the overall fold

of the protein. The HSQC gives a 2-dimensional spectrum with the first being proton and the second being a heteroatom, predominantly nitrogen (or carbon). Backbone amides are represented by a single peak observed in the spectrum (with the exception of prolines which due to their lack of protons directly attached to nitrogens do not show a signal) whilst for side chain NH_2 groups (Asn and Gln) two peaks are observed at the same nitrogen frequency, corresponding to the two protons attached to a single nitrogen. In effect to ascertain whether all the expected peaks were observed the following equation was used:

Peaks expected = [(Number of amino acids – prolines) + 2(Asn + Gln)]

Qualitative assessment was also used in which as long as a good signal to noise ratio was observed, proteins were classified into four main groups:

a) The “folded” globular proteins suitable for structure determination with >90% of peaks well dispersed between 7-10ppm, as was the case for the envoplakin PRD (see figure 5.3).
b) The “flexible and/or partially folded” proteins exhibiting <90% of expected peaks but with dispersion limited to the 8-8.5ppm region, as in the case of SR56 (see figure 3.72).
c) The “unstructured / highly disordered” protein with >90% peaks visible, relatively good dispersion of peaks (particularly for small proteins) in the ^{15}N dimension but poor dispersion in the ^{1}H dimension (usually between 8-8.5 ppm), as in the case of the desmoplakin SH3 domain (see figure 3.71).
d) The “unresolvable” protein where few peaks were observed and no interpretation of overall fold could be reliably inferred, owing to possible aggregation or intermediate timescale dynamics occurring, as in the case of SR8-CT (see figure 3.74).

2.66 NMR titrations

In order to characterise protein-protein interactions, NMR titrations were conducted with one protein ^{15}N labelled and the other unlabelled. Initially an $^{1}H,^{15}N$-HSQC was taken of the labelled protein. These were compared and superimposed onto $^{1}H,^{15}N$-HSQC spectra of ^{15}N labelled protein in the presence of unlabelled protein ligand. Depending on stock concentrations of labelled and unlabelled proteins, NMR titrations were conducted in one of two ways. If stock concentrations were low, one NMR tube was used with the unlabelled protein added in gradual amounts, increasing the overall volume and lowering the labelled protein's concentration. The resulting decrease in labelled protein's concentration was artificially compensated during analysis by altering the threshold levels of the signal to noise ratio. Where stock concentrations were ample, two NMR tubes were used with one tube containing an excess of labelled protein and the other tube an excess of unlabelled protein. These were then analysed by $^{1}H,^{15}N$-HSQC and taken as the start and end points of the titration. An equal volume was then taken from each tube and added to the other and analyzed by 2D-HSQC at different points until a saturation point was reached.

2.67 Data processing and analysis

Before analysis, FIDs were processed in NMRPipe (Delaglio et al., 1995) using a shifted sine bell apodization . This is a post-experimental processing step in order to optimize signal to noise ratios. Superpositions of spectra collected with similar parameters were carried out using the program CCPN analysis (Vranken et al., 2005).

2.7 Small Angle X-Ray Scattering (SAXS)

2.71 History and the use of SAXS

SAXS was used to characterise the structures and dynamics of constructs in solution with conditions similar to those used for NMR. It is becoming an increasingly important and complementary tool for structural biologists across a wide spectrum of fields. It is a fundamental technique that allows for the study of proteins, nucleic acids and complexes in solution. It can give information on size, shape and flexibility amongst other things (Putnam et al., 2007). While SAXS has a low resolution of 1-2 nm it has an advantage over crystallography in terms of the physiological relevance and lack of the need for crystals, being in solution. It also has an advantage over NMR in terms of the almost limitless size of protein it can characterize. The technique is especially useful where proteins are refractory to structure elucidation by NMR or X-ray crystallography as was the case with many proteins in this project.

The technique was originally pioneered by Guinier in the late 1930s when working on metallic alloys (Guinier, 1939). The work was later refined in the 1950s for use on easily purified proteins such as haemoglobin (Guinier, 1955). Breakthroughs occurred in *ab initio* modeling by Dmitri Svergun and co-workers in the 1990s (Semenyuk and Svergun, 1991, Svergun et al., 1995). This, together with advances in instrumentation, the use of synchrotron radiation and high-throughput automation have resulted in SAXS becoming accessible to the wider scientific community and not just restricted to structural biologists.

Since SAXS was one of the main structural techniques for the experiments presented in this thesis a detailed scope of the theory and methods involved is given below.

2.72 SAXS theory

X-rays are plane waves which have properties of amplitude and wavelength. As they pass through matter they interact with their atoms and secondary wavelets are generated which constructively or destructively interfere with one another in a coherent fashion. Diffraction patterns are thus created which if detected appropriately (i.e. by pixel detectors) are exploited by X-ray crystallography and similarly in SAXS. During the SAXS experiment a sample is exposed to a highly collimated X-ray from a very small distance. Because biomolecules scatter weakly and concentrations are relatively dilute the method often requires high intensity X-ray photon beams only available at synchrotron facilities. The resulting scattering pattern is collected at small angles (typically ≤ 3 Å) by a detector and the direct exposure of the beam is absorbed by a beamstop (figure 2.2). The scattering pattern is radially averaged to give a scattering profile which is calculated by the intensity (I) versus the scattering angle (q) (where $q=4\pi\sin\theta/\lambda$, 2θ is the scattering angle and λ is the wavelength of the X-ray).

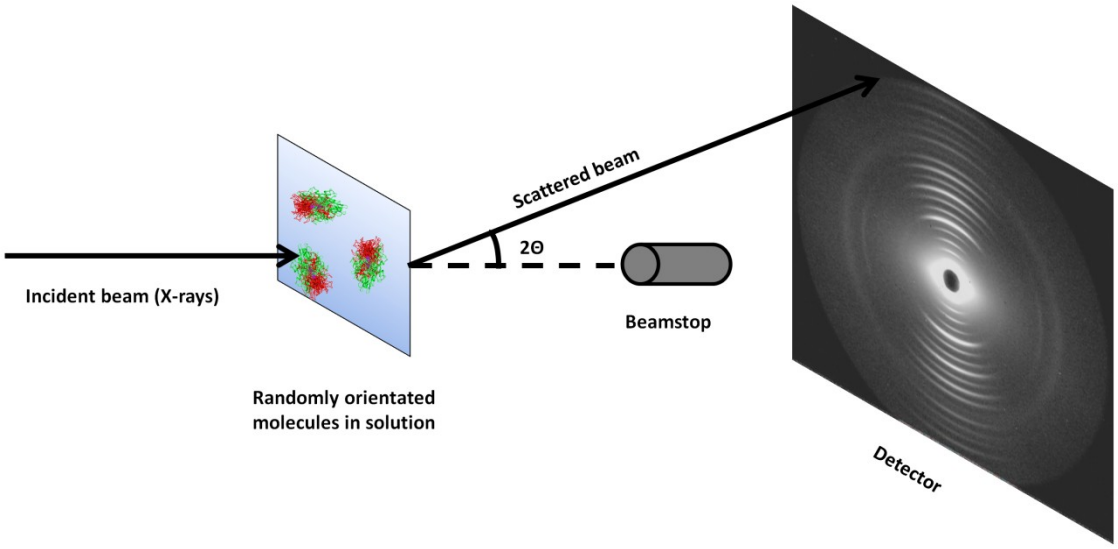


Figure 2.2 The experimental setup of SAXS

A schematic diagram for the experimental setup of SAXS. Samples are exposed to X-rays which scatter at small angles and are collected by detector (see text for details).

2.721 SAXS Data collection and reduction

Protein measurements were collected from three runs on the EMBL X33 beamline of the storage ring DORIS III (Deutsches Elektronen Synchrotron, Hamburg, Germany). All measurements were conducted at room temperature using an automated SAXS sample changer. The data was recorded using a single photon counting pixel detector system (PILATUS 500 K) at a sample detector distance of 2675 mm covering the range of 0.06-6 nm using a wavelength of 1.5 Å. Proteins were measured between 0.5 and 10 mg /ml in 20mM sodium phosphate (pH 7.2) and 100mM NaCl unless otherwise stated. Using the in-house automated processing software available, buffer scattering was automatically subtracted from the protein scattering. The software also indicated possible radiation damage to the sample; such samples were excluded from further analysis. The same software gave basic analysis of the protein sample which included auto-subtracted scattering curves to allow for manual checking of the data quality and parameters such as: R_g and D_{max} (see 2.73). Both sets of information gave indications as to whether the sample should be re-run. Given that runs at synchrotron facilities are infrequent, it was important to maximize quality of data before the experiment had ended.

2.73 Ascertaining data quality

The correct interpretation and processing of SAXS data is the most critical bottleneck in data analysis. Mis-interpretation of the data can lead to incorrect processing and erroneous biological conclusions. There are numerous parameters that need to be checked prior to continuing on to data analysis. Using the program PRIMUS (Konarev et al., 2003) the main parameters were checked as described below and are based on checkpoints suggested at the European Molecular Biology Organization (EMBO) SAXS workshop (ESRF, Grenoble, May 2011) and S4SAS Workshop (Cardiff University, September 2010):

a) Are inter-particle effects observed?

   This can be checked by firstly making sure the Guinier region is linear. If an upward trend is observed this suggests that at that particular concentration aggregation is occurring. If a downward trend is observed it suggests that repulsive interactions are occurring. If either event is observed at the very least the scattering at low angles should not be used in the merged data set, or the entire data set not used at all.

b) Have alternate concentrations of the protein been merged efficiently?

   Acquisition of at least three different concentrations is encouraged at synchrotrons. One of the main benefits of this strategy is based upon the fact that differing concentrations can benefit the scattering graph. High concentrations give the benefit of increased signal to noise, especially at higher angles, although inter-particle effects are much more likely to be observed at lower angles. Low concentrations give the benefit of a decreased likelihood of inter-particle effects observed at lower angles and would therefore give a more accurate radius of gyration (R_g), although signal to noise is decreased at higher angles. Once scaled, the merged data sets of a combination of low concentration at low angles and high concentration at higher angles would be used for further data analysis.

c) Have mutants / homologues been scaled and superimposed correctly?

   Due to a normalization process of the scattering patterns by the in house auto processing software, two similar proteins may superficially appear different. This can be verified by scaling the scattering patterns using PRIMUS. Superimposable scattering patterns suggest that no global changes have occurred owing to the low resolution of SAXS. Differences may be apparent at atomic resolution, however, would need to be verified by alternative structural methods.

In addition to the above mentioned checkpoints, further processing suggestions were taken from an excellent review which discusses guidelines on SAXS analysis and is increasingly becoming the gold standard for SAXS analysis interpretation (Jacques and Trewhella, 2010).

2.74 SAXS Data analysis

Once the data was judged to be of high quality, each merged data set was analysed using a suite of programs called "ATSAS" (Konarev et al., 2006). PRIMUS was used to scale, merge and evaluate R_g. The evaluation of the R_g was conducted manually using the Guinier approximation (Guinier, 1939) assuming that at small angles $s < 1.3/R_g$. The R_g was also automatically evaluated using the program GNOM which also provided the distance distribution function p(r) and the maximum particle dimension (D_{max}). Typically the R_g value from GNOM was used for this thesis as it has an algorithm to automatically detect the region most representational of the R_g from the scattering curve (as long as it corroborated with the manual R_g value). D_{max} was found using a method suggested by the S4SAS Workshop (Cardiff University, 2010) in which 3 x R_g was taken as the initial value and a larger value was input to gauge where the distribution function became negative. Furthermore, a variety of values corresponding approximately to where the distribution function became negative were tried until the slope did not have a y-axis value of 0 preceded by a sharp decline.

2.75 Ab initio shape reconstruction and comparison

The output files from GNOM were used as the input files for the *ab inito* reconstruction program DAMMIF (Franke and Svergun, 2009) which creates envelopes of the protein based on the SAXS scattering profile. Ten low resolution *ab initio* bead models were generated using its default parameters. The program works using a simulated annealing process to fit a curve, initially one of a theoretical spherical shape, to one more representational of the scattering curve. Spatial discrepancy was minimized in the 10 low resolution *ab initio* models by DAMSEL

which gives an normalized spatial discrepancy (NSD) value to each model with the lowest value given to the model most representational of all the models. The DAMSEL output values were input into DAMAVER (Volkov and Svergun, 2003) to give an overall averaged model and finally DAMFILT to filter non overlapping dummy residues. In some cases bead models were converted into a volumetric map using Situs 2.5 (Wriggers and Chacon, 2001). Models were visualized using either PYMOL (Schrodinger, 2010) or CHIMERA (Pettersen et al., 2004).

2.76 Comparison of SAXS models to atomic resolution models

Atomic resolution models were either taken from the Protein Data Bank (PDB), from structures solved in this project or by using the I-TASSER (Roy et al., 2010) or PHYRE (Kelley and Sternberg, 2009) model prediction servers. A number of methods were used to compare SAXS models to atomic resolution models. Firstly, CRYSOL (Svergun et al., 1995) was used to back-calculate theoretical scattering curves and fitted to experimental scattering data from atomic resolution models. A χ^2 value was given, with values between 0.8-1.2 deemed to be highly similar. Secondly, SUPCOMB (Kozin and Svergun, 2001) was used to compare a SAXS model to an atomic resolution model. A NSD value (a unit-less value due to the units involved in the calculation cancelling out) was given with values under 1.5 deemed to be highly similar. Unlike Root Mean Squared Deviation (RMSD) values which require atomic resolution coordinates between two similar structures, NSD values calculated by SUPCOMB have the benefit of allowing comparisons between two models created by two differing structural techniques (i.e. an *ab-initio* envelope created by SAXS and a crystal structure). For visualization purposes some models required manual fitting.

2.77 Flexibility assessment

SAXS scattering curves from multi-domain proteins were subjected to flexibility assessment by the Ensemble Optimization Method (EOM) suite (Bernadó et al., 2007). This incorporates two

separate programs: RANdom CHain (RANCH) and Genetic Algorithm Judging Optimization of Ensembles (GAJOE) (Bernadó et al., 2007). EOM assumes that for a given SAXS scattering data set, multiple conformations of a flexible protein exist in solution giving an averaged final curve. It can be used to quantitatively characterize flexibility and analyze size distribution between a number of different conformers. EOM was conducted using either the ATSAS Online server (Petoukhov et al., 2007) or manually. If conducted manually RANCH was used to generate 10000 structures with chains between globular domains denoted as “random”. GAJOE was then used to create a subset of 50 conformers most representational of the SAXS scattering curve using default parameters. GAJOE then gives the fit to the SAXS scattering curve represented visually and by a mathematical χ^2 value, of which 1 would be optimal. The ATSAS online server uses the same protocol (as the manually conducted EOM parameters) except the final subset was 20 conformers. The advantage of using ATSAS online is that it is computationally (processor speed and memory demands) less demanding yet retains the reliability of conducting it manually. It should be noted that fits that do not match the experimental SAXS scattering curve do not represent the protein in solution. Further details can be found in the paper by Bernadó et al (Bernadó et al., 2007). The subset of models most representational of the SAXS scattering curve (called the selected ensemble) was taken for further analysis. An in-house program designed to gauge inter-domain distance distributions between two different amino acids from the random pool and selected ensemble was kindly created by Dr Pau Bernado (Code given in appendix C). Using this program, two geometric points (amino acids) were selected and the distance between them were calculated. A distribution graph comparing the two amino acid distances of each model were then created of the random pool and the selected ensemble.

2.8 X-ray crystallography

2.81 Introduction

X-ray crystallography is one of the oldest and most widely used techniques for determining the three-dimensional structure of macromolecules. The first stage in this process is to crystallise the protein of interest by reducing protein solubility and hence forcing the protein molecules in solution to interact with another in a specific and controlled manner. To aid in the crystallisation screening process of proteins, there are numerous commercial crystallisation screens available which manipulate the protein in different ways such that it promotes protein-protein interactions in solution. The purity of the protein solution is critical as the presence of contaminants can disrupt crystal formation or yield crystals that diffract X-rays weakly. Typically the "hanging-drop" vapour diffusion method is used to crystallise proteins in which a small drop of protein solution is mixed with crystallization condition in an enclosed chamber with a well of crystallization buffer. Due to differences in osmolarity between the drop and the well reagent, water diffuses out of the drop by vapour diffusion until equilibrium is achieved. During this process the protein concentration in the drop steadily increases reaching supersaturated levels from which nucleation and crystal formation may take place. Following several days of incubation, crystallization plates are viewed using a light microscope to identify initial conditions that have led to crystal formation. These preliminary hits are subsequently optimised in an attempt to grow diffraction-grade crystals. Once crystals of suitable size and morphology are grown they are flash-cooled in liquid N_2 and subjected to X-ray diffraction experiments using a focused X-ray beam. The crystal upon exposure to an X-ray beam will diffract X-rays in various directions. These diffracted X-rays are recorded by an X-ray detector in the form of array of spots (or reflections) that are collectively referred to as a diffraction pattern. More specifically, it is the electrons that surround the protein atoms within the crystal that diffract X-rays. The crystal is

then rotated with respect to the X-ray beam to allow all possible diffracted X-rays to be recorded and permit a complete data set to be collected. The diffraction data is then processed which can lead to solving the three dimensional structure of the protein. This involves determining the amplitude, phase and frequency for each reflection (observed as a “dot”) on the diffraction pattern. The amplitude of each reflection can be calculated directly from the intensity. The frequency is simply the wavelength of the X-ray beam, and typically 1.54 Å for an in-house CuKα X-ray generator. The phase angles of each reflection cannot be determined experimentally (known as the ‘phase problem’) but can be acquired using homologous structures (molecular replacement) or soaking with heavy atoms (multiple isomorphous replacement and single wavelength anomalous dispersion). The combination of amplitudes and estimated phase information allows for an electron density map to be calculated which permits a molecular model to be built. The model is then improved in an iterative fashion in a process called refinement, whereby it is altered to fit the electron density map. The progress of refinement is continually monitored through a reduction in two statistics, the R_{work} and R_{free}.

2.82 Crystallization screens

A variety of commercially available sparse-matrix based screens from Hampton Research (Index), Emerald Biosystems (Wizard 1 and II) and Molecular Dimensions Ltd (PACT and JCSG+) were used for crystallisation screening trials. Small-scale crystallisation screening trials were set up in 96 well plates (Iwaki) using a Mosquito nanolitre crystallization robot (TTP Labtech). Briefly, 100 nl of protein were mixed with 100 nl of crystallization buffer on a viewdrop2 plate seal and equilibrated against 100 µl of crystallization buffer. A variety of protein concentrations were used in the initial crystallisation screening trials. Once a number of crystallization buffers had been identified as providing promising “hits”, these were optimized by altering conditions such as pH of the crystallization buffer, precipitant concentration and drop

volume. In addition, further optimisations involved adjusting the protein concentration as well as attempting sitting drop experiments. Once these conditions had been optimized, large scale screens were set up using 24 well Linbro plates to grow diffraction-grade crystals. Larger size drops typically 1 µl of protein solution and 1 µl of crystallisation reagent were manually mixed on a cover slip which was inverted onto a greased chamber containing 1 ml of the crystallisation solution. The enclosed system allowed the hanging drop to equilibrate with the well solution until crystals were formed. All screening trials were incubated at 23 °C and examined for crystal growth at 3 days and then weekly thereafter using a light microscope (Leica).

2.83 Data collection and processing

Data collection and processing was conducted primarily by Dr Fiyaz Mohammed so only a brief overview will be given. Crystals deemed to be worthy of X-ray data collection were soaked in a number of solutions containing crystallization buffer and increasing concentrations of the cryoprotectant, ethylene glycol. Crystals were then flash cooled at 100 K in N_2 (Oxford Cryosystems). X-ray data were collected using an in-house MicroMax 007HF rotating anode X-ray generator (Rigaku) with a Saturn CCD detector (University of Birmingham Macromolecular X-ray facility). Diffraction data sets were integrated, scaled and merged using programs of the XDS suite (Kabsch, 2010). Molecular replacement was attempted using the PHENIX suite (Adams et al., 2010) with varying lengths of different models or homologues.

2.84 Single Anomalous wavelength Dispersion (SAD) phasing with Iodine

An alternative approach to molecular replacement utilizes a halide such as iodide, whereby initial phase information was determined through their anomalous scattering signal. An approach pioneered by Dauter et al (2000) . Large scale screens using optimized crystallization buffer in the presence of no greater than 1 M sodium or ammonium Iodide were attempted. Crystals already grown in native conditions were soaked in mother liquor (and cryoprotectant)

supplemented with varying concentrations of sodium iodide concentrations. Structure determination was conducted by Dr Scott White (University of Birmingham) using programs of the PHENIX suite (Adams et al., 2010) and will therefore not be discussed in detail. The molecular model was refined with PHENIX interspersed with rounds of model building using COOT (Emsley et al., 2010).

2.91 Data analysis, presentation and molecular docking modelling

Proteins were visualized using either PYMOL (Schrodinger, 2010) or CHIMERA (Pettersen et al., 2004). Electrostatic surfaces were calculated using an online server for PDB2PQR at: http://kryptonite.nbcr.net/pdb2pqr/ and APBS using default parameters (Dolinsky et al., 2004). The program High Ambiguity Driven DOCKing (HADDOCK) was used to model potential complexes using default parameters (Dominguez et al., 2003).

2.92 Vimentin co-sedimentation assays

Vimentin co-sedimentation assays were used to ascertain whether protein constructs from the C-terminal tails could bind to vimentin in the polymerized state. Using previously published protocols (Choi et al., 2002) and manufacturer's instructions (Cytoskeleton, Inc) desmoplakin PRD C and envoplakin were co-sedimented with vimentin using an Optima Max ultracentrifuge (Beckman-Coulter).

Chapter 3 – Biophysical analysis of constructs from plakin domains

Many members of the plakin family bind to both desmosomes and IFs. These include desmoplakin, periplakin, envoplakin and plectin. It is known that the N-terminal head domain is responsible for tethering the plakin protein to the desmosome while the C-terminal tail region binds to IFs (Choi et al., 2002, Bornslaeger et al., 2001, Kowalczyk et al., 1997, Meng et al., 1997a, Stappenbeck et al., 1993). The desmoplakin N-terminal region is also mandatory for effective assembly of the desmosome in cooperation with plakoglobin and plakophilin (Bornslaeger et al., 2001). However, the binding regions on the desmoplakin N-terminal region have not been mapped more precisely. The N-terminal region of plakin proteins typically consist of a presumably unstructured extreme N-terminal region and a globular plakin domain.

Previous studies revealed insights into the plakin domain architecture. In 1992 the plakin domain was proposed to consist of 5 α-helical regions termed NN, Z, Y, X, W, V (Virata, 1992). It was not until 2007 that a study on BPAG1 revealed that the plakin domain was actually part of the spectrin superfamily (Jefferson et al., 2007). Further studies then showed greater details for the proposed architecture of the entire plakin domain based on sequence homology (Sonnenberg et al., 2007). These studies suggested that each plakin domain typically consisted of 6 SRs and a SH3 domain. However, how the SRs organized in combination with the SH3 domain within the plakin domain architecture in solution was unclear. Ascertaining how the plakin domain SR modules orientate may clarify why certain regions are involved in assembly of the desmosome. Understanding how the SR modules orientate would corroborate ligand binding studies to other desmosomal proteins conducted by other members of the lab.

During the preparation for this chapter, two independent groups solved the structures of the first 4 spectrin repeats of desmoplakin (Choi and Weis, 2011) and a homologous region of plectin (Ortega et al., 2011) with both encompassing the SH3 domain. The structures reveal a series of

canonical SRs connected by rigid α-helical linkers and a non-canonical SH3 domain which has an intimate association with SR4. The interactions between the SH3 domain and SR4 are conserved between desmoplakin and plectin which consists predominantly of hydrophobic residues. Both sets of authors propose that this interaction works to rigidify the SR3-6 region of the globular plakin domain. The architecture of the more divergent C-terminal region of the plakin domain which was previously proposed to contain SR modules was unknown and was of interest to this study. In addition to this it was hoped that a solution based study using NMR and SAXS would give an indication of the dynamics of the SR modules and linking regions. It was hoped by acquiring such information it would help understand how the plakin domain's architecture could be involved in desmosome assembly and/or binding to its known ligands.

The aim of the work in the following chapter was therefore to elucidate the modular organization of the desmoplakin plakin domain by CD, AUC, thermal stability assays, NMR and SAXS. It was hoped that the structure of these modules could be elucidated either by NMR or X-ray crystallography. In order to elucidate the structural and functional units of desmoplakin's plakin domain, a variety of constructs spanning this entire region were made. As stated in the methods sequence alignments , solubility prediction tools, and identification of regions predicted to be folded were used to aid construct design.

3.1 Expression and purification of subdomains

In order to structurally characterize the subdomains they were first purified to homogeneity. SR56 will be used as an example since each subdomain was purified the same way (with the exception of SR78 and SR8-CT which had an extra histag cleavage step). Since each of the subdomains was his-tagged they were purified by single step nickel ion chromatography using a Histrap column. Figure 3.1A depicts the purification of SR56 from lysis of the bacterial pellet to elution from the Histrap column. The protein was soluble as it can be seen in the soluble fraction

(size is ~30 kDa). It can also be seen in the wash fraction suggesting that not all the SR56 was able to bind to the column. However, once eluted from the column in 1 ml fractions it was the most abundant protein albeit with contaminants still remaining. To eliminate the remaining contaminants fractions with the most SR56 were pooled and concentrated for further purification by size exclusion chromatography using an S75 superdex column. Figure 3.1B shows the pooled fraction, concentrated pooled fraction and the elutions from the size exclusion column. The purity of the SR56 protein was estimated to be >95% and so was considered suitable for structural characterization. Other desmoplakin subdomains were also purified to >95% homogeneity (not shown).

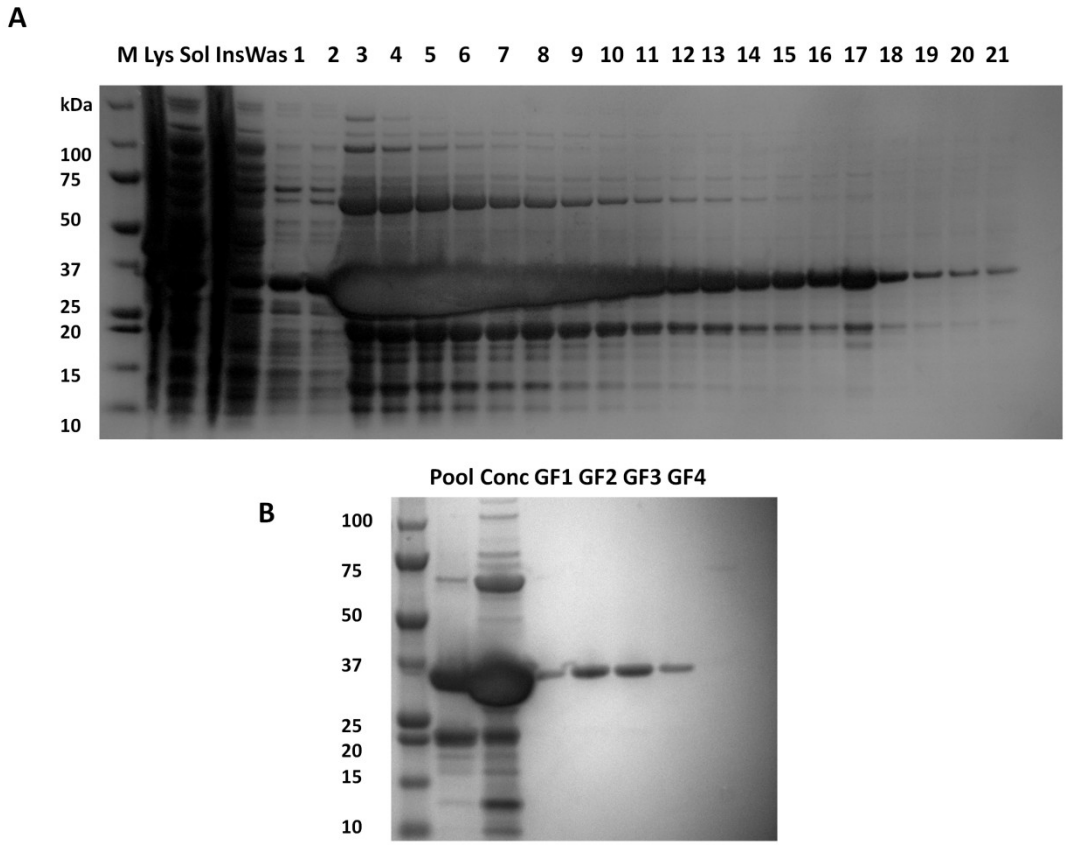


Figure 3.11 Purification and SDS-PAGE analysis of SR56

SDS-PAGE analysis of purification of SR56. A. Shows Lysate (Lys), Soluble lysate fraction (Sol), Insoluble lysate fraction (Ins), Histrap wash elution (Was) and 1ml Histrap elutions numbered from 1-21. B. Shows the pooled 1ml Histrap elutions (pool), this was then concentrated (Conc), and further purified by size exclusion chromatography with the resulting gel filtration elutions numbered GF1-GF4. M: molecular weight marker.

3.2 Expression trials of SR34

Two different temperatures and induction times were attempted to optimise solubility of SR34 as initial trials suggested it was insoluble upon expression in *E.coli*. Initially his-tagged SR34 was grown for 4 h at 20 °C and 1H at 37 °C. SR34 was observed in the insoluble fraction at approximately 25 kDa (figure 3.2A). The construct was then optimized to contain a Small Ubiquitin-like MOdifier (SUMO) fusion solubility tag known to increase solubility of difficult to express proteins (Marblestone et al., 2006). Even with the addition of a SUMO fusion tag the protein could not be expressed in the soluble fraction (figure 3.2B). The lack of a soluble desmoplakin SR34 is in stark contrast to the homologue in BPAG1 that was expressed as a SUMO-fusion and eventually crystallized. Despite the inability to produce soluble SR34 protein, other constructs from the desmoplakin plakin domain were then attempted.

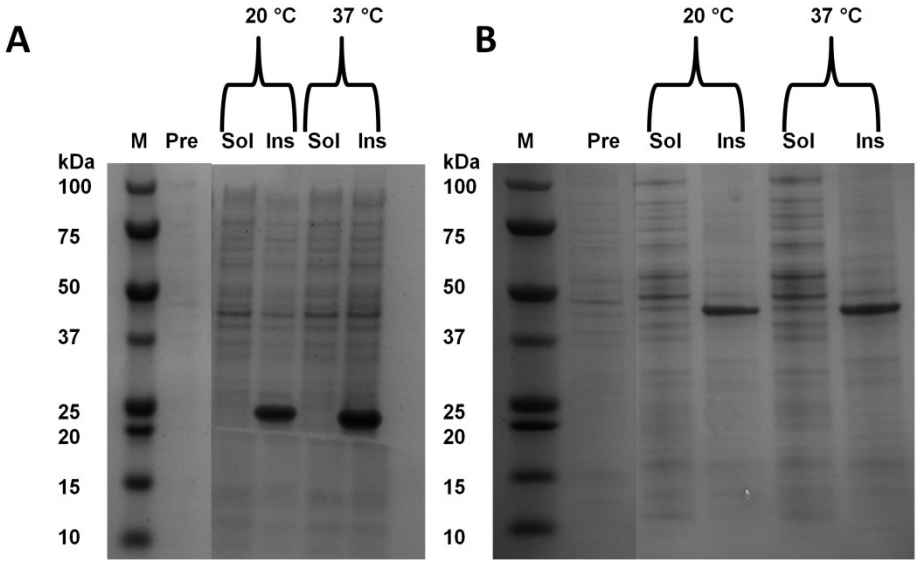


Figure 3.2 Expression trials of SR34

SR34 was expressed as two constructs and two different temperatures to optimise solubility. A: Expression of the his-tagged construct reveals that SR34 remains in the insoluble fraction. (M: marker, Pre: Pre induction sample, Sol: Soluble fraction post lysis, Ins: Insoluble fraction post lysis). B: Expression of the SUMO fusion construct seen at approximately 45kDa in the insoluble fraction (M: marker, Pre: Pre induction sample, Sol: Soluble fraction post lysis, Ins: Insoluble fraction post lysis). Despite the expression with a his-tag and solubility enhancing SUMO tag, SR34 could not be expressed in the soluble fraction.

3.3 Circular dichroism of desmoplakin plakin domain subdomains

The plakin domain of desmoplakin was predicted to be predominantly composed of α-helical subdomains. This was investigated using CD analysis of the subdomains of the desmoplakin plakin domain.

SR56 was subjected to CD at room temperature within the range of 240 nm to 190 nm. As expected a predominantly α-helical structure was observed as shown by the minima at 222 nm and 208 nm (figure 3.3). Deconvolution using the DICHROWEB server (Whitmore and Wallace, 2004) revealed estimated secondary structure of SR56 as: 73% α-helical, 10% β-sheet, 6% turn and 11% disorder. This is consistent with amino acid derived secondary structure predictions of: 61% α-helical, 11% β-sheet, 4% turn and 24% disorder using SOPMA (Geourjon and Deléage, 1995).

The SH3 domain was not subjected to CD analysis as 2D-HSQC revealed a completely unfolded state (see 3.71).

SR78 was subjected to CD at room temperature within the range of 240 nm to 190 nm. Once again as expected for a SR tandem, a predominantly α-helical structure was observed as shown by the minima at 222 nm and 208 nm (figure 3.3). Deconvolution using the DICHROWEB server revealed estimated secondary structure as: 88% α-helical, 0% β-sheet, 3% turn and 8% disorder. This is consistent with amino acid derived secondary structure predictions of: 88% α-helical, 1% β-sheet, 0% turn and 11% disorder using SOPMA. Furthermore, unlike the CD spectrum of SR56, SR78 had a positive elliptical value more than double its negative minima values. This suggested that the α-helical content is greater in SR78 than that of SR56 (Personal communication: Dr Raul Pacheco-Gomez, University of Birmingham) and is reflected in the DICHROWEB predictions.

SR8-CT was subjected to CD at room temperature from 240 nm to 190 nm. A predominantly α-helical structure was observed as shown by the minima at 222 nm and 208 nm (figure 3.3), however deconvolution using the DICHROWEB server revealed estimated secondary structure as: 39% α-helical, 16% β-sheet, 17% turn and 28% disorder. This is inconsistent with the secondary structure of a typical spectrin repeat tandem. It is also inconsistent with amino acid derived secondary structure predictions of: 86% α-helical, 0% β-sheet, 0% turn and 14% disorder using SOPMA. This therefore implies that SR8-CT has a structure anomalous to that predicted for it by sequence homology.

The results indicated that of SR56, SR78 and SR8-CT, SR56 and SR78 are the most folded. SR8-CT revealed a distinct CD spectrum which suggested a need for further characterization to understand the anomalous result. Thermofluor thermal stability assays were then used to probe the folded state and melting temperatures (T_m) of the SRs.

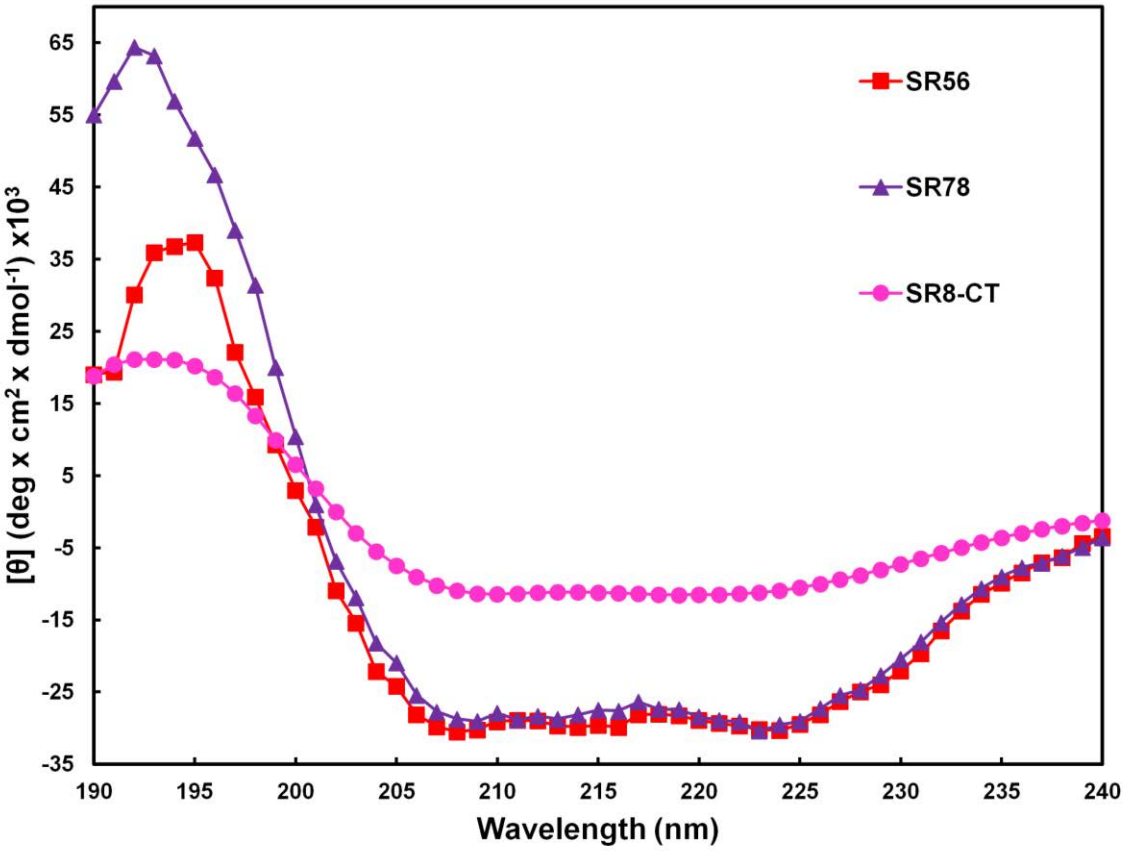


Figure 3.3 Circular dichroism analysis of each subdomain

CD spectra of SR56 (red, squares), SR78 (purple, triangles), and SR8-CT (pink, circles) were collected at room temperature using a Jasco J-810 spectrometer and a protein concentration of 0.1 mg/ml in 20 mM sodium phosphate (pH 7.2). The scanned wavelength range was 190-240 nm. Spectra were analyzed by DICHROWEB to ascertain their secondary structure content. Spectra of SR56 and SR78 were indicative of α-helical proteins. SR8-CT reveals a structure of some α-helical content. See text for further details. Adapted with permission from Elsevier, (Al-Jassar et al, 2011) under license number: 2862440887646.

3.4 Thermofluor stability assays

One of the main aims of the chapter was to elucidate the solution structure of the core structures of the desmoplakin plakin domain. Using a custom prepared 96 well plate of various buffer conditions and additives, it was hoped that an optimized condition could be found for structural studies. An optimized condition would be classed as a buffer and/or additive which gives a positive thermal shift of at least 2 °C (Personal communication: Professor Michael Overduin, based on unpublished findings from the Structural Genomics Consortium, Oxford). Previous studies reported this high-throughput method to successfully optimize crystallization of *E.coli* proteins (Ericsson et al., 2006). In addition to this our lab has used Thermofluor to find improved conditions for NMR structural determination (unpublished results: Overduin lab). Qualitatively, Thermofluor can also be used to ascertain whether the protein is folded, as typically a sigmoidal curve is observed. Unfortunately a significantly optimized condition was not found for any of the constructs used. Instead sigmoidal curves obtained in 20 mM sodium phosphate and 100 mM NaCl will be used to compare the thermal stabilities of all subdomains (figure 3.4).

SR56 gave a typical sigmoidal curve expected of a folded protein, which also suggested that the spectrin repeats and SH3 domain acted as cooperatively folded units. In 20 mM sodium phosphate (pH 7.2), 100 mM NaCl a T_m of 50 °C was recorded.

The SH3 domain did not give the sigmoidal denaturation curve expected for a stable domain in any buffer condition (data not shown). This is based on the premise that a hydrophobic core of the protein is present for which the dye can bind. This is discussed further in 3.71.

SR78 gave a typical sigmoidal denaturation curve expected of a folded protein, which also suggested that both spectrin repeats 7 and 8 act cooperatively as folded units as shown by the

single transition. In 20 mM sodium phosphate (pH 7.2), 100 mM NaCl a T_m of 59 °C was recorded.

The SR8-CT construct gave a sigmoidal curve expected of a folded protein, which also suggested that both spectrin repeats 8 and the CT region acted cooperatively as a folded unit. The optimal condition included 20 mM sodium phosphate (pH 7.2), 100 mM NaCl and yielded a T_m of 52 °C was recorded.

The thermal stability results are in agreement with the CD data as all canonical spectrin repeat constructs yielded spectra indicating that they were folded. To gauge their oligomerization states all four subdomains were subjected to analytical ultracentrifugation.

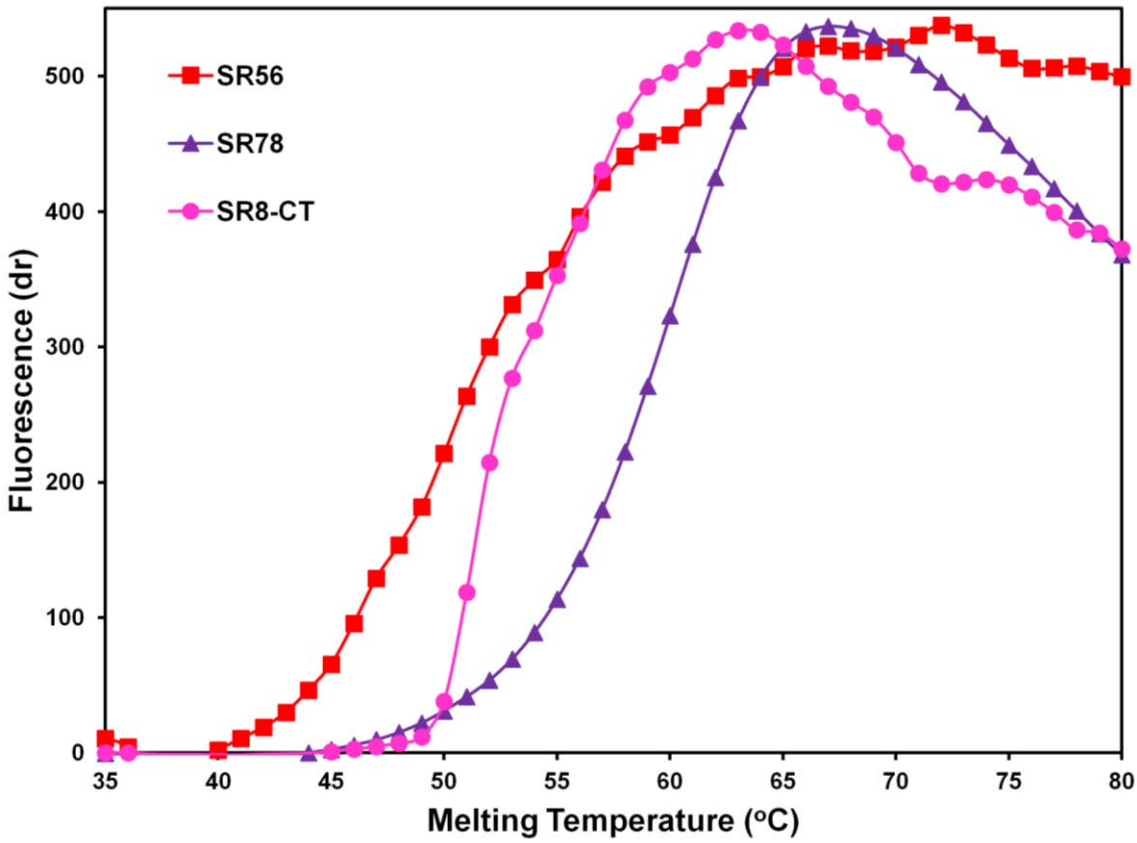


Figure 3.4 Raw thermofluor thermal stability profiles of each subdomain

Raw thermal stability profiles of SR56 (red, squares), SR78 (purple, triangles), and SR8-CT (pink, circles) were calculated using a Stratagene MX-3005 with temperatures ranging from 25 °C to 90 °C. Proteins were in 20 mM sodium phosphate (pH 7.2) and 100 mM NaCl at a final concentration of 2 µM. The midpoint reflects the melting temperature (T_m) (see text). It shows that whereas SR56 and SR8-CT have similar T_m's of 50 °C and 52 °C respectively, SR78 is more thermally stable and this is reflected by its higher T_m of 59 °C. Adapted with permission from Elsevier, (Al-Jassar et al, 2011) under license number: 2862440887646.

3.5 Analytical ultracentrifugation of SH3, SR56, SR78, SR8-CT

Although the rod domain of the desmoplakin full length protein had been proposed to mediate homo-dimerization, the oligomeric states of the plakin domain and its subdomains were unknown. Particularly of interest was the fact that the N-terminal head domain (where the plakin domain is present) is shown as 2 lobes in close proximity by electron microscopy studies which implied a possible association (Green et al., 1990). AUC was used to investigate whether the subdomains of the desmoplakin plakin domain self associate.

AUC analysis of the SH3 domain revealed a single species corresponding to a molecular weight of 8.0 kDa (figure 3.5). This compared favourably with the theoretical sequence based molecular weight of 9.8 kDa. This suggested that the desmoplakin SH3 domain remains monomeric in solution.

AUC analysis of SR56 revealed a single species corresponding to a molecular weight of 31.1 kDa (figure 3.5). This compared favourably with the theoretical molecular weight of 31.3 kDa. In common with the SH3 domain, it suggested that SR56 is monomeric in solution.

AUC analysis of SR78 revealed a single species corresponding to a molecular weight of 25.8 kDa (figure 3.5). This compared favourably with the theoretical molecular weight of 27.8 kDa. SR78, again like SR56 and the SH3 domain is therefore monomeric in solution.

AUC analysis of SR8-CT domain revealed a single species corresponding to a molecular weight of 31.1 kDa (figure 3.5). This compared favourably with the theoretical molecular weight of 30.6 kDa. This confirmed that SR8-CT is monomeric.

The results confirm that all subdomain constructs investigated from the plakin domain are monomeric. This is consistent with previous reports suggesting that dimerization of the desmoplakin full length protein is mediated by the adjacent rod domain (Green et al., 1990).

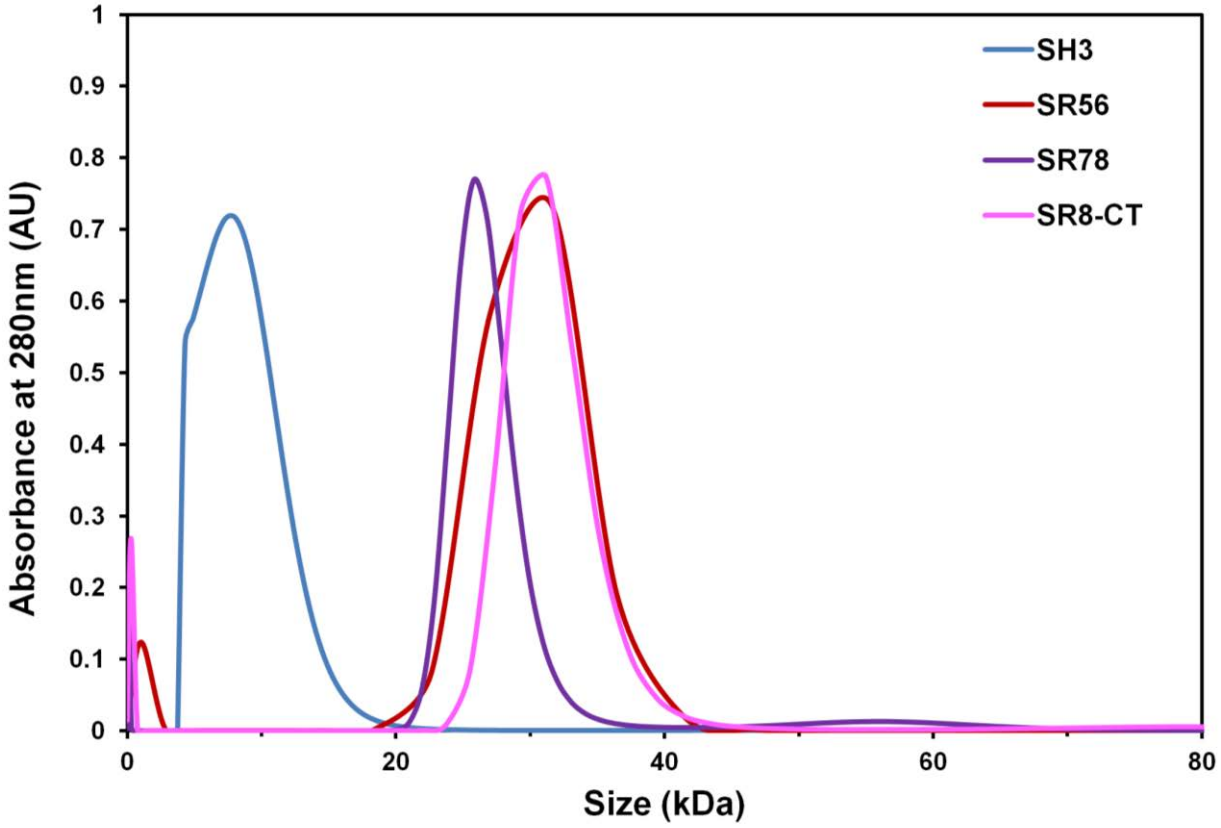


Figure 3.5 Analytical Ultracentrifugation (AUC) profiles of each subdomain

Each subdomain was subjected to AUC experiments at room temperature at a speed of 25000 rpm for 20 h on a Beckman XL-1 analytical ultracentrifuge using an eight cell 50Ti rotor and detected at 280 nm. Samples were run at between 0.05 and 4 mg/ml in 20 mM sodium phosphate (pH 7.2) and 100 mM NaCl. The results were analysed by the program SEDFIT (Brown and Schuck, 2006) in which the continuous c(S) analysis method was used to calculate estimated molecular masses. The peaks were then standardised. SH3 (blue), SR56 (red), SR78 (purple), and SR8-CT (pink) each reveal a monomeric profile. No other higher order species were observed. A residual buffer signal is observed at very low molecular weight. Adapted with permission from Elsevier, (Al-Jassar et al, 2011) under license number: 2862440887646.

3.6 Crystallization trials of the SH3 domain, SR56 and SR78

One of the aims of the work herein was to obtain the structure of a desmoplakin plakin domain subdomain. The SH3 domain, SR56 and SR78 were subjected to crystallization trials using a variety of commercially available screens. Although a few preliminary hits were identified, these were subsequently identified as salt crystals and no further crystals were found. Further crystallization attempts were abandoned due to the lack of crystal formation. Instead, NMR was used an alternative strategy. X-ray crystallography is heavily reliant on the bottleneck of protein crystallization. However, NMR analysis in contrast can be conducted rapidly on small monomeric proteins in solution. The sizes of all the subdomains used in the experiments of this chapter were within the size limitations of NMR (<30 kDa).

3.7 NMR analysis

At the time the experiments for this chapter were undertaken the structure of the desmoplakin plakin domain was unknown. As previously described, structures of homologues of SR34 from plectin (Sonnenberg et al., 2007) and BPAG1 (Jefferson et al., 2007) are available. 2D-HSQC NMR analysis of ^{15}N labelled SH3 domain, SR56, SR78 and SR8-CT were conducted to gauge whether these constructs were viable for NMR structure determination. Expression and purification was conducted in the same way as their WT counterparts and no significant differences were recorded in either yield or purity.

3.71 2D-HSQC NMR analysis of the desmoplakin SH3 domain

The SH3 domain was analysed by 2D-HSQC NMR. A total of 85% of the expected peaks were observed (figure 3.71). The $N^{\varepsilon}H$ resonance of the single tryptophan 493 was resolved at 10.2 ppm. The lack of dispersion of backbone amide resonances indicated a fully disordered protein. The lack of dispersion is indicated by the peaks residing predominantly within the range of 8-8.5 ppm on the ^{1}H chemical shift axis.

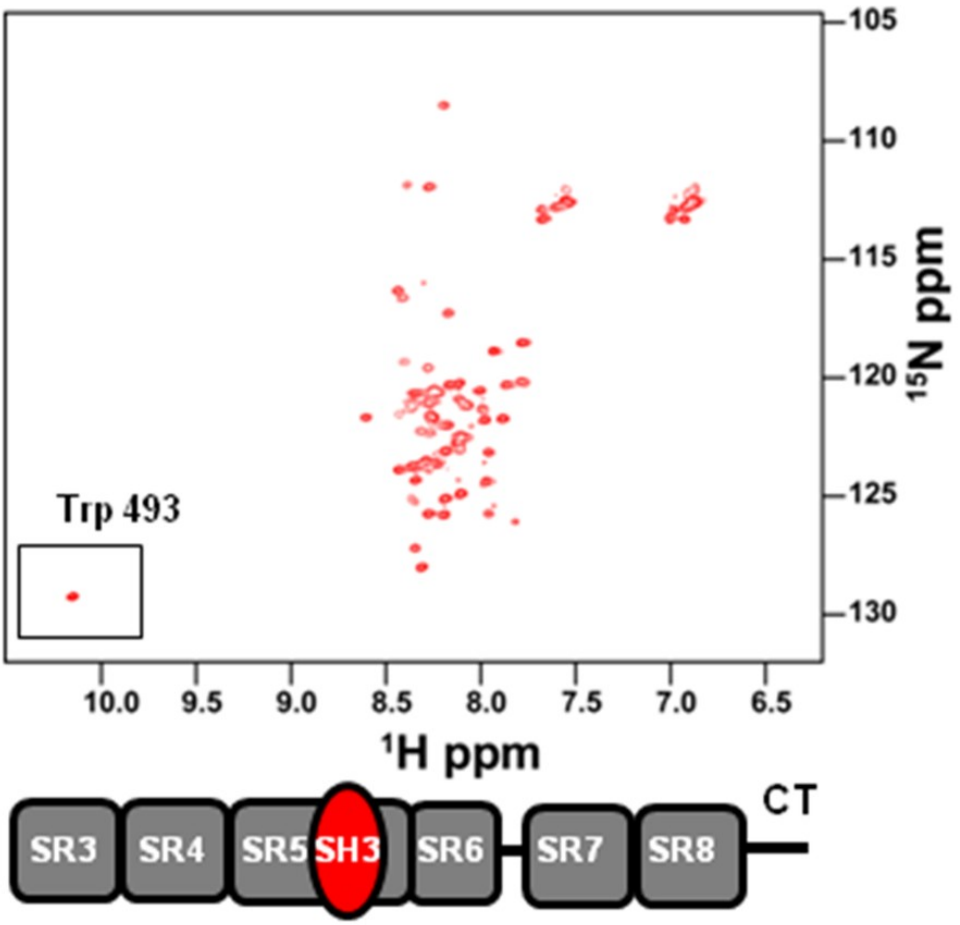


Figure 3.71 2D-HSQC of the SH3 domain

A 2D-HSQC spectrum was acquired for the SH3 domain. The protein was run at 25 °C on an Agilent 600 MHz spectrometer in 20 mM sodium phosphate (pH 7.2) and 100 mM NaCl at 100 µM. Boxed is the sole trypophan 493 present in the SH3 domain construct. The spectrum is typical of an unfolded protein as shown by the lack of dispersion of peaks located between 7.8-8.5 ppm. Adapted with permission from Elsevier, (Al-Jassar et al, 2011) under license number: 2862440887646.

3.72 2D-HSQC NMR analysis of SR56

The SR56 protein was analysed by 2D-HSQC NMR. Of the 185 peaks expected only 75 were observed (figure 3.72). The $N^{\varepsilon}H$ resonances of tryptophan residues 493, 537 and 550 were viewed as two overlapping peaks at ~10.2 ppm. Although CD confirmed the presence of a secondary structure, a globular tertiary structure is unlikely to be resolvable by NMR. This is based firstly on the few observable peaks residing predominantly in the random coil region of the spectrum. (8-8.5 ppm) and secondly on the lack of sufficiently intense and dispersed NH peaks. An increase in temperature from 25 to 30 °C did not improve the quality of the spectrum (data not shown).

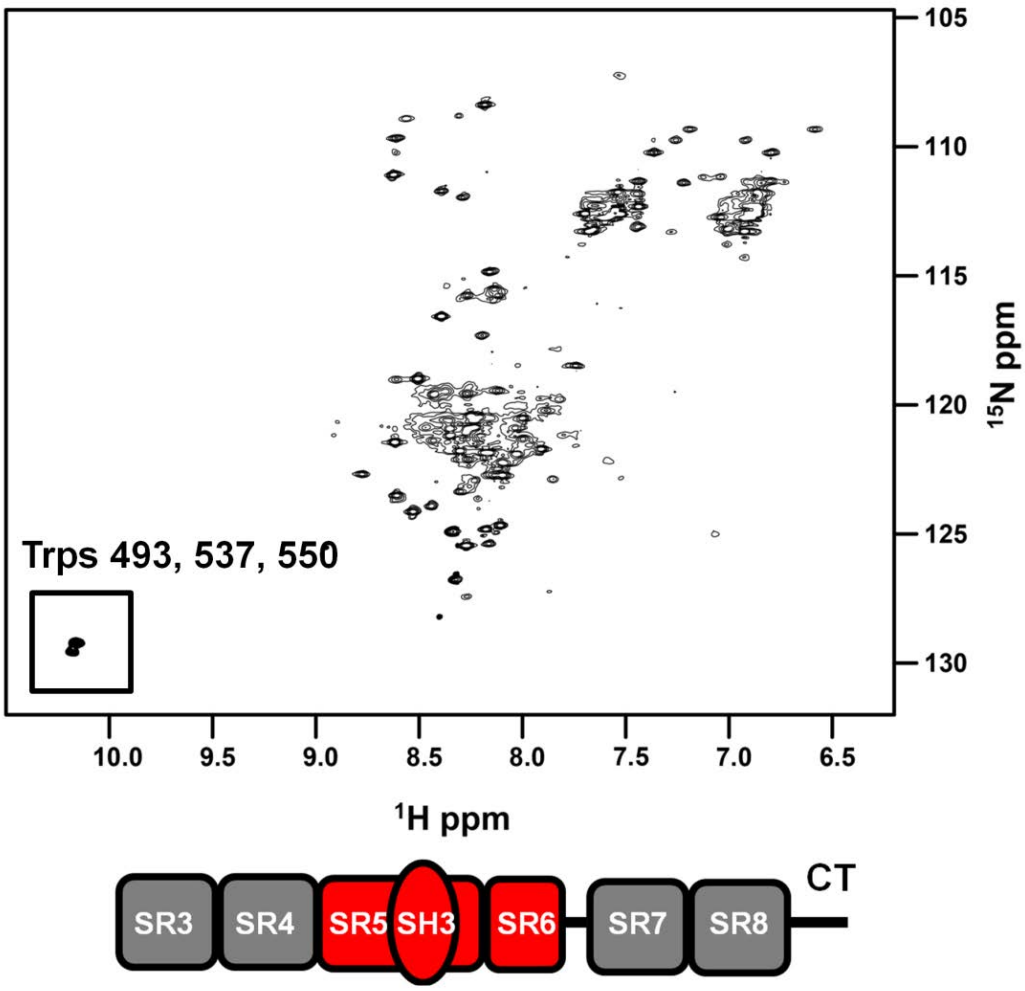


Figure 3.72 2D-HSQC of SR56

A 2D-HSQC spectrum was acquired for SR56. The protein was run at 25 °C on an Agilent 600 MHz spectrometer in 20 mM sodium phosphate (pH 7.2) and 100 mM NaCl at 100 µM. Boxed are tryptophans 493, 537 and 550 present in the SR56 construct seen as 2 overlapping peaks. Secondary structure and CD predictions suggested that the spectrum would be indicative of a mixed α-helical and β-sheet protein with well dispersed peaks. However, the general lack of dispersion of peaks indicates that SR56 is predominantly flexible although some α-helical elements may be present as some dispersion of peaks occurs between 8 and 8.75 ppm. The spectrum thus indicates that although there may be some α-helical and β-sheet content (as suggested by secondary structure and CD predictions), there is unlikely to be an overall tertiary fold. Adapted with permission from Elsevier, (Al-Jassar et al, 2011) under license number: 2862440887646.

3.73 2D-HSQC NMR analysis of SR78

The SR78 protein was analysed by 2D-HSQC NMR. At 25 °C approximately 40% of the expected peaks were observed (figure 3.73A). However, increasing the temperature to 30 °C increased this to 90% of the expected peaks observed (figure 3.73B). The spectrum is typical for predominantly α-helical proteins as observed by the spread from 7-9 ppm. Each of the $N^{\varepsilon}H$ resonances of tryptophan residues 667, 867 and 879 were observable. This confirmed that the SR78 construct had a tertiary fold and is viable for further NMR structure determination. The requirement of an increased temperature at 30 °C is likely due to the prolate shape of the protein typical of SR tandems previously observed (Yan et al., 1993). Models of SR78 from the PHYRE server (Kelley and Sternberg, 2009) suggested an extended helical bundle fold, typical of SRs. Combined with the size of the protein (25kDa), at 25 °C it is highly likely that the protein tumbles in solution at a nonisotropic rate in a manner unfavourable for NMR determination. This leads to a faster relaxation time, which ultimately leads to a decreased signal to noise ratio. At 30 °C however, the tumbling is increased due to the increased energy in the system and therefore relaxation time is decreased, and finally increased signal to noise ratio. To date, the spectrum is the most dispersed recorded by NMR for a tandem spectrin repeat construct in the literature. Attempts to determine the backbone assignments are discussed in 3.77

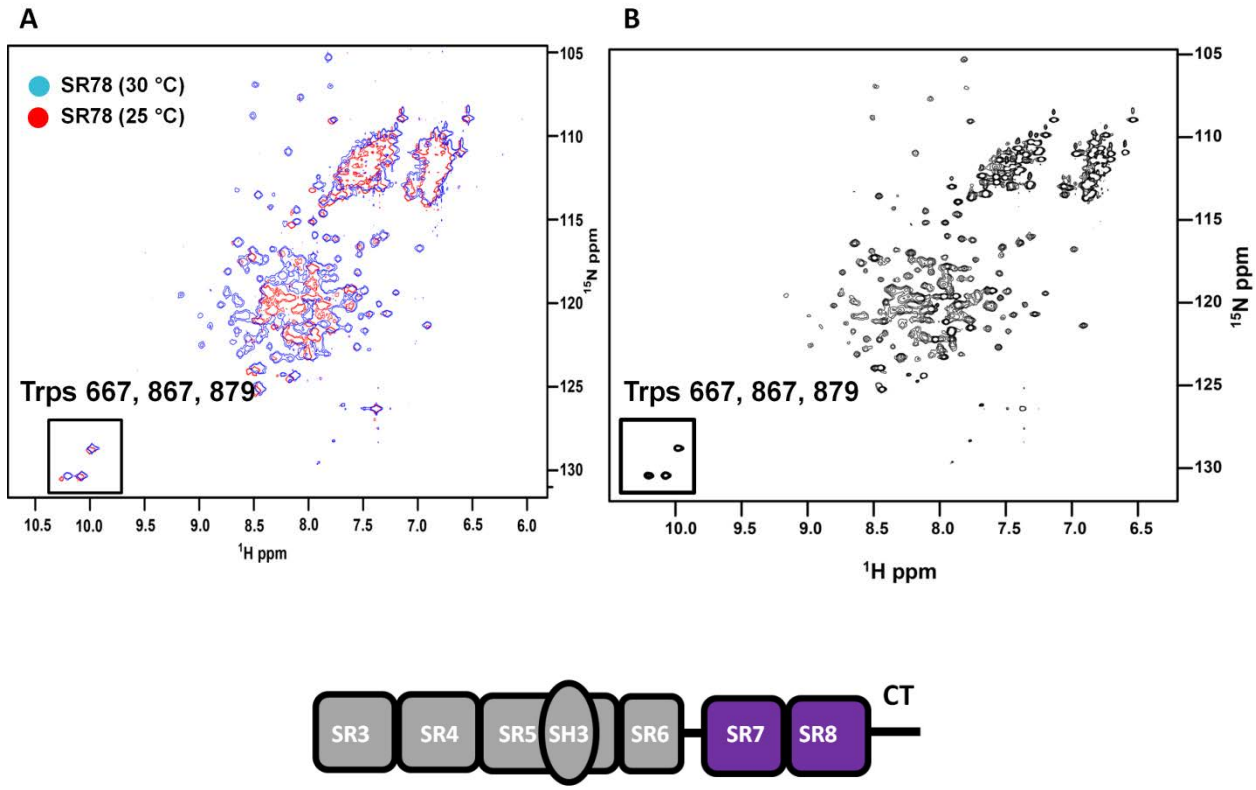


Figure 3.73 2D-HSQC of SR78

2D-HSQC spectra were acquired for SR78. The protein was run at two different temperatures (25 °C and 30 °C) on an Agilent 800 MHz spectrometer in 20 mM sodium phosphate (pH 7.2) and 100 mM NaCl at 100 µM. A) A comparison of SR78 2D-HSQC at two different temperatures superimposed (blue, 30 °C, red 25 °C). Boxed are tryptophans 667, 867 and 879 present in the SR78 construct. The spectrum indicates that SR78 is predominantly α-helical with a tertiary fold as shown by the high dispersion between 7.0 and 9.0 ppm. B) 2D-HSQC of SR78 at 30 °C only.

3.74 2D-HSQC NMR analysis of the SR8-CT

The SR8-CT protein was analysed by 2D-HSQC NMR. Very few of the expected number of peaks were observed (figure 3.74). The $N^{\varepsilon}H$ resonances of tryptophans 867, 879 and 901 also overlap as one superimposed peak at 10.2 ppm. Although CD confirmed the presence of some secondary structure, a globular structure observable by NMR was unlikely to exist. This suggested that SR8-CT contains flexible elements, which in turn make the folded elements unresolvable on a NMR spectrum. Like SR56 an increase in temperature from 25 to 30 °C did not improve the quality of the spectrum (data not shown).

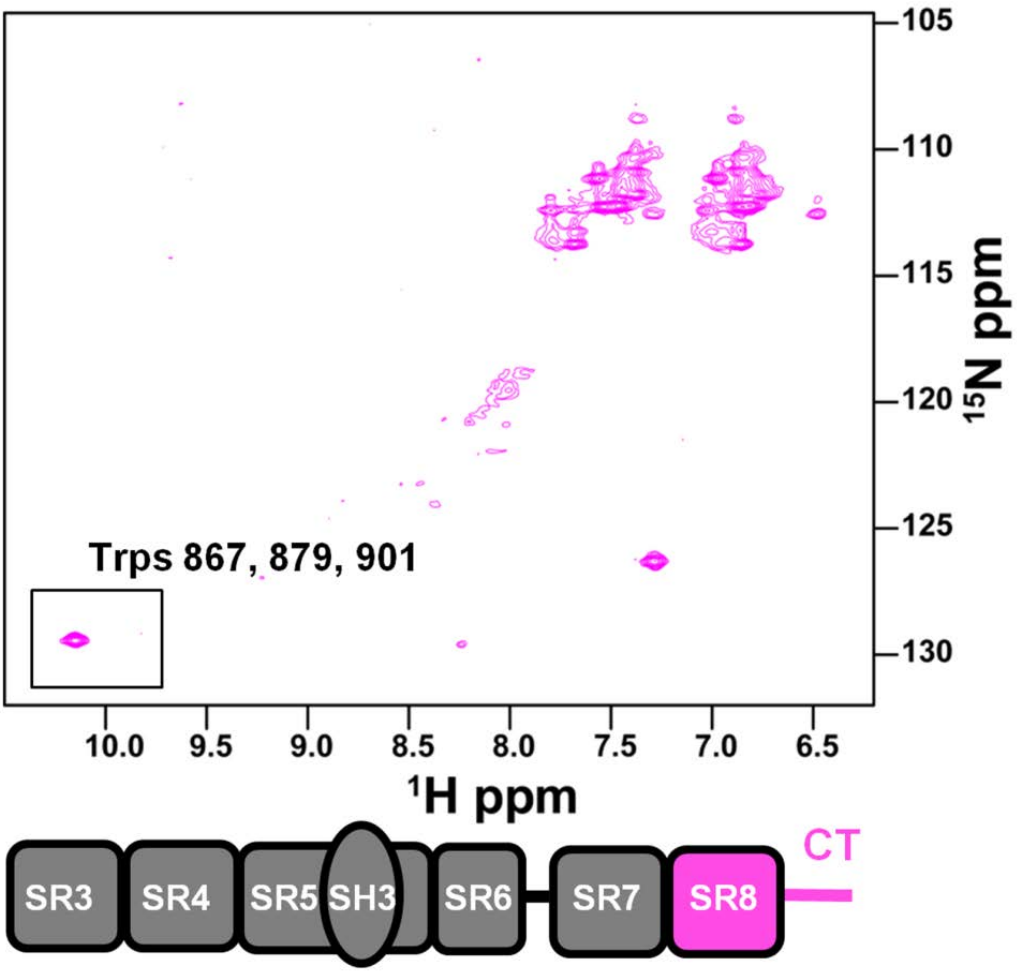


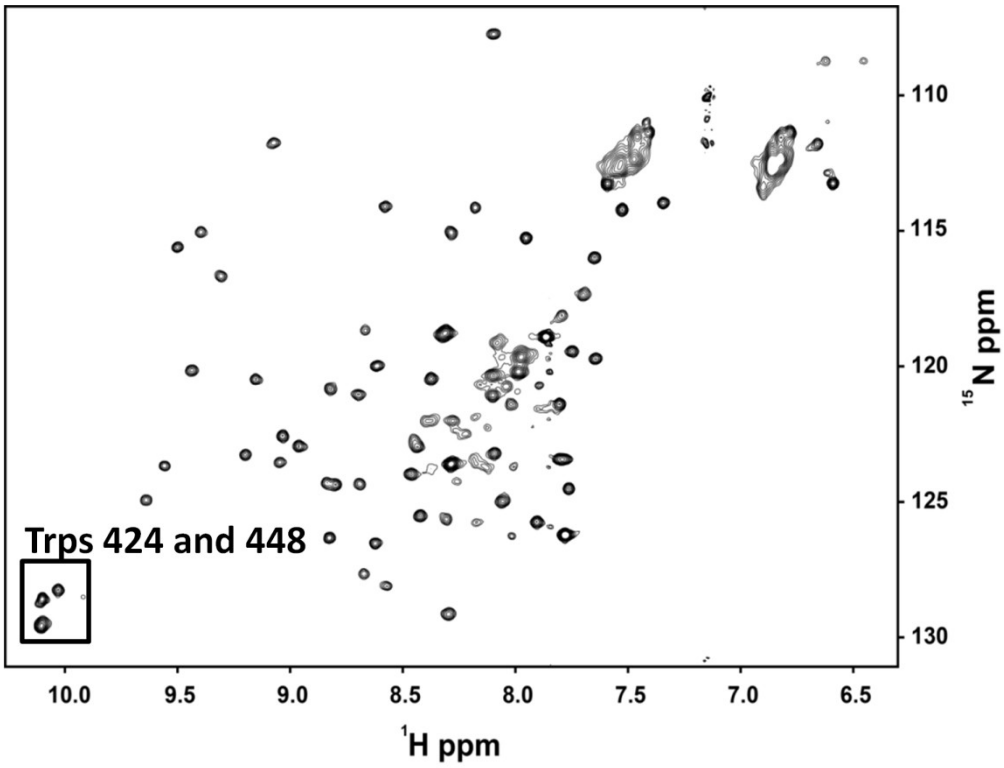
Figure 3.74 2D-HSQC of SR8-CT

A 2D-HSQC spectrum was acquired for SR8-CT. The protein was run at 25 °C on an Agilent 600 MHz spectrometer in 20 mM sodium phosphate (pH 7.2) and 100 mM NaCl at 100 µM. Boxed are tryptophans 867, 879 and 901 present in the SR8-CT construct. The spectrum indicates that S8-CT was predominantly flexible or unfolded due to the few peaks observable. Adapted with permission from Elsevier (Al-Jassar et al, 2011) under license number: 2862440887646.

3.75 2D-HSQC NMR analysis of the envoplakin SH3 domain

It was hoped that since the desmoplakin SH3 domain in isolation was disordered that other plakin SH3 domains may harbour more favourable folding properties for structural determination. The envoplakin and periplakin SH3 domains were thus investigated.

The envoplakin SH3 domain was analysed by 2D-HSQC NMR. Overall the spectrum yielded a typical dispersion indicative of a folded predominantly β-sheet containing protein (figure 3.75). This is based on the dispersion being spread between 7.5 and 10.2 ppm. The $N^{\varepsilon}H$ resonances of tryptophans 424 and 448 were observed as 3 peaks between 10.0 and 10.2 ppm. The presence of an extra peak may reflect a chemical shift of a backbone residue that is particularly downshielded. An alternative explanation may be that the envoplakin SH3 domain may have a tryptophan that is dynamic. The observation of a folded plakin SH3 domain was a surprising result given the high sequence homology to the desmoplakin SH3 domain which itself was shown to be disordered (figure 3.71). As the NMR spectrum indicated that the envoplakin SH3 domain was folded, it was taken for further structural elucidation, as discussed in 3.77.



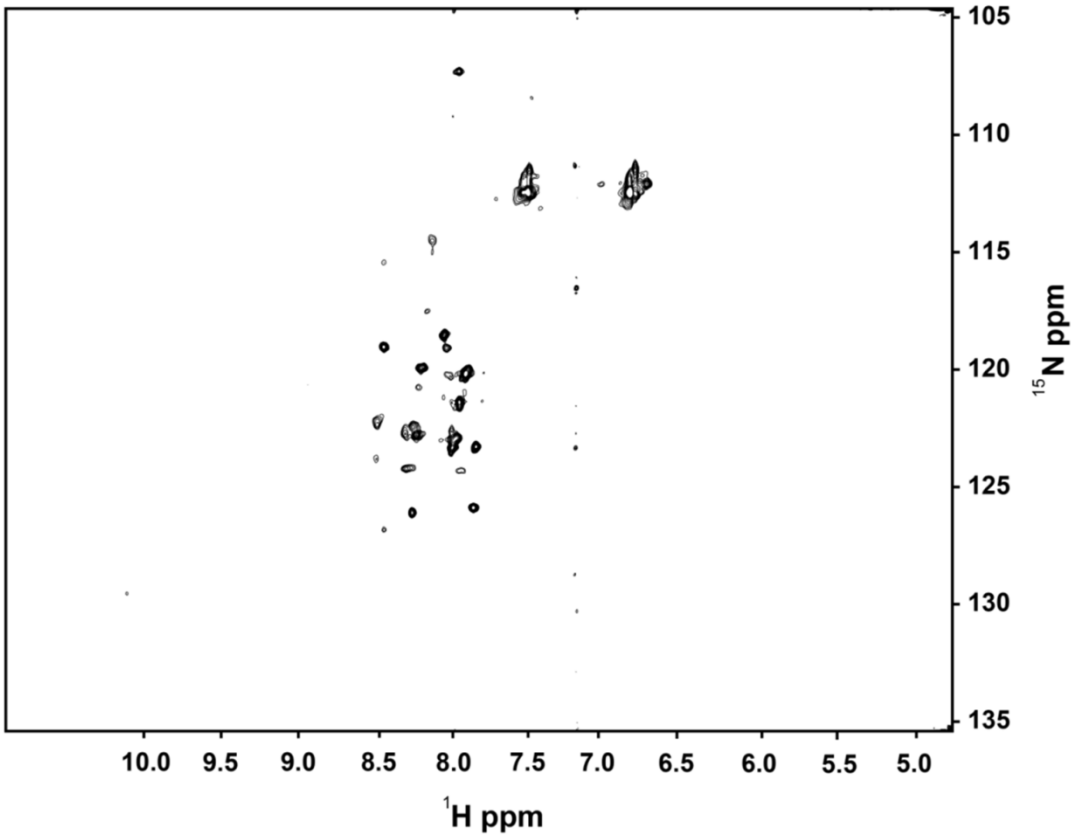
3.75 2D-HSQC of the envoplakin SH3 domain

A 2D-HSQC spectrum was acquired for the envoplakin SH3 domain. The protein was run at 25 °C on an Agilent 600 MHz spectrometer in 20 mM sodium phosphate (pH 7.2) and 100 mM NaCl at 200 µM. Boxed are tryptophans 424 and 448 present in the envoplakin SH3 domain construct. Three peaks are present because of the possibility of the protein being observed in 2 possible conformations. The spectrum indicated that the envoplakin SH3 domain is a folded, predominantly β-sheet containing protein.

3.76 2D-HSQC of the periplakin SH3 domain

The periplakin SH3 domain was analysed by 2D-HSQC NMR spectroscopy.Very few peaks were observed for the periplakin SH3 domain (figure 3.76). The $N^{\varepsilon}H$ resonances of its tryptophans were not observed in the spectrum. Overall the spectrum yielded dispersion indicative of a disordered or aggregated protein. This is based on the dispersion being spread narrowly between 7.7 and 8.5 ppm. The lack of peaks and the narrow spread of those present suggested the protein was either disordered, highly dynamic or self associating.

Overall, the 2D-HSQC NMR data showed that each plakin SH3 domain exhibited differential conformations in isolation. Again, given the high sequence similarity between each plakin SH3 domain this is a surprising result.



3.76 2D-HSQC of the periplakin SH3 domain

A 2D-HSQC spectrum was acquired for the periplakin SH3 domain. The protein was run at 25 °C on an Agilent 600 MHz spectrometer in 20 mM sodium phosphate (pH 7.2) and 100 mM NaCl at 300 µM. No $N^{\varepsilon}H$ resonances were observed in the usual position at the bottom left corner of the HSQC spectrum. The spectrum indicated that the periplakin SH3 domain was predominantly disordered and/or oligomerizes as shown by the lack of dispersion, the lack of tryptophan $N^{\varepsilon}H$ resonances and the low number of expected peaks.

3.77 Feasibility of assigning SR78 and the envoplakin SH3 domain

To assign the backbone of any protein by NMR a variety of 3D spectra are acquired from samples that are ^{13}C and ^{15}N labelled. It is only economically feasible to conduct further 3D NMR experiments if the HSQC suggested that the protein is folded since ^{13}C labelling of the protein is expensive. Both SR78 and the envoplakin SH3 domain were therefore good candidates to take for further structural elucidation based on their resolved HSQC spectra (figures 3.73 and 3.75 respectively). The "gatekeeper" to 3D NMR spectra is the HNCO which acquires the i-1 of each hydrogen bound to a ^{15}N labelled amide. It is the most sensitive of all 3D spectra and therefore determines whether further 3D spectra acquisition is worthwhile. As each HNCO experiment was conducted by Dr Mark Jeeves (University of Birmingham) only a brief overview will be given.

3.78 HNCO of SR78 and the envoplakin SH3 domain

A HNCO of ^{13}C and ^{15}N labelled SR78 was conducted on an Agilent 800 MHz spectrometer at 30 °C at 500 µM in 20mM sodium phosphate (pH 7.2) and 100mM NaCl. Unfortunately, very few peaks were observed in the HNCO suggesting that NMR assignment was going to be challenging (data not shown). Therefore, it appeared that although the protein was folded, there may be other factors preventing collection of data via the HNCO experiment. Since the HNCO is the "gatekeeper" to conducting further NMR experiments, this route was not continued. Previous NMR studies have solely investigated single SRs (1 SR) as opposed those in tandem (2 SRs) which are more physiologically relevant (Park et al., 2003, Pascual et al., 1997). Therefore the failure to acquire 3D NMR data for the SR78 construct may reflect a problem faced by other researchers studying non globular prolate structures in solution.

An HNCO of the ^{13}C and ^{15}N labelled envoplakin SH3 domain was conducted on an Agilent 600 MHz spectrometer at 25 °C at 700µM in 20mM sodium phosphate (pH 7.2) and 100mM NaCl. Unfortunately, very few peaks were observed in the HNCO suggesting that further NMR

structural determination was again going to be challenging (data not shown). Therefore, in common with SR78 it appears that although the protein is folded there may be other factors preventing the HNCO and therefore other 3D NMR experiments from being acquired. One possible explanation may involve the observation of an extra peak in the range usually expected for the $N^{\varepsilon}H$ resonances of tryptophans. It implies that if the extra peak is that of a $N^{\varepsilon}H$ resonance from a tryptophan then one of the tryptophans is being observed in 2 orientations. If true, it implies that the envoplakin SH3 domain could be dynamic and exists in two alternating conformations.

Given that none of the subdomain constructs were suitable candidates for atomic structural determination by NMR or X-ray crystallography an alternative approach was required to ascertain further structural properties.

3.8 SAXS of the desmoplakin plakin domain subdomains

Although SAXS is a low resolution technique, it allows for basic structural properties, foldedness and similarity to atomic resolution models to be probed. One of the main benefits of SAXS is that it is a relatively low risk technique often used for proteins refractory to conventional structural elucidation experiments such as NMR or X-ray crystallography.

The SR56, SR78 and SR8-CT constructs were subjected to SAXS analysis at the DESY X33 beamline (EMBL, Hamburg). A key element of SAXS structural determination is the importance of monodispersity. Since each subdomain was shown to be monomeric in solution by AUC this suggested they were amenable for structural characterization by SAXS. Prior to exposure to the beamline, each sample was dialyzed in 20mM sodium phosphate (pH 7.2), 100mM NaCl and a variety of concentrations ranging from 1-7 mg/ml were used. Each protein was analyzed in the

same fashion as one another and SR78 will be used as a guide to show how SAXS analysis was conducted.

3.81 SAXS analysis of SR78

A buffer subtracted SAXS scattering curve was acquired for SR78 using previously published methods (figure 3.81A). SR78 was then checked for interparticle effects. As shown in figure 3.81B the Guinier region was linear. The guinier region is usually/nearly always at lower angles as long as S<1.3 for globular proteins. The R_g value can then be taken manually via linear regression which was calculated to be 4.2 nm. Another parameter to calculate was the D_{max}. This was conducted using the program GNOM (Semenyuk and Svergun, 1991) and a P(r) distribution graph created as shown in figure 3.81C. The D_{max} was calculated to be 13 nm. An automated value for the R_g was also calculated by GNOM which was 4.3 nm - in good agreement with the manually obtained Rg value. Ten *ab-initio* DAMMIF (Franke and Svergun, 2009) models were created and then manually superimposed on one another. As shown in appendix D little heterogeneity is shown between the models suggesting that the protein is a rigid structure (Personal communication: Dr Katsuaki Inoue, Diamond light source, Oxford, September 2009). Of the 10 models, the program DAMSEL selects the one most representational of all 10 and the program DAMAVER (Volkov and Svergun, 2003) creates an average based on these statistics. Finally DAMFILT is used to filter any extraneous regions not aligned with the average model. Since DAMMIF's output is bead orientated, it must be stated that each bead does not depict an individual amino acid and is representational of space instead. Although non essential, for easier visual inspection the bead model was then converted to a volumetric map by the program SITUS (Wriggers and Chacon, 2001). Since the atomic-resolution structure of SR78 was unobtainable, its sequence was fed into the PHYRE server. This yielded a model 14% identical to the crystal structure of erythroid spectrin SR8 and SR9

(Protein Data Bank ID: 1S35). Although 14% is a relatively low figure, this is regarded as an acceptable percentage by those in the spectrin repeat field (Mirza et al., 2010). The model was input into the envelope automatically using SUPCOMB (Kozin and Svergun, 2001) and a normalized spatial discrepancy value of 0.456 was obtained (figure 3.81D). This indicated that both SR7 and SR8 exist as canonical SRs with three-helix-bundles arranged in a linear extended rod. This also validated the results of the 2D-NMR HSQC of SR78 which suggested that the construct was likely to be extended and contain an α-helical structure. The program CRYSOL (Svergun et al., 1995) was then used to create a theoretical SAXS scattering profile for the SR78 model which when superimposed onto the experimental SAXS scattering profile gave a good fit (appendix D). Although not perfect, it suggested some discrepancies between the atomic model and the envelope. This may be due to flexible termini which would skew the result created by uncertainty as to where SR7 and/or SR8 begin and end respectively. This is understandable given the low identity between the sequence and the template. Nonetheless, given that the atomic model fits well into the envelope it is highly likely the orientation of the core region (i.e. non-flexible elements) is a reasonable approximation.

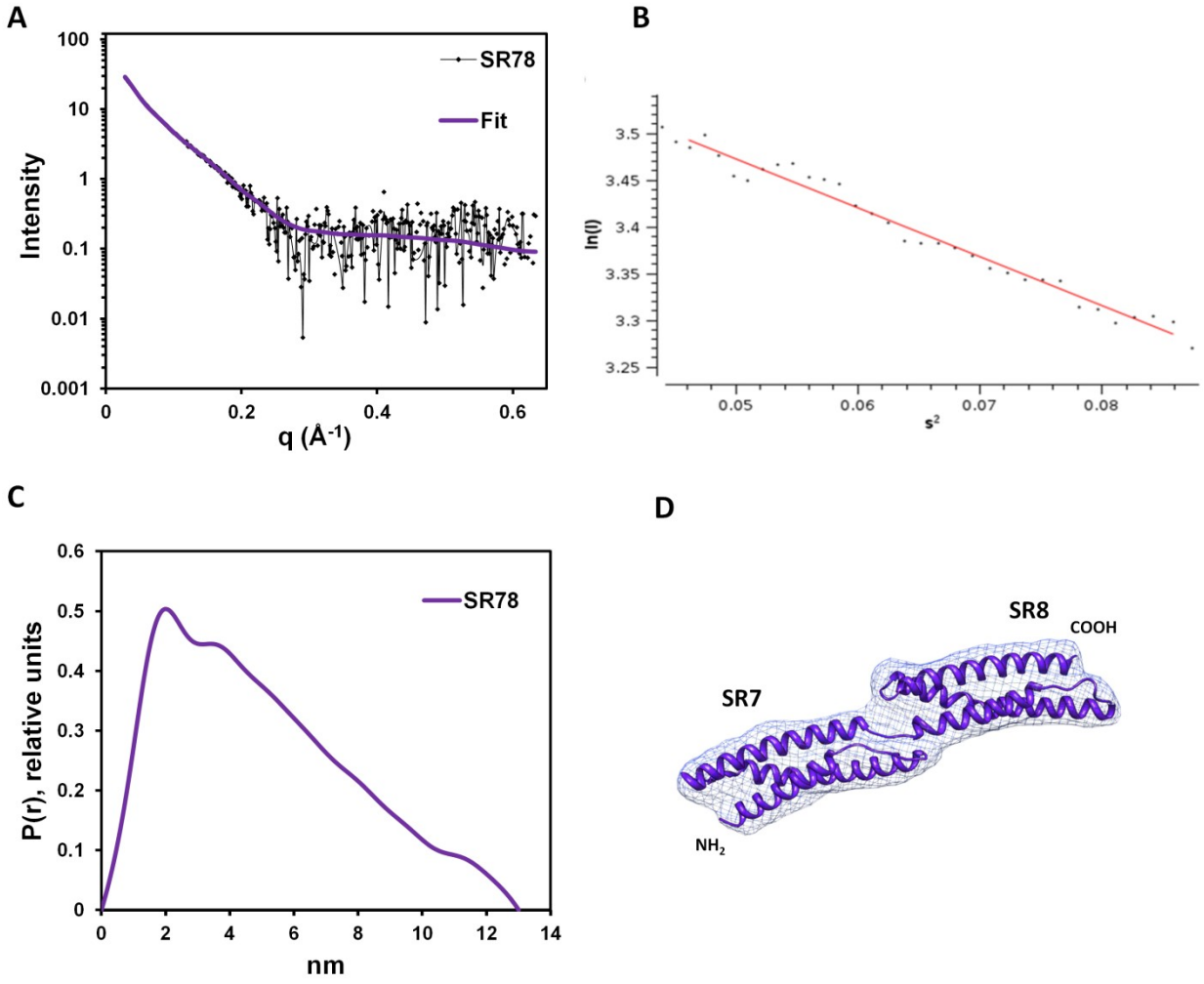


Figure 3.81 SAXS analysis of SR78

SR78 was analyzed at 1, 2 and 5 mg /ml in 20 mM sodium phosphate (pH 7.2) and 100 mM NaCl. The SAXS scattering curves were scaled and merged to create a final SAXS scattering curve in accordance with published methods. A: Buffer subtracted SAXS scattering curve with a DAMMIF fit. B: Guinier analysis showing that the fit is good indicating a reliable R_g value of 4.2 nm (visualized in PRIMUS). C: P(r) distribution curve indicating the D_{max} (13 nm). D: Final averaged and filtered envelope with a model inserted automatically (see text). Adapted from Al-Jassar et al (2011) with permission from Elsevier under license number 2862440887646.

3.82 SAXS analysis of SR56

As with SR78, no inter-particle effects were observed so SAXS analysis was undertaken. GNOM revealed a R_g value of 3.9 nm and a D_{max} of 12.5 nm. A total of 10 *ab-initio* models were created by the program DAMMIF and superimposed manually to gauge flexibility. As shown in appendix E the 10 superimposed models revealed a degree of heterogeneity. This is in keeping with the suggestion that SR56 contains flexible elements based on the 2D-NMR HSQC in figure 3.82. A final averaged and filtered envelope revealed a bent-rod shape containing two dissimilar lobes. Given that how and if the protein was correctly folded was unknown, an estimation of the model orientation is given in figure 3.82B using a small hinge between SR5 and SR6 (as postulated by Choi and Weis, 2011). A theoretical SAXS scattering profile of the crystal structure derived for the region encompassing SR56 by Choi and Weis (2011) was superimposed onto the experimental SAXS scattering curve using CRYSOL (appendix E). This gave a poor fit and therefore suggested that the SR56 construct in isolation and solution did not resemble its crystallized counterpart.

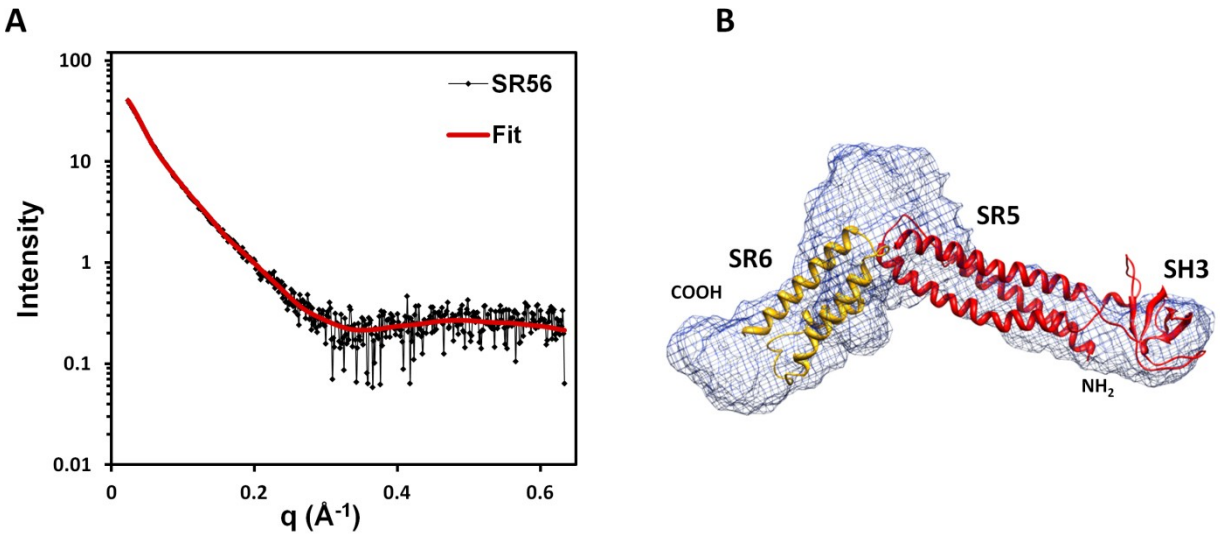


Figure 3.82 SAXS analysis of SR56

SR56 was analyzed at 1, 2 and 6 mg /ml in 20 mM sodium phosphate (pH 7.2) and 100 mM NaCl. The SAXS scattering curves were scaled and merged to create a final SAXS scattering curve in accordance with published methods. A: Buffer subtracted SAXS scattering curve with a DAMMIF fit. B: Final averaged and filtered envelope with crystal structure derived atomic data (pdb code: 3R6N) manually inserted (see text). Adapted from Al-Jassar et al. (2011) with permission from Elsevier under license number 2862440887646.

3.83 SAXS analysis of SR8-CT

No interparticle effects were observed for SR8-CT so this construct was progressed for further SAXS analysis. GNOM revealed a R_g value of 4.3 nm and a D_{max} of 12.5 nm. Ten *ab-initio* models were created by the program DAMMIF and superimposed. In common with SR56 the 10 superimposed models of SR8-CT revealed a degree of heterogeneity (appendix F). This is in keeping with the CD data which showed a relatively higher level of disorder compared to SR56 and SR78 and the NMR data which revealed no discernible peaks. Together this suggested that like SR56 there was a degree of flexibility within the SR8-CT construct. This is more evident with the final averaged and filtered envelope which revealed a shape highly dissimilar to that of a canonical spectrin repeat shown in figure 3.83B. A tandem spectrin repeat model predicted by the PHYRE server (calculated from the 1S35 template) could not be inserted by either automated or manual fitting (data not shown). A SR8 model was inserted manually leaving the remainder of the envelope to the flexible CT element (figure 3.83B). A theoretical SAXS scattering profile of the model was superimposed onto the experimental SAXS scattering profile and a reasonable fit was observed (appendix F). The reasonable fit like that observed for a SR78 model suggests that the protein in solution may behave somewhat like a canonical spectrin repeat but has elements that do not. Combine this data with the 10 superimposed DAMMIF models and CD and NMR data and it is likely that SR8-CT is at best spectrin repeat like. It is possible that the CT region may still harbour a spectrin repeat-like domain but the data remains inconclusive and required further investigation.

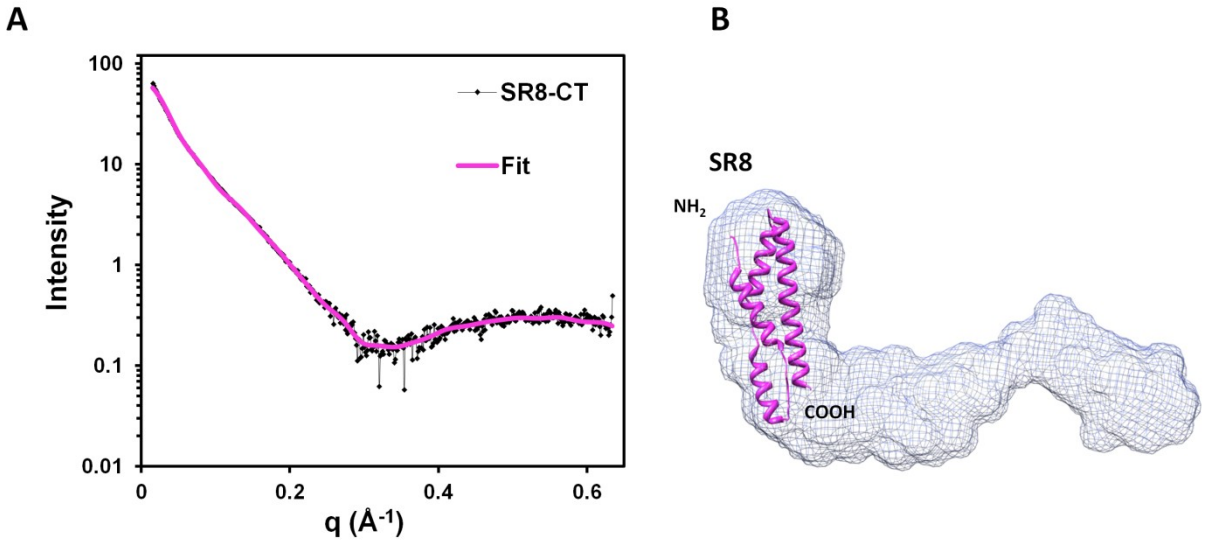


Figure 3.83 SAXS analysis of SR8-CT

SR8-CT was analyzed at 2, 5, and 8 mg /ml in 20 mM sodium phosphate (pH 7.2) and 100 mM NaCl. The SAXS scattering curves were scaled and merged to create a final SAXS scattering curve in accordance with published methods. A: Buffer subtracted SAXS scattering curve with a DAMMIF fit. B: Final averaged and filtered envelope with a model for SR8 manually inserted (see text).Adapted from Al-Jassar et al. (2011) with permission from Elsevier under license number 2862440887646.

3.9 Chapter 3 conclusions

Overall it appears that the modular organization of the desmoplakin plakin domain is complex. SR78 appears to be the most independently stable and well defined of all 4 subdomain constructs analyzed. SR8-CT appears to be the most anomalous of all soluble subdomains, possibly requiring the adjacent rod domain to acquire its fold. In light of the recent publications on the N-terminal regions of the desmoplakin (Choi and Weis, 2011) and plectin (Ortega et al., 2011) plakin domains it is also likely SR56 requires SR4 to acquire its stable fold.

The SR domains of desmoplakin's plakin domain were investigated to gauge their modular organization in solution. Each SR was analyzed separately, with SR34 and SR8-CT behaving anomalously and SR56 and SR78 folding independently. The SR34 construct appeared to require the neighbouring SR5/SH3 element for stability and therefore did not form an independent structural unit shown by its insolubility. This was confirmed by two studies on desmoplakin and plectin on the same region in which both suggested that SR4 had an intimate association with the SH3 domain (Choi and Weis, 2011, Ortega et al., 2011). The SH3 domain of desmoplakin was shown by 2D-HSQC to be unfolded despite being soluble. This was also likely to be the case for the periplakin SH3 domain. In contrast however, the envoplakin SH3 domain appeared to be independently folded and a well dispersed 2D-HSQC indicated a protein with high β-sheet content. It is unusual to observe an SH3 domain incapable of independent folding since they are often folded (Kaneko et al., 2008, Musacchio et al., 1992, Musacchio et al., 1994). It is also interesting to observe the variation of folding capabilities of the SH3 domains from different plakin proteins. The reason why the envoplakin SH3 domain remained independently folded whereas the others were not remains uncertain but may be due to the subtle differences in the sequences between the plakin SH3 sequences and their respective interfaces with SR4. The uniqueness of the independent fold of the envoplakin SH3 fold is supported by alteration of

its key contacts with SR4. In particular tryptophan 424 in envoplakin's RT loop is bulkier than the typical tyrosine at this position (which is known to be crucial for stability of the SH3 fold). In particular its Ala-Cys sequence just prior to β-strand 5 which replaces the Val-Gly found here in desmoplakin that was shown to engage SR4 (Choi and Weis, 2011). The desmoplakin SR4 residues glutamic acid 290, tryptophan 360 and isoleucine 364 that provide critical contacts with its SH3 are substituted in envoplakin by Gln, asparagine and leucine residues respectively, and hence the electrostatic and hydrophobic interdomain contacts could again be compromised (alignment shown in appendix B). These subtle alterations may make envoplakin distinct from other plakin SH3 domains in that it is able to adopt an independent fold without the need of a bound SR4. Since the envoplakin SH3 domain is independently folded it is possible that plakin SH3 domains may have differential abilities to bind other as yet unknown ligands. However, given that the primary function of a SH3 domain is ligand binding, the presence of an independent fold (in contrast to those observed for desmoplakin and periplakin and presumably for plectin) suggests the envoplakin SH3 domain may be more available for ligand binding. Unfortunately, this hypothesis cannot be validated until a ligand can be found for the envoplakin SH3 domain. Both studies that investigated the crystal structure of plectin and desmoplakin suggested that the SH3 domain is solely required for "rigidifying" the SR3-6 element (Choi and Weis, 2011, Ortega et al., 2011). While it is still possible that the SR4-SH3 interface may be a crystal artefact this is unlikely since both report the same residues involved in the interface. However, it is possible that this may represent a "snapshot" of the true dynamics between the SR4-SH3 interface that is favoured in crystallization conditions. Further solution based analysis such as NMR could clarify this question.

A construct encompassing SR5, the SH3 domain and SR6 (SR56) behaved anomalously. The crystal structure of the desmoplakin plakin domain encompassing this region suggested a rigid

rod like structure. However, the SAXS envelope of this construct revealed two lobes at different angles (figure 3.82). Subsequently the SR56 crystal structure did not fit the data and when fit to the envelope required manual fitting of SR6 at an obtuse angle. This suggests that a flexible hinge may exist between SR5 and SR6 where in the crystal structure a more favourable rigid rod is adopted. Indeed the authors of the paper implied a hinge may exist between SR5 and SR6 in addition to the possibility that SR6 may only be "spectrin repeat like" (Choi and Weis, 2011). In combination with the NMR data, the lack of a fully dispersed 2D-HSQC spectrum also reflects the possibility of internal dynamics only when the SR4 element is not included in the construct.

The presence of SR7 was identified contrary to previous reports (in tandem with SR8) (Sonnenberg et al., 2007). It forms an independently folded and stable structure as well as being the most thermally stable of each soluble SR construct analyzed. This was identified as having a canonical SR fold by SAXS which also fit a sequence derived model with a NSD value of 0.46 (figure 3.81D). This clarifies the legitimacy of SR7 since it had previously been questioned (Sonnenberg et al., 2007). It appears that SR7 and SR8 therefore cooperate as a tandem structured unit as often observed in other SR constructs (Broderick and Winder, 2005, Grum et al., 1999, Legardinier et al., 2009, Yan et al., 1993, Ylänne et al., 2001). The 2D-HSQC of the SR78 construct was, as far as I am aware, the most dispersed ever recorded for a SR tandem in the literature. However, the lack of intense peaks in the HNCO spectra of SR78 and the relatively high D_{max} observed for the construct suggested a tumbling nature refractory for NMR analysis.

The SR8 and the CT region tandem, behaved particularly anomalously. In combination with the NMR and CD data it suggested that the CT region destabilizes the tandem. It is therefore likely that the CT region is disordered or at best partially folded. Furthermore its tendency to precipitate compared to the other SR tandems corroborated this argument since disordered

regions often expose hydrophobic patches which promote nonspecific protein to protein interactions. This may reflect its rather specialized role at the C-terminal end of the plakin domain adjoining the rod domain.

Altogether dissection of the desmoplakin plakin domain reveals a variety of specialized units which originates in their inter-dependant nature. This included the rigid SR36 region, a thermally stable SR78 and an anomalous CT region. This leaves a region of unknown significance between SR6 and SR7 and is discussed further in the next chapter.

Chapter 4 - Plakin domains are unexpectedly non-linear and flexible

The N-terminal head region of plakin proteins are important for tethering to the desmosome (Bornslaeger et al., 2001). They are typically comprised of a presumed unstructured N-terminal region and a globular plakin domain. The importance of this region is highlighted by its importance in binding to multiple desmosomal partners (Bornslaeger et al., 2001, Kami et al., 2009). In light of this modular organization discussed in chapter 3, it was felt that a structural perspective on full length plakin domain constructs of desmoplakin, envoplakin and periplakin may give further insights into how the SRs collectively orientate in solution. It was anticipated that gauging the overall orientation of a plakin domain would shed light into how its structure translates into function (e.g. binding ligands). Desmoplakin is incorporated into desmosomes and is essential for desmosome function whereas envoplakin and periplakin act as accessory proteins and are not essential for desmosome function. Therefore by gauging their plakin domain structures it would allow assessment for why this is the case.

Plakin proteins are members of the spectrin superfamily, which includes spectrin and dystrophin, by virtue of their plakin domains which consist of a variable number of spectrin repeats. Proteins such as spectrin and dystrophin typically consist of many spectrin repeats (>10) (Broderick and Winder, 2005). Chains of up to four serial spectrin repeats from a number of different proteins have been shown to have a linear organisation by X-ray crystallography (Ortega et al., 2011, Ylänne et al., 2001, Choi and Weis, 2011). It was therefore assumed that serial chains of SRs are linear elements with limited propensity for flexibility. The relevance of this is that functionally; SRs are known to withstand mechanical stress which has been characterized extensively (Law et al., 2003). However, serial chains of SR proteins (≥2) had not been investigated by NMR or SAXS and their orientation in solution remains unknown. It is unclear how the orientation of a

serial chain of SRs in solution correlates with their ability to withstand mechanical stress. One possible explanation is that flexibility exists within and between SRs.

Grum et al proposed a mechanism in which non α-helical linkers within SRs allow limited flexibility (Grum et al., 1999). However, as shown in the protein dystrophin there are regions between SRs that are thought to be linkers or “hinges” (Koenig and Kunkel, 1990). The contribution of these “hinges” to the overall orientation of a serial chain of SRs in solution remains unknown. The presence of a “hinge” or linker was suggested to exist in the region separating SR6 and SR7 in the desmoplakin plakin domain as described in chapter 3. It is possible that these “hinges” may provide some form of flexibility to the serial chain of SRs. SAXS is one of the few experimental techniques that can provide information on the presence of flexibility in proteins and was therefore employed in the following chapter to explore flexibility in plakin domains. The plakin domains from desmoplakin, envoplakin and periplakin were purified and analyzed using a battery of biophysical techniques, including CD, AUC, limited proteolysis and SAXS in order to determine the structure of the plakin domains in solution and ascertain whether they exist as rigid and/or flexible domains.

4.1 Purification of plakin domains

All plakin domain constructs used contained a cleavable N-terminal GST tag and a C-terminal 6x HisTag and were purified in the same way. Purification of the desmoplakin plakin domain is described in detail as an example. Bacterial cells expressing GST fused desmoplakin plakin domain were grown in culture, harvested and lysed. The GST fused protein was not visible in the cell lysate or in the soluble fraction following centrifugation (figure 4.1). Nevertheless the soluble fraction was loaded on to a GSTrap column. Following elution with GST a peak of 125 kDa was observed in the GST elution fraction (figure 4.1). This band is the GST-desmoplakin plakin domain fusion protein. Overnight cleavage by PreScission protease resulted in two bands, one at 100 kDa and one at 25 kDa. The 100 kDa band is the desmoplakin plakin domain while the 25 kDa band is free GST. GST was removed from the solution using an additional single step his trap purification. The his trap elution is observed in the “Elu2” fraction in figure 4.1. The purity of the final desmoplakin plakin domain product was estimated to be > 90 % and was deemed suitable for further biophysical analysis. Final purity of the plakin domains of envoplakin and periplakin were similar (i.e. >90 %) (data not shown).

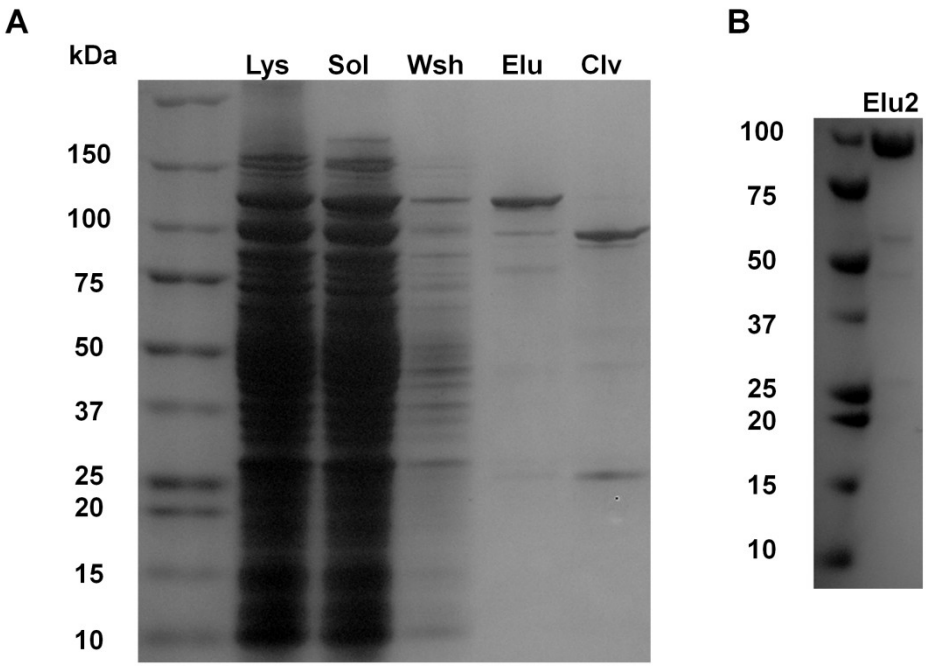


Figure 4.1 SDS-PAGE analysis of purification of the desmoplakin plakin domain.

A. Shows: molecular weight marker (M), Lysate (Lys), Soluble lysate fraction (Sol), GSTrap wash elution (Wsh), single step glutathione elution (Elu) and cleavage of the eluted protein by PreScission protease (overnight at 4 °C) (Clv). B. Single step his trap purification step, final elution is shown (Elu2). The SDS-PAGE gels reveal that the desmoplakin plakin domain is suitable for further biophysical analysis as shown by its high purity.

4.2 Circular dichroism of the plakin domains from desmoplakin, envoplakin and periplakin

The desmoplakin plakin domain was analysed by CD to gauge its secondary structure. It was anticipated that the secondary structure would be predominantly α-helical based on sequence based secondary structure predictions and previous reports (Sonnenberg et al., 2007). The spectrum was recorded at wavelengths between 240 nm and 190 nm. The CD spectrum shows minima at 208 and 222 nm indicative of predominantly α-helical structure. The data was fitted using DICHROWEB (Whitmore and Wallace, 2004) and a predicted secondary structure of 67% α-helices, 8% β-sheet, 12% turn and 13% disorder obtained. This is consistent with an amino acid derived secondary structure prediction using SOPMA (Geourjon and Deléage, 1995) of 75% α-helices, 5% β-sheet, 2% turn and 18% disorder.

The envoplakin and periplakin plakin domains were also analysed by CD to gauge their secondary structure in the same way as the desmoplakin plakin domain. For envoplakin, a secondary structure content of 67% α-helices, 9% β-sheet, 10% turn and 14% disorder was predicted by DICHROWEB. This is consistent with an amino acid derived secondary structure prediction using SOPMA of 74% α-helices, 4% β-sheet, 1% turn and 21% disorder. For periplakin, a secondary structure content of 62% α-helices, 4% β-sheet, 11% turn and 23% disorder was predicted by DICHROWEB. This is consistent with an amino acid derived secondary structure prediction using SOPMA of 81% α-helices, 2% β-sheet, 1% turn and 16% disorder.

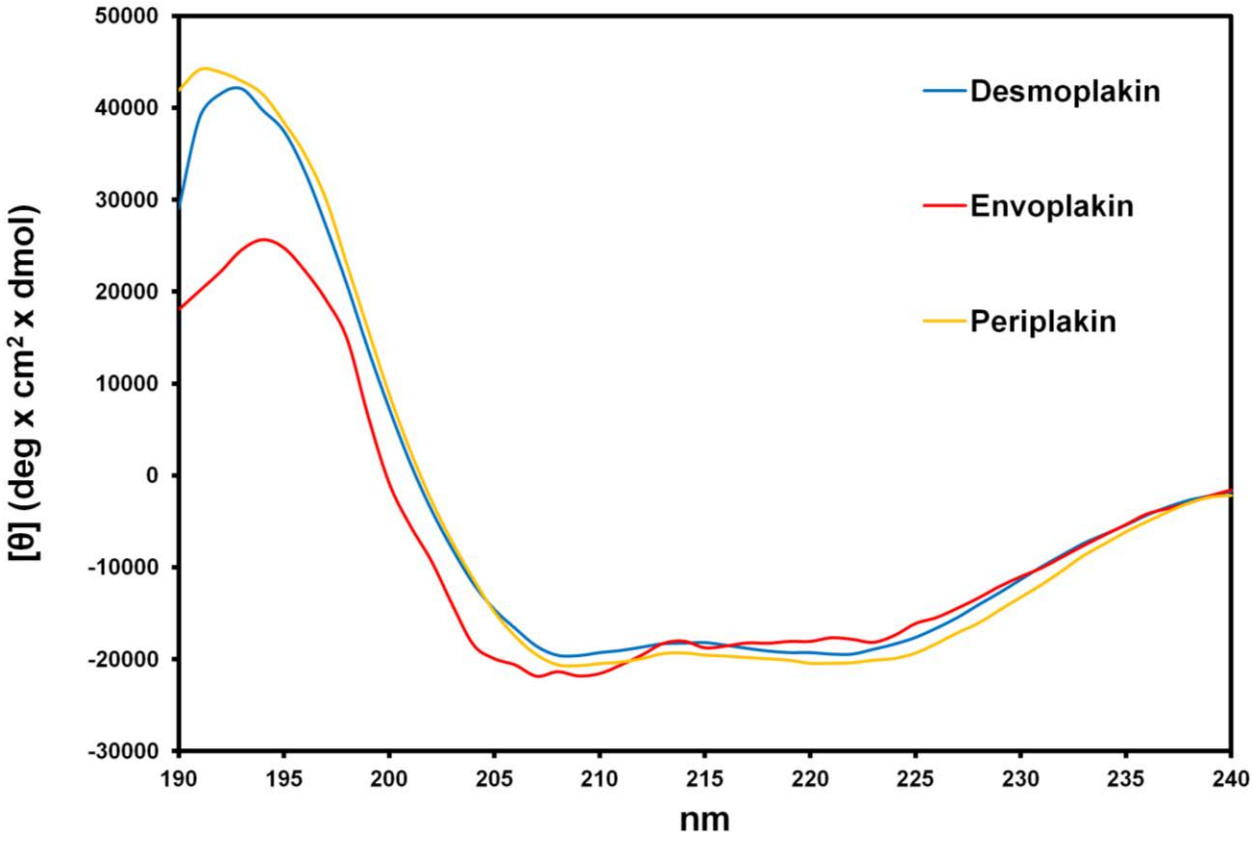


Figure 4.2 Circular dichroism analysis of the plakin domains of desmoplakin, envoplakin and periplakin

Circular dichroism spectra of desmoplakin (blue), envoplakin (red) and periplakin (gold) were collected at room temperature using a Jasco J-810 spectrometer and a protein concentration of 0.1 mg/ml in 20 mM sodium phosphate (pH 7.2). The scanned wavelength range was 190-240 nm. Spectra were analyzed using DICHROWEB to ascertain their secondary structure content. Each spectra is indicative of an α-helical protein (see text).

4.31 Limited proteolysis of the desmoplakin plakin domain

The purified desmoplakin plakin domain protein was subjected to limited proteolysis in order to identify protease resistant fragments since they are more amenable for structural determination. Limited proteolysis can be also be used to facilitate in future construct design in addition to qualitatively gauging the location of solvent exposed loops or hinges that are more accessible to proteases. Three predominant bands were observed by SDS-page analysis following limited proteolytic digestion with trypsin (figure 4.3). Bands at approximately 26 kDa, 53 kDa and 85 kDa were seen. The band at 27 kDa was subjected to further analysis as a candidate for structural determination by NMR due to its relatively small size (in chapter 3). N-terminal sequencing (Alta Bioscience, University of Birmingham) of the first 6 amino acids revealed a sequence of ^{654}VIETNR^{659} which is preceded by a lysine in the desmoplakin sequence. This resides in the C-terminal of the desmoplakin plakin domain. Given that plakin domains are made up of serial chains of SRs it suggested the presence of a tandem pair of SRs. Since SRs are often found as tandem 25 kDa species, this corroborates Sonnenberg et al's (2007) prediction of the presence of a SR8 but also a potential SR7, which was not predicted. This was shown to be a tandem SR named SR78 in chapter 3.

The presence of a 53 kDa band (figure 4.3) corroborated Weis et al's findings in which they also found a protease resistant fragment of the same size that turned out to be SR3-6 and was subsequently crystallized (Choi and Weis, 2011). Finally this leaves the region separating the two protease resistant fragments which was cleaved by trypsin, indicating a potentially solvent exposed loop. This coordinates to residues P627-K653, the relevance of which is discussed later in the chapter.

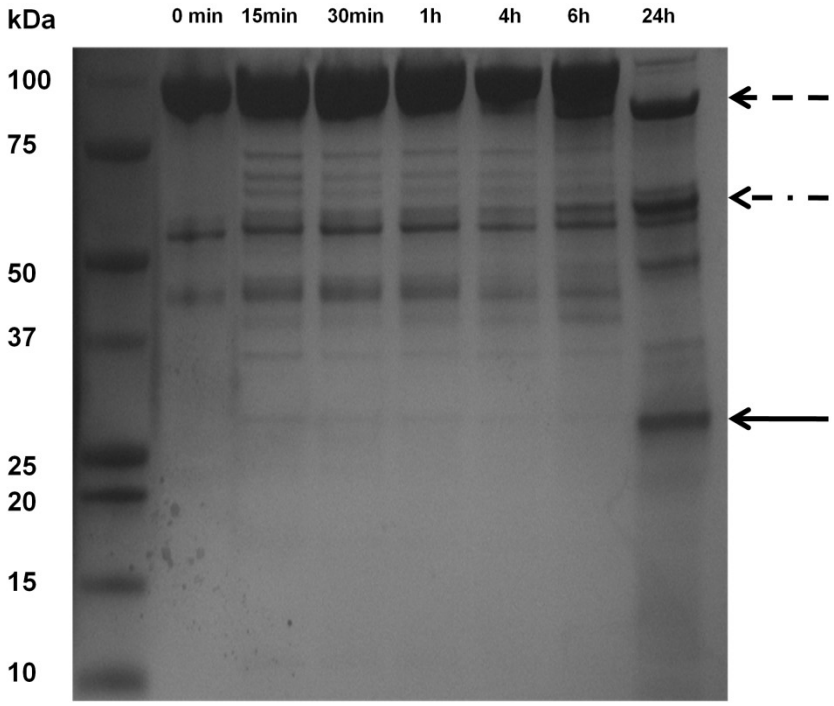


Figure 4.31 Limited proteolysis of the desmoplakin plakin domain

The desmoplakin plakin domain (10 mg/ml) was subjected to limited proteolysis at 37 °C by 100 ng of trypsin (a 1:1000 protease to protein ratio) at various times points (indicated). Three bands were observed, one at approximately 85 kDa (dashed arrow), one at approximately 55 kDa (dashed, dot, dashed arrow) and one at approximately 27 kDa (solid arrow).

4.32 Limited proteolysis of the envoplakin and periplakin plakin domains

Based on the limited proteolysis of the desmoplakin plakin domain and the high homology between all plakin domains, it was assumed that other plakin domain homologues adopt a similar organization. The envoplakin plakin domain was subjected to limited proteolysis with trypsin (0.001 mg/ml) for 60 mins and sampled at different time points. This yielded 3 prominant bands : one at ~48 kDa, one at ~85 kDa and one at 27 kDa. This yielded a similar limited proteolysis profile to that of desmoplakin. N-terminal sequencing of the ~48 kDa band revealed a sequence of GPLGS (Alta Bioscience). These amino acids are vector derived and located at the extreme N-terminal end of the envoplakin plakin domain construct. The data suggest that in common with the desmoplakin plakin domain the envoplakin plakin domain also has a rigid/protease resistant SR3-6 region.

Limited digestion of the periplakin plakin domain was also conducted in a similar manner to that of envoplakin. This yielded a limited proteolysis profile dissimilar to those observed for desmoplakin and periplakin in which only a single ~48 kDa band was seen. N-terminal sequencing of this band revealed a sequence of ^{131}EVDP^{134} (Alta Bioscience) corresponding to the first 4 residues of the periplakin plakin domain. Hence the protease resistant fragment is likely to be SR3-6. Interestingly, it also appears that at least qualitatively, the SR3-6 fragment of periplakin is the most resistant of all plakin domain SR3-6 protease resistant fragments. This can be observed by the fact that the band is the most prominent of all 3 plakin domains analyzed by limited proteolysis (figure 4.32).

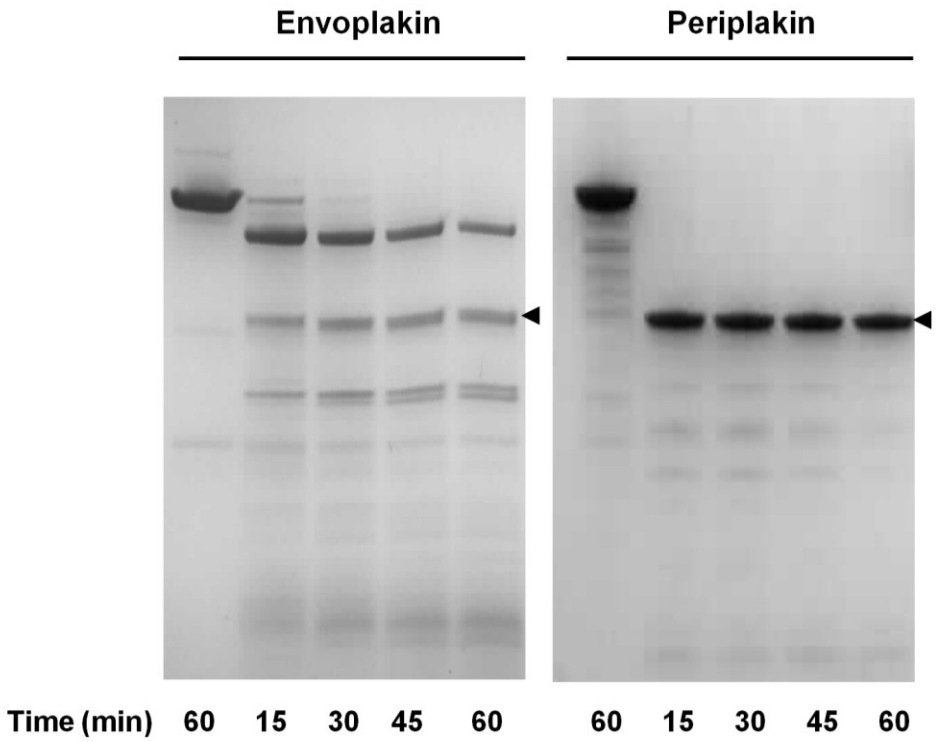


Figure 4.32 Limited proteolysis of the envoplakin and periplakin plakin domains

The envoplakin and periplakin plakin domains (10 mg/ml) were subjected to limited proteolysis at 37 °C by 0.001 mg/ml of trypsin. Left: in envoplakin three bands were observed, one at approximately 85 kDa, one at approximately 48 kDa (arrowhead) and one at approximately 27 kDa. Right: a different limited proteolysis profile was observed for the periplakin plakin domain in which only a single prominent band was observed at ~48 kDa (arrowhead). Experiment conducted in collaboration with Dr Martyn Chidgey.

4.33 Identification of a potential solvent exposed loop/hinge in the plakin domains of desmoplakin, envoplakin and periplakin

A solvent exposed loop/hinge appeared to present in all plakin domains as ascertained by the limited proteolysis results. To more accurately address the position of the solvent exposed loop/hinge the plakin domains of desmoplakin, envoplakin and periplakin were aligned using CLUSTALW (Thompson et al., 1994) (figure 4.33). It revealed the presence of a sequence separating the protease resistant fragments of the SR3-6 region and the (desmoplakin) SR78 fragment from in all 3 plakin domains. This coordinates to the region immediately after the SR6 region which was suggested to be a solvent exposed loop/hinge by the limited proteolysis data shown in 4.31 and 4.32. Interestingly, the sequence present in desmoplakin appeared to be longer. Unfortunately, time did not allow for the modular organization of envoplakin and periplakin to be assessed. However, that said, this region was taken for further analysis by SAXS flexibility programs (discussed further in 4.53 and 4.62).

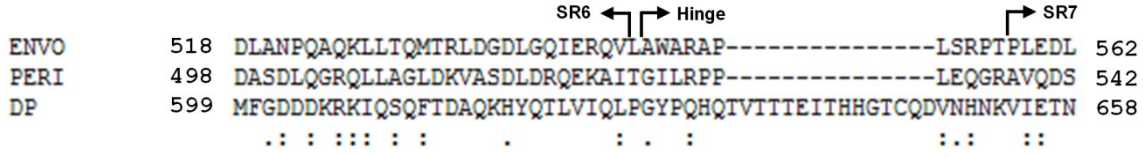


Figure 4.3 Identification of a hinge region in the plakin domains of desmoplakin, envoplakin and periplakin

The plakin domain sequences of envoplakin, periplakin and desmoplakin were aligned using CLUSTALW (Thompson et al., 1994). It revealed that a sequence separating SR6 and SR7 was present in each of the plakin domains which coordinated to the solvent exposed loop/hinge. Of interest was the fact that the desmoplakin solvent exposed loop/hinge was longer than those observed for envoplakin and periplakin.

4.4 Analytical ultracentrifugation of the desmoplakin, envoplakin and periplakin plakin domains

The oligomeric state of the desmoplakin plakin domain was gauged by analytical ultracentrifugation. A single peak of which approximated to 89.8 kDa (as calculated by SEDFIT) (Brown and Schuck, 2006) was observed (figure 4.4). This indicates that the desmoplakin plakin domain is monomeric in solution and corroborates previous work suggesting the homo-oligomerization of desmoplakin is facilitated by the adjacent rod domain (Green et al., 1990). The oligomeric state of the envoplakin and periplakin plakin domains was also ascertained by AUC. For envoplakin and periplakin single peaks which approximated to 89.5 kDa and 84.1 kDa respectively were observed (figure 4.4). The data indicate that both the envoplakin and periplakin plakin domains are also monomeric in solution.

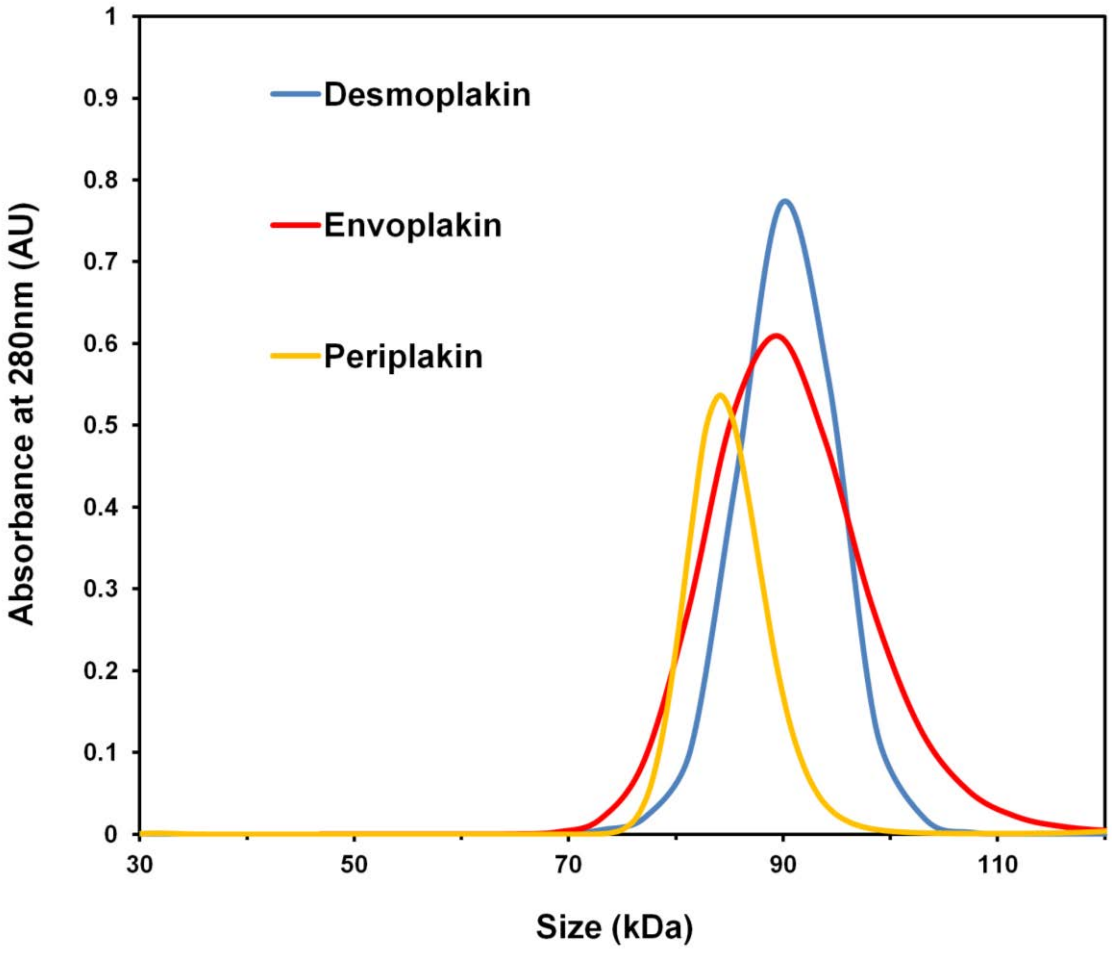


Figure 4.4 Analytical Ultracentrifugation (AUC) profiles of each subdomain

Each plakin domain was subjected to AUC at room temperature at a speed of 40000 rpm for 20 h on a Beckman XL-1 analytical ultracentrifuge using an eight cell 50Ti rotor and detected at 280 nm. Samples were run at between 0.05 and 4 mg/ml in 20 mM sodium phosphate (pH 7.2) and 100 mM NaCl. The results were analysed by the program SEDFIT (Brown and Schuck, 2006)in which the continuous c(S) analysis method was used to calculate estimated molecular masses. Desmoplakin (blue), envoplakin (red) and periplakin (gold) each reveal a single peak. Molecular masses of 89.8, 89.5 and 84.1 kDa were estimated for desmoplakin, envoplakin and periplakin respectively. No other higher order species were observed.

4.5 SAXS analysis of the desmoplakin plakin domain

4.51 Acquiring biophysical data of the desmoplakin plakin domain by SAXS

Purified desmoplakin plakin domain protein was analysed by SAXS to determine its structure in solution (figure 4.51A). The Guinier region indicated lack of aggregation as indicated by a linear regression (figure 4.51B). This implies that the protein is of sufficient quality to continue with SAXS analysis. GNOM analysis revealed an R_g value of 9.4 nm. GNOM was also used to gauge the D_{max} and using the $P_{(r)}$ function it was calculated as 33.5 nm (figure 4.51C). Using *ab-initio* modelling based on the program DAMMIF 10 bead models were created and superimposed manually in an attempt to ascertain the convergence of the models (figure 4.51D). When the models were superimposed manually a large amount of heterogeneity was observed which implied the model may require some extra modelling (discussed further in 4.52). However, a final *ab-initio* envelope was acquired using DAMAVER and DAMFILT which revealed an “L” shape with the longer arm being 24.0 nm and the shorter arm being 17.9nm long. This is in stark contrast to a linear model expected for a serial chain of SRs.

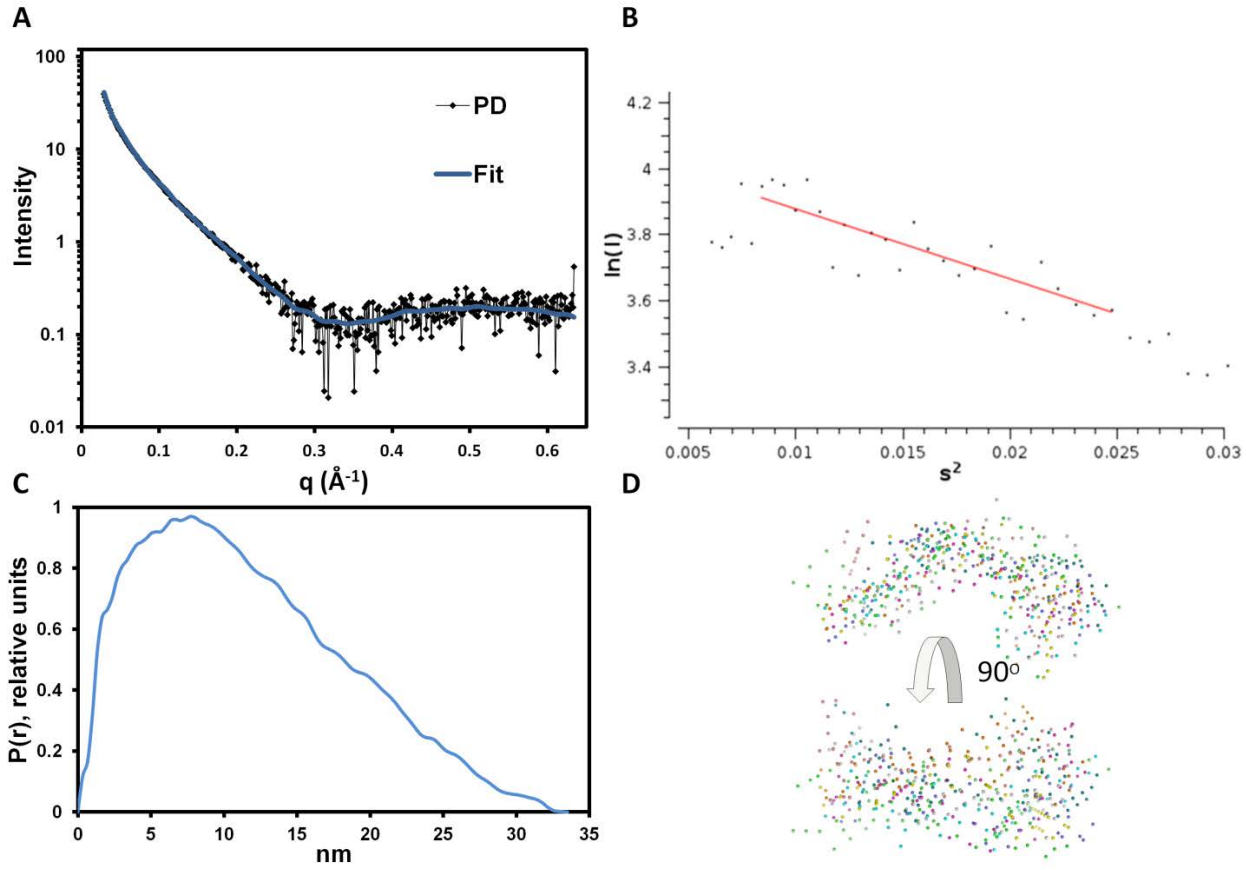


Figure 4.51 SAXS analysis of the desmoplakin plakin domain

The desmoplakin plakin domain was analyzed at 0.5, 1, 2 and 4 mg /ml in 20 mM sodium phosphate (pH 7.2) and 100 mM NaCl. The SAXS scattering curves were scaled and merged to create a final SAXS scattering curve in accordance with published methods. A: Buffer subtracted SAXS scattering curve with a DAMMIF fit (blue). B: Guinier analysis showing that the fit is good indicating a reliable R_g value (visualized in PRIMUS). C: P(r) distribution curve indicating the D_{max} to be 33.5 nm. D: Superimposition of the 10 output files of DAMMIF highlighting the heterogeneity of each pdb file. Adapted with permission from Elsevier from Al-Jassar et al. (2011) under license number 2862440887646.

4.52 Gauging non-linearity of the desmoplakin plakin domain

Unexpectedly, the *ab-initio* envelope of the desmoplakin plakin domain was revealed to be non-linear. Hence linear models of the entire plakin domain were created by I-TASSER (Roy et al., 2010) and BUNCH. If the SAXS scattering data was truly representational of a non-linear structure, one would expect to observe a poor fit of a linear model to the SAXS scattering curve. Using the program CRYSOL linear models from the two independent modelling programs were superimposed to the SAXS scattering curve (figure 4.52A). This gave a poor fit for both modelling programs suggesting a full length plakin domain cannot be linear in solution. When the known structured regions of the desmoplakin plakin domain (SR3-6 and SR78) were manually input into the *ab-initio* envelope only one orientation appeared likely (figure 4.52B). This left a region separating the SR6 and SR7 and the likely flexible CT region. This corroborates the limited proteolysis data which suggests that the plakin domain is comprised of two structured regions separated by a solvent exposed loop or hinge. Since a bend is observed between SR6 and SR7 it implies there may be some flexibility between the two regions. This leaves an unoccupied (and non-cylindrical) region proceeding SR8. An anomalous construct encompassing SR8-CT was shown to be flexible or partially folded in chapter 3. Therefore it is highly likely that given the non-cylindrical shape of this region observed in the envelope, the CT region is likely to move relatively freely in solution (i.e. it is flexible).

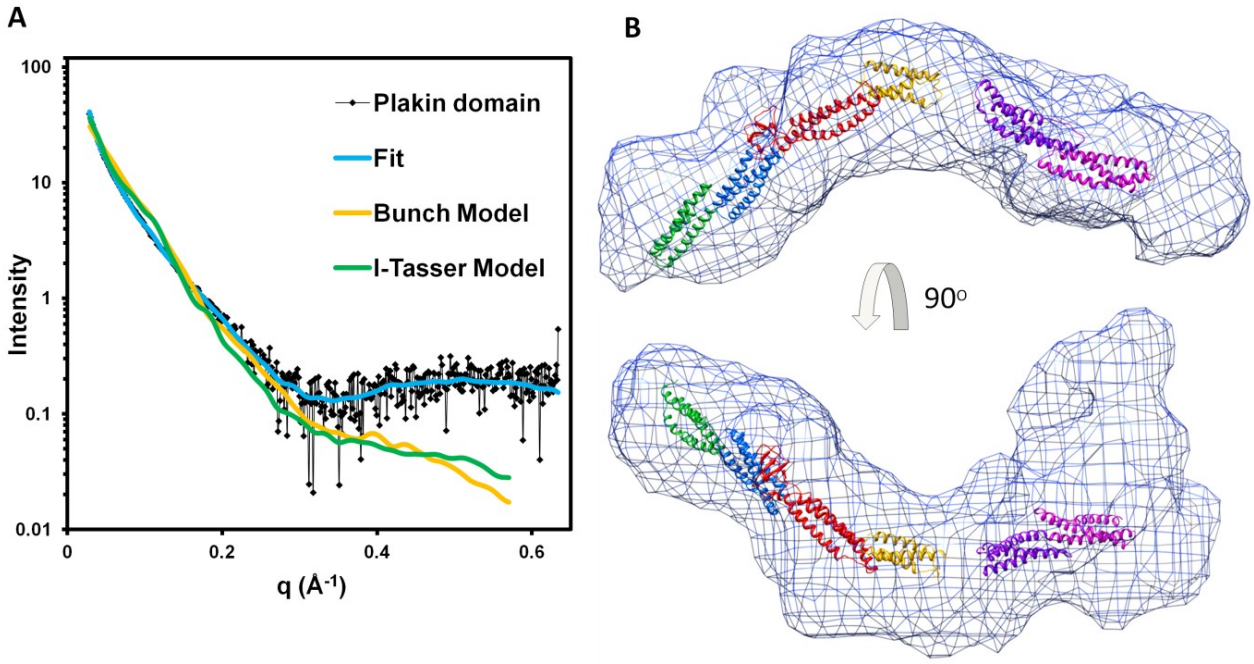


Figure 4.52 Non-linearity of the desmoplakin plakin domain

A: Buffer subtracted SAXS scattering curve for the desmoplakin plakin domain (black), DAMMIF fit (cyan) and CRYSOL created back-calculated theoretical curves using a linear BUNCH model (gold) and a linear I-TASSER model (green). B: A final averaged and filtered *ab-initio* envelope shown in two orthogonal views viewed by CHIMERA. The recently derived crystal structure of SR3-6 (Choi and Weis, 2011) was inserted manually (SR3: green, SR4: blue, SR5+SH3: red, SR6: gold). A gap is left intentionally for a likely hinge region. The model created for SR78 was also placed manually (SR7: purple, SR8: magenta). The likely flexible CT region is left unoccupied. Collectively, it suggests that the desmoplakin plakin domain is non-linear and contains loop/hinge regions between SR6 and SR7 and proceeding SR8. Adapted with permission from Elsevier from Al-Jassar et al. (2011) under license number 2862440887646.

4.53 Assessing flexibility of the desmoplakin plakin domain by SAXS

Given the non-linearity of the desmoplakin plakin domain, the presence of a solvent exposed loop/hinge between SR6 and SR7 and the anomalous behaviour of the CT region, it was thought that the desmoplakin plakin domain contained flexible elements. Using the program EOM, two models based on anticipated flexible regions were calculated to investigate whether specific regions of the plakin domain contributed to the proposed flexibility. As stated in the methods the EOM suite is based on two algorithms. The first program (RANCH) created 10000 randomly orientated models based on the structured and unstructured regions input. The second (GAJOE) then used these models to select an ensemble of random models that fits the SAXS scattering data named “the selected ensemble”.

Since the crystal structure of desmoplakin SR3-6 and a similar homologous region in plectin were shown to be rigid and linear (Choi and Weis, 2011; Ortega et al, 2011) this region was marked as structured in both theoretical models. On one model the solvent exposed and likely hinge at amino acid region P627-K653 was designated as a rigid and linear α-helix predicted from the I-TASSER model (named Plakin Domain Rigid or PDR model), while the second model had the same region assigned as flexible (named plakin domain flexible or PDF model). This is based on the limited proteolysis data in which SR3-6 and SR78 were both protease resistant with remaining regions as likely solvent exposed and/or flexible. In both models the CT region was assigned as flexible based on previous biophysical analysis of this region (discussed in the chapter 3).

X^2 values of 3.1 and 0.8 were calculated for PDR and PDF models respectively (figure 4.53). X^2 values indicate how valid the proposed model is representational of the protein in solution. Given that a χ^2 value close to 1 represents a fit of a selected ensemble matching closely to the actual behaviour of the protein in solution (as long as the curve fits the experimental data perfectly); a

difference is observed between the PDR and PDF models (figure 4.53). The higher value for the PDR model implied that it is less representational of the protein in solution than that of the PDF model. The PDR model also gave a poorer fit to the SAXS scattering curve compared to the PDF model. In contrast, the PDF model which additionally contained a flexible hinge region at position P627-K653 is significantly contributing to the flexibility observed for the desmoplaklin plakin domain. It must be noted that given the low resolution of SAXS the exact stretch of amino acids that are responsible for the flexibility cannot be identified. This also has implications on the DAMMIF generated *ab-initio* envelope which implied a static model. Since the SAXS scattering data is an average of all orientations in solution, DAMMIF in turn creates an averaged representation of the protein. While the DAMMIF model is not incorrect, the ensemble of models generated by the inclusion of flexibility within the plakin domain is more valid and is discussed further in 4.7. To investigate whether the flexibility is an inherent feature of all plakin domains, the plakin domains of envoplakin and periplakin were also analyzed by SAXS.

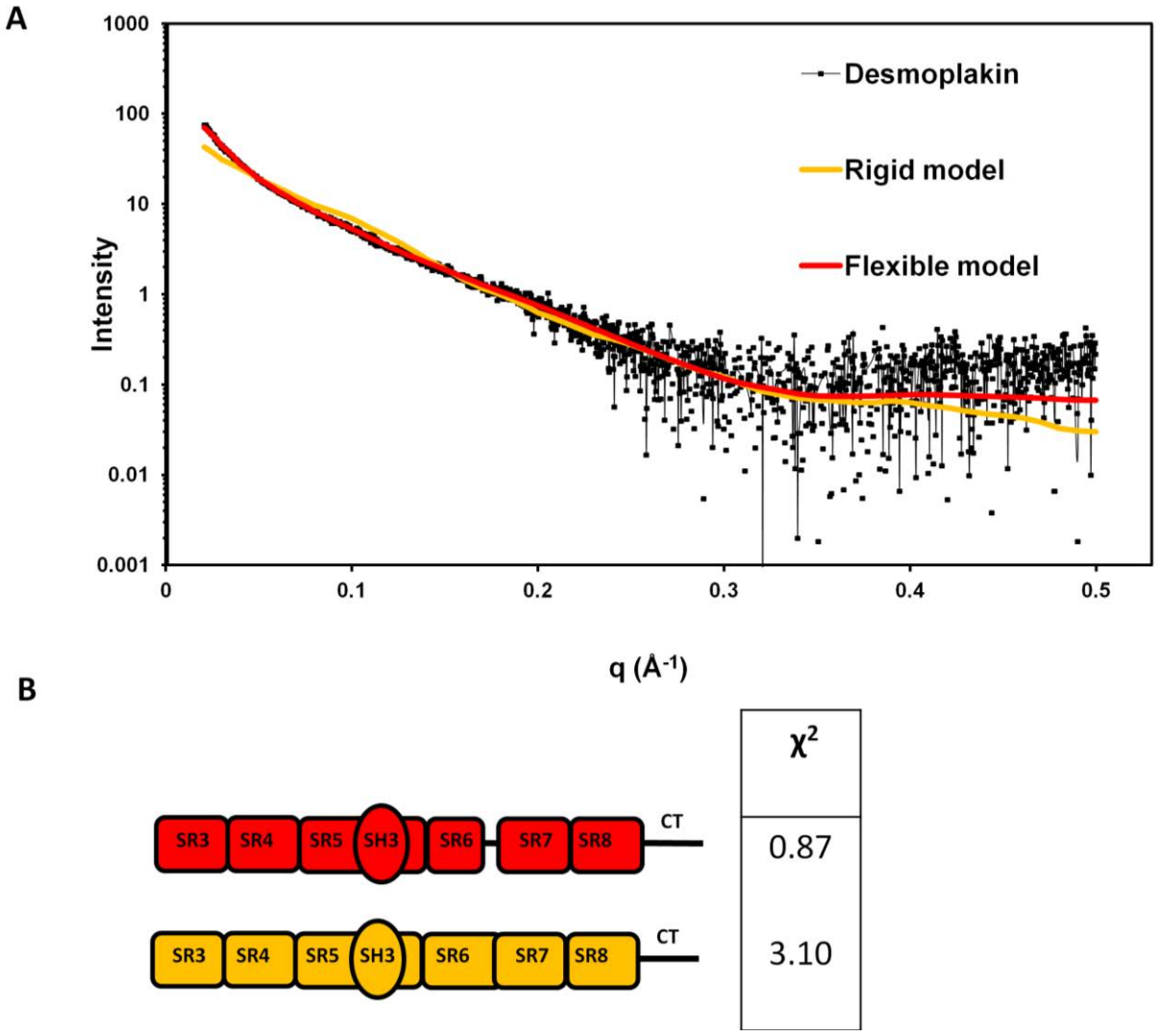


Figure 4.53 Assessing flexibility of the desmoplakin plakin domain using the Ensemble Optimization Method (EOM).

Two models differing in their flexible regions were created to gauge the possible presence of a flexible hinge within each of the plakin domains using EOM (see methods for further details). (A) Buffer subtracted SAXS scattering curve of desmoplakin (black) including EOM fits for a flexible (red) and rigid (yellow) model. (B) χ^2 values for the flexible (PDF) (red) and rigid (PDR) (yellow) models.

4.61 SAXS analysis of the envoplakin and periplakin plakin domains

The envoplakin plakin domain was analyzed by SAXS. The Guinier region indicated lack of aggregation with a manually estimated R_g value of 8.2 nm. GNOM analysis revealed an R_g value of 7.7 nm corroborating the manually acquired Guinier R_g value. A D_{max} value of of 23.1 nm was obtained by GNOM.

Similarly the periplakin plakin domain was analyzed by SAXS. The scattering curve indicated lack of aggregation with a manually estimated R_g of 8.9 nm. GNOM analysis revealed a similar Rg value of 9.3 nm and a D_{max} value of 35.0 nm

Since envoplakin and periplakin are homologous to desmoplakin it was assumed that they also do not adopt rigid structures and therefore was not be investigated by DAMMIF for an averaged model. Using the program CRYSOL, a theoretical SAXS scattering curve for a linear model created by I-TASSER was superimposed to the experimental SAXS scattering curve of the envoplakin plakin domain. In common with the desmoplakin plakin domain it revealed a poor fit confirming the non-linearity of the envoplakin plakin domain (figure 4.61A). Linear models were also created of the periplakin plakin domain by I-TASSER. CRYSOL was used to superimpose a theoretical SAXS scattering curve on the experimental SAXS scattering data. Again a poor fit was obtained (figure 4.61B). Thus it appears that all plakin domains analyzed are non-linear in solution.

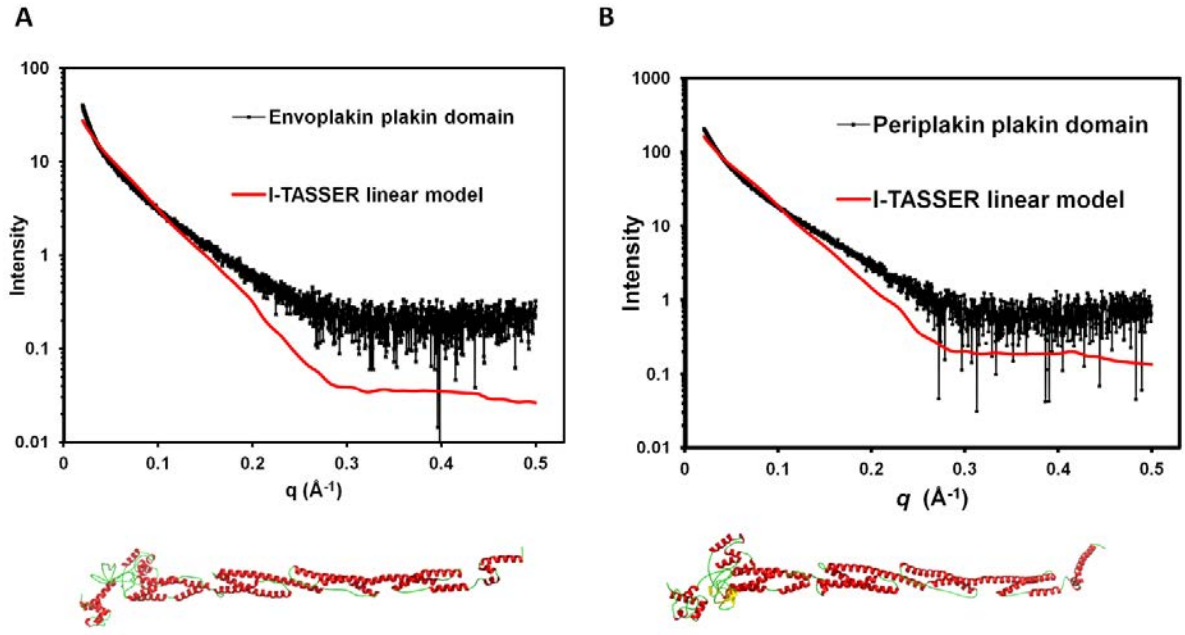


Figure 4.61 The envoplakin and periplakin plakin domains adopt non-linear conformations in solution.

A: the envoplakin plakin domain (black) was analyzed at 1, 2 and 5 mg /ml in 20 mM sodium phosphate (pH 7.2) and 100 mM NaCl. A theoretical SAXS scattering curve (red) was superimposed on to the experimental data for a linear model created by I-TASSER (below graph). B: the periplakin plakin domain (black) was analyzed at 2, 4 and 10 mg /ml in 20 mM sodium phosphate (pH 7.2) and 100 mM NaCl. A theoretical SAXS scattering curve (red) was superimposed on to the experimental data for a linear model created by I-TASSER (below graph). The SAXS scattering curves were scaled and merged to create a final SAXS scattering curve in accordance with previously published methods. The poor correlation between the experimental data and the I-TASSER linear models suggest that plakin domains are non-linear in solution.

4.62 Gauging flexibility of the plakin domains of periplakin and envoplakin

Knowledge of flexible/disordered regions within multidomain proteins for EOM analysis is highly advantageous. Unlike the desmoplakin plakin domain which had subdomains spanning the entire region subjected to biophysical analysis in chapter 3, only limited information was available on the modular architecture for periplakin and envoplakin. This included alignment data present in the literature (Sonnenberg et al, 2007) and in section 4.33, limited proteolysis data (figure 4.32) and models created by I-TASSER which collectively suggested a SR3-6 protease resistant rigid domain, a sovent exposed loop connected to a SR78-CT region (although SR78 of periplakin was shown to be protease sensitive in 4.32). SAXS data for the envoplakin and periplakin plakin domains were subjected to EOM analysis to gauge flexibility. Hence, a strategy similar to that used to investigate the possibility of a hinge in desmoplakin was applied to envoplakin and periplakin. In the flexible models both SR3-6 and SR78 were set as "rigid" while the solvent exposed loop/hinge and CT region were set as flexible in the EOM program. The rigid models contained only a flexible CT region with the solvent exposed loop/hinge set as a rigid linear α-helix based on I-TASSER predictions. Based on the previously described sequence alignment in 4.33, the exposed loop/hinge was judged as being V545-P558 in envoplakin while in periplakin this was I525-A538. In addition to the CT region proceeding SR8 of envoplakin and periplakin the solvent exposed loop/hinge was set as flexible regions in the program GAJOE while the remaining regions were set as rigid. Using GAJOE the χ^2 values of envoplakin and periplakin were also obtained for each of the rigid and flexible models and are given in figures 4.62 and 4.63 respectively. Like desmoplakin, the insertion of a flexible hinge between SR6 and SR7 significantly lowers the χ^2 value in both envoplakin and periplakin, arguing for the existence of a flexible hinge. Again due to the low resolution of SAXS, the exact amino acid stretch that is responsible for the flexibility cannot be determined.

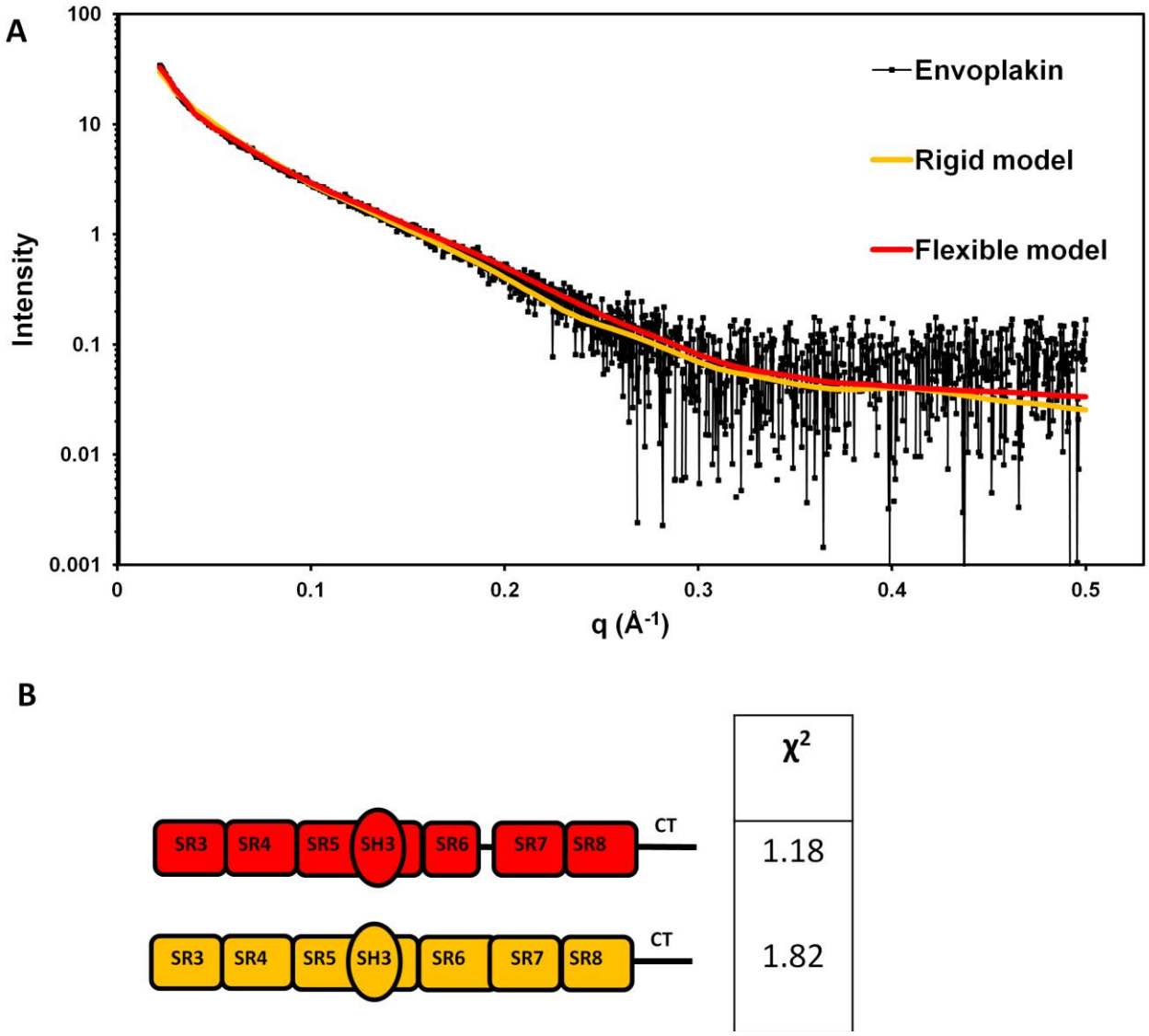


Figure 4.62 Assesing flexibility of the envoplakin plakin domain using the Ensemble Optimization Method (EOM).

Flexible and rigid models were created to gauge the possible presence of a flexible hinge within each of the plakin domains using EOM (see methods for further details). (A) Buffer subtracted SAXS scattering curve of envoplakin (black) including EOM fits for a flexible (red) and rigid (yellow) models. (B) χ^2 values for the rigid and flexible models. The χ^2 values indicate that the flexible model is more representative of the envoplakin plakin domain in solution.

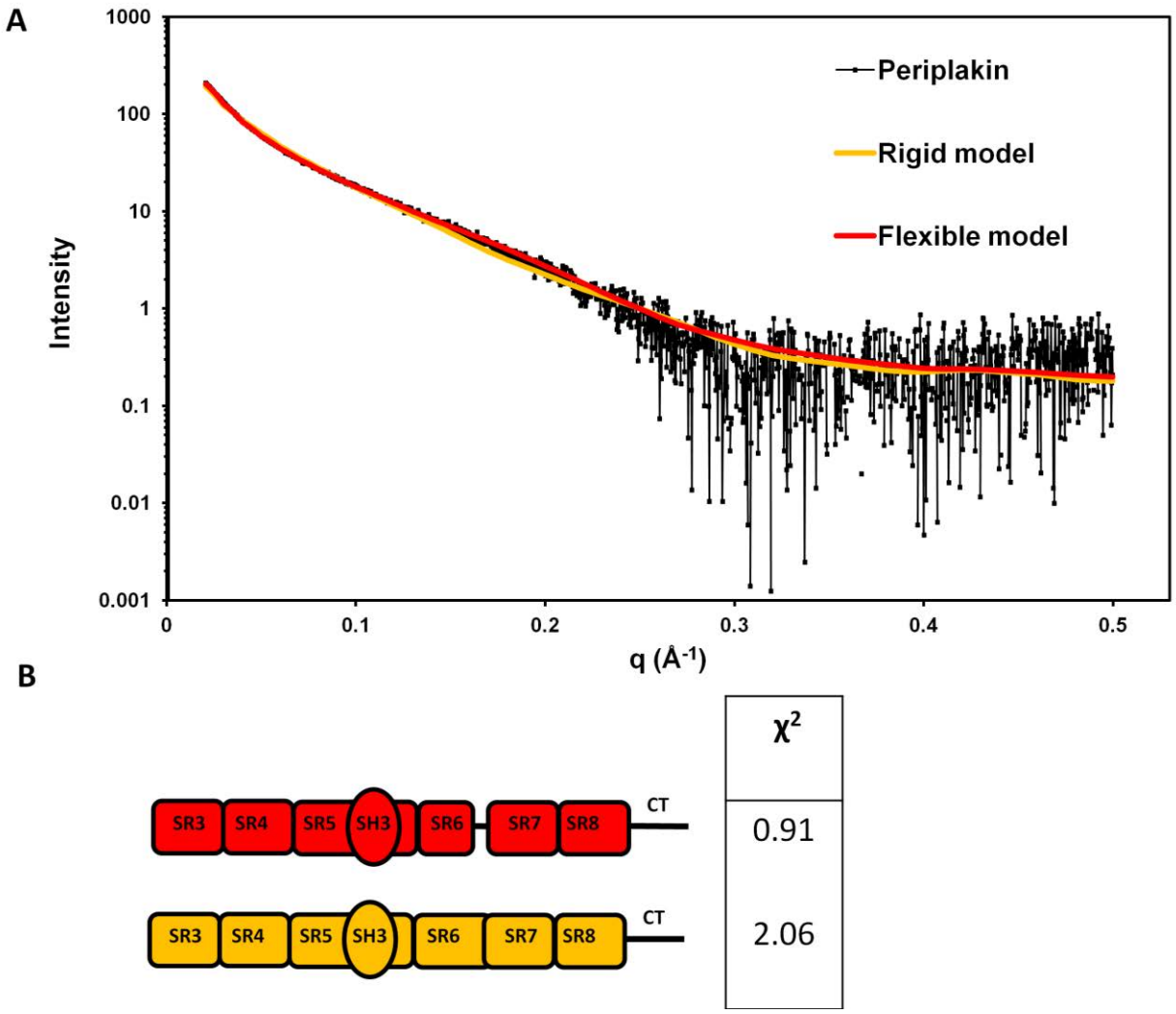


Figure 4.63 Assesing flexibility of the periplakin plakin domain using the Ensemble Optimization Method (EOM).

Flexible and rigid models were created to gauge the possible presence of a flexible hinge within each of the plakin domains using EOM (see methods for further details). (A) Buffer subtracted SAXS scattering curve of periplakin including EOM fits for a flexible (red) and rigid (yellow) model. (B) χ^2 values are given for both the rigid and flexible models. The χ^2 values indicate that the flexible model is more representative of the periplakin plakin domain in solution.

4.7 Assessment of the flexible models created by EOM for the plakin domains

The “selected ensemble” of 50 models that GAJOE selects to fit to the data can be visualized. In the following sections the 50 models selected for desmoplakin, envoplakin and periplakin were both qualitatively (visually), and quantitatively analyzed.

4.71 Qualitative assessment of the selected ensemble from desmoplakin, envoplakin and periplakin

A visual inspection was conducted using the selected ensemble of 50 models selected by GAJOE from the pool of the original 10000 random models created by RANCH. A representation of each of the plakin domain selected ensembles is shown (figure 4.71). For ease of inspection the SR3-6 region was superimposed for each model with the remaining SR78 and CT free in solution. It is important to note that the 50 models do not necessarily reflect the only orientations the protein is able to adopt (Personal communication: Dr Pau Bernadó, Barcelona, 2011). With that said, qualitative comparisons to the pool and general orientations can be examined. In all three selected ensembles the existence of a subset of linearized models was not observable. The hinge in desmoplakin more so than envoplakin and periplakin appears to create a “U-bend” in which SR78 almost becomes anti-parallel with the SR36 rigid entity. Envoplakin, more so than periplakin and desmoplakin, appears to contain a SR78 domain orientating around SR36 in a “helicopter” propeller fashion. Envoplakin’s CT region also appears to point in the direction of SR36. Periplakin has a more random orientation of both the SR78 domain and CT region.

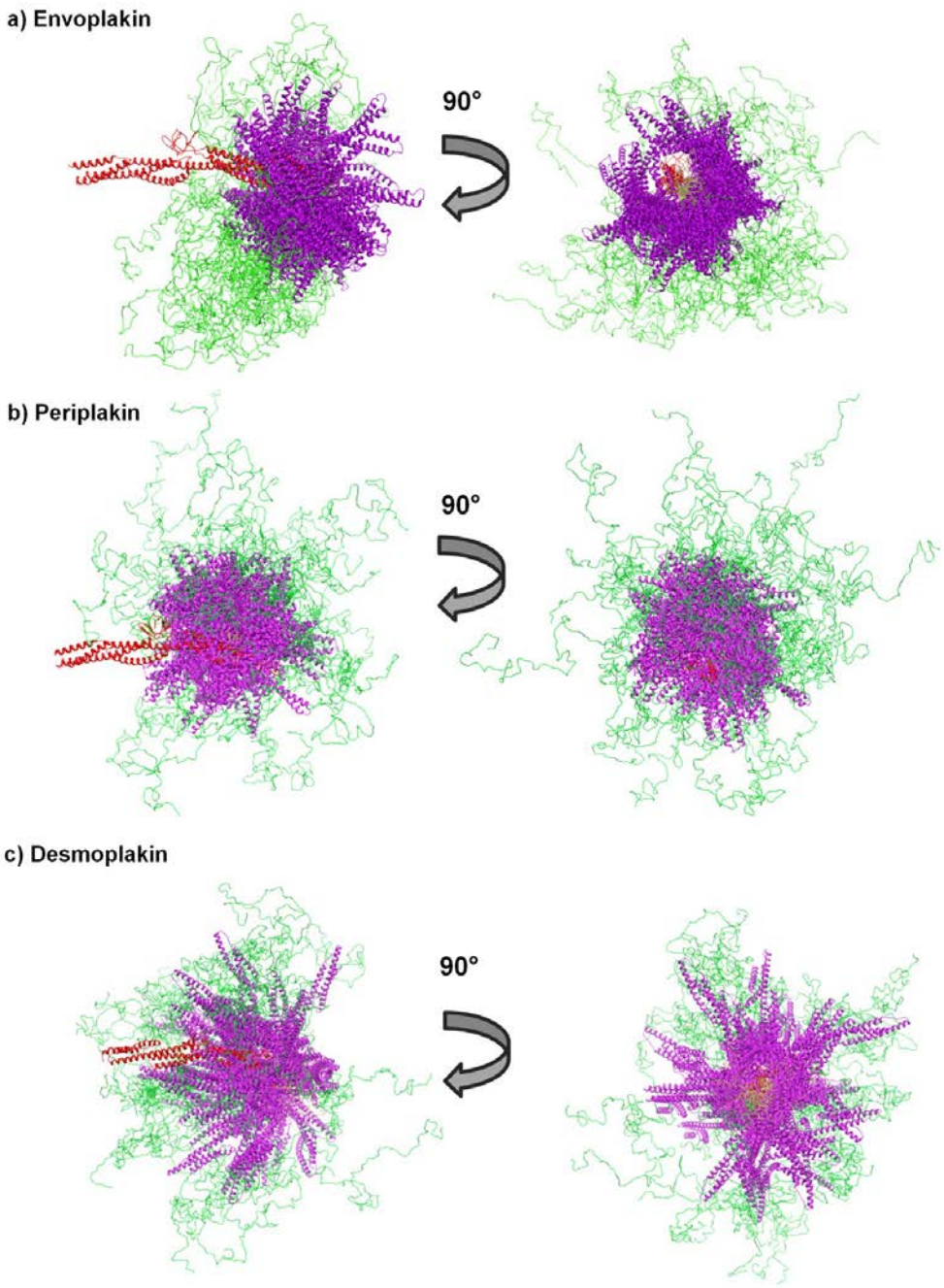


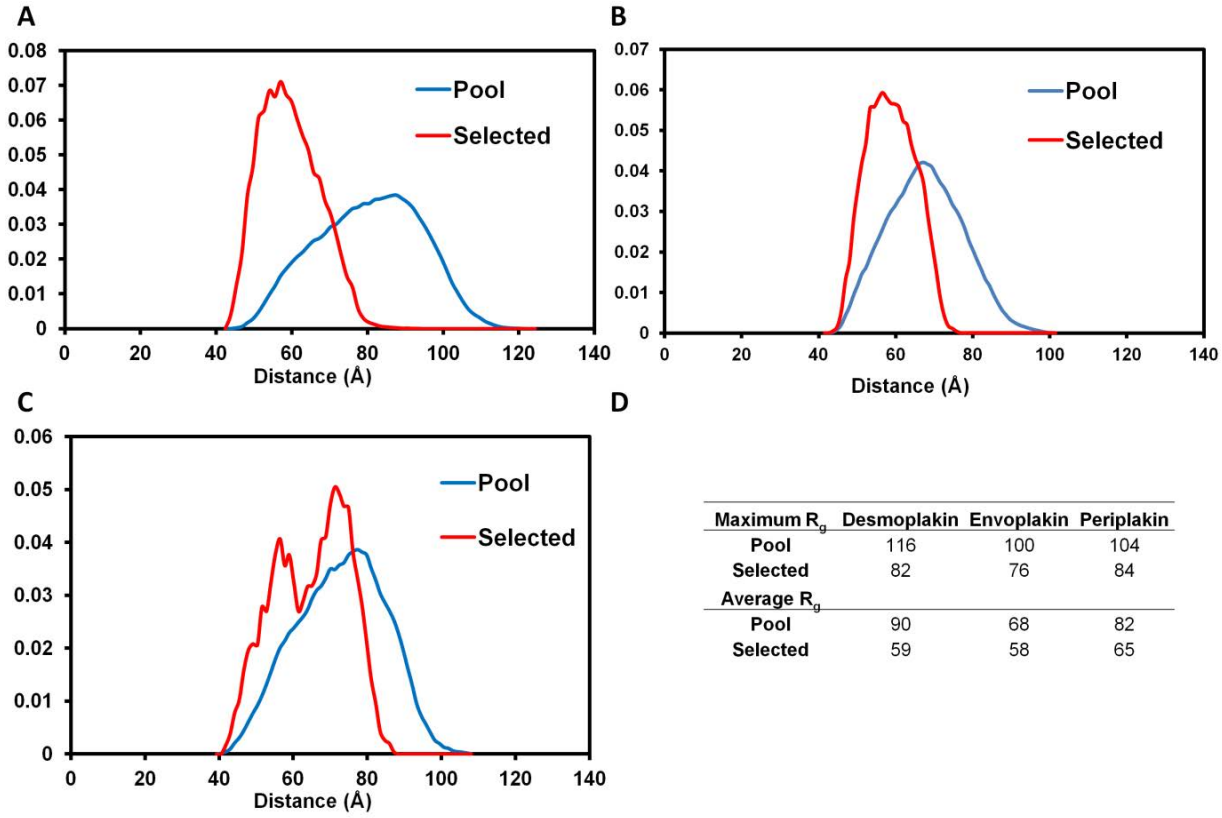
Figure 4.71 Molecular representation of the selected ensemble.

The 50 models selected by the program GAJOE most representative of their respective SAXS scattering curves were superimposed. In each case SR3-6 (red) was fixed in position and SR7-8 (purple) and the CT region (green) were allowed to rotate freely about the hinge (gold). A: envoplakin B: periplakin C: desmoplakin.

4.72 Size distribution graphs of all plakin domains

Distribution graphs based on outputs of GAJOE include R_g and D_{max} size distribution graphs of the 10000 randomly created models and the final 50 selected ensemble most representative of the SAXS scattering curve are given in figure 4.72. As one would expect, a randomly generated pool of models for each plakin domain gives a broad Gaussian distribution (figure 4.72). The selected ensemble for each plakin domain gave an average and maximum R_g lower than that of the random pool of models (figure 4.72). The same was true for the D_{max} (figure 4.73).

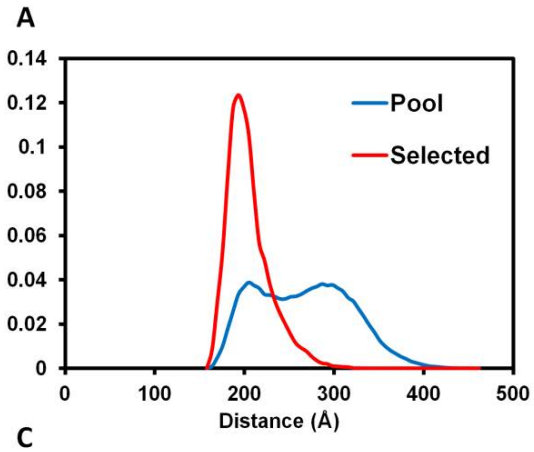
An overview of the rigidity and flexibility of the plakin domains based on the distribution graphs can be ascertained. If the protein was entirely disordered, the selected ensemble would superimpose over the distribution of randomly generated models. Since the plakin domain selected ensembles did not superimpose on the random pool, this can be discounted. Conversely if the protein was entirely rigid a very narrow and sharp peak would be observed solely for the selected ensemble distribution. Although qualitatively the selected ensemble peaks were sharper than the random pools, it is unlikely the proteins are entirely rigid given the diversity of the 50 selected models shown in figure 4.71. This would be consistent with the idea that plakin domains contain two rigid linear domains separated by a flexible hinge. Interestingly for all plakin domains the selected ensemble distribution of R_g values appears to be broader than that of the D_{max} values. This would suggest for a motion in which a central domain (taken as the centre of origin) is fixed and another domain orientates around it relatively freely. This appears to be the case upon closer inspection of the 50 selected ensembles for each plakin domain (see figure 4.71) in which an “umbrella” or “helicopter” type action can be observed.

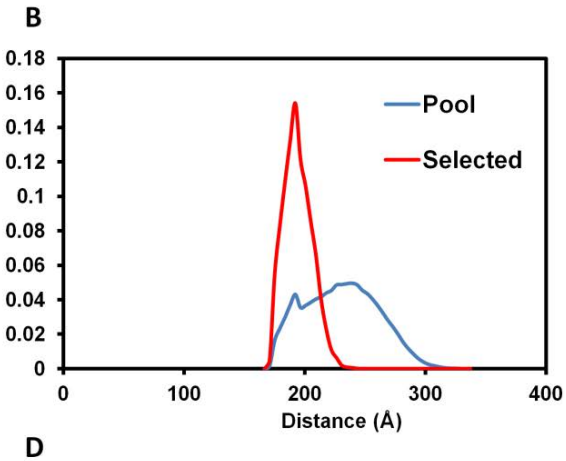


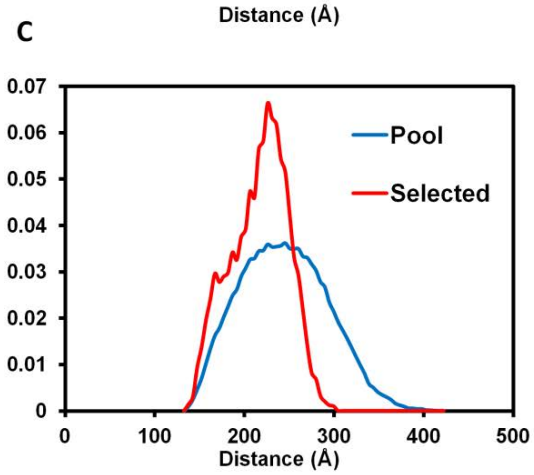
Maximum R_g	Desmoplakin	Envoplakin	Periplakin
Pool	116	100	104
Selected	82	76	84
Average R_g			
Pool	90	68	82
Selected	59	58	65

Figure 4.72 Radius of gyration (R_g) distributions between the pool and selected ensemble of each plakin domain.

The program GAJOE was used to calculate R_g for each model from the random and selected pools and a distribution was created. Comparisons between the starting 10000 conformations, the random pool (blue) and selected ensemble (red) R_g values are given for each plakin domain. A: desmoplakin, B: envoplakin, C: periplakin D: Table of maximum and average R_g values of each plakin domain. Pool and selected values are compared for each plakin domain.







D

Maximum D_{max}	Desmoplakin	Envoplakin	Periplakin
Pool	411	308	393
Selected	299	228	302
Average D_{max}			
Pool	301	254	238
Selected	197	194	229

Figure 4.73 Maximum particle distance (D_{max}) distributions between the pool and selected ensemble of each plakin domain.

The program GAJOE was used to calculate D_{max} for each model from the random and selected pools and a distribution was created. Comparisons between the starting 10000 conformations, the pool (blue) and selected ensemble (red) D_{max} values are given for each plakin domain. A: Desmoplakin, B: Envoplakin, C: periplakin D: Table of maximum and average D_{max} values of each plakin domain. Pool and selected values are compared for each plakin domain

4.8 Quantitative analysis of inter-domain distribution distances between two amino acids

To mathematically determine the distance distributions between two amino acids on the protein chain from each model from the random and selected pools of each plakin domain, a custom built program kindly created by Dr Pau Bernado was used (code can be seen in appendix C).The program calculates the distance between two geometrical positions (in this case, amino acids) at different regions of the protein and calculate their distance apart in angstroms. The program creates a distribution of the distances from all 10000 randomly generated models and compares it to the 50 models from the selected ensemble most representational of the experimental SAXS data by GAJOE.

Using the first and last amino acids of each plakin domain, inter-domain distribution graphs were created (figure 4.8). A few general trends are observable between the plakin domain graphs. Firstly that the selected ensemble is negatively skewed in each plakin domain meaning that the end amino acid of each plakin domain is likely to be closer to the beginning amino acid of the protein when compared to a random pool of 10000 structures. Secondly periplakin and envoplakin have similar inter-domain distribution graphs for their selected ensembles suggesting that the inter-domain distributions of the beginning to end amino acids are similar. Thirdly and most interestingly, the desmoplakin inter-domain distribution between the beginning and end amino acids is significantly more negatively skewed than compared to periplakin and envoplakin. This suggests that the end amino acid is much more likely to occupy a region closer to the beginning amino acid when compared to the desmoplakin pool and the selected ensembles of periplakin and envoplakin.

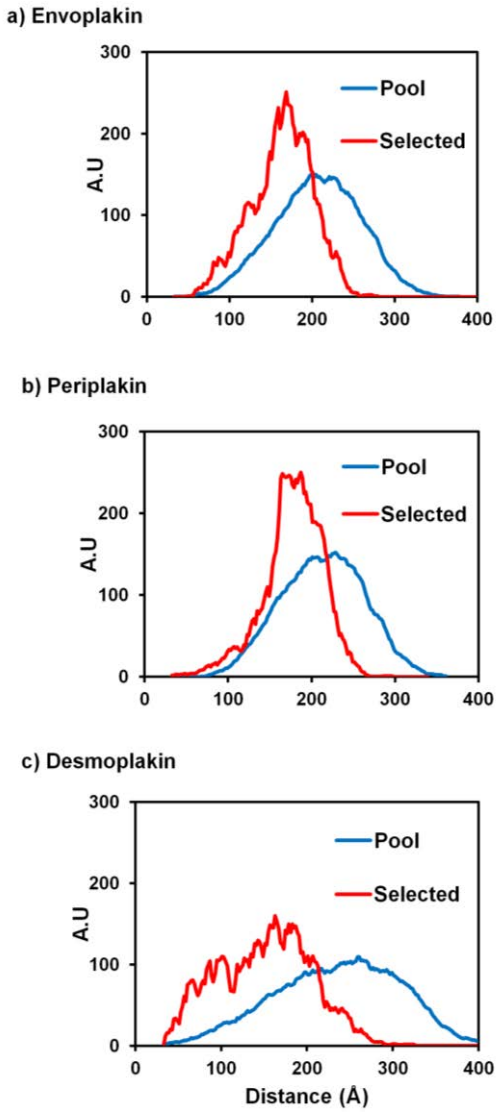


Figure 4.8 Inter-domain distance distribution graphs between the first and last amino acid of each plakin domain.

Distance distributions from each of the 10000 models and the 50 selected ensemble were calculated for the first and last amino acids only using a program written by Dr Pau Bernado for this project (appendix C). Distributions from the random pool are shown as blue while distributions from the selected ensemble are shown as red. (A) Comparisons between the pool and selected ensemble distance distribution between envoplakin's first (P144) and last (S891) amino acids. (B) Comparisons between the pool and selected ensemble distance distribution between periplakin's first (E131) and last (P871) amino acids. (C) Comparisons between the pool and selected ensemble distance distribution between desmoplakin's first (W180) and last (K1022) amino acids.

4.9 Chapter 4 conclusions

The experiments presented within this chapter describe the biophysical properties of entire plakin domains from desmoplakin, envoplakin and periplakin in solution. They reveal that plakin domains are predominantly α-helical as predicted from sequence alignment, and that plakin domains are monomeric, suggesting that dimerization is facilitated by their respective rod domains as previously described (Green et al., 1990, Ruhrberg et al., 1997). Furthermore, this is the first time that a serial chain of 6 spectrin repeats has been investigated in solution by SAXS analysis.

When linear models were created for each of the plakin domains investigated, none fit the SAXS scattering curves of the proteins in solution. Thus, it appeared that each of the three plakin domains investigated was non-linear. This was in stark contrast to previously published crystal and NMR studies which have all shown that chains of up to 4 spectrin repeats are linear (Choi and Weis, 2011, Ortega et al., 2011, Ylänne et al., 2001). While the findings described herein do not invalidate in any way previously published findings, what they do indicate is that there may be a crystallization bias towards rigid and linear SRs. This is because proteins with flexible regions are unlikely to be crystallized. When the potential dynamics caused by the likely hinge regions of desmoplakin, envoplakin and periplakin were investigated by SAXS using the suite of programs from EOM, differences were revealed between the three plakin domains. For example the C-terminal region of the desmoplakin plakin domain is much more likely to lie anti parallel with its N-terminal region. Envoplakin's C-terminal region of its plakin domain is likely to orientate in a "helicopter" fashion. Finally, periplakin's orientation appeared to be much more random than that of desmoplakin and envoplakin. Interestingly, subsets of linear models within the selected ensemble were either very limited or non-existent suggesting that in isolation plakin domains cannot form linear rods further corroborating the idea that plakin domains are non-linear.

Initially, the desmoplakin plakin domain was investigated as it was the primary target for structural elucidation given desmoplakin's critical role in the desmosome and cardiomyopathy related disease. The desmoplakin plakin domain was subjected to limited proteolysis to identify protease resistant fragments. It revealed the presence of two predominant bands (figure 4.31), one which was N-terminal sequenced and was subsequently identified as SR78. The other 52 kDa band (previously identified by Choi and Weis, 2011) was revealed to be SR3-6. These protease resistant domains were separated by an exposed proteolytically accessible loop region, the hinge sequence (figure 4.33). Given the presence of a loop between the two domains it is interesting to note that this is the first time a protein with a serial chain of spectrin repeats has indicated such a possibility by the use of limited proteolysis. Serial chains of SRs are often observed with small (3-5 amino acids) α-helical linkers between tandem spectrin repeats (Law et al., 2003). Atomic force microscopy showed that the protein dystrophin has a rod domain consisting of a serial chain of SRs interrupted by two hinge regions (Bhasin et al., 2005). To further ascertain the structural organization of the desmoplakin plakin domain SAXS was conducted.

The SAXS envelope of the desmoplakin plakin domain was the first of its kind, encompassing a serial chain of multiple SRs. It revealed a two-lobed structure, one 24.0 nm and the other 17.9 nm consisting of the SR3-6 and SR78-CT regions respectively. The lobe appeared to have an elbow as seen by the kink between the two lobes. The kink between the SR3-6 and SR78-CT domains suggests either a potential hinge or flexibility whereby the SR3-6 domain acts as a rigid entity and SR78 (being the most thermally stable) acts as a structural unit. This corroborates the previous limited proteolysis data which implied a loop region between SR3-6 and SR78. It therefore suggests that proteins harbouring serial chains of spectrin repeats may not in fact be static rods as originally thought. Whether this was a desmoplakin plakin domain specific

phenomenon was also investigated by exploring the plakin domains of envoplakin and periplakin.

Limited proteolysis results also revealed a 48 kDa protease resistant fragment for each of the envoplakin and periplakin plakin domains (figure 4.32). When N-terminal sequencing of these fragments was conducted it revealed a homologous region similar to that found by Choi and Weis (2011) from desmoplakin which corresponds to SR3-6. This suggested that the SR3-6 fragment in each of the plakin domains studied in this thesis was a rigid entity resistant to proteolysis. Based on the crystal structure of desmoplakin SR3-6 it is highly likely that SR3-6 of envoplakin and periplakin are also linear which was corroborated by models created by the I-TASSER server. Interestingly, no SR78 or similar protease resistant fragment was observed for periplakin. This therefore suggests that SR78 of desmoplakin is more proteolytically resistant than that of periplakin and may have a specialized role. That said, a similarly sized protease resistant fragment of 25 kDa was observed for envoplakin but was not N-terminal sequenced (figure 4.32).

Using the SAXS data for envoplakin and periplakin the non-linearity observed in desmoplakin was also investigated in envoplakin and periplakin. Linear models of the entire periplakin and envoplakin plakin domains were created using the I-TASSER server. Theoretical SAXS scattering patterns for each linear model were superimposed onto the experimentally derived patterns using the program CRYSOL. Both revealed a very poor fit confirming they were not linear. This suggests that like desmoplakin, the plakin domains of periplakin and envoplakin adopt a non-linear conformation.

Alignment data of envoplakin, periplakin and desmoplakin suggested the presence of a hinge located between SR6 and SR7 (figure 4.33). No structural motif was suggested between SR6

and SR8 in a previous alignment in desmoplakin (Sonnenberg et al, 2007), which failed to take into account the recently characterized SR7 in chapter 3. This still leaves a region of 27 amino acids (P627-K653). A stretch of amino acids aligned with this region in envoplakin (V545-P558) and periplakin (I525-A538) was also observed. To investigate the possibility of a flexible hinge within the plakin domains of envoplakin, periplakin and desmoplakin, the above mentioned regions were designated as flexible and analyzed by the program EOM. Two models were created for each plakin domain which both contained a CT region designated as flexible. The 2 models differed at the region between SR6 and SR7. One contained a linearized and rigid α-helical hinge (rigid model) while the other contained a flexible hinge (flexible model). Using the program RANCH a pool of 10000 randomly generated models were created for each model. Using the program GAJOE a selected ensemble of 50 models from the 10000 randomly generated pool that were most representational of the SAXS scattering curve, were fitted. For all 3 plakin domains the rigid model fit the data reasonably well, while the flexible model gave an excellent fit. This is especially the case at higher angles which gives the most significant effect on shape determination. Although χ^2 values are not applicable to data that do not fit excellently, there is a significant drop in the flexible model χ^2 value suggesting that the added flexibility in the hinge is more representational of the SAXS scattering curve than the rigid model. This suggested that a flexible linker or hinge is present in all 3 plakin domains, but given the low resolution of SAXS, the precise location cannot be obtained. However, based on the alignment data it is likely to reside between SR6 and SR7.

When R_g and D_{max} values were compared between the selected ensemble and the random pool of each plakin domain, a trend is observed. Each plakin domain has a selected ensemble average R_g and D_{max} value lower than that of the pool (figures 4.72 and 4.73). The selected ensemble peaks are also narrower and sharper when compared to the pool. Together this

suggests the added flexibility is allowing the protein to compact, with a very low propensity to linearize in isolation. That said, the ensemble of proteins selected for desmoplakin and envoplakin in particular do not form entirely random ensembles. Furthermore, virtually no linear models were selected in the ensemble of any plakin domains. Closer inspection of all R_g, D_{max} and selected ensembles data also suggests relative differences in the degrees of flexibility.

Envoplakin appears to contain the least flexible hinge element as shown by the relatively low number of different orientations assumed for the SR7 to CT region and the sharp R_g and D_{max} values, compared to that of desmoplakin and periplakin. Desmoplakin's hinge appears to create a "u-bend" in the SR7 to CT region which can almost become anti-parallel with SR3-6 (figure 4.71). The CT region of desmoplakin also appears to reside more closely to the beginning of the protein when compared to envoplakin and periplakin. Periplakin appears to adopt the most random orientations between all three plakin domains shown by the large heterogeneity in the selected ensemble (figure 4.71).

Inter-domain distribution graphs were then created for the start and end amino acids of each plakin domain (figure 4.8). Both envoplakin and periplakin shared similar profiles indicating that the distances between the beginning and end amino acids of each protein were similar. Desmoplakin however, adopted a broader and more negatively skewed selected ensemble when compared to both the selected ensembles of envoplakin and periplakin and that of the desmoplakin pool. This suggested that the end residue of desmoplakin is much more likely to reside nearer to the beginning residue when compared to envoplakin and periplakin and these findings were further corroborated by qualitative visual assessment of the superimposed 50 selected ensemble models.

Overall it appears that as well as identifying a flexible hinge, each plakin domain has varying degrees of flexibility and behaviour in solution. How this translates to function is less clear, although is likely to relate to the ability of SRs to withstand mechanical stress. Each plakin protein is expressed in tissues subjected to high mechanical stress (e.g. heart and skin). Spectrin repeats are well known to resist mechanical stress as shown by a variety of atomic force microscopy studies although plakin domains have to yet been shown to directly resist mechanical stress. It can however, be inferred that a component of the desmosome is likely to resist mechanical stress based on their presence in tissues experiencing mechanical stress such as the epidermis and heart. The likely (and probably only) candidate for this is the plakin domain given the presence of serial SRs. How these newly identified flexible hinges could relate to biological function is based on previous studies conducted on the hinges found in the serial SR containing rod domain of dystrophin. As far as I am aware the only direct experimental evidence for the function of flexible hinges in serial spectrin repeat domains comes from two papers based on dystrophin's H2 and H3 hinges in their serial SR chain containing rod domain (Banks et al., 2010, Bhasin et al., 2005). It was originally proposed by sequence based modelling that dystrophin had four hinges, with two in the serial SR rod domain (Koenig and Kunkel, 1990). Patients with deletions of H3 in dystrophin suffer from Becker Muscular Dystrophy (BMD) suggesting that H3 is likely to have an important functional role (Bhasin et al., 2005). Since desmoplakin's hinge has ~40% homology to H3 of dystrophin (data not shown) it is likely to share a similar functional role. H2 and H3 of dystrophin contain proline residues suggested to be critical to their function. No such proline residues exist in the proposed hinges of any of the plakin domains. However, H3 and the proposed desmoplakin hinge both contain threonines which align well and it is possible that these residues are important for desmoplakin. The relatively high abundance of threonines in the desmoplakin hinge (40%) implies the possible presence of secondary structure elements such as threonines can be incorporated into both

structured and disordered stretches of sequences. Banks et al (2010) proposed that the enhanced flexibility of H3 lead to an improvement of artificial mini-dystrophins to alleviate dysfunction caused by BMD. It is likely therefore that the hinge present in the plakin domains is also critical to the function of the plakin domain. Bhasin et al (2005) proposed that H3 of dystrophin as well as acting as a mechanical separator between SRs, is also likely to limit extensible unfolding. This is also likely to be the case in plakin domains, although since plakin domains have not been directly tested to withstand mechanical stress this would require further exploration.

No ligands are yet known for the envoplakin and periplakin plakin domains – although ligands are known for the periplakin N-terminal unstructured region (Boczonadi et al., 2007, Groot et al., 2004). As a result it would be difficult to speculate on the biological significance of the flexibility of the periplakin and envoplakin plakin domains. The desmoplakin N-terminal head region is known to bind to the intrinsically disordered desmoglein specific cytoplasmic region (DSCR) (Kami et al, 2009), plakophilin1 and plakoglobin (Bornslaeger et al., 2001, Kowalczyk et al., 1997) although their specific binding regions have not been localized (with the exception of PKP1 which is likely to be localized to SR34 (Al-Jassar et al., 2011)). Why the desmoplakin SR78-CT region folds back anti-parallel to SR3-6 in relation to its involvement in these interactions is unclear. It may reflect an orientation favourable for binding any of the above mentioned ligands. Given the flexibility of the desmoplakin plakin domain it is possible that upon induction of mechanical stress an alternative orientation may present a more preferable conformation of one ligand over the other. Indeed, ligand binding upon conformational change caused by mechanical stress has recently been shown in a protein with a similar role to desmoplakin (del Rio et al., 2009). Although this has not been directly shown in SRs the up and coming field of “intracellular mechanotransduction” has identified a number of candidate proteins

and complexes with such qualities including cell-cell junctions (Gomez et al., 2011, Chen, 2008). In combination with the thermal stability assays of the desmoplakin plakin domain subdomains, it suggests that there may be further subdivisions of functional properties within a plakin domain related to mechanical stress. SR78, being the most thermally stable of all constructs may act as the predominant “shock-absorber” in the desmoplakin plakin domain given that thermal stability correlates well with SRs ability to withstand mechanical stress (Law et al., 2003). The SH3 domain would rigidify SR3-6 (Choi and Weis, 2011) and SR34 shown to bind to PKP1 may act as the “mechanotransducer” allowing ligand binding in the unstressed state (Al-Jassar et al., 2011) while releasing PKP1 in the stressed state. This may also occur in the other ligands especially plakoglobin. Evidence for this possibility also comes from the fact that (in addition to desmosomally located catenin family members) there are nuclear and cytosolic stores of these catenin family members (Garcia-Gras et al., 2006, Palka and Green, 1997, Zhurinsky et al., 2000, Mertens et al., 2001) and it is unknown how/why the location of these proteins is so varied. Therefore the non-linear and flexibility of plakin domains described in this chapter may reflect a function related to mechanotransduction which has an implication towards altered gene expression and/or subsequent desmosome integrity.

Chapter 5 – Biophysical analysis of the envoplakin and desmoplakin plakin repeat domains

The plakin family proteins typically serve as cytolinkers that tether the intermediate filament (IF) cytoskeleton to the desmosome, with a crucial role of developing and maintaining tissue integrity (Garrod and Chidgey, 2008, Sonnenberg and Liem, 2007b). Their N-terminal head domains serve to bind to the desmosome, while their C-terminal tails bind to the IF proteins vimentin, desmin and keratin (Bornslaeger et al., 2001, Fontao et al., 2003, Kazerounian et al., 2002, Kouklis et al., 1994, Stappenbeck et al., 1993). Disruption of the desmoplakin-IFs linkages in desmosomes leads to skin and/or heart disease (Awad et al., 2008, Russell et al., 2004).

This chapter focuses on the C-terminal tail and more specifically the PRDs, although the C-terminal linker domains are also discussed in greater detail in chapter 6. In general plakin family members have one highly homologous “linker domain” and variable numbers of PRDs. The functional significance of the different number of PRDs in these plakin proteins is unknown, although their PRD sequences share high sequence similarity. It has been previously shown that binding occurs across multiple elements of the plakin c-terminal tails with IFs (Fontao et al., 2003, Kazerounian et al., 2002, Meng et al., 1997b, Nikolic et al., 1996, Stappenbeck et al., 1993, Steinböck et al., 2000). More specifically there appears to be a complex interplay between the PRDs, the linker domain and the extreme C-terminal tail. Furthermore, desmoplakin has a serine residue 23 amino acids from its carboxyl terminus (S2849) that plays a critical role in regulating cytoskeletal binding (Stappenbeck et al., 1994). Phosphorylation of S2849 decreases binding affinity to keratin 9-fold and desmin/vimentin 2-fold. While desmoplakin binding to desmin or vimentin is predominantly dependent on its PRD region, its binding propensity for keratin is dependent on both the PRDs and the extreme C-terminal tail (Meng et al., 1997b, Stappenbeck et al., 1993, Lapouge et al., 2006). It has been suggested that desmoplakin has

preference for binding keratin (K8/K18 hetrodimer) over desmin or vimentin. Plectin in contrast, has a higher affinity for desmin and vimentin compared to cytokeratins 5/14 (Steinböck et al., 2000). Thus desmoplakin and plectin appear to have distinct binding preferences for cytoskeletal proteins. The same also applies to envoplakin and periplakin in which desmin/vimentin is preferred over cytokeratins (DiColandrea et al., 2000, Karashima and Watt, 2002, Kazerounian et al., 2002). How this is achieved at the molecular level is unclear, and a more thorough understanding of the structural determinants present in the C-terminal tails of plakin family members may offer potential explanations.

The crystal structures of the PRDs B and C of desmoplakin were determined in 2002 and revealed a novel 38 amino acid motif repeated 4.5 times consisting of a β-hairpin linked to anti-parallel α-helices (Choi et al, 2002). Prior to their structure determination, PRDs were thought to possess an extended structure similar to that of the ankyrin repeat motif (Li et al., 2006). However, pilot structural studies revealed that the PRDs were unexpectedly compact due to multiple hydrophobic contacts between and within the PR motifs of PRDs. Further analysis of the PRD structure revealed a long positively charged groove on the surface area encompassed between the two halves of the PRD formed by PR1-2 and PR3-4. This was proposed as a putative binding site for the cytoskeletal protein vimentin. Interestingly, since vimentin also possesses multiple highly negatively charged patches, binding to the desmoplakin PRDs was proposed to occur via electrostatic interactions (Choi et al, 2002). Due to sequence homology it is presumed that PRDs of other plakin family members utilize similar mechanisms to bind to cytoskeletal proteins. Interestingly, two opposing theories suggest how the desmoplakin and plectin C-terminal tails form globular structures which allow interactions with cytoskeletal proteins. Firstly, the “beads on a string” theory based on studies of the desmoplakin C-terminal tail by Weis and colleagues suggests that each PRD is a separate unit capable of binding

vimentin individually (Choi et al, 2002). Secondly, a “compacted cylindrical” theory by Wiche and colleagues, based on studies of the plectin C-terminal tail, suggests an intricate interplay is required between each PRD where disulphide bonds are critical for inter-PRD binding (Janda et al., 2001). The “beads on a string” theory is based on the fact that PRDs B and C of desmoplakin independently fold (Choi et al, 2002). Conversely the compacted cylindrical theory of plectin is based on electron microscopy studies of the C-terminal tail of plectin which revealed a single globule 9 nm in diameter (Foisner and Wiche, 1987). While there is some evidence to support both theories there is currently no conclusive answer. This raises questions about what molecular mechanisms are involved in cytoskeletal binding and what governs cytoskeletal preferences. It was hoped that a structural approach would help answer these questions since the only X-ray crystallographic information available to date is on desmoplakin PRDs. A PhD thesis by Radavan Spurny reported an inability to crystallize plectin PRDs 4 and 5 (homologous to PRD B of desmoplakin), which may have been compromised by inclusion of flexible His tags and partial linker domains (Spurny, 2008). Hence there is still a need to characterise PRD-ligand and linker structures and interactions in order to address the opposing hypotheses of desmoplakin and plectin C-terminal tail assembly mechanisms. Overall, several questions remain unanswered including:

a) What key structural features are conserved and divergent between different PRDs?
b) Is the molecular surface and groove implicated in vimentin binding conserved between different PRDs and how does this contribute to specific binding of cytoskeletal ligands?
c) What is the behaviour of the PRDs in solution?

Less attention has been given to the interactions of cytoskeletal proteins with the plakin family members envoplakin and periplakin. This study aimed to determine why envoplakin possesses only a single PRD, whereas desmoplakin has 3, plectin has 6 and periplakin has none (see

figure 1.3). It was intended that by characterizing the structure of the envoplakin PRD the project would shed light on the mechanisms involved in cytoskeletal binding and whether certain specific PRDs were functionally unique. By understanding these mechanisms it would give a clearer picture as to how the desmosome and IF system are linked. Therefore it would allow for a better understanding of the molecular details of the mechanisms could lead to the cardio-cutaneous diseases of the desmosome.

This chapter will therefore discuss the biophysical properties of the envoplakin and desmoplakin PRDs and their binding propensity to vimentin.

5.1 Construct design, expression and purification of the envoplakin PRD

5.11 Construct design

The extreme C-terminal tail of the envoplakin sequence is known to harbour the envoplakin PRD. This sequence was aligned with the homologous solved structure of the desmoplakin PRD C domain (PDB code: 1LM5). As observed in Figure 5.1 the PRD C of desmoplakin and envoplakin share 45% sequence identity and 65% sequence homology. Based on this alignment 3 envoplakin PRD constructs were created; a long, medium and small construct. The long construct (F1812-R2030) was based on the exact sequence of the construct used to crystallize the desmoplakin PRD C (Choi et al, 2002). The medium construct (D1822-A2014) removed sequences at the N and C termini thought to be least homologous to the desmoplakin PRD C and therefore not critical to the core PRD structure. The small construct was based on the premise that the extra residues located in the medium construct at the N and C termini did not contribute to the PR motifs. The small construct therefore represented the smallest of all constructs with respect to sequences not contributing to a canonical PR motif.

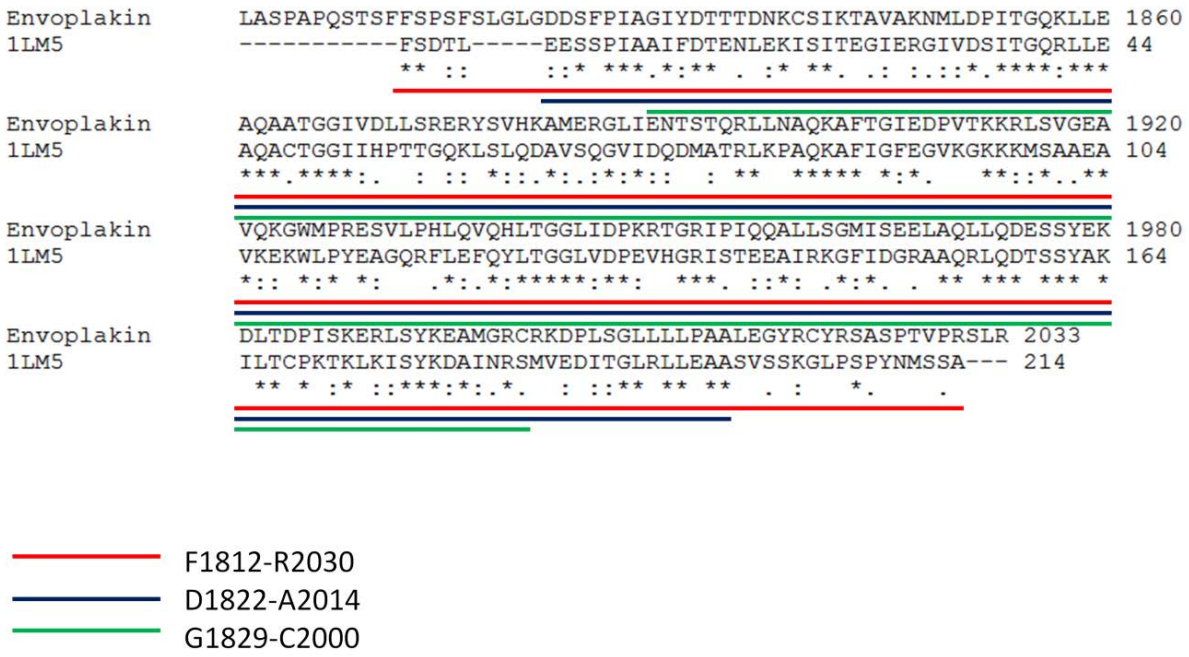


Figure 5.11 Construct design of the envoplakin PRD

The desmoplakin PRD C sequence which was used to crystallize the construct (1LM5) (Choi et al, 2002) was aligned against that of the full human envoplakin sequence. The alignment was intended to determine similarities/divergence between the two sequences. This alignment was also used to aid in the design of three envoplakin PRD constructs for structural studies. The final long, medium and small constructs are underlined in red, blue, and green respectively.

5.12 Expression and purification of envoplakin PRD constructs

Each of the three envoplakin PRD constructs were subjected to standard cleavable his-tagged purification methods. The construct that provided the highest yields of protein and purity were selected for subsequent large-scale production of envoplakin PRD for downstream biophysical analyses.

Firstly, the envoplakin PRD small construct (incorporating residues 1829 to 2000; 22kDa, including a vector derived cleavable his-tag) was investigated. Purification beyond Ni^{2+}-NTA chromatography was not possible owing to the fact that multiple species were observed following analysis by SDS-PAGE (Figure 5.21A). In fact a variety of protein species were present, one at the correct molecular weight for the small construct and multiple others at slightly lower molecular weights of equal quantity, suggesting non-specific cleavage or spontaneous degradation. In an attempt to remove the majority of contaminants, it might have been possible to subject the Ni-NTA eluted samples to additional purification steps such as size exclusion or ion exchange chromatography. However, this was deemed impractical as use of such conventional purification steps would most likely adversely affect the final protein yield and prevent subsequent biophysical analysis. Hence, the short construct was considered unsuitable for structural studies and was not utilised further.

Secondly, the envoplakin long construct was investigated (incorporating residues 1812 to 2030; 27 kDa including a vector derived cleavable his tag) for its ability to produce high levels of pure and stable protein. However, SDS-PAGE analysis of the long construct following purification with Ni-NTA and size exclusion chromatography revealed that the protein was either degrading spontaneously or cleaved by a non-specific protease post elution from the size exclusion column (Figure 5.21B). Interestingly, the degraded or smaller fragment appeared to be similar in size to the medium construct, suggesting that the termini of the protein were susceptible to cleavage.

As a result, it was decided that the envoplakin long construct was not viable for further biophysical analysis.

Lastly, the envoplakin PRD medium construct was investigated (incorporating residues 1822 to 2014; 24 kDa including a vector derived cleavable his tag). SDS-PAGE analysis following Ni^{2+}-NTA purification revealed a high quantity of relatively pure protein (Figure 5.22A). This suggests that the medium construct is more stable than the small construct. The protein was incubated with TEV protease to cleave off the his-tag and loaded onto the Ni^{2+}-NTA column. The cleaved His-tag and TEV protease were bound to the Ni^{2+} column allowing the protein to be collected in the flow through and wash steps. Finally, the protein was subjected to size exclusion chromatography on a S75 column (Figure 5.22B). The protein eluted at 12.6 ml, and its purity was gauged as ~95% by SDS-PAGE analysis (Figure 5.22C). This construct was regarded as most suitable for biophysical analysis. Therefore the envoplakin medium construct was the optimal construct for further biophysical analysis and from hereon will be referred to as the envoplakin PRD. As a control, similar strategies were employed to purify desmoplakin PRD C using previously described methods (Choi et al, 2002) (data not shown).

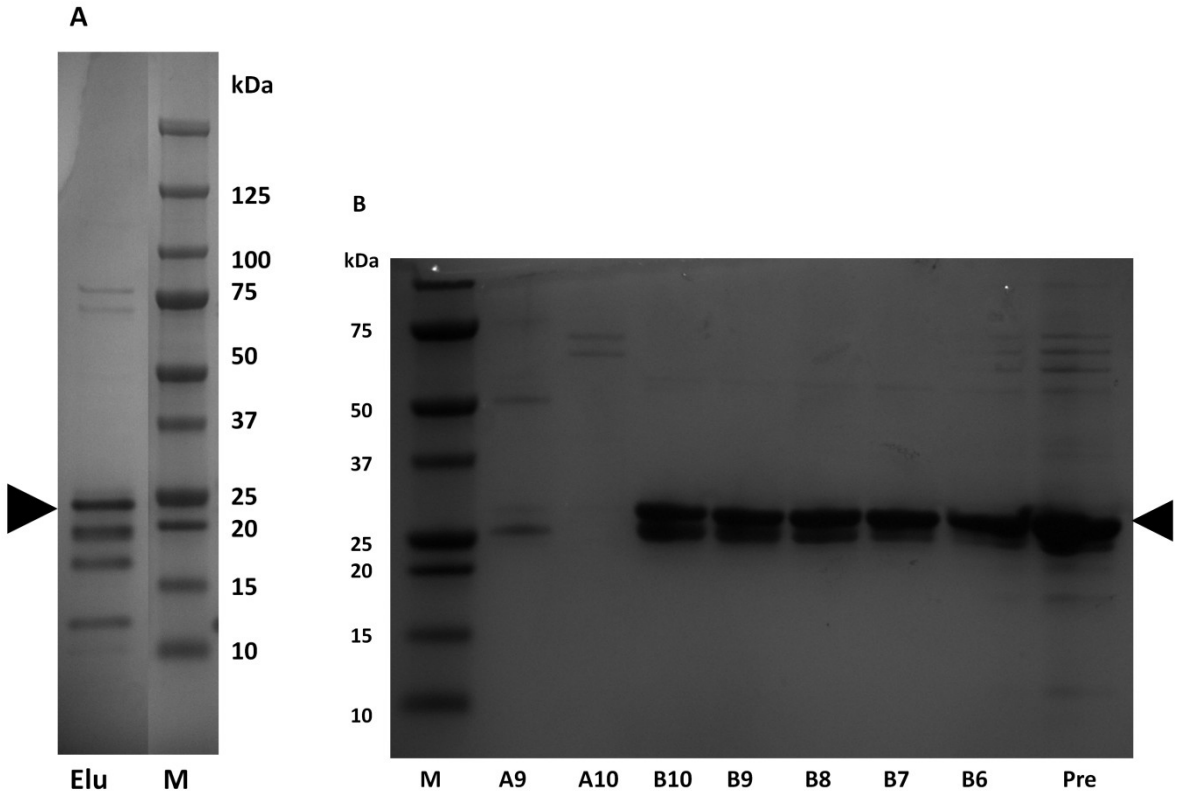


Figure 5.12 Purification of the short and long envoplakin PRD constructs

A: single-step His-trap purification of the small envoplakin PRD construct. Elu: Elution, M: marker. An arrowhead indicates the position of the envoplakin small construct (with the his tag retained) at ~22 kDa. B: Post size exclusion purification of the long envoplakin PRD. Pre: envoplakin long construct (post his cleavage) prior to size exclusion purification, A9-B6: fractions eluted from the S75 size exclusion column, M: marker. The arrowhead indicates the envoplakin long construct at approximately 27 kDa. The data shows that the envoplakin short and long constructs are susceptible to spontaneous cleavage and/or degradation.

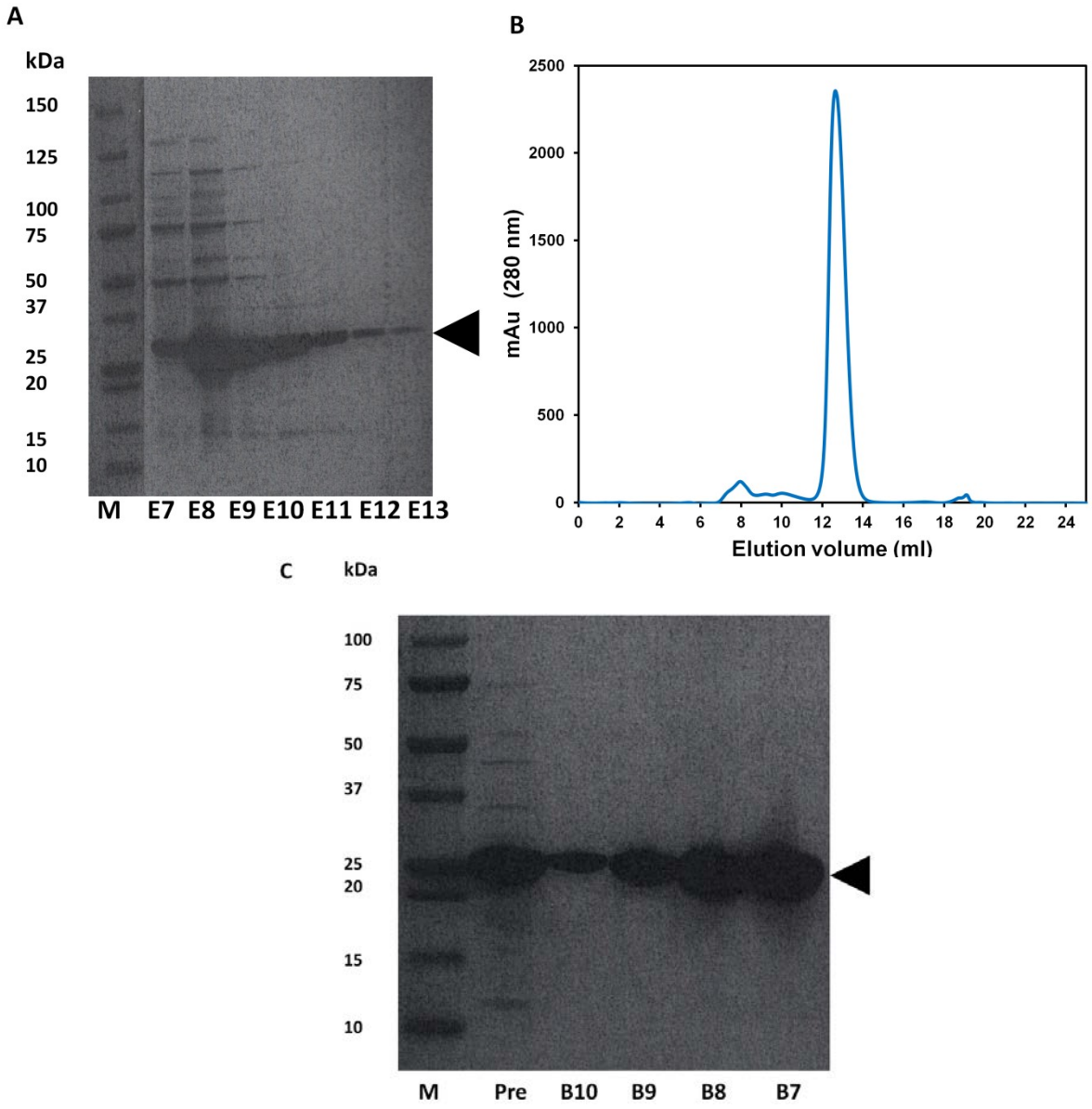


Figure 5.13 Purification of the medium envoplakin PRD construct

A: His-trap purification of the medium envoplakin PRD construct (pre his cleavage eluted in 1ml fractions). M; marker, E7-E13; elution fraction number. B: A S75 size exclusion profile of the medium envoplakin PRD construct (post his tag cleavage; ~24 kDa) eluted in 20 mM HEPES (pH 7.5) and 50 mM NaCl at 12.6 ml. C: SDS-PAGE analysis of envoplakin PRD fractions following purification by size exclusion chromatography. M: marker, Pre: pre gel filtration sample, B10-B7: fractions eluted from the S75 size exclusion column, Arrowheads indicate the position at which the envoplakin medium construct migrates. The final envoplakin medium construct was pure and deemed suitable for further biophysical analysis.

5.2 AUC of the envoplakin and desmoplakin C PRD

Although both known crystal structures of the PRDs of desmoplakin B and C crystallized as dimers (Choi et al, 2002), it was unclear whether the dimer represented a physiologically relevant species. Indeed, to date a comprehensive analysis of the oligomeric state of any PRD in solution is undefined and such studies may provide insights into the molecular function of PRDs. Intriguingly, the desmoplakin PRD B was observed as a dimer in solutions up to 100 mM NaCl by size exclusion chromatography (Choi et al, 2002). However at salt concentrations above 100 mM a monomeric species was observed (Choi et al, 2002). This suggested the possibility that under physiologically relevant conditions, PRDs may exist as dimers in solution. To address this issue the envoplakin PRD (his-tag cleaved) and desmoplakin PRD C (his-tag retained) were subjected to sedimentation velocity AUC to assess their oligomeric state in solution. Analysis by SEDFIT (Brown and Schuck, 2006) demonstrated a single species at 18.1 kDa for the envoplakin PRD and a single species at 21.3 kDa for desmoplakin PRD C. This is in good agreement of the sequence derived molecular weight determinations of 22.0 kDa and 26.0 kDa for envoplakin and desmoplakin PRDs, respectively and clearly demonstrates each protein is monomeric in solution.

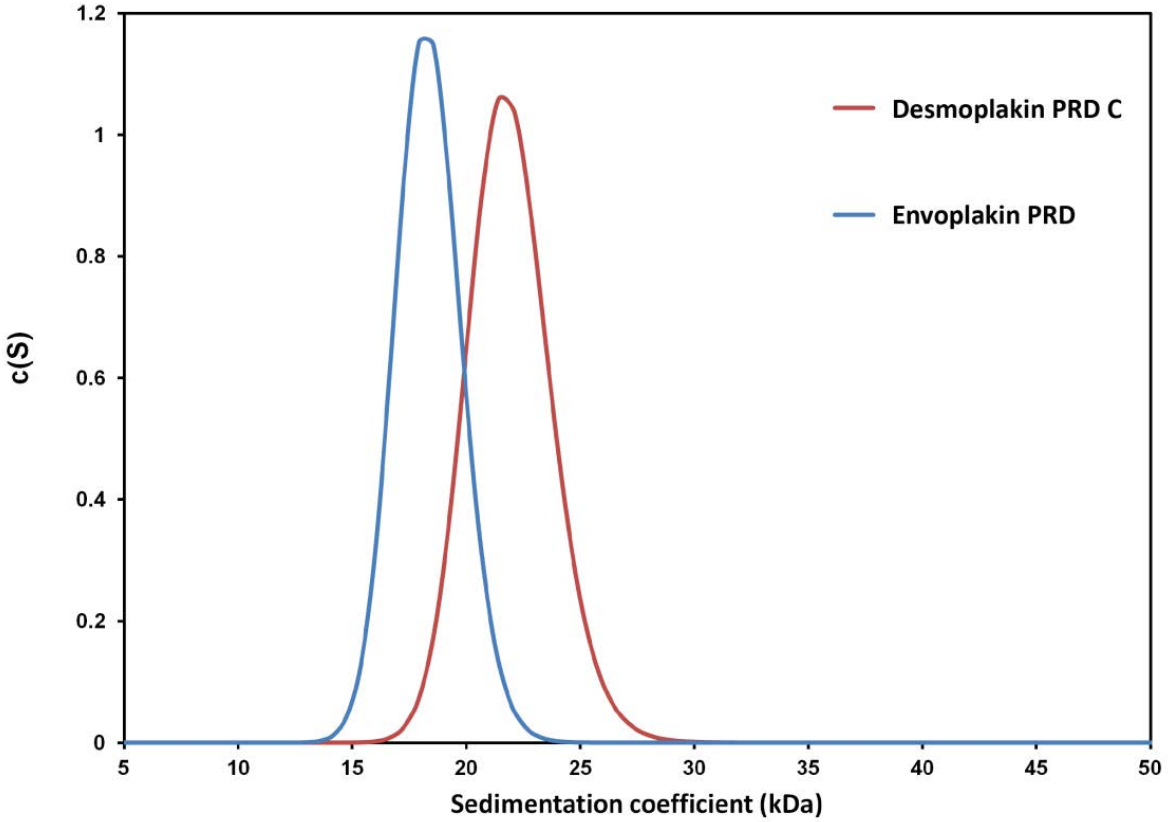


Figure 5.2 Analysis of the envoplakin and desmoplakin PRD C by AUC

The envoplakin PRD (his-tag cleaved) and desmoplakin PRD C (his-tag retained) were subjected to AUC analysis to gauge their oligomeric states in solution. The AUC experiments were carried out in 20 mM Hepes (pH 7.5) and 100 mM NaCl at 0.1 mg/ml using a Beckman XL-1 at 20000 rpm at 20 °C. Sedimentation coefficient profiles were converted to estimated molecular weights. Envoplakin and desmoplakin PRDs reveal single peaks at 18.1 kDa (blue) and 21.3 kDa (red), respectively. Although the values observed deviate slightly from the sequence derived molecular weights of envoplakin (22 kDa) and desmoplakin (26 kD) PRDs, both are within acceptable error ranges for AUC of a monomeric protein in solution.

5.3 Thermofluor thermal stability assay of the envoplakin PRD

The envoplakin PRD was subjected to thermofluor thermal stability assay to identify an optimal buffer condition that provided the highest thermal stability. This method also allows qualitative assessment of the foldedness of the protein as folded proteins typically give rise to sigmoidal curves. Unfortunately, no sigmoidal curves were observed for the envoplakin PRD in any of the conditions tested which included various buffers at different pHs, salts and additives. There are three possible explanations for the lack of a sigmoidal curve during the thermofluor thermal stability assay. Firstly, the envoplakin PRD may not possess a stable hydrophobic core and therefore no tertiary fold is present to unfold. Secondly, the PRD may include a hydrophobic patch on its surface which is spuriously binding the fluorescent dye. Finally, the protein sample may have a tendency to precipitate or aggregate in solution. Since the protein exhibits high sequence homology to a crystallized homologue, it is unlikely that the protein is unfolded. Moreover, based on the observation that the envoplakin PRD is very stable at high concentrations for several weeks at 4 °C, suggests that it is unlikely to precipitate or aggregate at the low concentrations used. Hence, the most likely scenario is that a surface exposed hydrophobic patch is sufficiently exposed, permitting the dye to bind envoplakin and giving rise to an anomalous signal.

5.4 2D-HSQC NMR analysis of the envoplakin PRD

Although the protein was expected to be folded the thermofluor results suggested that it could be unfolded. A more rigorous way of testing foldedness would be to use 2D-HSQC NMR analysis of ^{15}N labelled envoplakin PRD, as the peak dispersion represent a much more sensitive and detailed indicator of foldedness. Hence a ^{15}N labelled version of the envoplakin PRD was expressed and purified in the same manner as its unlabelled counterpart. No significant differences were observed in final purity or yield. As shown in Figure 5.3 a very well dispersed

HSQC is observed with a spread between 7-10 ppm which indicates its well folded nature. The data clearly show that the protein is folded and has a stable tertiary structure. If the protein was unfolded a spectrum similar to that observed for the desmoplakin SH3 domain (figure 3.71) would have been expected. Given the spread of the peaks, the spectrum is indicative of a mixed α-helical β-sheet containing protein, consistent with the secondary structural predictions of the envoplakin PRD. Collectively, this information is contrary to the results obtained from the thermofluor thermal stability assay which suggested it may be unfolded or aggregated. This highlights the power of NMR spectroscopy using HSQCs and the limitations of thermofluor in terms of gauging foldedness. Given the stability, yield and high dispersion of the protein by NMR a choice between either NMR assignment or X-ray crystallography was needed for further structural elucidation at atomic resolution. A realistic time-frame for full NMR assignment of the envoplakin PRD was estimated to be 1 year (Personal communication: Dr Mark Jeeves, 2011), which was unfortunately beyond the time constraints of this PhD study given the other ongoing experiments. Based on the tight time constraints, it was decided that although there was a higher element of risk, due in part to the inherent challenges of protein crystallization, the X-ray crystallography route was to be initiated first, with NMR spectra being used predominantly to monitor interactions.

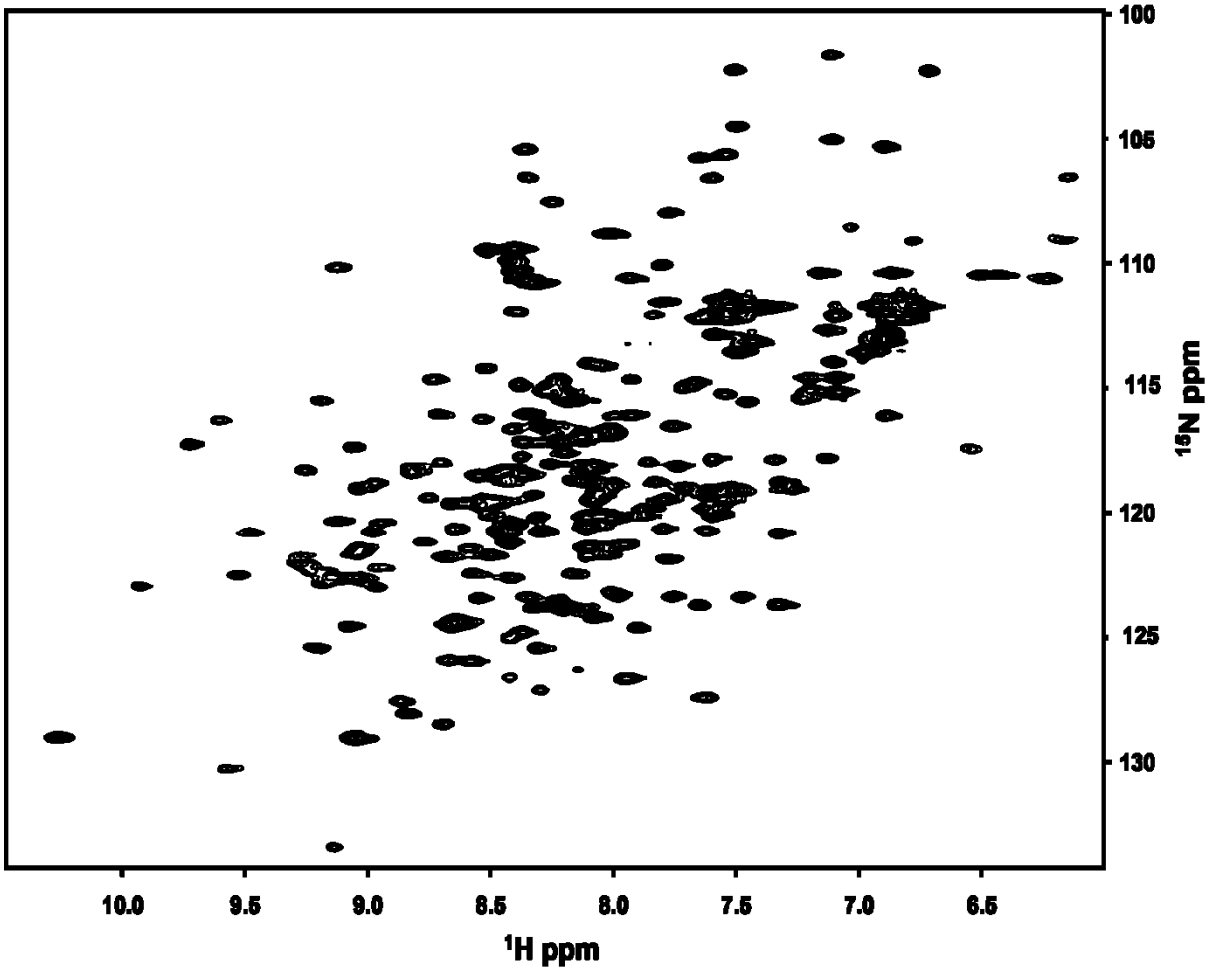


Figure 5.3 2D HSQC of the envoplakin PRD

^{15}N labelled envoplakin PRD was subjected to 2D HSQC analysis by NMR. This was conducted on an Agilent 600 MHz spectrometer at 298 K in 20mM Hepes (pH 7.2) and 100 mM NaCl. The data was processed using NMRPIPE and visualised (Delaglio et al., 1995). The HSQC spectrum displays well dispersed ^{1}H, ^{15}N peaks and is indicative of a folded protein.

5.5 Structural studies of the envoplakin PRD using X-ray Crystallography

The determination of the high resolution three dimensional structure of the sole envoplakin PRD was important in an attempt to address some of the key questions posed in the introduction of this chapter. These questions included whether the structure of the envoplakin PRD was similar to those already crystallized from desmoplakin and if the putative vimentin binding groove was retained. Purified envoplakin PRD was tested at a variety of concentrations in 20 mM Hepes (pH 7.5) and 50 mM NaCl in various crystallization screening buffers in order to crystallize the protein for eventual X-ray data collection and structure determination.

5.51 Crystallization of the envoplakin PRD using sparse matrix commercial screens

The initial crystallization hits were obtained with the hanging drop vapour diffusion method using the Mosquito nano-litre crystallisation robot. It was evident that the sparse matrix based commercial crystal screens yielded a large number of initial crystallisation leads for the envoplakin PRD (Table 5.1). Interestingly, the precipitant PEG 3350 was a common reagent in many of the preliminary crystallisation hits identified for envoplakin PRD. The most promising crystal hit was grown in with the Index condition (42) 0.1 M Bis Tris (pH 5.5), 25% PEG 3350 at 6.6 mg/ml (figure 5.41A). No crystals were observed in the negative control drops (in the absence of protein) equilibrated against index condition 42 (figure 5.41B). This suggested that the initial crystallisation hits were protein rather than salt derived and therefore worthy of further optimization for subsequent X-ray diffraction experiments.

Crystallization screen	Protein concentrations (mg/ml)	Hits
Index	6.6 and 5.0	5
JCSG+	11.0 and 8.5	3
PACT	11.0 and 6.0	13

Table 5.1 Preliminary crystallization screen hits

A general overview of the number of envoplakin PRD crystal hits observed from each crystallization commercial screen. A variety of protein concentrations were used to ascertain optimal protein concentrations. Crystals were grown at 298 K and appeared after 3 days.

5.52 Optimization of initial crystallization hits

The condition that yielded the preliminary hit was subjected to a variety of modifications to generate diffraction grade crystals. This is because the dimensions of the original crystals were too small and of too poor a morphology for high resolution X-ray data collection. The optimization strategy consisted of varying several factors. Initially, altering the concentrations of the constituent precipitant (PEG 3350 25%-10%) were trialled in order to reduce the number of nucleation sites. Although reducing the PEG 3350 concentrations from 25% to 23% yielded slightly larger sized crystals, they were still deemed unsuitable for X-ray diffraction experiments (Figure 5.41C). Additional optimization trials for the envoplakin PRD consisted of varying the pH of the crystallization buffer (pH 4-9) and further reducing the PEG 3350 concentration (from 23%-18%). This yielded single but larger crystals in each drop and were grown in the presence of 0.1 M Bis Tris (pH 6.5) and 20 % PEG 3350 with a protein concentration of 6.6 mg/ml (figure 5.41D).

Finally, large scale trials were performed during which the protein to crystallization buffer ratio was adjusted (0.5 µl-3.5 µl) as well as attempting the sitting drop method. Indeed, the most favourable crystallization condition was identified when utilising a 2:1 ratio of protein to

crystallisation buffer with the sitting drop method. The final optimised condition that reproducibly provided diffraction quality envoplakin PRD crystals consisted of 0.1M Bis Tris (pH 6.5), 20% PEG 3350 at 6.6 mg/ml after 3 days (figure 5.42).

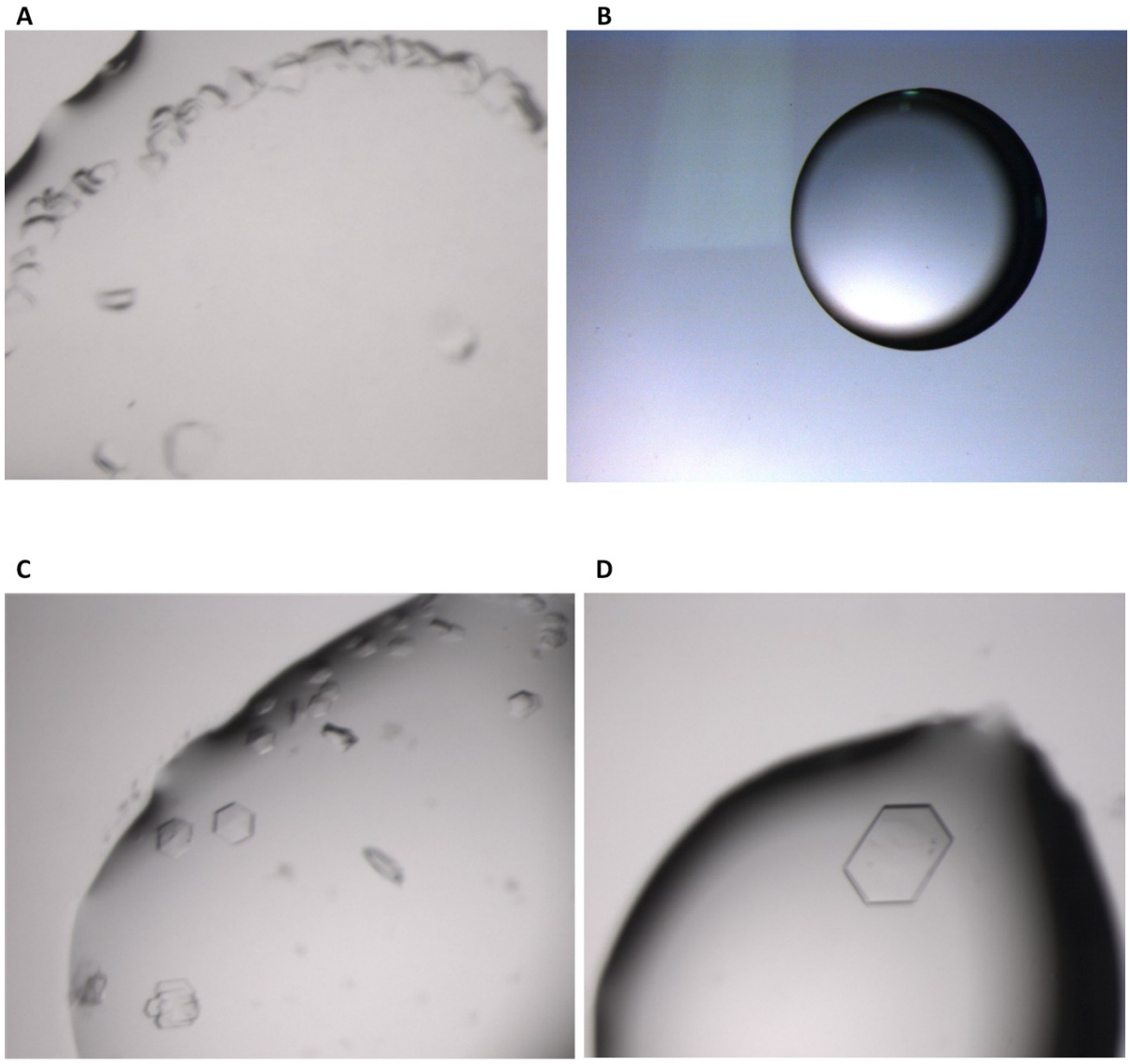


Figure 5.41 Preliminary and optimized hit for the envoplakin PRD

A: initial crystal hit obtained from the Index screen condition 42 (0.1 M Bis-Tris pH 5.5, 25% PEG 3350) at 6.6 mg/ml. B: Negative control with buffer alone (in the absence of protein) highlights no crystal formation suggesting that the crystals are derived from protein and not salt C: Optimized crystals grown in 0.1 M Bis Tris (pH 5.5) and 23% PEG 3350 at 6.6 mg/ml. Crystals were deemed to be 65 x 55 µm. D: Final optimized condition yielding a single relatively large crystal consisted of 0.1M Bis Tris (pH 6.5), 20% PEG 3350, at 6.6 mg/ml. Crystal was deemed to be 226 x 154 µm. All crystals were grown at 298 K and appeared after 3-5 days.

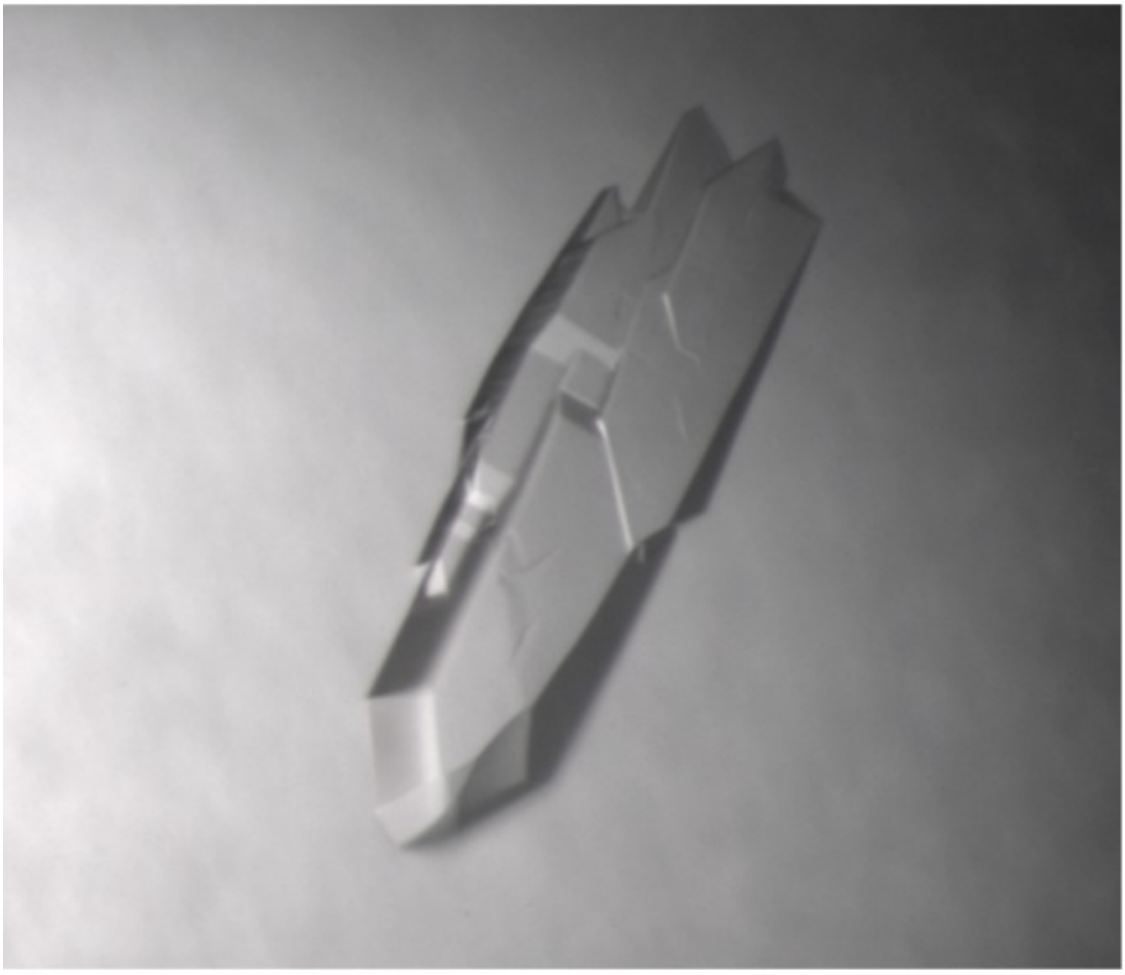


Figure 5.42 Crystals of the envoplakin PRD grown on a large scale

Crystals were grown using the sitting drop vapour diffusion method at a volume ratio of 2:1 (µl) protein to buffer in 0.1M Bis Tris (pH 6.5), 20% PEG 3350 at 6.6 mg/ml and appeared after 3 days. Crystal was deemed to be 650 x 250 µm.

5.53 Data collection and processing

Crystals were soaked in the reservoir crystallisation buffer supplemented with increasing concentrations of ethylene glycol (at 5, 10, 15 and 18%) as a cryoprotectant in preparation for X-ray data collection. These crystals were then flash cooled at 100 K under a nitrogen gas stream (Oxford Cryosystems). X-ray data was collected to 1.6 Å resolution (Figure 5.43) by Dr Fiyaz Mohammed on an in-house MicroMax 007HF rotating anode X-ray generator (Rigaku) using a Saturn CCD detector (Macromolecular X-ray facility, University of Birmingham). The envoplakin PRD crystallises in the orthorhombic space group $P2_12_12_1$ with unit cell parameters (a = 42.4 Å, b= 68.8 Å, c = 112 Å, α = 90=β = γ = 90 °). Envoplakin PRD diffraction data were integrated, scaled and merged using programs of the XDS suite (Kabsch, 2010). The relevant data processing statistics are listed in Table 5.2.

Resolution Range (Å)	**20-1.6**
Total Reflections	**540914 (19433)**
Unique reflections	**78068 (94130**
Completeness (%)	**92.9 (97.2)**
R_{sym} (%)	**5.8 (38.5)**
I/σ(I)	**27.9 (2.1)**
Multiplicity	**6.9 (2.1)**

Table 5.2 Data processing statistics for the envoplakin PRD

X-ray data processing statistics of the envoplakin PRD crystal. Numbers in parentheses represent data from the highest resolution shell.

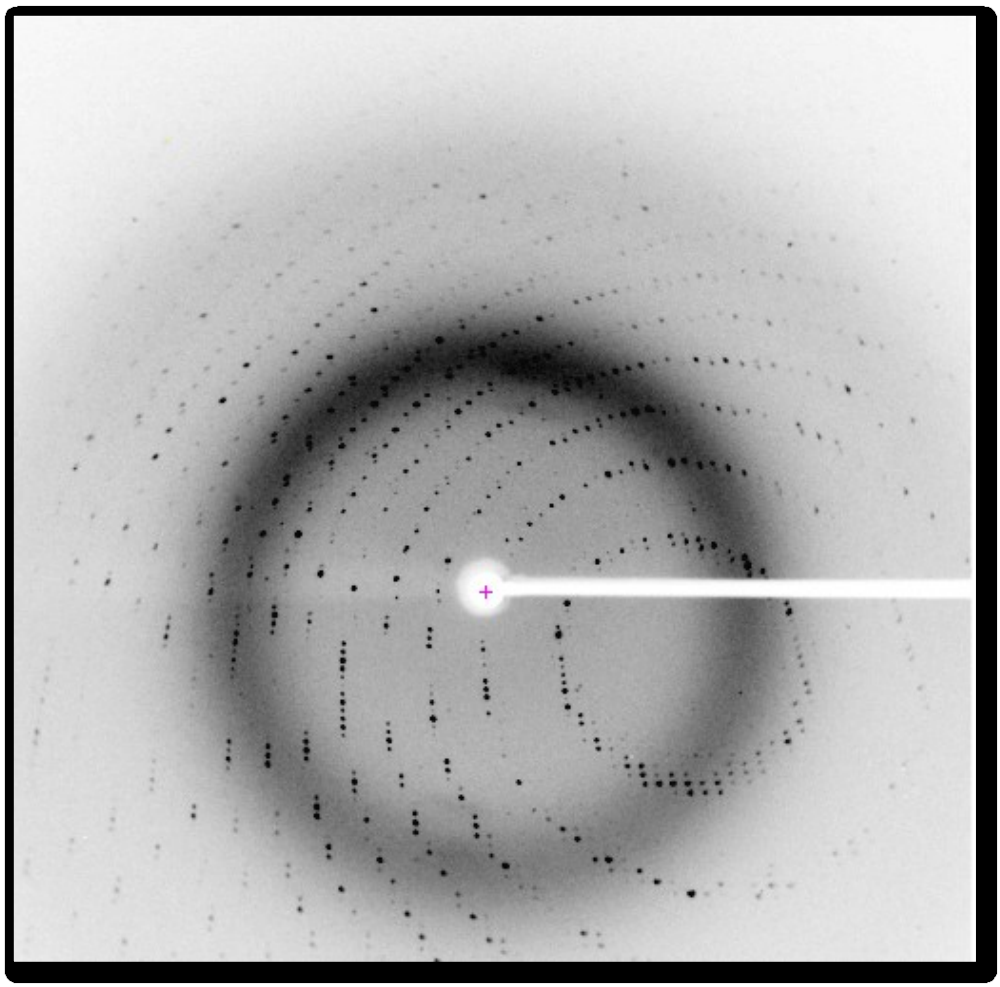


Figure 5.43 Diffraction pattern for the envoplakin PRD crystal

A 0.5° oscillation image of the envoplakin PRD crystal collected at 100 K using the University of Birmingham Macromolecular X-ray diffraction facility. The crystal diffracted X-rays to a resolution of 1.6 Å.

5.54 Structure determination using Molecular Replacement

Assuming the envoplakin PRD crystal contains two molecules per asymmetric unit, a solvent content of 32% was calculated, which is similar to that of the desmoplakin PRDs (Choi et al, 2002). Using the program PHENIX (Adams et al., 2010), a variety of search models were used to calculate initial phase information by automated molecular replacement. Models included full length desmoplakin PRD B (PDB ID code; 1LM7), desmoplakin PRD C (PDB ID code: 1LM5) and a sequence derived model of the envoplakin PRD calculated by I-TASSER (Roy et al., 2010). Surprisingly, despite a very high sequence identity between the desmoplakin and envoplakin PRDs (45%), no rotation function peaks were observed. Therefore, the search models were improved by truncating the termini as well as deleting the internal flexible loops. Unfortunately, none of these truncated models provided any rotation function solutions, suggesting that molecular replacement could not be used to acquire initial phase information.

5.55 Single wavelength anomalous dispersion (SAD)

Due to the failure of molecular replacement to obtain initial phase information, an alternative method was required to calculate phases and hence determine the structure of the envoplakin PRD. One approach was to use multi-wavelength anomalous dispersion (MAD), in which selenium substitutes for methionine. However, this required expressing the protein in an *E.coli* methionine auxotroph strain, purifying and crystallising the protein and collecting X-ray data through the use of a tuneable X-ray source e.g. synchrotron radiation. Due to time constraints the MAD approach was not feasible. Therefore, an alternative strategy was to attempt 'in house' X-ray data collection on envoplakin crystals pre-soaked or co-crystallised in high concentrations of sodium iodide and determine the initial phase information using SAD (Dauter et al., 2000). This strategy works on the premise that iodide ions bind to positively charged residues on the protein and are located easily due to their large anomalous signal at Cu Kα radiation

(wavelength = 1.5418 Å). In addition, the SAD method has the extra benefit of being relatively inexpensive, rapid and less toxic compared to other techniques such as MAD and multiple isomorphous replacement.

Using the previously described crystallization condition, attempts were undertaken to co-crystallise envoplakin PRD with sodium iodide. Unfortunately, despite constantly striving to optimise the concentration of sodium iodide added to the crystallisation buffer, none of these conditions generated envoplakin PRD crystals. Therefore, an alternative approach was employed which involved soaking existing envoplakin PRD crystals in mother liquor supplemented with various concentrations of sodium iodide (0.1-1 M) and cryoprotectant. Initial soak attempts with higher concentrations of sodium iodide (0.6 M) led to the envoplakin crystals dissolving (2/2 crystals). Therefore, the soaking steps were repeated at lower iodide ion concentrations. Through trial and error the maximum sodium iodide concentration was optimized to 0.45 M allowing a number of crystals to be soaked for up to 45 minutes where 3/5 were soaked successfully and vitrified for subsequent data collection. The optimised soaking procedure for generating iodine derivatised envoplakin PRD crystals are listed in Table 5.3

Soak (#)	Derivative	Cryoprotectant	Time (m)
1	Sodium iodide (0.20 M)	Ethylene glycol (5%)	10
2	Sodium iodide (0.30 M)	Ethylene glycol (10%)	15
3	Sodium iodide (0.40 M)	Ethylene glycol (15%)	30
4	Sodium iodide (0.45 M)	Ethylene glycol (18%)	45

Table 5.3 Soaking procedure used to generate iodine derivatised crystal of envoplakin PRD.

To maximize uptake of sodium iodide into the crystal, the optimal soaking procedure is given. Mother liquor of 0.1 M Bis tris (pH 6.5) and 20 % PEG 3350 was supplemented with increasing concentrations of derivative (0.2-0.45 M) and cryoprotectant (5-18%).

X-ray data collection for the envoplakin derivatised crystal was performed as described previously for the native crystal in section 5.53. X-ray data was collected to 2.3 Å resolution (figure 5.44). The derivatised envoplakin crystal was isomorphous to the native crystal belonging to the orthorhombic space group $P2_12_12_1$ with unit cell parameters (a = 42.3 Å, b= 68.5 Å, c = 113.7 Å, α = 90 =β = γ = 90 °). This suggested that soaking of the heavy atom iodide had little effect on the internal crystal packing and hence space group.

Resolution Limit (Å)	**20-2.3**
Total Reflections	**743113 (129679)**
Unique reflections	**31129 (5527)**
Completeness (%)	**99.8 (99.9)**
Rsym (%)	**10.6 (37.9)**
I/σI	**32 (10.9)**
Multiplicity	**23.9 (23.5)**
SigAno	**1.7 (1)**

Table 5.4 Data processing statistics for the iodine derivatised crystal of envoplakin PRD.

Numbers in parentheses represent statistics from the highest resolution shell.

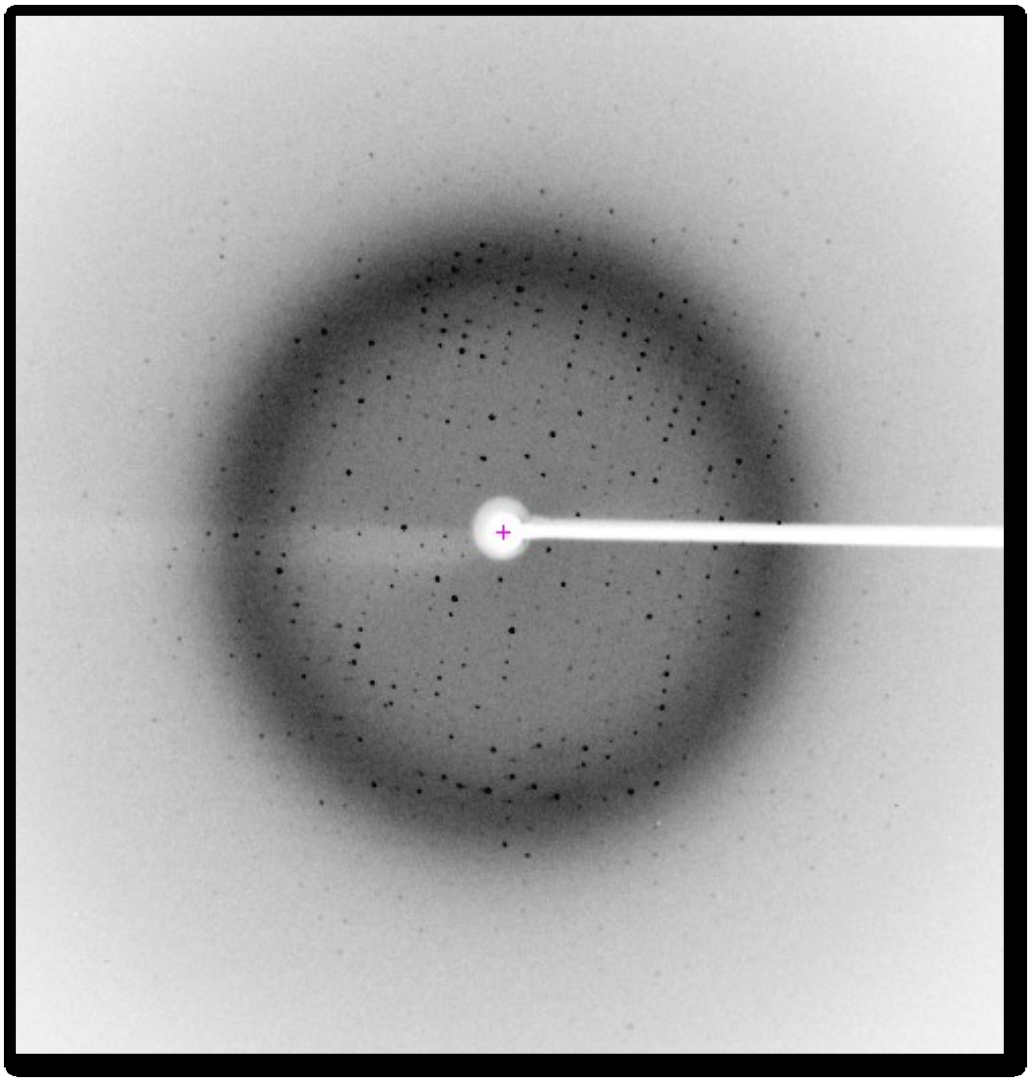


Figure 5.44 Diffraction pattern for the iodine derivatised crystal of envoplakin PRD.

A 0.5° oscillation image of the envoplakin derivatised crystals collected at 100 K using the University of Birmingham X-ray facility. The crystal diffracted X-rays to a resolution of 2.3 Å. The native envopakin crystal was pre-soaked in 0.45 M sodium iodide and 18% ethylene glycol for up to 45 minutes.

5.56 Structure determination of the envoplakin PRD

Since the phase information could not be acquired by molecular replacement techniques, SAD phasing using iodide ions was used. To acquire the phase information, firstly the iodide ions must be located. Since these steps were carried out by Dr Scott White (University of Birmingham) only a very brief description will be given. As mentioned earlier, data sets were collected in house using a Rigaku 007-HF X-ray generator and a Saturn CCD detector. Data were reduced with XDS (Kabsch, 2010) and iodide sites were located using PHENIX (Adams et al., 2010). The anomalous substructure was refined and extended, and phases were estimated using PHASER (McCoy et al., 2007) followed by density improvement in RESOLVE. Initial models were built using automated building in AUTOBUILD (Terwilliger et al., 2007), followed by model extension/rebuilding in PHENIX. The final model was produced after several iterative rounds of manual re-building in COOT (Emsley et al., 2010) and refinement in PHENIX. An example of the quality of electron density is shown in figure 5.45. Final refinement statistics are listed in Table 5.5. The stereochemical quality of the protein structure was verified using the program PROCHECK (Laskowski, 1993) and demonstrates that the bond lengths and angles are acceptable (Table 5.5). Furthermore, all non-glycine residues are found in the allowed regions of the Ramachandran plot (figure 5.46). The final model encompasses two molecules in the asymmetric unit; residues 1822-2014 in molecule A (and 2 vector derived amino acids at the N-terminal end) and residues 1822-2014 in molecule B.

Parameter	Value
Resolution (Å)	19.9-1.6
Total reflections used	41953
Free value test reflection set	4207
R_{cryst} (%)[a]	17.1
R_{free} (%)[b]	20.4
Protein residues	383
Water molecules	405
RMS deviations Bond lengths (Å) Bond angles (°)	 0.006 1.038
Ramachandran plot Most favoured (%) Additionally allowed (%) Generously allowed (%)	 94 5.4 0.6
Average B factors (Å^2) Main chain (A) Side chain (A) Overall (A) Main chain (B) Side chain (B) Overall (B)	 13.0 19.9 16.3 12.6 18.4 15.4

Table 5.5 Final refinement and stereochemical statistics for the envoplakin PRD

[a]R_{cryst} (F) = $\Sigma_h||F_{obs}(h)|$ - $|F_{calc}(h)||/\Sigma_h|F_{obs}(h)|$ and $|F_{calc}(h)|$ are the observed and calculated structure factor amplitudes for the h, k, l reflection. [b]R_{free} is calculated over reflections in a test set not included in atomic refinement.

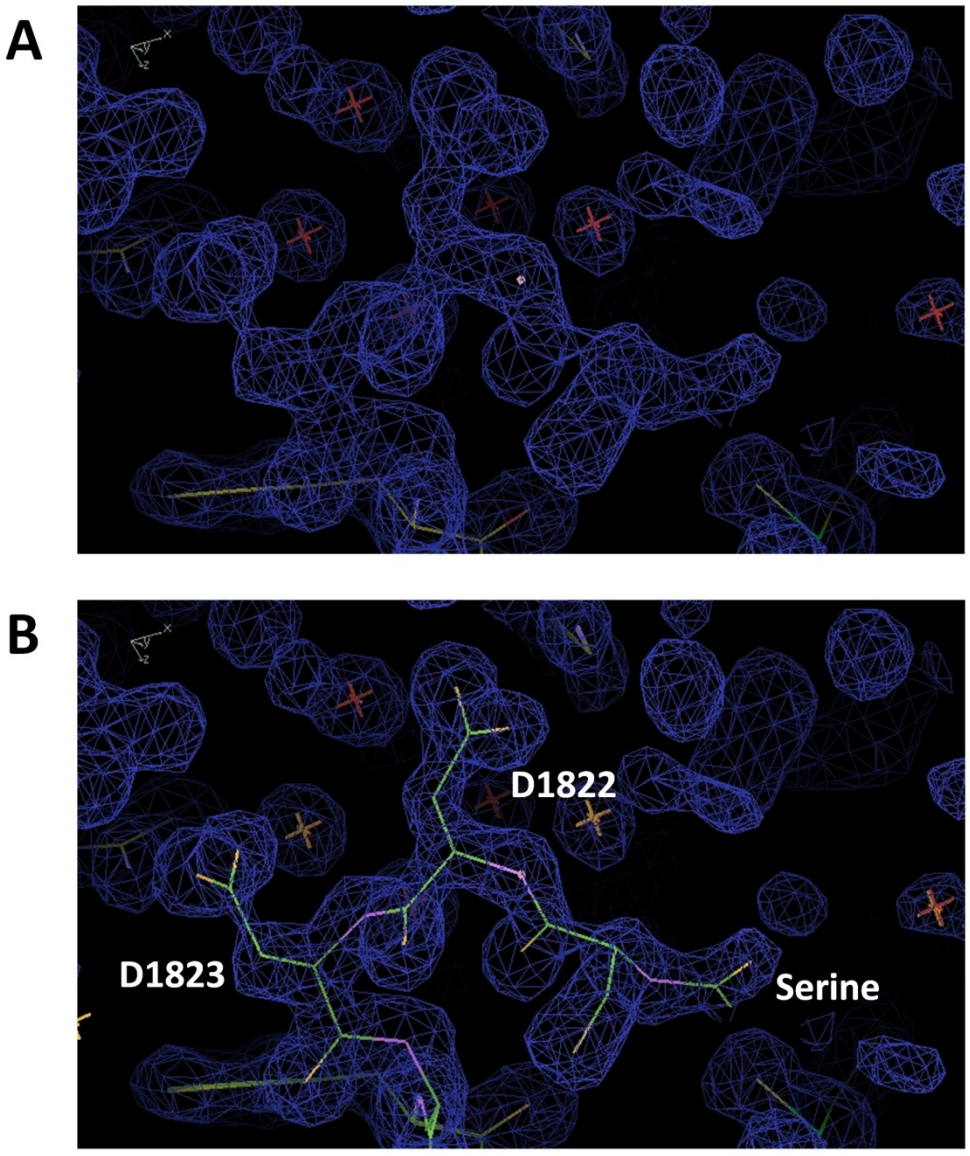


Figure 5.45 Electron density for the envoplakin PRD Molecule A.

2Fo-Fc electron density map contoured at 1σ (blue mesh) is shown for molecule A. A: Electron density at the extreme N-terminal end. B: The same view as A except with the protein chain model built in using COOT (Emsley et al., 2010). Highlighted are residues D1822 and D1823 at the N-terminal end of the envoplakin PRD and a vector derived serine. Spheres with crosshairs represent water molecules.

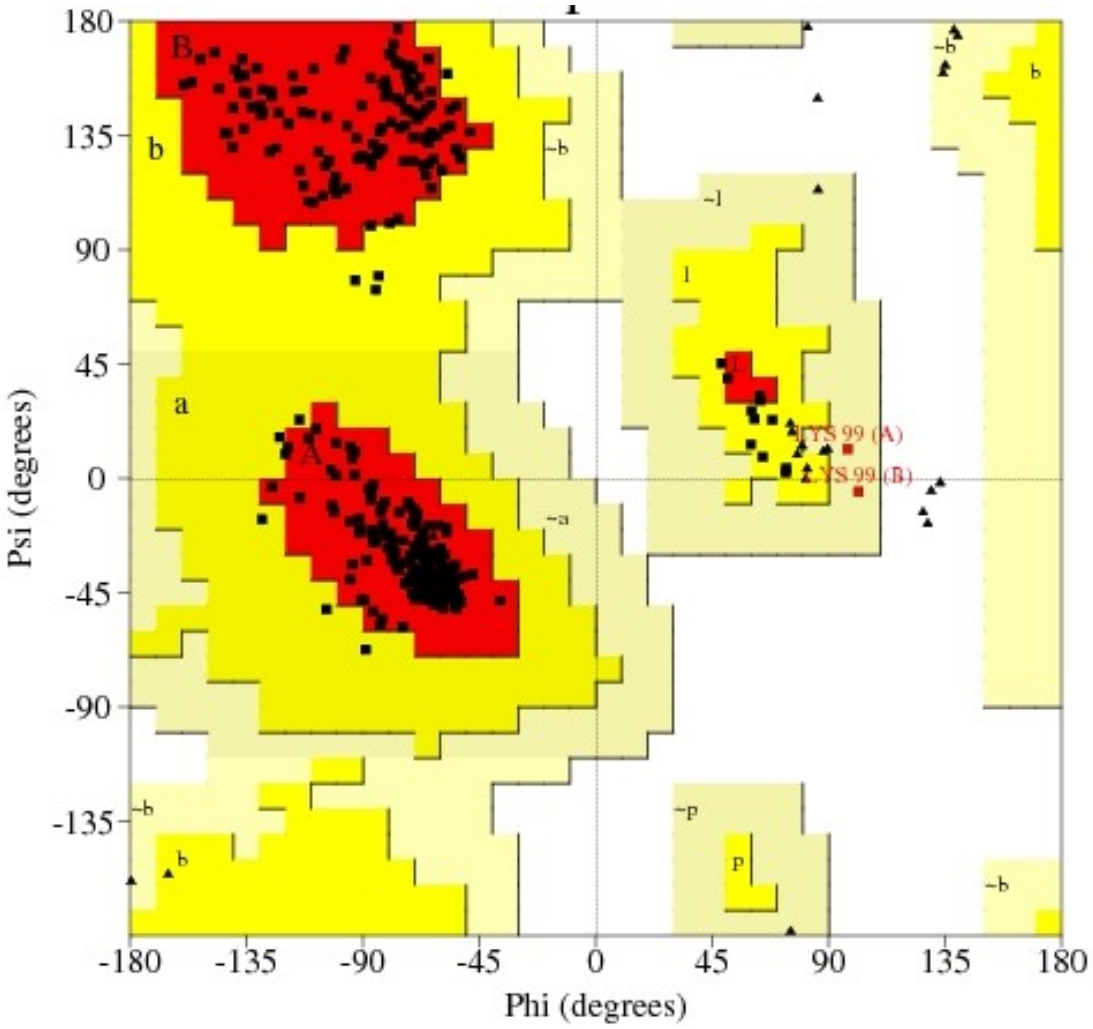


Figure 5.46 A Ramachandran plot of the envoplakin PRD structure

The Phi and Psi dihedral angles for the peptide bond of each residue in the envoplakin PRD structure are depicted as squares. Regions of the Ramachandran plot include the most favoured (red), additionally allowed (yellow), generously allowed (cream) and disallowed (white). Since glycine residues (depicted as triangles) exhibit additional flexibility and freedom they can occupy any region.

5.6 Crystal structure of the envoplakin PRD

5.61 Overall structure of the envoplakin PRD

The overall structure of the envoplakin PRD is largely similar to previously determined PRD structures (Choi et al, 2002), forming a globular structure with the same topology and fold (figure 5.51). The envoplakin PRD is characterised by a 38 amino acid motif (plakin repeat (PR)) that is repeated 4.5 times. Primary sequence alignments of the envoplakin PRs 1-5 reveals high degrees of conservation, as documented for desmoplakin PRDs B and C (Choi et al, 2002). The structural PR is comprised of an 11 amino acid beta hairpin (encompassing strands S1 and S2) followed by a pair of anti-parallel helices (designated H1 and H2), typically 8 and 12 residues long, respectively (although subtle variations are evident). The secondary structure elements of the envoplakin PRD for molecules A and B are listed in table 5.6.

PR1	**PR2**	**PR3**	**PR4**	**PR5**
1829-1832 S1	1868-1870 S1	1906-1908 S1	1944-1954 L^{+}	1982-1984 S1
1833-1836 L	1871-1874 L	1909-1911 L	1955-1962 H1	1985-1987 L
1837-1839 S2	1875-1877 S2	1912-1915 S2	1963-1964 L	1988-1991 S2
1840-1848 H1	1878-1886 H1	1916-1924 H1	1965-1974 H2	1992-2000 H1
1849-1850 L	1887-1892 L*	1925-1926 L	1974-1978 3/10 helix	
1851-1866 H2	1893-1905 H2	1927-1942 H2	1979-1981 L	
1867 L		1943 L		

Table 5.6 Secondary structure elements of the envoplakin PRD

Secondary structure elements from the 38 amino acid motif of each PR were determined using the structure of the envoplakin PRD. S1: strand 1, S2: strand 2, H1: helix 1, H2: Helix 2, L: loop. * indicates a longer loop than usually observed between H1 and H2 of PR2 suggesting that H2 is unusually shorter than a typical H2. $^{+}$ indicates that the S1+loop+S2 arrangement (β-hairpin forming regions) is replaced with a random coil in PR4. Furthermore, another novel feature of PR4 is the presence of 3/10 helix beyond H2.

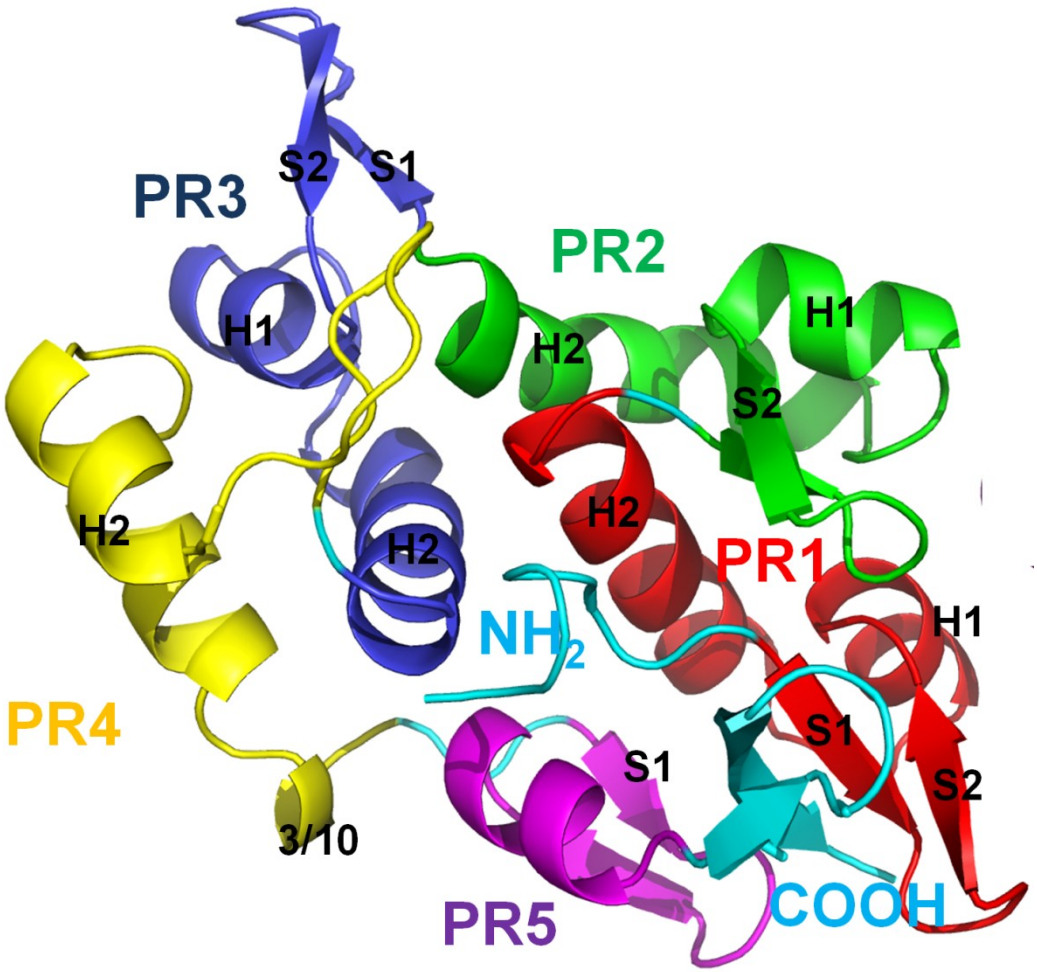


Figure 5.51 Crystal structure of the envoplakin PRD Molecule A

The final refined crystal structure of the envoplakin PRD visualised by PYMOL. The PR motifs are colour coded as follows: PR1: red, PR2: green, PR3: dark blue, PR4: yellow, PR5: magenta. S1: stand 1, S2: strand 2, H1: helix 1, H2: helix 2, 3/10: 3/10 helix only observed in PR4. The N and C termini do not contribute to a PR motif and are shown in cyan.

5.62 Assessing the possibility of envoplakin PRD dimerization – an academic exercise

Since two molecules are found in the asymmetric unit (figure 5.52A), the structure raised the possibility that the crystal dimer may represent a physiologically relevant form of the envoplakin PRD. Indeed, previous desmoplakin PRD structures, also determined by X-ray crystallography, demonstrated dimeric species within the asymmetric unit (Choi et al, 2002). However in the earlier cystallographic study from Weis' group, the physiological relevance of the distinct crystal dimer was not discussed, suggesting that the lattice contact may be an artefact of the crystal packing rather than representing an interaction surface. Nevertheless, the envoplakin PRD dimer was subjected to a series of structural analyses to clarify whether this was a bona fide functional subunit-subunit interface or a non-specific interaction.

Firstly, the number and size of inter-chain contacts, usually measured by the buried surface area were calculated, to discriminate between homodimeric and monomeric proteins in the crystalline state. For the envoplakin PRD dimer interface, a buried surface area of 589.2 $Å^2$ was calculated using PISA (Krissinel and Henrick, 2007), a relatively less extensive contact area relative to the PRD C (673.8 $Å^2$) of desmoplakin. These surface areas are considerably below values that are considered physiologically relevant (Krissinel and Henrick, 2007). Upon closer inspection, the envoplakin dimer interface is predominantly stabilised by a series of polar interactions that are listed in Table 5.7.

Molecule A	Location	Molecule B	Location	Distance
$K1847^{N}$	PR1 (H1)	$S1824^{OG}$	(*)	3.08
$N1837^{OD1}$	PR1 (L)	$E1975^{OE1}$	PR4 (H2)	3.14
$N1837^{ND2}$	PR1 (L)	$K1994^{NZ}$	PR5 (H1)	3.05
$K1842^{NZ}$	PR1 (H1)	$G1998^{O}$	PR5 (H1)	3.07
$D1870^{OD1}$	PR2 (S1)	$R1999^{NH1}$	PR5 (H1)	3.17
$D1870^{OD2}$	PR2 (S1)	$R1999^{NH2}$	PR5 (H1)	2.89
$K1842^{NZ}$	PR1 (H1)	$C2000^{O}$	PR5 (H1)	2.95

Table 5.7 Envoplakin PRD residues involved in stabilising the dimer interface.

The side chain atom that is involved in forming molecular contacts is shown in superscript. The majority of the residues that contribute to the dimer interface are located in PR1 and PR2 in molecule A and PR5 in molecule B. * denotes that S1824 is not associated with the PR motif.

If dimerization is important for the function of PRD molecules, one might expect that interface residues have evolved to optimize this interaction and these residues are conserved among PRD family members during evolution. Analysis of the desmoplakin PRD C dimer interface revealed that the majority of the residues were located in PR2 and PR3 of molecule A and PR4 of molecule B. In contrast, the envoplakin PRD dimer is stabilised by residues from the PRs 1 and 2 in molecule A and PR5 of molecule B (table 5.7). Given that the dimerization interface is not conserved between envoplakin and desmoplakin C PRDs it is unlikely that the dimer interface is physiologically relevant.

To assess this further the geometric parameter, shape complementarity (Sc), was determined to quantify the “goodness of fit” between the two envoplakin PRD surfaces. Large values of Sc usually indicate a close and good shape complementarity between surfaces (Lawrence and Colman, 1993). Sc values of 0.67 and 0.72 were calculated for the envoplakin and desmoplakin PRD C dimers, which was consistent with typical Sc values (0.65-0.80) determined for protein-

protein interfaces in oligomeric proteins. However, despite the relatively large Sc values, the biophysical data obtained for the envoplakin PRD and desmoplakin PRD C does not support the notion that the envoplakin PRD forms a physiologically relevant dimer. The data include a sharp peak eluting in the monomeric range from size exclusion chromatography (figure 5.13B), the AUC data which indicated a monomeric species (figure 5.2) as does analysis of SAXS results (discussed further in 5.7).

Superimposition of molecules A and B of the envoplakin PRD revealed that they are structurally similar as reflected by the Root Means Square Deviation (RMSD) of 1.04 Å (Figure 5.52B). The small differences observed between the two envoplakin molecules particularly in PR4, the C-terminus and the β-hairpin in PR3 were attributed to crystal packing effects. However, since those regions are divergent in the two asymmetric units (indicated in figure 5.52B) they may also be due to genuine flexibility or multiple conformers. Since the two envoplakin PRD molecules are almost identical (based on the RMSD value), molecule B (herein referred to as envoplakin PRD) was selected in the remaining structural analysis.

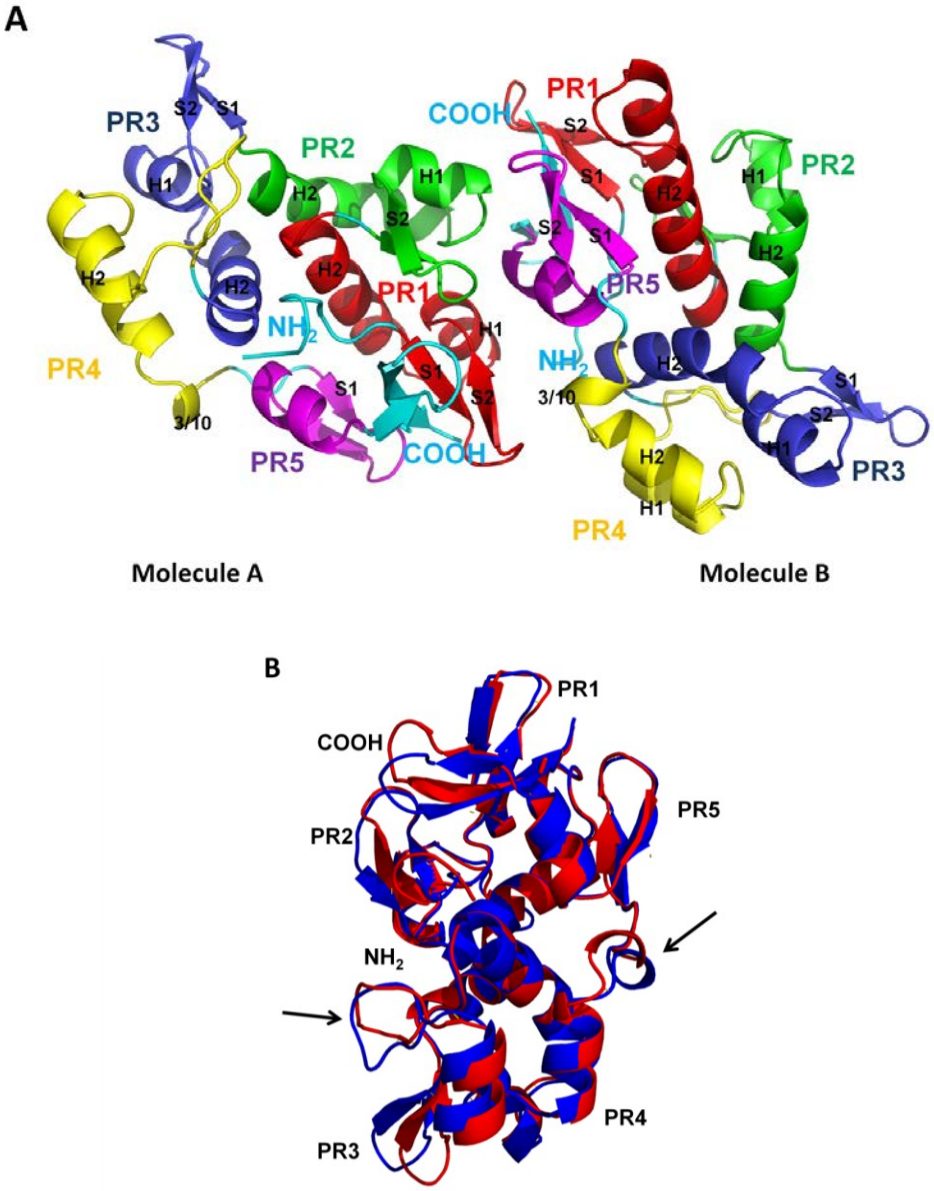


Figure 5.52 Crystal structure of the envoplakin PRD dimer

A) The crystal structure of the envoplakin PRD contains two molecules in the asymmetric unit (molecule A and molecule B). The PR motifs are colour coded as follows: PR1: red, PR2: green, PR3: dark blue, PR4: yellow, PR5: magenta. S1: stand 1, S2: strand 2, H1: helix 1, H2: helix 2, 3/10: 3/10 helix only observed in PR4.The crystal dimer interface is stabilised by residues protruding from PRs 1 and 2 in molecule A and PR5 of molecule B. B) Superimposition of envoplakin PRD molecule B (red) onto molecule A (blue). Black arrows indicate two of the regions (one in the β-hairpin of PR4 and the other at the end of H2 in PR4 – discussed further in 5.63) that are structurally divergent.

5.63 Comparison of the envoplakin PRD to PRD C of desmoplakin

Superimposition of envoplakin PRD to desmoplakin PRD C reveals some degree of heterogeneity (Figure 5.53A). When either molecule A or B were superimposed to the individual molecules of desmoplakin PRD C, RMSD values were calculated to be 1.68 and 1.95 Å for molecules A and B, respectively. This is not surprising given the high level of sequence similarity between the two proteins of 65% homology.

When the PR motifs from envoplakin and desmoplakin are superimposed on top of one another it reveals a striking similarity in orientation between all 5 PR motifs (figure 5.53B). These are also reflected in the RMSD values calculated for PRs1, PRs2, PRs3 and PRs4 (data not shown) between the PR motifs of envoplakin and desmoplakin. This implies envoplakin PR motifs adopt conformations similar to their desmoplakin PRD C counterparts. The hydrophobic core interactions that stabilise the PRs in desmoplakin PRD C are generally conserved in envoplakin. The residues involved in intra and interdomain interactions can be seen between the helices of H1 and H2 of envoplakin with the numerical positions within each PR motif conserved similarly to those seen for desmoplakin PRD (figures 5.54 and 5.55) Hydrophobic interactions between the extreme C and N terminals (absent in the envoplakin PRD small construct) and PR1 are also observed (figure 5.56), similar to observations in PRD C of desmoplakin, suggesting these interactions between these regions are required for folding or structural stability. Indeed, this may reflect why the envoplakin PRD small construct was insoluble/unstable.

Close inspection of the superimposition does reveal some differences between the PRDs of envoplakin and desmoplakin. Firstly, the majority of the differences are restricted to the flexible loops within each PR motif (figure 5.53A). These differences may provide a molecular rationale for the lack of success in determining initial phases with molecular replacement. Secondly, the PR2 motif is unique in that it exhibits a distinctive “kink” in the envoplakin PRD which is also

observed in PR2 of desmoplakin PRD C (figure 5.53B). Interestingly, the kink appears to be created by two different mechanisms in each PRD. In desmoplakin PRD C, a proline at position 2690 contributes to the kink at the beginning of H2, whereas the corresponding residue in envoplakin PRD consists of Asn 1898. Nevertheless, the envoplakin PRD contains a shorter H2 element due to a longer loop region proceeding H1, which typically separates the two anti-parallel α-helices (table 5.6). This may allow for greater flexibility in the backbone to adopt this conserved "kinked" conformation in envoplakin. Despite the differing mechanisms for generating the kink, the precise molecular role of the kink is undefined. Thirdly, in the envoplakin PRD PR4 contains a 3_{10} helix (an α-helix created by 3 residues per turn rather than 3.6) at the end of H2 – which itself is shortened. In the corresponding region in the desmoplakin PRDs B and C no such helix is observed and instead H2 is a canonical length. Its presence on the surface may present a functionally relevant binding position divergent from desmoplakin.

Finally, in PR4 the typical β-hairpin is replaced with a random coil whereas in desmoplakin PRD C the same region contains a canonical β-hairpin. In envoplakin this region contains subtle shifts in the backbone that places S1 and S2 sufficiently apart to prevent normal hydrogen bonding interactions. Therefore a canonical beta hairpin is not observed. Ambiguity in the electron density is also not observed, and is well define further in both asymmetric molecules arguing that the random coil formation is real. Closer inspection of the random coil region reveals that although a negatively charged residue is observed at position 4 (D1946) of the motif, a positively charged residue usually present at position 19 is absent and is instead replaced by a serine (S1961). Since an electrostatic interaction is typically observed between position 4 and 19 to "fix" the β-hairpin into place, this may explain why in envoplakin PR4 this region is replaced with a random coil. As shown in figure 5.52B both asymmetric units PR4 random coil region is divergent, arguing that the region is also flexible. This may have a functional role since

interestingly; a residue conserved in the putative vimentin binding groove is located in this loop region (R1952) and is discussed further in 5.85.

Overall, while the fold, conformation and secondary structure elements are similar between the envoplakin PRD and desmoplakin PRD C, there are some subtle differences that may have as of yet uncharacterized functional significance.

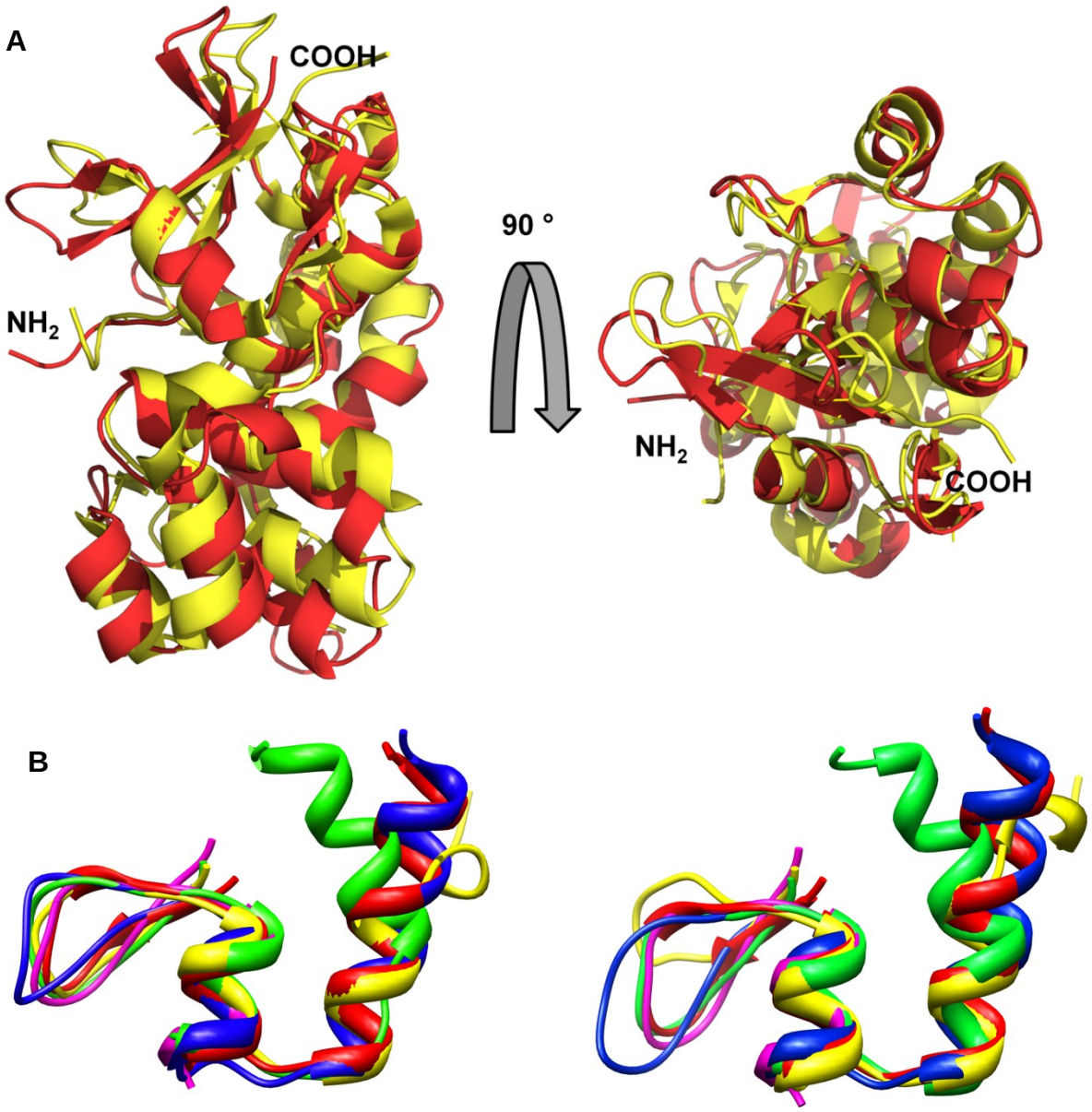


Figure 5.53 Structural comparison of the PRD C of envoplakin and desmoplakin

A: The crystal structures of envoplakin PRD molecule B (red) and desmoplakin PRD C (yellow) were superimposed using PYMOL. Two orthogonal views are shown with their NH_2 and COOH termini highlighted. B: Superimposition of the PR motifs using CHIMERA. Left: envoplakin PR motifs, right: desmoplakin PR motifs. PR1: red, PR2: green, PR3: blue, PR4: yellow, PR5: magenta. The similarities in the overall fold shown in A and the PR motif orientations in B show that overall the fold is similar, although closer inspection reveals subtle differences (discussed in text).

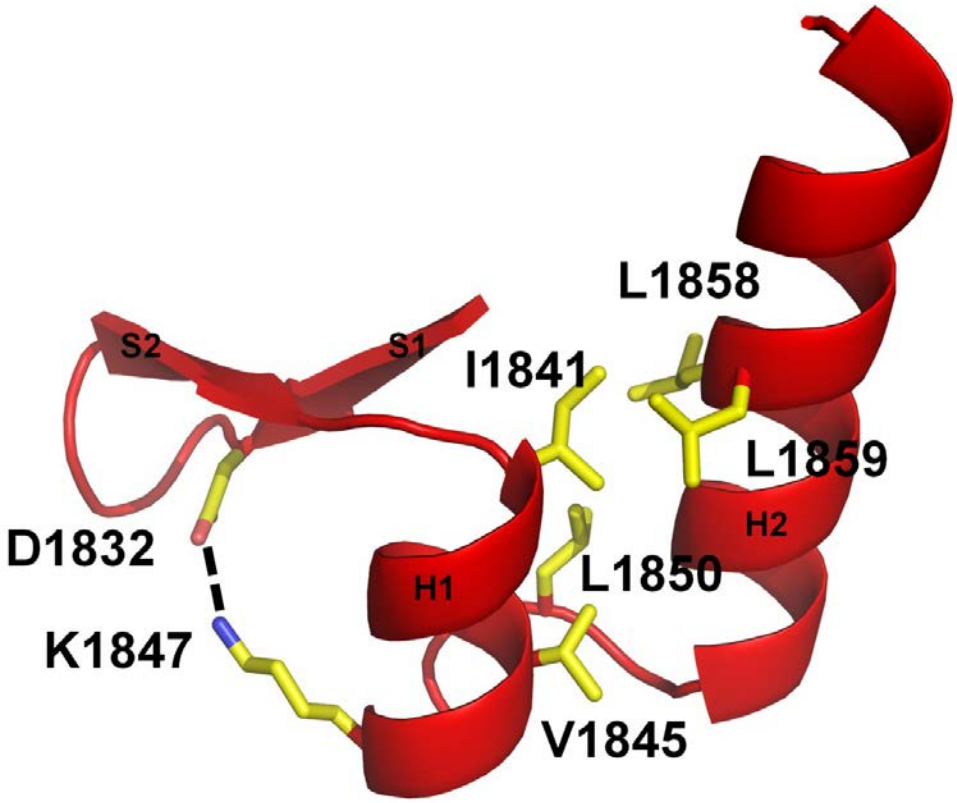


Figure 5.54 Intradomain interactions in PR1 of the envoplakin PRD

The two main categories of typical intradomain interactions between PR motifs are highlighted using PR1 from the envoplakin PRD. The first type of interaction is observed between position 4 from S1 of the PR motif (in this case D1832) and position 19 from H1 (in this case K1847). This involves an electrostatic interaction which ensures that the β-hairpin is “fixed” in solution. The second interaction involves multiple hydrophobic contacts from H1, the loop separating the two anti parallel helices and H2. Five hydrophobic residues are involved in PR1 which in effect creates a mini hydrophobic core between the two helices ensuring that the two anti parallel helices are orientated and fold correctly.

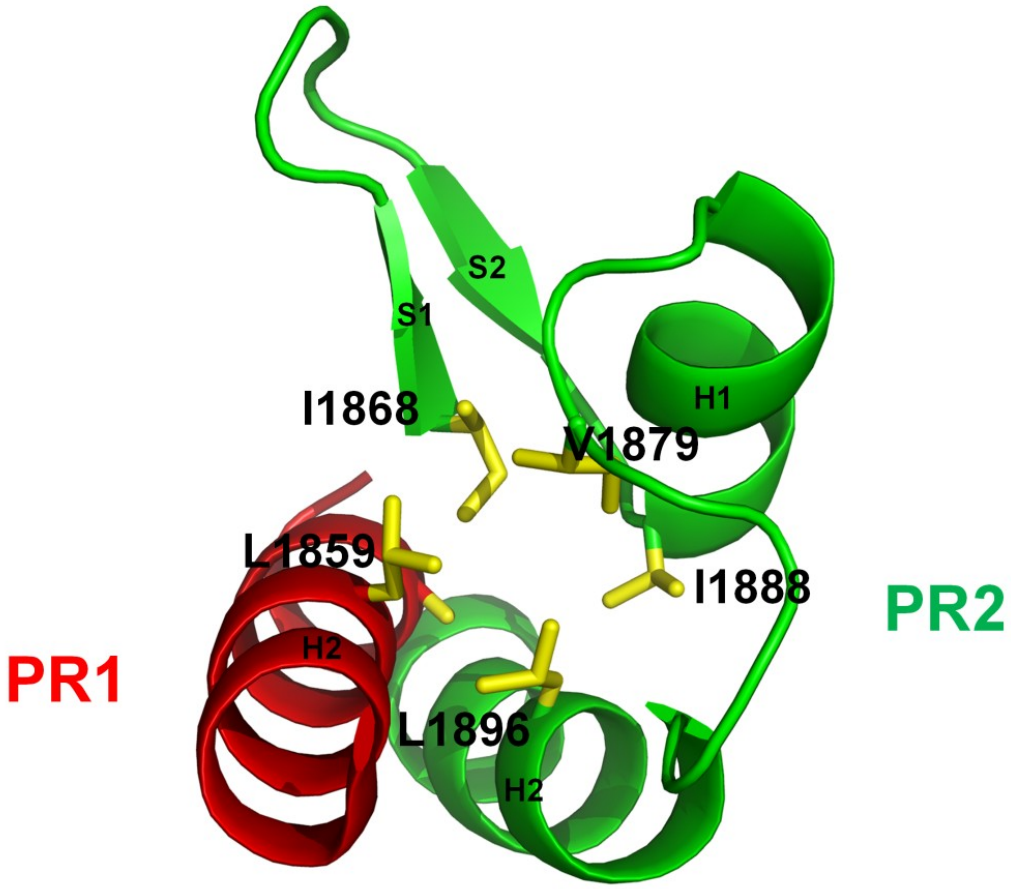


Figure 5.55 Interdomain interactions between PR1 and PR2 of the envoplakin PRD

The main type of interdomain interaction is highlighted in which multiple hydrophobic contacts are required to stabilize the fold of the protein using two different PRs. In this case a hydrophobic residue (L1859) from H2 of PR1 (red) interacts with multiple hydrophobic residues from PR2 (green) to create a mini hydrophobic core.

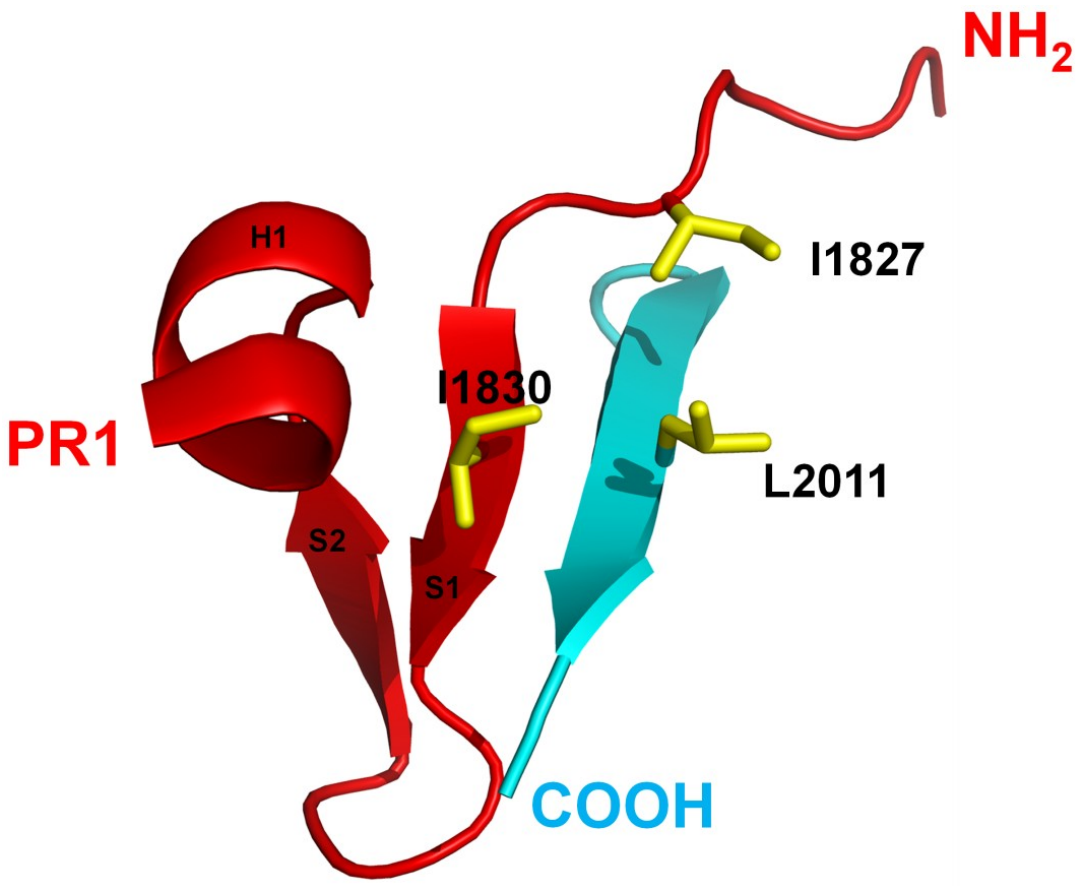


Figure 5.56 Interdomain interactions between PR1 and the extreme C-terminal strand

An interdomain interaction is observed between the extreme C-terminal of the protein (cyan), PR1 and the extreme N-terminal (both red). The extreme C-terminal orientates in a way in which it becomes parallel with S1 of PR1 and is stabilized utilizing an interaction of three hydrophobic residues (yellow) including I1827 (from the extreme N-terminal), I1830 (from S1 of PR1) and L2011 (from the extreme C-terminal). Although this was also observed in the desmoplakin PRDs B and C, the novel fact that deletion of both extremes of the N-terminal and C-terminal (that do not contribute to a PR motif) caused insolubility upon expression of the envoplakin PRD small construct implies that this region may act as an essential element to ensure complete and stable folding of the domain.

5.7 SAXS analysis of the envoplakin PRD

The structures of the desmoplakin PRDs B and C are unusually compact given the presence of repeat motif structures. Envoplakin PRD is also similarly compact based on the RMSD values when compared to desmoplakin PRDs B and C. Crystallization artifacts can occur whereby a non-physiologically relevant conformation is favoured for structural determination. To exclude crystallisation artefacts, solution state biophysical techniques such as SAXS can be used.

The envoplakin PRD was analyzed by SAXS to corroborate the envoplakin PRD crystal structure. Firstly, the R_g and D_{max} values for the envoplakin PRD were calculated as 1.9 and 5.6 nm respectively using GNOM. Using CRYSOL theoretical SAXS scattering curves for the monomeric and dimeric crystal structures of the envoplakin PRD was superimposed onto the experimentally acquired data. The theoretical curve for the monomeric crystal structure gave excellent agreement with the experimental data (Figure 5.61) Furthermore a χ^2 value of 1.1 was recorded which highlights the high level of corroboration between the theoretical and experimentally derived data sets. In contrast the theoretical curve of the dimeric crystal structure gave a poor fit with a χ^2 value of 4.9. This suggests that in solution the protein is monomeric solution. To further corroborate the findings of CRYSOL an *ab-initio* envelope model was created utilizing the SAXS scattering curve and the program DAMMIF. The crystal structure of the envoplakin PRD was then automatically superimposed on to the bead model using SUPCOMB (figure 5.62). A NSD value of 1.375 was recorded which indicated a good fit. Interestingly a non-overlapping section of the bead model is also revealed. This most likely represents the small flexible vector derived sequence located at the N-terminal end of the protein. Collectively, this suggests that the monomeric crystal structure of the envoplakin PRD is representative of its orientation in solution and therefore molecular and functional insights gleaned from the structure are physiologically valid.

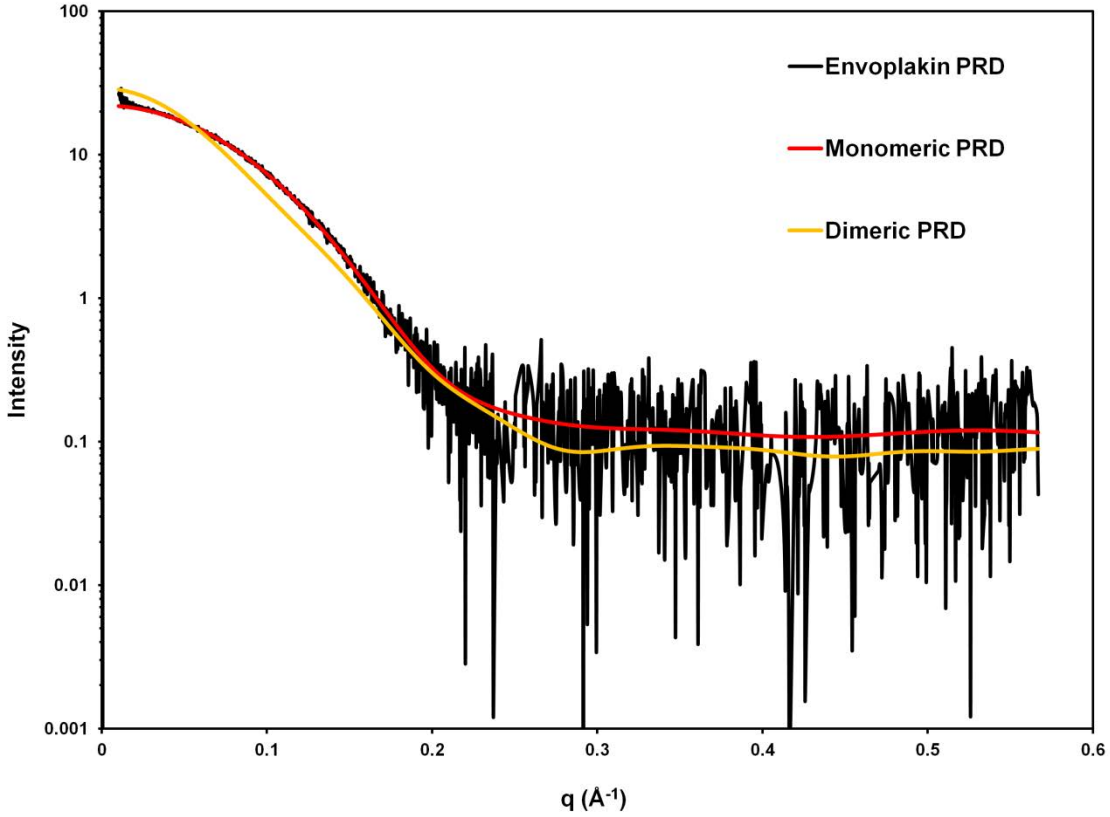


Figure 5.61 SAXS analysis of the envoplakin PRD

An experimental SAXS scattering curve for the envoplakin PRD was acquired and buffer subtracted (black).Protein was run at room temperature in 20 mM Hepes (pH 7.5) and 100 mM NaCl. Concentrations of 2, 4 and 8 mg /ml were used and scaled and merged accordingly to previously described methods. A theoretical SAXS scattering curve was created from the monomeric envoplakin PRD structure (red) and superimposed onto the experimental scattering curve using CRYSOL. A theoretical SAXS scattering curve was also generated for the dimeric envoplakin PRD structure (yellow) and superimposed onto the experimental scattering curve. A high level of similarity is observed between the theoretical monomeric PRD and experimental curves (χ^2 = 1.1). The dimeric PRD did not fit as well (χ^2 = 4.9). This confirms that a monomeric crystal derived model of the envoplakin PRD is physiologically relevant.

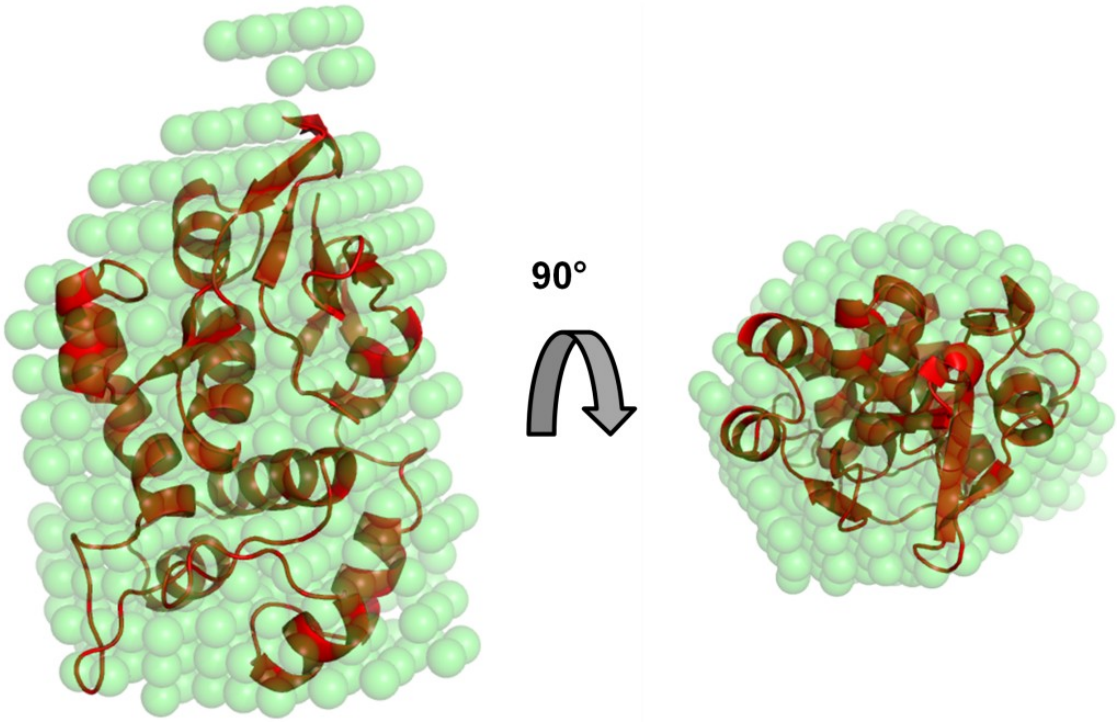


Figure 5.62 Superimposition of a monomeric envoplakin PRD into a SAXS ab-initio envelope

An *ab-initio* model was created using previously described methods and was superimposed on to the monomeric crystal derived structure of the envoplakin PRD by SUBCOMB and visualized by PYMOL. The crystal structure fits well in the SAXS derived *ab-initio* envelope (NSD = 1.375), highlighting the physiological relevance of the crystal.

5.8 Binding of the envoplakin PRD to the cytoskeletal protein vimentin

5.81 Expression and purification of vimentin

Expression and purification of full length vimentin was conducted for ligand titration experiments. The protein was expressed in *E. coli* where it typically forms insoluble inclusion bodies. Previously published purification protocols for vimentin were attempted but failed (Quinlan et al., 1986). Therefore a previously documented *in vitro* dilution refolding method for immunoglobulin domain containing proteins (Garboczi et al., 1992) was used to renature vimentin. However, this protocol failed to produce soluble vimentin in sufficient levels for titration experiments. To circumvent this problem, an optimized version of the Garboczi protocol was used. Following lysis of the bacteria vimentin is located in the insoluble fraction (figure 5.71). It remains in the remaining insoluble fractions during the wash steps. Upon solubilisation with 6 M urea it was finally purified to ~85% purity in a final buffer of 20mM Tris-HCl pH 7, as demonstrated by the predominant band migrating at ~53 kDa (Figure 5.71). Since the biophysical properties of vimentin have previously been well characterized by a variety of techniques (Herrmann and Aebi, 2004, Herrmann et al., 1996), it was instead examined as a potential ligand for the envoplakin PRD in NMR titrations (discussed further in 5.82. Due to time constraints the refolding of vimentin was not tested however, it was assumed that vimentin had refolded based on previously described methods (Quinlan et al., 1986). In addition to recombinant “in-house” purified vimentin, a commercially available source (Cytoskeleton, Inc) was used in co-sedimentation pull down assays (5.82).

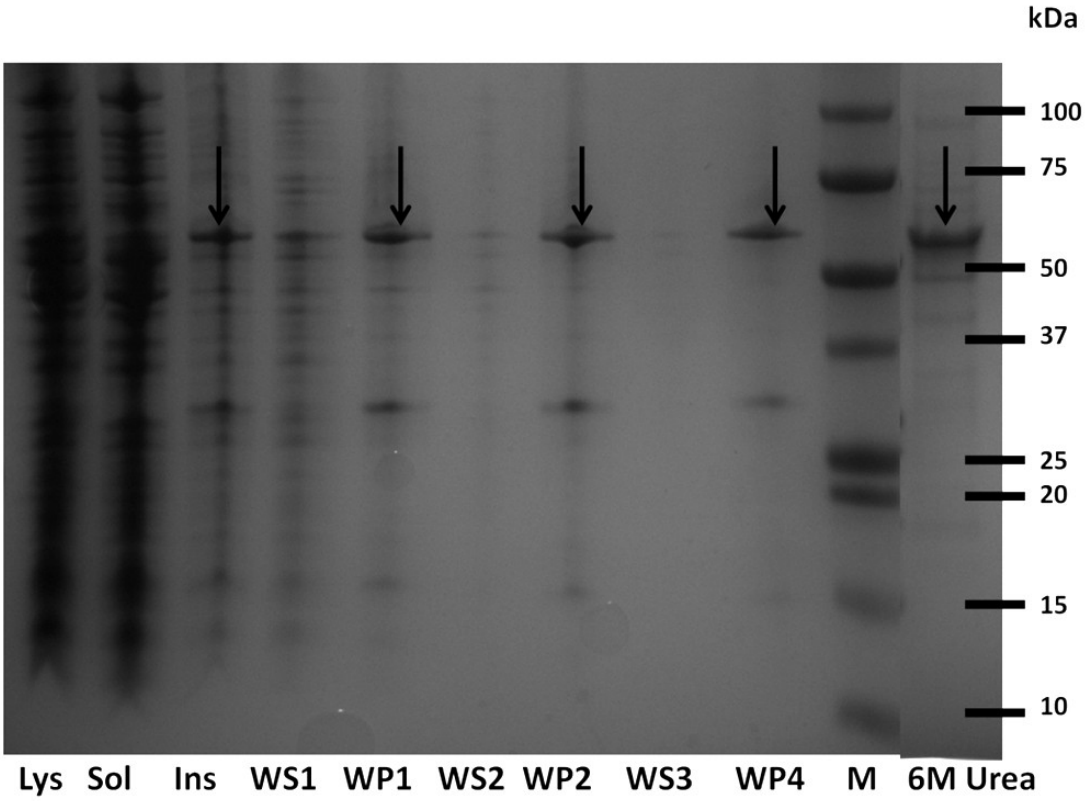


Figure 5.71 Renaturation of full length vimentin using an optimized Garboczi method.

Lys: bacterial lysate, Sol: bacterial soluble fraction, Ins: bacterial insoluble fraction, WSx: wash step soluble fraction, WPx: wash step insoluble fraction, M: marker, 6M Urea: solubilised vimentin in 6M urea. The arrows indicate the position at which vimentin migrates (~53 kDa), and this is in good agreement with its sequence derived molecular weight. The contaminant observed at approximately 30 kDa has been removed upon solubilisation of vimentin with 6M urea. Urea was then removed in a stepwise fashion until it had been completely removed. Vimentin protein was estimated to be 85% pure and was used for NMR titration studies with the envoplakin PRD.

5.82 Titration of unlabelled vimentin to ^{15}N labelled envoplakin PRD

Vimentin is known to bind to a variety of plakin family C-terminal domains including the linker domain of periplakin and the PRDs of desmoplakin (Kazerounian et al., 2002, Meng et al., 1997b, Stappenbeck and Green, 1992). A previous study suggested that the envoplakin PRD fails to bind vimentin *in vivo* (DiColandrea et al., 2000), but since it exhibits high homology to the desmoplakin PRD C, verifying this with NMR titration experiments was of considerable interest. If binding was observed, it was hoped that future backbone assignments could be used to identify residues critical for vimentin binding. Notably, vimentin has not previously been examined by NMR, possibly due to its large size and elongated shape presenting technical challenges. This is especially true for vimentin as it displays a propensity to oligomerize under certain conditions. To keep the experiment straightforward and prevent or limit oligomerization, the NMR experiment was conducted in the absence of NaCl and therefore vimentin was in an "unpolymerized" state. In this state, vimentin is typically observed as a mixture of dimeric and tetrameric oligomers (Quinlan et al., 1986).

To investigate the possibility of vimentin binding to the envoplakin PRD, unlabelled vimentin was titrated against ^{15}N labelled envoplakin PRD. Strikingly, the titration experiments unequivocally demonstrated vimentin binding to the envoplakin PRD (Figure 5.72). For example, at a 10:1 ratio (envoplakin to vimentin), approximately 40% of the peaks broaden out. At a higher ratio of 2:1 (envoplakin to vimentin), the majority of the envoplakin PRD peaks have broadened out. This suggests that at a ratio of 2:1 (envoplakin to vimentin), the interaction between the envoplakin PRD and vimentin has reached a saturation point. Unfortunately, given the high number of peaks broadening out at low vimentin concentrations, it is unlikely that the identification of specific residues critical to vimentin binding can be identified. The significant level of peak broadening raises two additional possibilities. Firstly, that the entire protein surface of the

envoplakin PRD is required for binding to vimentin. Secondly, the significant broadening is merely a reflection of the relatively large molecular weight of the vimentin-PRD complex being larger than the 35 kDa limit for NMR. The latter scenario represents a more likely explanation since the envoplakin PRD-vimentin complex would tumble in a manner that is unfavourable for NMR analysis. Unfortunately, the foldedness and function of vimentin was not tested during experiments in this chapter. However, the foldedness and function was validated in the following chapter based on an observed difference in binding between the periplakin and envoplakin linker domains.

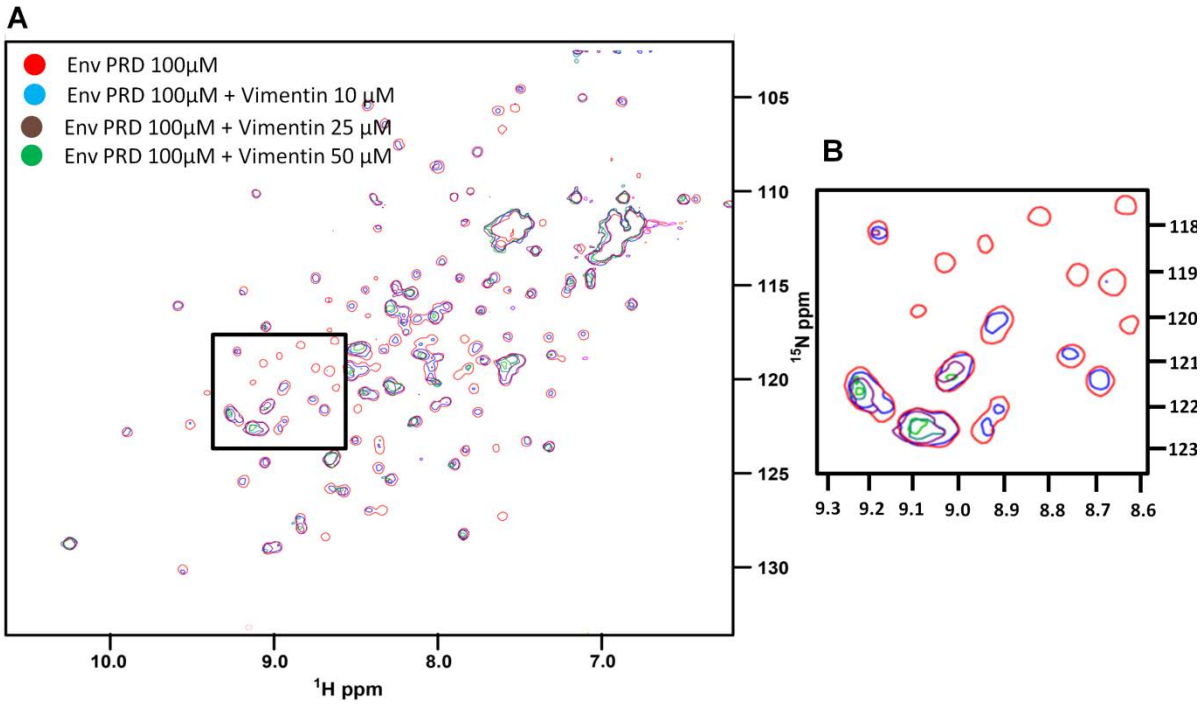


Figure 5.72 Vimentin titration to ^{15}N labelled envoplakin PRD

^{15}N labelled envoplakin PRD was subjected to 2D HSQC analysis in a titration series with unlabelled recombinant vimentin. This was conducted on an Agilent 600 MHz spectrometer at 298 K in 20mM Tris (pH 7). A variety of different sample points were collected: firstly envoplakin PRD alone and then a variety of different concentrations of vimentin were added while the volume was kept constant. A: overall HSQC with ratios between the envoplakin PRD and vimentin highlighted in their respective colours in the top left corner. B: Boxed section of the HSQC so that closer inspection reveals the extent to which peaks begin to broaden out. Peaks begin to broaden out when envoplakin PRD is in 10 fold excess and saturation occurs at two fold excess. This can be observed more clearly in B as many of the peaks in red (apo) do not contain any overlapping peaks from points when vimentin is added (e.g. in blue, purple or green).

5.83 Analysis of binding polymerized vimentin to desmoplakin PRD C and the envoplakin PRD using co-sedimentation pull down assays

Previously the envoplakin PRD was shown to interact with non-polymerized vimentin using NMR spectroscopy (section 5.82). As polymerized vimentin is the main physiologically relevant species in the cell, an interaction between vimentin and the envoplakin PRD was investigated. Co-sedimentation pull down assays were used to test the interaction between the envoplakin PRD and polymerized vimentin since they are the gold standard for probing interactions of soluble proteins to polymerized cytoskeletal IFs (Bishr Omary and Coulombe, 2004). A previous study using this method on the desmoplakin C-terminal tail showed that all three desmoplakin PRDs were capable of independently binding to vimentin (Choi et al, 2002)..

As expected, polymerized vimentin was observed solely in the pellet (figure 5.73). When incubated with desmoplakin PRD C it is able to pull down a fraction of the desmoplakin PRD C although the majority of the desmoplakin PRD C remains in the supernatant. Similarly when vimentin is incubated with the envoplakin PRD, the latter is pulled down into the pellet with the majority of the PRD remaining in the supernatant (figure 5.73). This suggests that both desmoplakin and envoplakin PRDs bind to vimentin in the polymerized state with similar affinity. These observations, together with the vimentin tritration experiments, suggest that the envoplakin PRD is capable of binding both unpolymerized and polymerized vimentin.

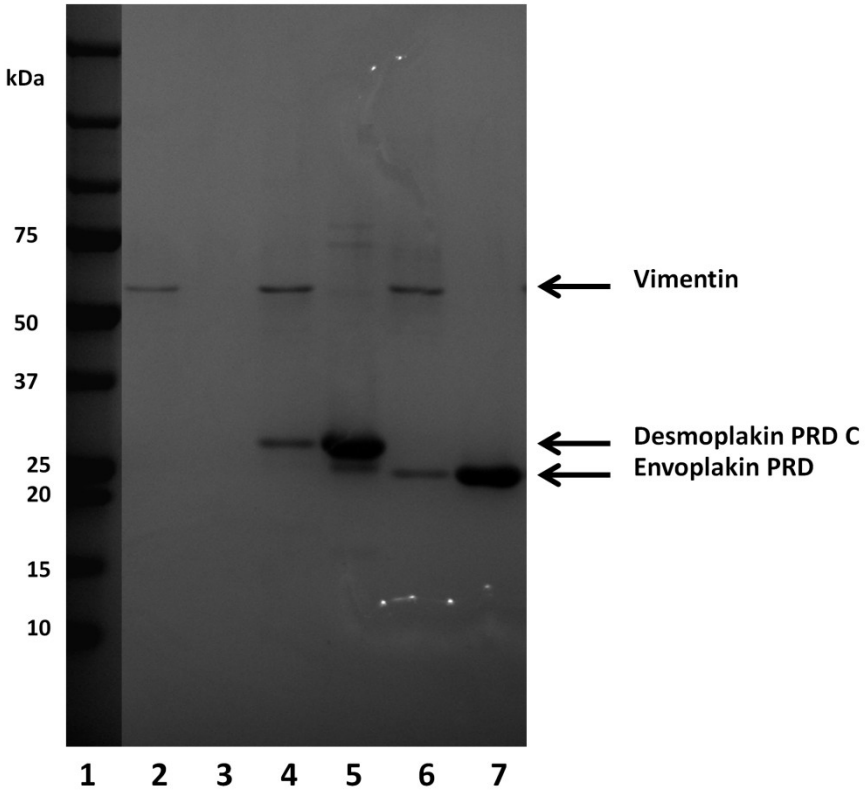


Figure 5.73 Analysis of binding polymerized vimentin to desmoplakin PRD C and the envoplakin PRD using co-sedimentation pull down assays

Commercially bought vimentin (Cytoskeleton, Inc) (stock concentration 0.5 mg/ml) was polymerized according to manufacturer's instructions and incubated with desmoplakin PRD C or envoplakin PRD (stock concentration 10 mg/ml) for 1 h at room temperature (final concentrations were at a ratio of 2 to 1 PRD to vimentin). Tubes were then centrifuged at 100 000g for 30 min on an Optima Max ultracentrifuge (Beckman-Coulter). The pellet was washed thoroughly three times with PBS to ensure that it was not contaminated with protein from the supernatant. 1: molecular weight marker, 2: vimentin (alone) pellet fraction, 3: vimentin (alone) supernatant fraction, 4: vimentin and desmoplakin PRD C pellet fraction, 5: vimentin and desmoplakin PRD C supernatant fraction, 6: vimentin and envoplakin PRD pellet fraction, 7: vimentin and envoplakin PRD supernatant fraction. It reveals that both the desmoplakin and envoplakin PRDs bound to vimentin at similar affinities.

5.84 Electrostatic analysis of the envoplakin PRD and modelling of vimentin binding

It was originally proposed that the PRDs of desmoplakin bind to cytoskeletal proteins such as vimentin through a conserved positively charged groove. The envoplakin PRD exhibits a groove (approximately 16 Å wide, 26 Å long and 11 Å deep) that is similar to that of the desmoplakin PRDs B and C (Choi et al, 2002). To compare their grooves the envoplakin PRD and desmoplakin PRD C were subjected to electrostatic surface analysis with the PDB2PQR server using default parameters (Dolinsky et al., 2004). The electrostatic surfaces are similar in some respects but different in others. Importantly the positively charged groove observed in desmoplakin is retained in the envoplakin PRD (Figure 5.74). For example, two previously identified positively charged side chains in desmoplakin (R2744 and K2668) are conserved in envoplakin (R1952 and R1876) and located in approximately the same orientation within the groove (Figure 5.74). In addition, the surface areas which could accommodate the proposed vimentin binding site are also similar.

Interestingly, the original study did not highlight that the desmoplakin PRD C possesses a negatively charged patch consisting of 4 proximal surface exposed residues consisting of E2626, E2629, E2635 and E2638 from PR1 (Figure 5.74A) (Choi et al, 2002). In contrast, the corresponding region in envoplakin displays a neutral patch as it encompasses residues T1834, N1837, T1843 and A1846 (figure 5.74B). Further comparisons of the molecular surfaces between envoplakin and desmoplakin revealed additional dissimilarities in their electrostatic properties. For instance, rotation of the envoplakin PRD approximately 180 ° from the proposed vimentin binding groove exhibits an electronegative charged patch (figure 5.74D). This is contributed by residues E1966, E1967, E1975 and E1979 from PR4 that is considerably larger than the patch observed in the corresponding region of desmoplakin PRD C (formed solely by

residues E2748 and E2749 from PR4) (figure 5.754C). Collectively, this suggests that although the positively charged putative binding groove is retained in each PRD solved to date (including the envoplakin PRD from this study) there appears to be subtle differences in other electrostatic surface regions. This may have an impact on cytoskeletal protein binding affinities, preferences or both since the structure of vimentin has provided hints as to how the cytoskeletal protein interacts with PRD containing proteins. Indeed, vimentin exhibits alternating charges on its surface and it is tempting to postulate that this could be an alternative / additive mode of binding in addition to the positively charged groove. The envoplakin-vimentin interaction warrants further investigation to validate whether the electrostatic patches play a key role in stabilising the binding ligands.

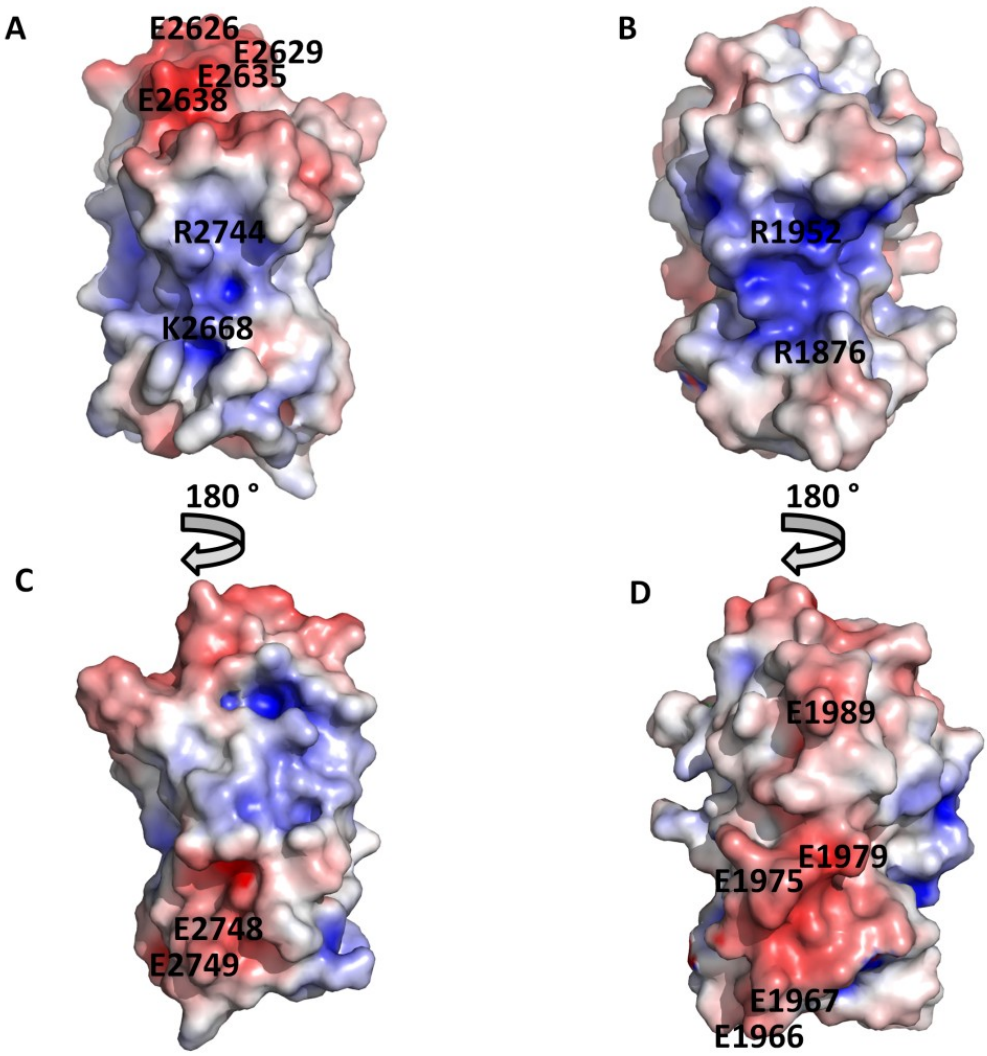


Figure 5.74 Electrostatic surface representation of the envoplakin PRD and desmoplakin PRD C

A: The electrostatic surfaces of both desmoplakin PRD C (A and C) and envoplakin PRD (B and D) were calculated using the PDB2PQR server (Dolinsky et al, 2004) and orientated in similar conformations. Residues conserved in the putative vimentin binding groove of desmoplakin PRD C (R2744 and K2688) are semi conserved in envoplakin (R1952 and R1876). Also indicated are other residues possibly involved in the interaction with vimentin such as the negative patches of desmoplakin PRD C (E2626, E2629, E2635 and E2638). These residues are altered to non charged residues in the envoplakin PRD. When the surface representations of both desmoplakin and envoplakin are rotated 180 ° an alternative negatively charged patch is observed, especially for envoplakin in which E1966, E1967, E1975, E1979 and E1989 create a large patch which may bind to a positively charged patch on vimentin.

5.85 Modelling of vimentin binding to the envoplakin PRD

The notion that vimentin binds to PRDs via their conserved positively charged groove was assessed by molecular modelling. The envoplakin PRD binding site on vimentin is unknown, although it has been proposed to reside in the rod domain based on previous studies (Favre et al., 2011, Choi et al., 2002). Interestingly, vimentin is dominated by a series of alternate negative and positive patches that may contribute to binding PRDs. The crystal structure of a fragment from the vimentin rod domain (PDB code: 1GK4) was used to assess the interaction between vimentin and the envoplakin PRD. Although it does not represent the entire rod domain, it was proposed that the rod domain contains homologous repeats of this crystallized fragment (Strelkov et al., 2002). This has since been proved by the very recent structural elucidation of three more fragments of the vimentin rod region (Chernyatina et al., 2012). Together with previous fragments elucidated, an atomic structure of the entire rod domain is now known. The 1GK4 fragment was manually docked to the proposed groove on the envoplakin PRD and provided a relatively snug fit (Figure 5.75). This is in agreement with previous manual docking experiments on the desmoplakin PRDs B and C (Choi et al, 2002) and suggested the interaction between the two proteins is sterically feasible.

To probe the interaction further, the High Ambiguity Driven DOCKing (HADDOCK) server was used to create a model of the envoplakin PRD-vimentin complex (Dominguez et al., 2003). As part of this process, the positively charged residues previously identified from the envoplakin PRD structure were selected as possible vimentin binding candidates based on manual modelling and previous suggestions (Choi et al, 2002). These included residues that contribute to the positively charged groove (R1952 and R1876) as well as side chains proximal to the groove (R1985 and R1914). Since the envoplakin PRD binding site on the vimentin rod domain has not been localized, two negative patches were identified on the crystallized fragment of

vimentin (PDB code: 1GK4) as possible binding sites. The most conserved negatively charged residues were present at E338, E341 and E348 since these residues were identical across vimentin full length sequences from multiple species from *Homo sapiens* to *Xenopus laevis* and semi conserved across other IF proteins desmin and keratin 8 when aligned by CLUSTALW (Thompson et al., 1994).

The results from the HADDOCK server suggested that an electrostatic interaction between the two is highly plausible. The proposed complex was then analyzed visually, by the program PISA (which calculates buried surface areas) and by the program CONTACT (which allows quantitative analysis of the interactions between certain atoms of an amino acid in a complex). Overall the putative model appears to accommodate 4 electrostatic interactions between the 3 of the 4 selected amino acids of envoplakin and all 3 of vimentin (figure 5.76 and table 5.8). In addition two separate hydrogen bond interactions are observed (figure 5.76 and table 5.8). Interestingly, the residue R1952 from the putative flexible random coil region of PR4 described in 5.63 is centrally located in this interaction. Furthermore, according to the program CONTACT it is involved in 2 electrostatic interactions between E338 and E341 of vimentin. This implies that R1952 may be critical to vimentin-envoplakin PRD complex formation. Finally, when the buried surface between the two proteins was analyzed by the program PISA (Krissinel and Henrick, 2007) a value of 889.2 $Å^2$ is given. Since this is significantly above the minimal level required of a physiologically relevant dimer it appears to suggest that the interaction between the two is entirely possible. Collectively, this suggests that multiple electrostatic interactions most likely play a significant role in stabilising an envoplakin PRD-vimentin complex.

Vimentin	Location	Envoplakin	Location
$K334^{NZ}$	N/A	$D1822^{OD1}$	(*)
$E338^{OE1}$	N/A	$R1952^{NH2}$	PR4 (L)
$E341^{OE1}$	N/A	$R1952^{NH2}$	PR4 (L)
$E341^{OE2}$	N/A	$R1876^{NH1}$	PR2 (S2)
$E348^{OE1}$	N/A	$R1895^{NH1}$	PR2 (H2)
$E348^{OE2}$	N/A	$Q1900^{NE2}$	PR5 (S2)

Table 5.8 Calculated contacts between envoplakin PRD and vimentin based on HADDOCK binding

A negative patch (E338, E341 and E348) from vimentin was selected as a potential binding to site for the envoplakin PRD using a crystallized fragment of the vimentin rod (PDB code: 1GK4). Amino acids selected from the putative positively charged groove of the envoplakin PRD were selected (R1876, R1985, R1914 and R1952) as a potential binding site for vimentin. Using HADDOCK a potential complex between the two proteins was predicted (figure 5.76) and using CONTACT potential interacting atoms from amino acid residues were calculated. It reveals that 3 of the 4 residues selected for envoplakin bind to all 3 selected for vimentin. R1952 appears to be an important residue since it binds to both E338 and E341. Side chain atoms involved in binding are shown in superscript.

This model is based on the premise that the envoplakin PRD has a positively charged binding groove, and since vimentin has multiple negatively charged patches, this proposed mechanism clearly requires further investigation. Due to time constraints this was not possible, but hopefully sets a foundation to address how PRDs interact with their cytoskeletal partners.

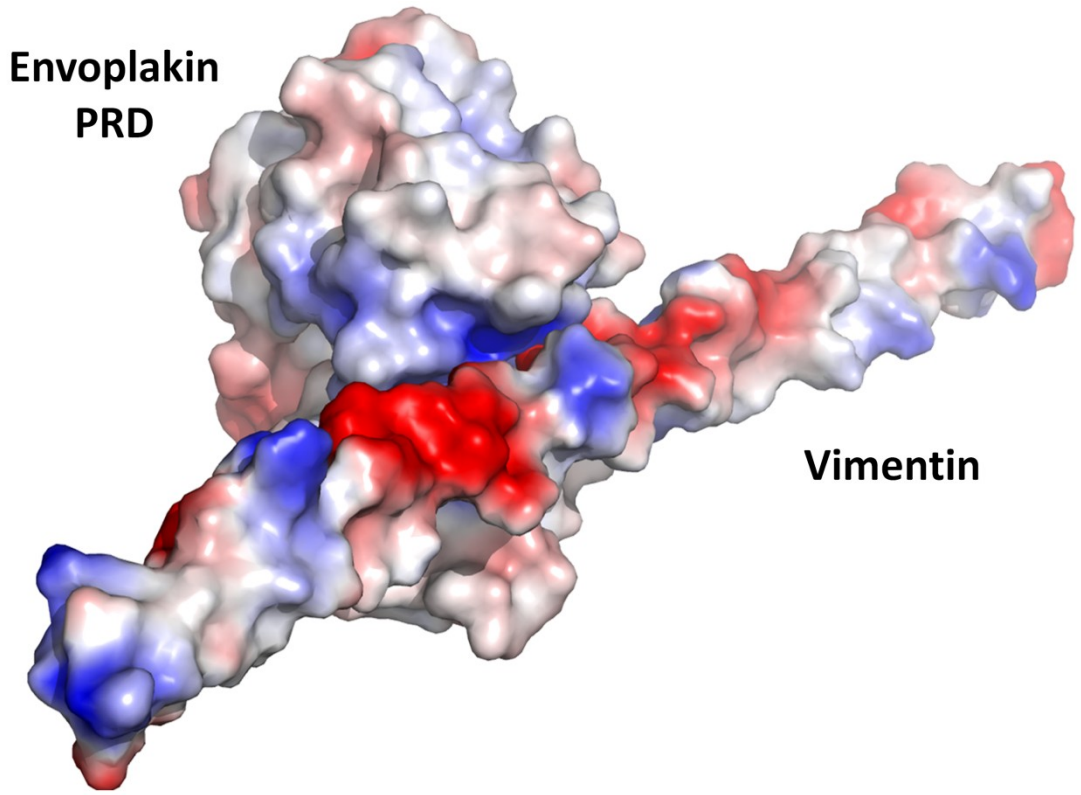


Figure 5.75 Manual docking of the envoplakin PRD-vimentin complex

An excellent fit was obtained following manual docking of vimentin (rod shape) to the envoplakin PRD (spherical shape) putative binding groove. Image visualized in PYMOL.

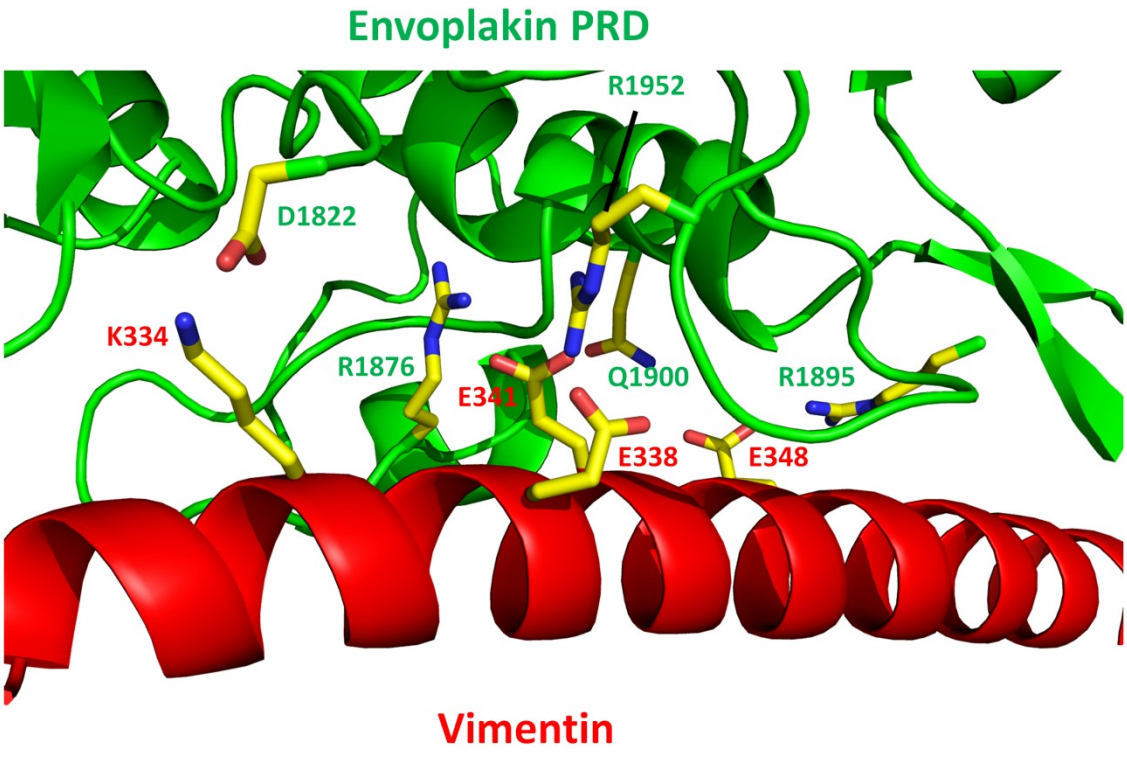


Figure 5.76 HADDOCK automated docking of the envoplakin PRD (green)-vimentin (red) complex

The complex reveals multiple electrostatic and hydrogen bond interactions. Vimentin side chains of K334, E338, E341, and E348 (highlighted in red text) interact with the envoplakin PRD side chains of D1822, R1876, R1985, Q1900 and R1952 (highlighted in green text) and are discussed further in the text. As shown in table 5.8 R1952 appears to be a critical residue in the formation of the complex since it is centrally located and interacts with both E338 and E341 of vimentin.

5.9 Chapter 5 Discussion

Since a vital function of plakin family members is the recognition of cytoskeletal protein partners, understanding the mechanism of PRD and vimentin complex formation was investigated. It is well known that the C-terminal tail of plakin family members is required for binding to cytoskeletal proteins. However, to date only two structures of PRDs from desmoplakin have been published (Choi et al, 2002). Understanding whether the structure and or function is conserved across other PRDs was therefore of interest. The envoplakin PRD was crystallized and its structure solved with the aim of gaining novel insights into vimentin binding and whether PRDs from different plakin proteins share similar structures and therefore binding properties. The experiments presented in this study revealed that the overall structures of PRDs are similar between different plakin family members. However, there appear to be minor differences in acquiring fold and potential binding mechanisms.

It was originally suggested that PRDs require 4.5 PR motifs to acquire a globular fold and any less were suggested to result in an unstable fold (Choi et al, 2002). The envoplakin PRD has 4.5 PR motifs and adopted a similar fold to those observed for PRDs B and C of desmoplakin. However, a smaller envoplakin construct retaining the minimal 4.5 PR motifs failed to fold due to likely instability or insolubility upon *E.coli* expression. This construct had its N and C termini shortened as these regions were more divergent than those observed in the “core” region giving rise to the 4.5 PR motifs (figure 5.1). Closer inspection of the N and C termini from the envoplakin PRD crystal structure suggests that they were proximal to one another (figure 5.56). Multiple hydrophobic residues from PR1, the N-terminus and the C-terminus appeared to create a mini hydrophobic core. This formed a parallel sheet which allows the domain to complete its globular fold (figure 5.56). This feature appears to be critical to the fold of all PRDs since these residues were semi conserved in similar regions of the crystal structures of desmoplakin PRDs B

and C. In effect these results indicated that the residues at the N and C-termini are indispensible for folding of all PRDs.

The mechanisms involved in acquiring the general fold therefore appear to be conserved between PRDs of different plakin family members. The same is true of individual PR motifs of the envoplakin PRD in which inter and intra domain hydrophobic and electrostatic interactions are required to acquire fold. Typically, amino acids present at position 4 (negatively charged amino acid) and position 19 (positively charged amino acid) of each PR motif engage in an electrostatic interaction. In addition to this interaction, multiple hydrophobic contacts are made between apolar residues situated between H1 and H2 (figures 5.54 and 5.55). These interactions are highly reminiscent to those observed in the desmoplakin PRDs B and C. In effect the envoplakin PRD utilizes similar mechanisms to those observed in the desmoplakin PRDs B and C for folding within the PRD domain and assembly of the PR motifs.

The biophysical properties of the envoplakin PRD and desmoplakin PRD C were also investigated in solution by AUC and SAXS. AUC analysis revealed that both proteins were monomeric in solution (figure 5.2). SAXS confirmed this assumption when theoretical SAXS scattering curves were created for dimeric and monomeric models of the envoplakin PRD and superimposed onto the experimentally derived curve (figure 5.61). This confirmed that the monomeric state of the crystal structure was physiologically relevant. Given the corroboration between the *ab-initio* and crystal structure models, the electrostatic surfaces were also investigated and appeared to be conserved (figure 5.62). This included the putative positively charged vimentin binding groove.

The putative binding groove of envoplakin's PRD retains similar dimensions and charged residues to those observed for the desmoplakin PRDs B and C. Unfortunately, the binding sites

between vimentin and any PRD are yet to be experimentally mapped. One of the main aims of the experiments described in this chapter was to investigate how binding between a PRD and vimentin could be achieved. Firstly, unpolymerized unlabelled vimentin was titrated against ^{15}N labelled envoplakin PRD and showed saturation at a ratio of 2:1 PRD to vimentin (figure 5.72). Similar findings were observed using ^{15}N labelled desmoplakin PRD C and unpolymerized unlabelled vimentin in which a saturation was observed at a ratio of 2:1 (Unpublished findings: Miss Penelope Rodriguez-Zamora, 2012). This implies that the two different PRDs are capable of binding unpolymerized vimentin at similar affinities. This is the first time an unpolymerized cytoskeletal protein has been shown to interact with a PRD *in vitro*. The finding is of interest because the main physiologically relevant cytoskeletal state is polymerized vimentin which forms a large fibrous network (Bishr Omary and Coulombe, 2004). However, studies have indicated the presence of a pool of unpolymerized soluble cytoskeletal proteins which form a heterogenous group of dimeric, tetrameric and higher order oligomeric states (Blikstad, 1983, Quinlan et al., 1986, Bishr Omary and Coulombe, 2004). These pools of unpolymerized cytoskeletal proteins are thought to form the precursor molecules of a larger cytoskeletal network. However, once cytoskeletal networks have been formed why these unpolymerized pools of cytoskeletal proteins remain in the cell is unclear. In addition, it is known that proteins bind to cytoskeletal proteins to aid their polymerization (Bishr Omary and Coulombe, 2004). Binding of the envoplakin PRD to unpolymerized vimentin may therefore aiding vimentin polymerization or sequestering soluble vimentin precursors, depending on the state of the cell. A previous study found that a sequence in the C-terminal tail from plectin could inhibit filament assembly in a dose dependent manner and conversely disassemble polymerized IFs at high concentrations (Steinböck et al., 2000). This was suggested as a possible mechanism for envoplakin in which a previous study reported bundling of intermediate filaments by envoplakin in the presence of periplakin (Karashima and Watt, 2002).

Regulatory mechanisms have been proposed for PRD interactions. Phosphorylation of a non PRD based amino acid in the desmoplakin C-terminal tail significantly disrupted binding to vimentin and keratin (Stappenbeck et al., 1994, Meng et al., 1997b). Therefore, it appears there may be a complex interplay in the way IFs and PRDs (or at least plakin family member C-terminal tails) interact with one another.

In an effort to ascertain the interaction with the more physiologically relevant vimentin species, the polymerized state, the envoplakin PRD and desmoplakin PRD C were analyzed for their ability to bind polymerized vimentin in a co-sedimentation assay. It was shown previously that desmoplakin PRDs can bind to polymerized vimentin (Choi et al, 2002). This study has now confirmed that the envoplakin PRD can also bind to polymerized vimentin, in a similar affinity to that of desmoplakin PRD C (figure 5.73). Comprehensive analysis of the binding affinities between the PRDs and vimentin were not conducted in this study due to time constraints. This is in agreement with the findings of the envoplakin PRD binding to unpolymerized vimentin when analyzed by 2D-HSQC NMR (figure 5.72). Interestingly, the results contradict the finding of a previous study which suggested that the envoplakin PRD was unable to bind to vimentin directly (DiColandrea et al, 2000). The envoplakin constructs these authors used tended to aggregate *in vivo*, and hence may not give a clear indication of whether the envoplakin PRD is able to bind directly to vimentin or not.

Ultimately, the results indicate no significant differences between the desmoplakin PRD C and envoplakin PRD in terms of their binding affinities to vimentin, either unpolymerized or polymerized. This may be reflected in the fact that their putative vimentin binding groove is similar both in terms of their charge and surface area. When molecular modelling was used to further characterize the envoplakin PRD-vimentin interaction it suggested that binding was feasible (figures 5.75 and 5.76). This included parameters such as interaction surface area and

the type of amino acid interactions involved, which corroborated previously published findings (Choi et al, 2002).

Regardless of the similarities between the envoplakin PRD and desmoplakin PRD C there are some subtle structural differences that may affect vimentin binding. A comparison of the molecular surfaces of desmoplakin PRD C and envoplakin PRD was investigated because the interaction between PRDs and vimentin were thought to be predominantly electrostatic (figure 5.75). It revealed that although the putative positively charged binding groove was conserved, electronegative patches were present in alternative positions on the surfaces of desmoplakin PRD C and the envoplakin PRD. The importance of this is that vimentin exhibits alternative charges on its surface (Strelkov et al., 2002, Chernyatina et al., 2012) and these may represent an alternative/additive mode of binding to a PRD. Furthermore, the recently elucidated structure of a fragment from keratin's rod domain, desmoplakin's preferred cytoskeletal binding partner (Lee et al., 2012), revealed that keratin has interesting surface charge properties. In contrast to vimentin, keratin exhibits one face which is entirely negatively charged while the other is entirely positively charged. The differences of surface charge properties between both the PRDs and cytoskeletal proteins may therefore reflect a mechanism of cytoskeletal preference which warrants further investigation. This is based on previous observations indicating that desmoplakin has a preference for keratin K8/K18 over desmin/vimentin in non phosphorylating conditions (Meng et al., 1997b) while plectin has a preference for desmin/vimentin over cytokeratins 5/14 (Steinböck et al., 2000).

At the amino acid level an interesting difference exhibited for the envoplakin PRD is the random coil feature (as opposed to the typical β-hairpin) of PR4. The electron density of this region in both asymmetric units was well defined which suggested this observation was real. In addition to this, the typical electrostatic interaction observed between the negatively charged residue on S1

and the positively charged residue on H1 was abolished in PR4. The positively charged residue at position 19 was instead replaced by a neutral serine (S1961). This is in contrast to other PR4 motifs from other PRDs where this interaction is retained. It is also possible that this feature exhibits flexibility based on the differences in its orientation between the two different molecules of the asymmetric unit (figure 5.52B). The importance of this putative flexible region comes from the fact that R1952 was proposed to be centrally located in the putative vimentin binding groove by HADDOCK (figure 5.76). In addition to this R1952 is proposed to bind to two negatively charged residues from vimentin (table 5.8). The flexibility of this region may allow extra freedom in orientating itself between the two negatively charged residues. To ascertain whether this proposed flexibility is functionally important, the electrostatic interaction could be artificially re-established by mutagenizing S1961 to a positively charged residue e.g. a lysine. The mutant construct could then be crystallized and its structure elucidated. If the β-hairpin motif is re-established with no alteration to binding properties it would suggest that the random coil feature is just a serendipitous observation. Regardless, it appears to warrant further investigation given R1952's putative central role in its binding to vimentin.

Overall, the structural elucidation of the envoplakin PRD has revealed novel insights in terms of general PRD fold with the observation of a C-terminal strand that completes the fold. Furthermore, it has established that although there are a variable number of PRDs at the C-terminal tails of plakin family members, they are highly likely to acquire similar folds. Regardless, there are subtle differences between desmoplakin PRD C and the envoplakin PRD to key regions such as those in PR4 and the alternative negatively charged surface areas. This suggests that differences between the two domains may explain differing cytoskeletal preferences between different plakin family members and therefore warrants further investigation.

Chapter 6 - Identification and flexibility of a structured domain in the C-terminal tails of plakin proteins

The C-terminal tails of plakin family members tether the desmosome (or in some cases IFs) to IFs and disruption of this link can lead to a variety of diseases. While some PRD to IF binding mechanisms were proposed in chapter 5, other studies have highlighted the importance of the highly homologous linker domains in these interactions (Steinböck et al., 2000, Fontao et al., 2003, Kazerounian et al., 2002, Meng et al., 1997b). These studies showed that different plakin family linker domains bind to a variety of IFs. Understanding the intricate interplay between linker domains and PRD binding to IFs requires further investigation in order to resolve the responsible domains, residues, and mechanisms underlying IF binding and preferences.

The linker domains, as they are known in the field, typically tether PRD B to PRD C at the C-terminal tails of plakin family member proteins. Recent work on plectin corroborates the suggestion that binding to cytoskeletal proteins can be conducted solely by the linker domains independently of the PRDs (Favre et al., 2011, Karashima et al., 2012). They are also thought to facilitate dimerization between full length envoplakin and periplakin proteins even with their rod domains deleted (Karashima and Watt, 2002). Furthermore, they are implicated in the disease paraneoplastic pemphigus (Kiyokawa et al., 1998, Mahoney et al., 1998). This disease is characterised by the presence of circulating autoantibodies directed against plakin family members. The envoplakin and periplakin linker domains harbour major antigenic sites but as yet, it is unclear whether the targeting of linker domains is the cause or consequence of the disease. The periplakin linker domain is also known to bind protein kinase B in which the linker domain is thought to be used as a scaffold (Heuvel et al., 2002).

As suggested by their name, linker domains are viewed as flexible/disordered elements in some reviews (Leung et al., 2002, Jefferson et al., 2004). Unfortunately, to date there is no available biophysical data on the linker domains. Given their critical role in cytoskeletal binding it was felt a structural approach would give a clearer indication as to how these interactions occur. Some authors have reported the presence of up to 2 PR motifs in these linker domains using the SMART database (Leung et al., 2002, Jefferson et al., 2004) but these observations have never been explored fully. Other motif identification servers do not corroborate the possible presence of 2 PR motifs. Indeed it is unusual to observe single PRs in isolation. It was shown in the previous chapter and the original study on PRDs (Choi et al, 2002) that a fully folded domain must contain a minimum of 4.5 repeats and some hydrophobic elements at the N and C termini. Indeed, linker domains may be disordered, since intrinsically disordered regions of proteins are capable of binding multiple partners such as that of the cytoplasmic tail of DSG our lab previously reported (Kami et al., 2009).

Three other groups involved in solving desmosomal protein structures (Weis, Sonnenberg/De Pereda/Wiche) typically use X-ray crystallography. The use of X-ray crystallography is predominantly suited to proteins with rigid globular structures and attempting to solve these linker domains by this method alone was presumably deemed too high a risk by these groups. This is based on the fact that if the linker domains were disordered it would have made their structural elucidation by X-ray crystallography refractory. However, alternative methods such as NMR and SAXS allow for basic structural analysis of folded or unfolded regions to be assessed with low risk. These include whether the protein is folded, basic structural properties and ascertain whether sequence derived models fit SAXS scattering data. Using NMR and SAXS it was hoped the following questions could be answered:

1) Are the linker domains folded or unfolded?

2) Are there conformational differences between the plakin family member linker domains?
3) What governs their binding preferences for cytoskeletal proteins?
4) How do envoplakin and periplakin linker domains interact to form the proposed heterodimer of their full length proteins?

By understanding these structural and functional properties it was hoped it would give a clearer insight into the role of the linker domains and their intricate interplay with the PRDs to bind to cytoskeletal proteins.

Due to the results of this chapter being collected towards the end of my PhD, collection of some experiments were carried out in collaboration with Miss Penelope Rodriguez-Zamora or Dr Claudia Fogl and will be stated as such within the text or figures legends.

6.1 Construct design and purification

6.11 Construct design

Construct lengths were based primarily on previous reports (Karashima and Watt, 2002). The alignment on which these linker domain constrcts were based is given in figure 6.11. As the linker domains were thought to be flexible or disordered, it was decided that all the constructs should be GST tagged to increase solubility. Constructs were therefore cloned into pGex6P-1 from their respective cDNA's.

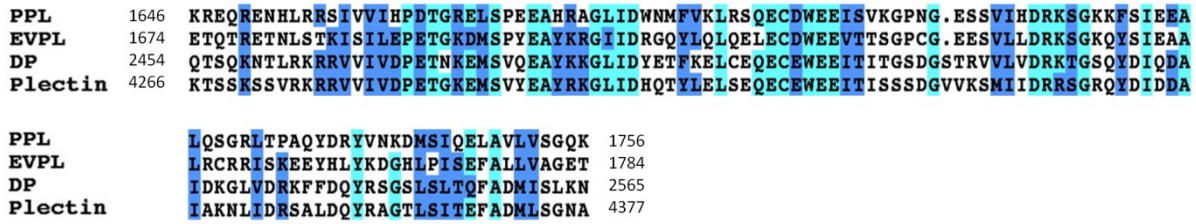


Figure 6.11 Linker domain alignment used for construct design

The linker domain is a well defined sequence and aligns well between all the plakin family members which include: periplakin (PPL), envoplakin (EVPL), desmoplakin (DP) and plectin. Dark blue residues represent semi conserved residues while light blue residues represent conserved residues. Each entire plakin linker domain sequence depicted here was used as the construct for each respective protein. Modified from Karashima and Watt (2002).

6.12 Purification of linker domains

Each linker domain contained a GST tag and in addition, the envoplakin linker domain contained a non-cleavable his-tag at its C-terminal end (LGHHHHHH). The GST fused envoplakin linker domain was purified using a GST column (GE Healthcare). The GST tag was cleaved by PreScission protease and the linker domain purified by gradient-step his-trap elution. The periplakin and desmoplakin linker domains were also purified using GST columns. The GST tag was cleaved by PreScission protease and finally a gradient step anionic Q-sepharose column (GE Healthcare) was used to separate the linkers domains and GST. Following purification, each linker domain was judge to be 95% pure as judged by SDS-PAGE analysis and therefore each protein was ready for biophysical analysis (figure 6.2).

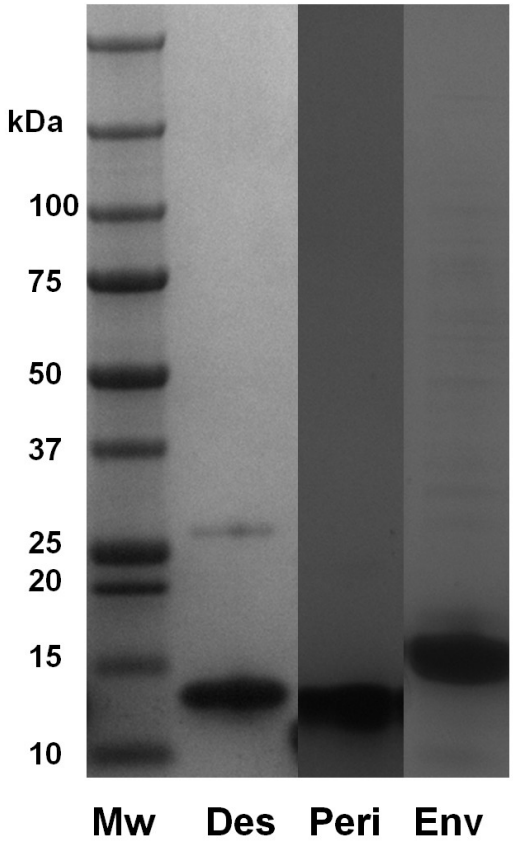


Figure 6.12 Final purity of linker domains assessed by SDS-PAGE analysis

Each domain is >90% pure. Des: desmoplakin linker domain, Peri: periplakin linker domain, Env: envoplakin linker domain.

6.2 Sedimentation velocity AUC profiles of the linker domains

The oligomeric state of the envoplakin and periplakin linker domains was unknown but both had been postulated to interact with one another (Karashima and Watt, 2002). Both the envoplakin and periplakin linkers were individually subjected to sedimentation velocity AUC to gauge their oligomeric states in solution in 20mM HEPES (pH 7.5) and 100 mM NaCl at 10 µM. In addition both were mixed at equimolar ratios to ascertain whether a complex (and therefore higher molecular weight species) could be observed. The results were analyzed by SEDFIT and shown in figure 6.2. It revealed two single peaks for both the envoplakin and periplakin linker domains at 11.5 and 11.8 kDa which suggested that each protein is monomeric. Although not shown, a complex of the two was not observed when the envoplakin and periplakin linkers were mixed together as only single species were observed which superimposed over the monomeric peaks of envoplakin and periplakin. This suggests that both linker domains are monomeric, and that a complex does not form at equimolar ratios unlike previous suggestions (Karashima and Watt, 2002). The desmoplakin linker domain was also monomeric in solution (Personal communication: Miss Penelope Rodriguez-Zamora). To further characterize the fold of each protein they were analyzed by 2D-HSQC NMR analysis.

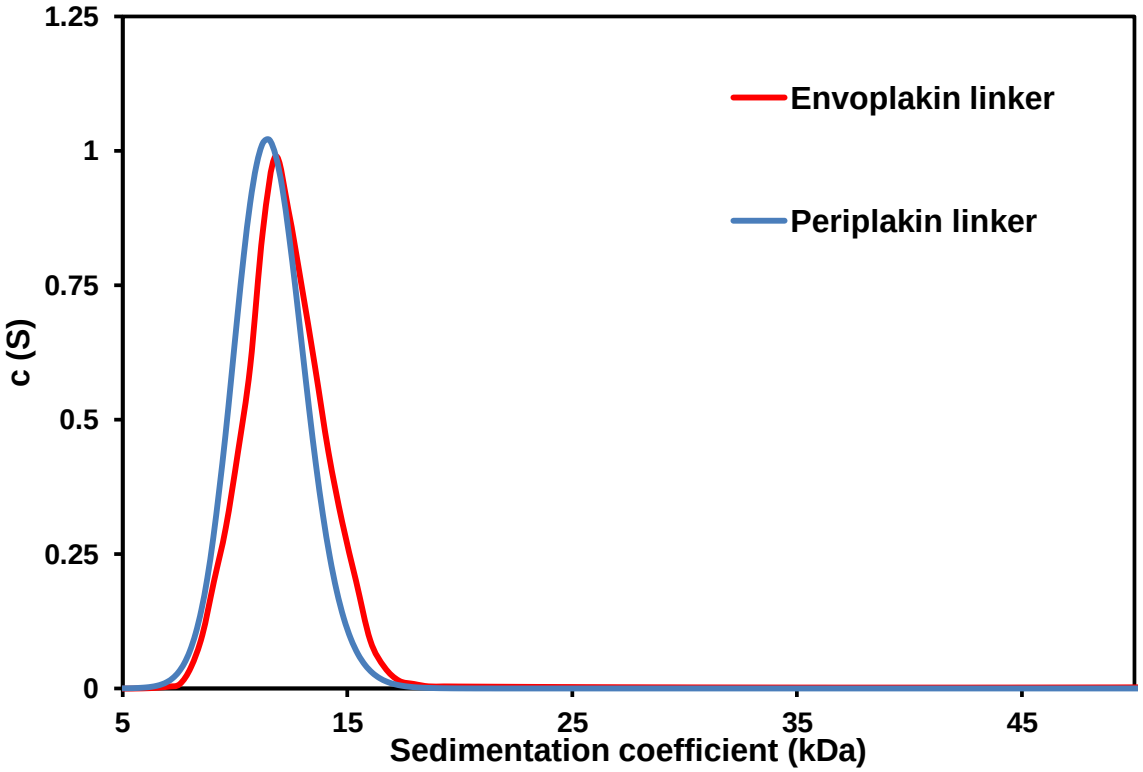


Figure 6.2 Sedimentation velocity AUC of the envoplakin and periplakin linker domains

The envoplakin and periplakin linker domains at 10 µM were subjected to AUC analysis in 20 mM Hepes (pH 7.5) and 100 mM NaCl at 10 µM using a Beckman XL-1 ultracentrifuge at 25000 rpm at 20 °C. A single species was observed for both proteins at 11.5 and 11.8 kDa for the periplakin and envoplakin linker domains respectively. The data reveal that both proteins are monomeric in solution. Molecular weights derived from AUC are within acceptable error ranges of sequence derived molecular weight estimations of 12.5 and 13.5 kDa for the periplakin and envoplakin linker domains respectively.

6.3 2D-HSQC NMR analysis of the linker domains

Bacteria transformed with linker domain constructs were grown and expressed in minimal media supplemented with ^{15}N labelled ammonium chloride. Expression and purification of the linker domains revealed no significant differences in yield or purity when compared to their unlabelled counterparts. Following purification they were analyzed by 2D-HSQC NMR analysis to gauge their overall fold.

6.31 2D-HSQC NMR analysis of the envoplakin linker domain

Since it was anticipated the linker domains were unstructured/disordered an HSQC was conducted in the presence of the GST tag to retain solubility in solution. Typically the combined size of monomeric GST and the envoplakin linker (38 kDa) is beyond the optimal size limit of NMR. NMR has previously been used as an effective strategy to deduce the structural integrity of a GST-tagged protein (Liew et al., 2008). The GST tag forms a homodimer independently of its fusion partner. The GST tag therefore tumbles at a rate unfavourable for NMR which results in most GST peaks broadening out. The remaining observable peaks from GST are often weaker in intensity than those of the fusion protein of interest. The vector sequence between GST and its fusion partner is elongated and flexible enough to allow independent tumbling of the fusion partner. The independent motion of the fusion protein can therefore tumble as if the GST was not present. This effectively means that the predominant peaks observed in the HSQC originate from the fusion partner of interest. The possibility exists that fusion proteins could be used for future structure determination of more challenging proteins.

When the fusion-protein was analyzed by 2D-HSQC NMR a well dispersed spectrum was observed indicative of a well folded protein (figure 6.31). Peak intensities are not isotropic which merely reflect the differences in observed origins of the peaks e.g. loop regions in GST (low intensities) and the linker domain itself (high intensities). To confirm that the high intensity peaks

observed in the dispersed regions originate from the envoplakin linker domain, the GST tag was cleaved, removed and run again by 2D-HSQC NMR and analyzed. Figure 6.32 shows the envoplakin linker domain in isolation (without GST) which again indicated that the original high intensity peaks observed in figure 6.31 originated from the envoplakin linker domain. Closer inspection by eye of the HSQC also facilitates the qualitative assessment of the secondary structure content. Proteins of high α-helical content contain peaks which disperse well between 7.5 ppm and 8.5 ppm on the ^{1}H chemical shift axis (see SR78 for an example – figure 3.73). Proteins of mixed α-helical and β-sheet content exhibit contain dispersion between 7 ppm and 9.5 ppm (see envoplakin PRD – figure 5.3). The envoplakin linker domain appeared to have the dispersion of a mixed α-helical and β-sheet containing protein. Overall this confirmed that the envoplakin linker domain was not a flexible/unstructured tether as previously presumed and was actually a well folded protein with a stable tertiary fold. Whether the periplakin and desmoplakin linkers also adopted a similar tertiary fold was also of interest.

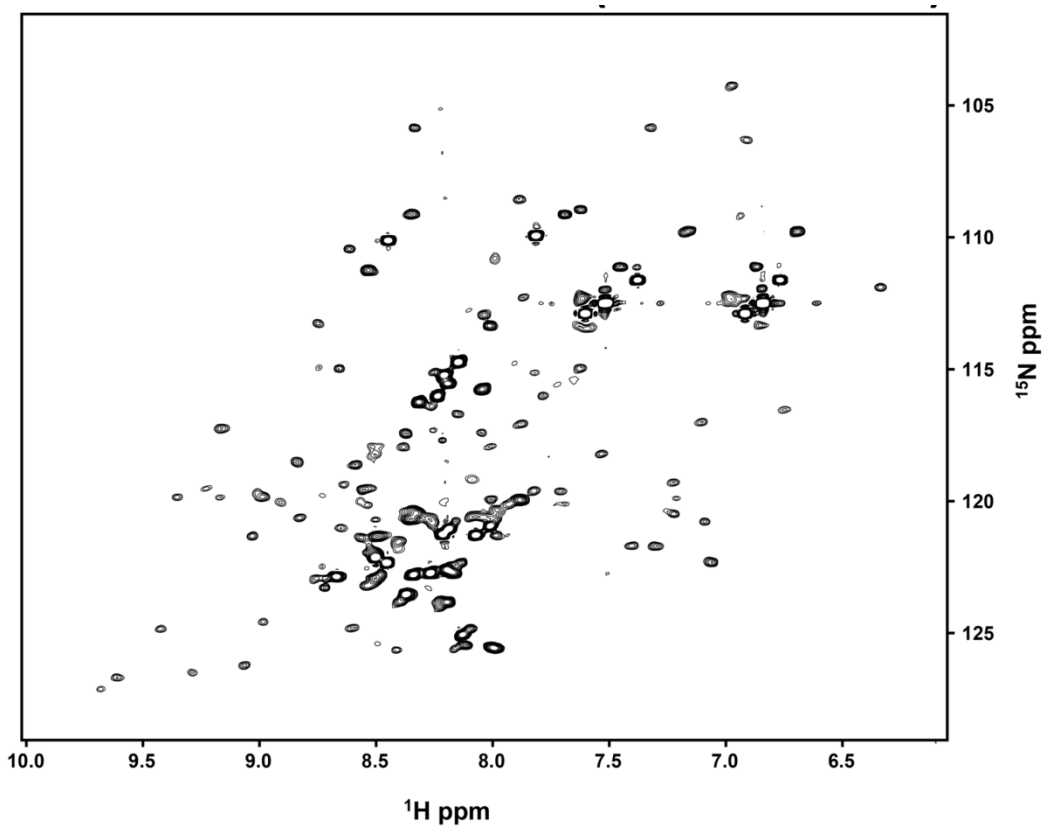


Figure 6.31 2D-HSQC of GST-fused envoplakin linker domain

The GST-fused envoplakin linker domain was analyzed by 2D-HSQC NMR analysis. The protein was analyzed using an Agilent 600 MHz spectrometer in 20 mM Hepes (pH 7.2) and 100 mM NaCl at 250 µM with a temperature at 298 K. The spectrum is indicative of a well folded protein as revealed by its well dispersed peaks between 7.5 ppm and 9.7 ppm. The peaks are not uniform in intensity; this is to be expected given that some low intensity peaks will be observed from GST loop regions as previously reported (Liew et al., 2008).

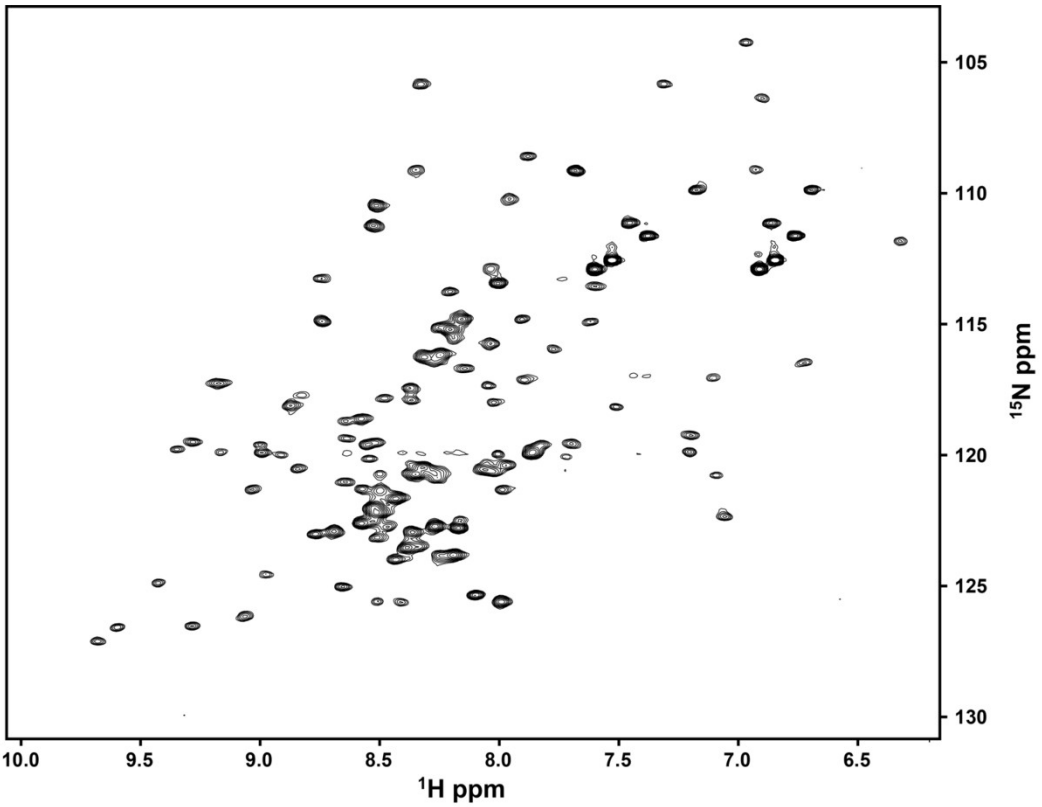


Figure 6.32 2D-HSQC of the envoplakin linker domain

The envoplakin linker domain in isolation was analyzed by 2D-HSQC NMR analysis. The protein was analyzed using an Agilent 600 MHz spectrometer in 20 mM Hepes (pH 7.2) and 100 mM NaCl at 200 µM with a temperature at 298 K. Since these peaks correspond with the high intensity peaks observed in the GST-fused variant of the envoplakin linker domain, it confirms that those peaks originate from the envoplakin linker domain.

6.32 2D-HSQC NMR analysis of the periplakin linker domain

Since the envoplakin linker domain was folded and all linker domains are highly homologous it was highly likely periplakin would similarly be folded without the presence of GST. It was therefore decided that the HSQC experiment would be conducted post GST cleavage. The periplakin linker domain was analyzed by 2D-HSQC NMR and a well dispersed spectrum was observed (figure 6.43). This suggested that the periplakin linker domain was well folded with a stable tertiary fold. Both the envoplakin linker domain and the envoplakin PRD have spectra indicative of mixed α-helical and β-sheet secondary structure content. Since the periplakin linker domain shares a similar HSQC to both the envoplakin linker domain and the envoplakin PRD it can be inferred each have a similar mix of α-helical and β-sheet secondary structure content.

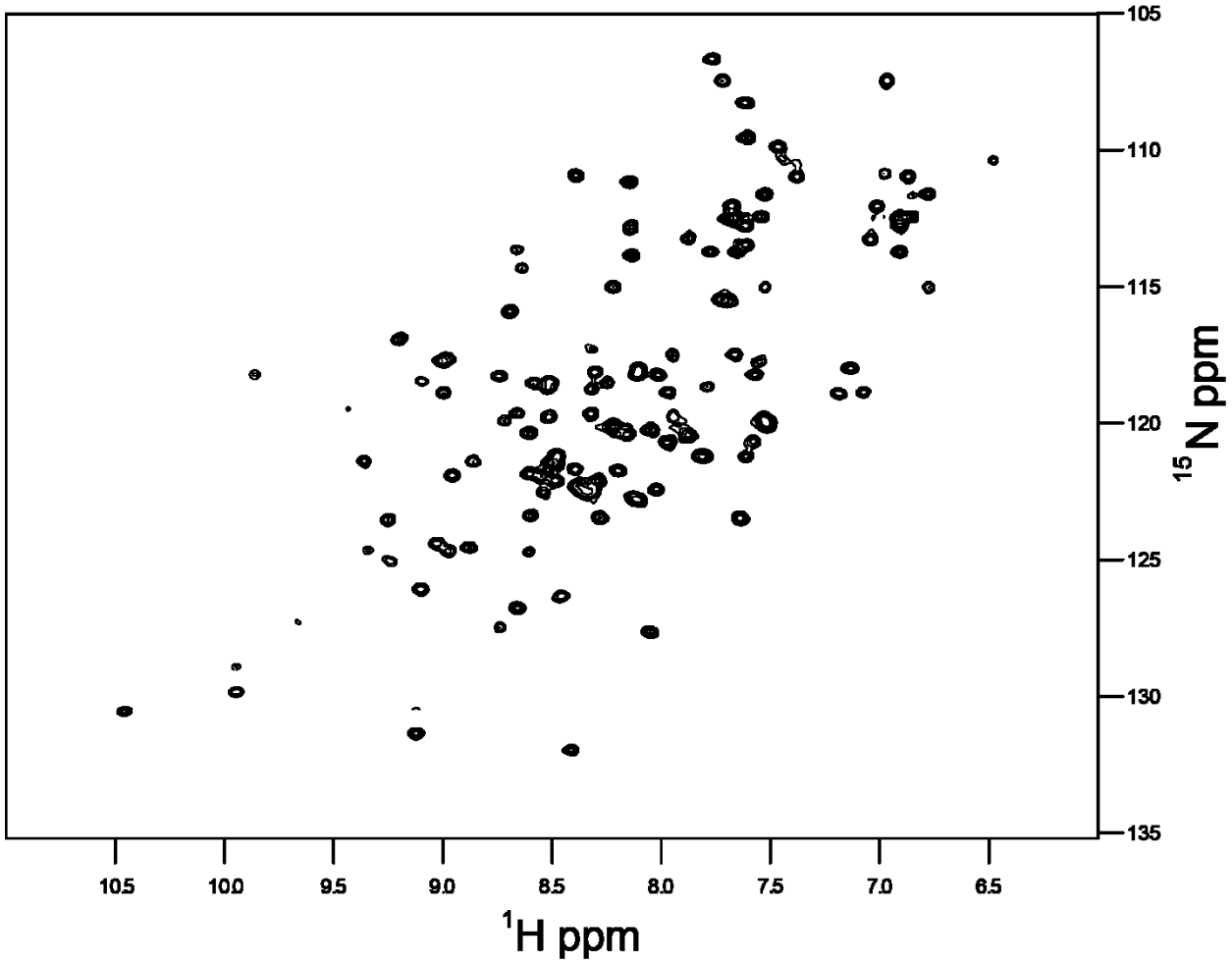


Figure 6.33 2D-HSQC NMR analysis of the periplakin linker

The periplakin linker domain was analyzed by 2D-HSQC NMR analysis. The protein was analyzed using an Agilent 600 MHz spectrometer in 20 mM Hepes (pH 7.2) and 100 mM NaCl at 250 µM at a temperature at 298 K. A well dispersed spectrum is observed indicative of a protein with a mixed α-helical and β-sheet type structure.

6.33 2D-HSQC NMR analysis of the desmoplakin linker domain

Since both envoplakin and periplakin linker domains were folded, it was highly likely the desmoplakin linker domain was similarly folded. The high sequence homology to other plakin family linker domains also suggested that the fold would be retained in the desmoplakin linker domain. This was investigated by 2D-HSQC NMR analysis. Once again a well dispersed spectrum was observed (figure 6.44). The spectrum was similar to those of envoplakin and periplakin also implying mixed α-helical and β-sheet secondary structure content. This therefore confirms that desmoplakin's linker domain is a well folded protein with a tertiary fold.

Figure 6.34 2D-HSQC of the desmoplakin linker domain

The desmoplakin linker domain was analyzed by 2D-HSQC NMR analysis. The protein was analyzed by an Agilent 600 MHz spectrometer in 20 mM Hepes (pH 7.2) and 100 mM NaCl at 100 µM at a temperature of 298 K. A well dispersed spectrum is observed indicative of well folded protein. Protein purification and NMR analysis conducted in collaboration with Miss Penelope Rodriguez-Zamora.

6.4 Modelling and SAXS analysis of the linker domains

Due to time constraints crystallization trials and further NMR experiments of the plakin linker domains were not possible. Since the linker domain fold of any plakin protein was unknown, it was decided that a modelling approach in combination with SAXS could be used to ascertain its structure. Based on the high homology between the linker domain plakin sequences it was highly likely that they share the same overall fold. Firstly, the sequences of each linker domain were fed to the I-TASSER server in order to acquire an atomic resolution model and in parallel, their sequences aligned to ascertain potential motifs. This data was then combined with SAXS scattering curves which were acquired for the desmoplakin and periplakin linker domains.

6.41 Sequence based modelling of the envoplakin, desmoplakin and periplakin linker domain

Modelling of the envoplakin linker domain was carried out first. The I-TASSER server selected desmoplakin PRD C and PRD B as templates (PDB codes 1LM5 and 1LM7 respectively) for the envoplakin linker domain with 14 and 11% identity respectively. Five models were created based on these templates. Although the sequence identity was low, it revealed the possible presence of two plakin repeat like motifs separated by a short flexible region which I have tentatively called the “WEE box” since it contains a highly conserved Trp, Glu, Glu, motif (figure 6.1). Being centrally located in the domain sequence and separating the two PR motifs it was envisaged that the WEE box would be an important feature of each linker domain. Superimposition of the 5 models generated by I-TASSER revealed a high level of convergence between the plakin repeat motifs of the different models. There was however a large degree of deviation observed between the WEE box of each model (figure 6.41). The convergence of the plakin repeat motif models suggests a measure of confidence in their general structure, however the ambiguity surrounding the modelling of the WEE box suggested it was potentially disordered. Models created for

desmoplakin and periplakin revealed a similar fold. To evaluate how similar these PR like motifs resembled actual plakin repeats motifs, all three linker domain sequences were aligned against canonical motifs of known structure (figure 6.42). Each of the proposed linker domain PR like motifs reveals a striking similarity to a canonical PR motif of known structure, which inferred a similar topology or fold.

In the first proposed PR motif the obligatory negatively charged residue at position 4 is conserved as Glu and Asp in envoplakin and desmoplakin respectively. At the same position a negatively charged residue is not observed in periplakin and consisted of a His. The negative charge is required for an interaction with a positively charged residue at position 19. The positively charged residue at position 19 is retained in envoplakin and desmoplakin as an Arg and Lys residue respectively, but is altered in periplakin to an Ala. This interaction is required to "fix" the β-hairpin into position. It suggests that periplakin, presumably devoid of this interaction, may allow the β-hairpin free movement. Alternatively, this putative PR motif β-hairpin may be observed as a random coil reminiscent to those of PR4 in the envoplakin PRD (see figure 5.52). The hydrophobic interaction between the two anti parallel α-helices is also retained whereby the Ile and Leu residues at positions 22 and 30 respectively are identical between the linker domains and the canonical plakin repeat motifs. This suggested that in all 3 linker domains the α-helical fold / topology is retained.

The second proposed PR motif appears slightly more divergent than the first although does retain some general properties that can still class this region as a PR like motif. At position 4 the negatively charged Asp residue is conserved entirely in each linker domain sequence. Interestingly the positively charged residue required at position 19 is conserved in desmoplakin as a lysine but is shifted two amino acids (position 21) along in envoplakin and periplakin as an

Arg. Both the envoplakin and periplakin models suggested that, like the desmoplakin model this shifted Arg retained the ability to interact with the Asp residue at position 4 (figure 6.53).

It therefore appears highly likely that the linker domains harbour two PR like motifs. This is consistent with the NMR data which suggests the presence of mixed α-helical and β-sheet secondary structure content. The WEE box which sits between the two PR motifs remains could not be modelled effectively. Whether this motif is disordered or contains secondary structure was of interest. SAXS was therefore used by building low resolution models to resolve the structure of these domains.

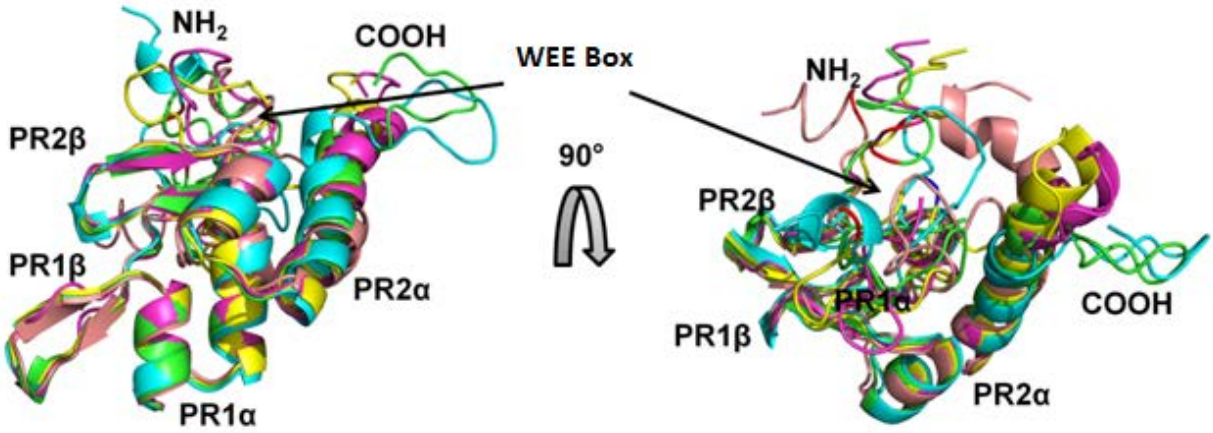


Figure 6.41 Superimposition of 5 models of the envoplakin linker domain by I-TASSER

Five models created by I-TASSER based on the envoplakin linker sequence were created using desmoplakin PRD B and PRD C as templates (pdb codes: 1LM7 and 1LM5 respectively). Highlighted are the N and C termini, the plakin repeat β-hairpin (PRxβ, where x is the plakin repeat it belongs to – see text) and the plakin repeat antiparallel α-helices (PRxα, where x is the plakin repeat it belongs to – see text). It reveals a strong convergence between the plakin repeat motifs but divergence between the WEE box separating the two repeats. The WEE box between the two repeats is buried within the hydrophobic core and its position indicated by arrows.

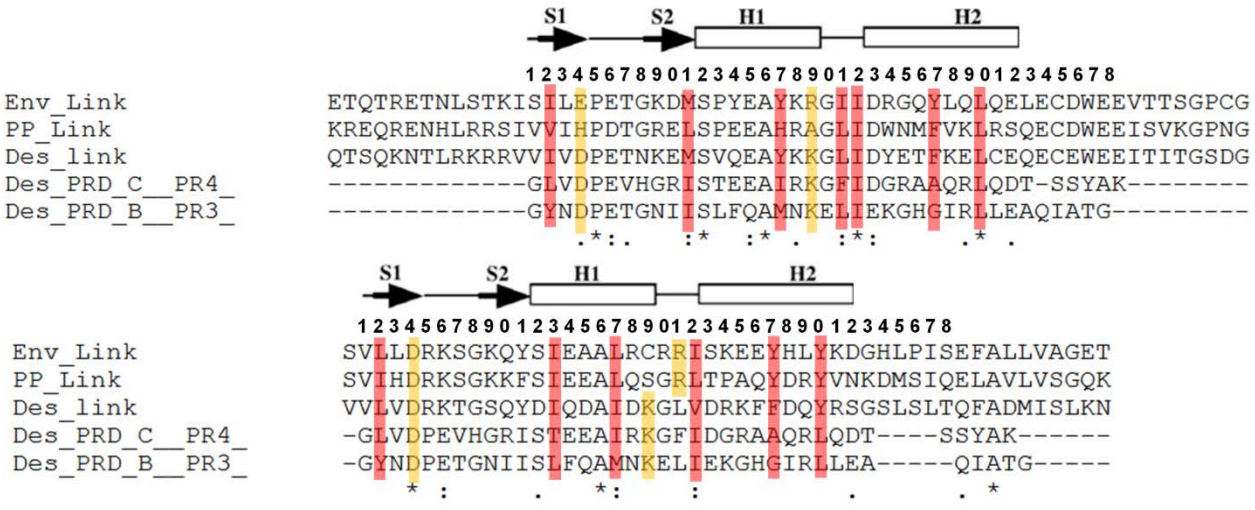


Figure 6.42 Identification of two plakin repeat motifs within the linker domain sequences of envoplakin, periplakin and desmoplakin

Linker domain sequences of envoplakin (Env_Link), periplakin (PP_Link) and desmoplakin (Des_link) were aligned with the fourth PR motif of PRD C (Des_PRD_C_PR4) and the third of PRD B (Des_PRD_B_PR3) of desmoplakin by CLUSTALW (Thompson et al., 1994) which had both been structurally elucidated. N.B The entire sequence of the linker domains is not shown (as in figure 6.1). Secondary structure elements expected at particular positions are shown (S1: strand 1, S2: strand 2, H1: helix 1, H2: helix 2). Furthermore the numerical position of each amino acid within the 38 amino acid motif is highlighted. Semi colons represent high homology, full stops represent some homology and asterisks represent identical residues. Highlighted in yellow are residues involved in “fixing” the β-harpin in position. Highlighted in red are hydrophobic residues involved in the correct orientation and folding of the anti parallel α-helices correctly.

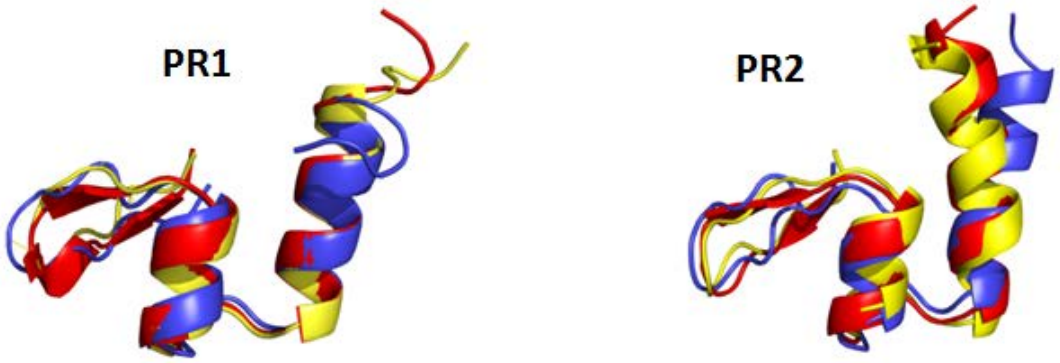


Figure 6.43 Superimposition of plakin repeats 1 and plakin repeats 2 of the desmoplakin, envoplakin and periplakin linker domains

PR motifs from sequence derived models of desmoplakin (yellow), envoplakin (red), and periplakin (blue) linker domains. PR1 of each linker domain is superimposed on the left and PR2 is superimposed on the right. The typical β-harpin and anti-parallel α-helices for canonical 38 amino acid PR motifs was apparent for both PRs for each linker domain.

6.42 SAXS analysis of the desmoplakin and periplakin linker domains

SAXS provides experimental low-resolution structural information in the absence of high resolution NMR or X-ray crystallography data. Given the high sequence similarity and well dispersed 2D-HSQC spectra of each linker domain, it was anticipated that the linker domains would be good candidates for SAXS analysis. For the SAXS analysis, purification was carried out in collaboration with Miss Penelope Rodriguez-Zamora and Dr Claudia Fogl. SAXS scattering curves were successfully acquired for the desmoplakin and periplakin linker domains. SAXS was attempted on the envoplakin linker domain but the protein experienced radiation damage and was subsequently omitted from the analysis.

Signal to noise ratios of buffer subtracted SAXS scattering curves for both the desmoplakin and periplakin linker domains were excellent and superimposed and compared in figure 6.44. R_g values as calculated by GNOM were 1.8 and 2.3 nm for desmoplakin and periplakin respectively. D_{max} values were calculated by GNOM as 5.8 and 8.3 nm for desmoplakin and periplakin respectively (table 6.1). Since the R_g and D_{max} values significantly differ between the desmoplakin and periplakin linker domains it is immediately clear that given the low resolutions permitted by SAXS, both domains exhibit considerable differences in their structures. This can be more clearly observed in the overlay of their respective SAXS scattering curves in figure 6.44. Interestingly the R_g and D_{max} values of desmoplakin and periplakin are similar to that shown for the envoplakin PRD (1.9 and 5.6 nm respectively) (section 5.7), which is twice the size of the linker domains by molecular weight. To ascertain the structures of both domains, *ab-initio* models were created by the program DAMMIF as previously described.

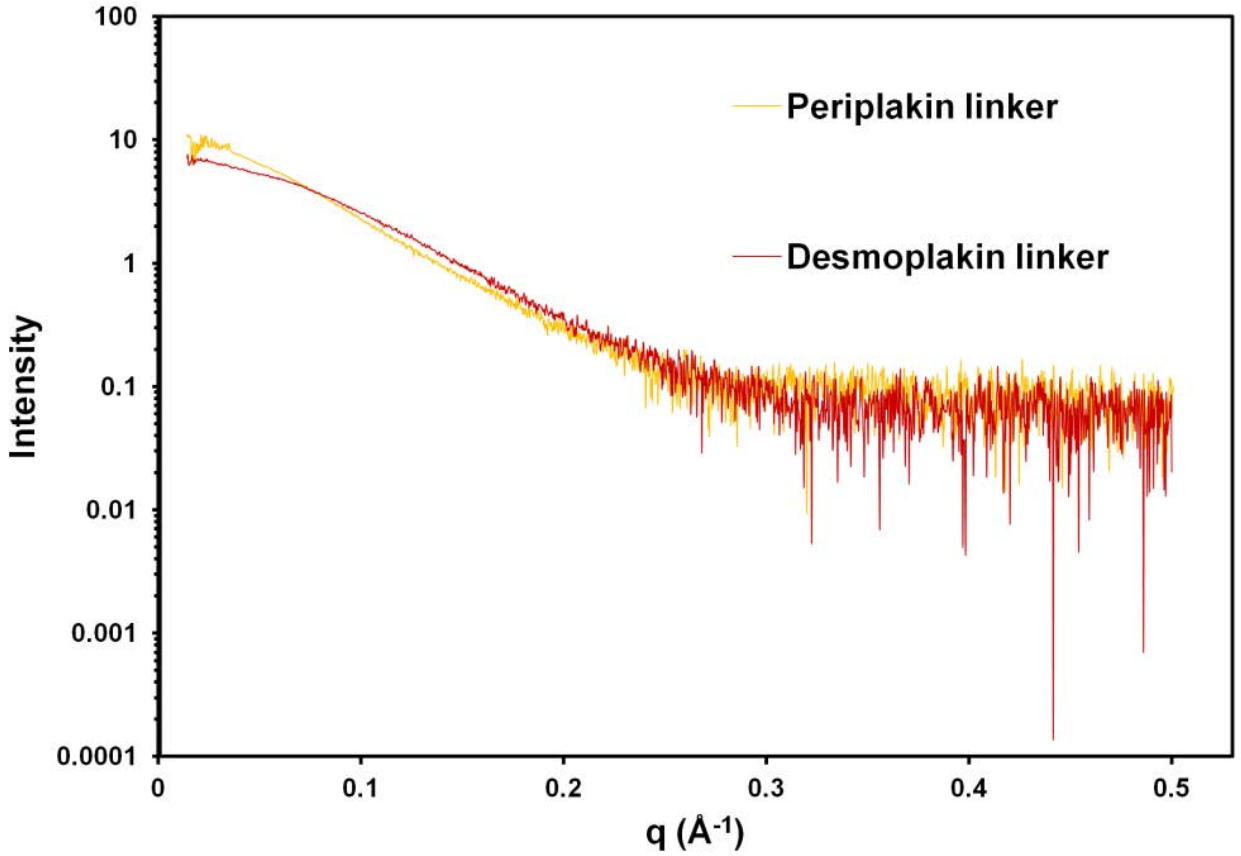


Figure 6.44 SAXS scattering curves of the periplakin and desmoplakin linker domains

Final scaled and merged SAXS scattering curves of the periplakin (yellow) and desmoplakin (red) linker domains are superimposed and shown (utilizing 10 mg/ml data sets). The scattering curves reveal a degree of heterogeneity suggesting their overall conformations in solution are different when analyzed by visual inspection.

Rigid body *ab-initio* modelling of the desmoplakin and periplakin linker domains were conducted. It revealed an extended conformation for both domains which corroborated the R_g and D_{max} values (figure 6.45). Both structures are similar except that periplakin is larger. When a sequence derived I-TASSER model was superimposed onto the *ab-initio* model by SUPCOMB it revealed a large proportion unoccupied for both periplakin and desmoplakin (data not shown). This is further highlighted by theoretical SAXS scattering curves achieving χ^2 values of 1.74 and 4.03 for desmoplakin and periplakin respectively (table 6.1). This suggested that rigid monomeric models were not valid for both desmoplakin and periplakin. However, since the linker domains appeared to adopt similar R_g and D_{max} values to the envoplakin PRD it was possible that in solution they may be forming dimers.

Using the *ab-initio* models as "starting points", two molecules from the model were manually inserted into their respective SAXS derived *ab-inito* models. A theoretical SAXS scattering curve was created based on the manually derived dimeric model for desmoplakin and was superimposed onto the experimental curve. Using CRYSOL a χ^2 value of 2.44 was achieved (figure 6.46). This suggested that the dimeric model is less likely to exist in solution compared to that of the monomeric model given the increase in the χ^2 value and poorer fit to the SAXS scattering curve. Collectively, this suggests that the desmoplakin linker domain is neither a rigid monomeric or dimeric protein and further modelling would be required to ascertain its fold (discussed further in 6.5).

A similar approach was applied to the periplakin linker domain. A rigid dimeric model achieved a significantly lower χ^2 value of 0.84 and also fit the data well (figures 6.47 and 6.48). This suggests that the periplakin linker domain in solution is highly likely to be a dimer. This is in stark contrast to the findings of section 6.2 where AUC analysis suggested the domain was monomeric. AUC analysis for this study was typically conducted at far lower concentrations than

those required for SAXS, NMR and X-ray crystallography. AUC could have been performed at higher concentrations however, was not conducted due to time constraints. However, it is possible to gauge whether this phenomenon was a concentration dependent effect using the SAXS scattering curves at a range of different concentrations and superimpose them on top of one another. When the desmoplakin SAXS scattering patterns were superimposed onto one another, it revealed no differences for desmoplakin at 3.2, 5 and 10 mg /ml (appendix G). Similarly when the SAXS scattering patterns of periplakin were superimposed onto one another it revealed no differences at 1, 5 and 10 mg/ml (appendix H). This suggests that the findings of the linker domain oligomeric states by SAXS analysis and modelling are valid. It also implies that for periplakin the dimerization process is concentration dependent whereby dimerization occurs at a concentration between 10 µM (0.1 mg/ml) and 80 µM (1 mg/ml). Therefore, the dimeric model of the periplakin linker domain appears to represent its structure in solution although the models produced for that of the desmoplakin linker domain did not fit the data and required further exploration.

Protein	R_g (nm)	D_{max} (nm)	CRYSOL monomer (χ^2)	CRYSOL dimer (χ^2)
Desmoplakin linker	1.8	5.3	1.74	2.44
Periplakin linker	2.3	8.3	4.03	0.84

Table 6.1 GNOM and CRYSOL derived SAXS parameters for the desmoplakin and periplakin linker domains

R_g and D_{max} values were calculated for the desmoplakin and periplakin linker domains by GNOM. The data revealed significant differences between the two domains. Sequence derived models were created by I-TASSER and were superimposed by CRYSOL. I-TASSER derived monomeric models gave poor fits with high χ^2 values which suggest these models do not represent the proteins in solution. Similarly a dimeric model fit poorly for the desmoplakin linker domain. A dimeric model however fit well for the periplakin linker with a low χ^2 value suggesting that the periplakin linker domain exists as a rigid dimer in solution.

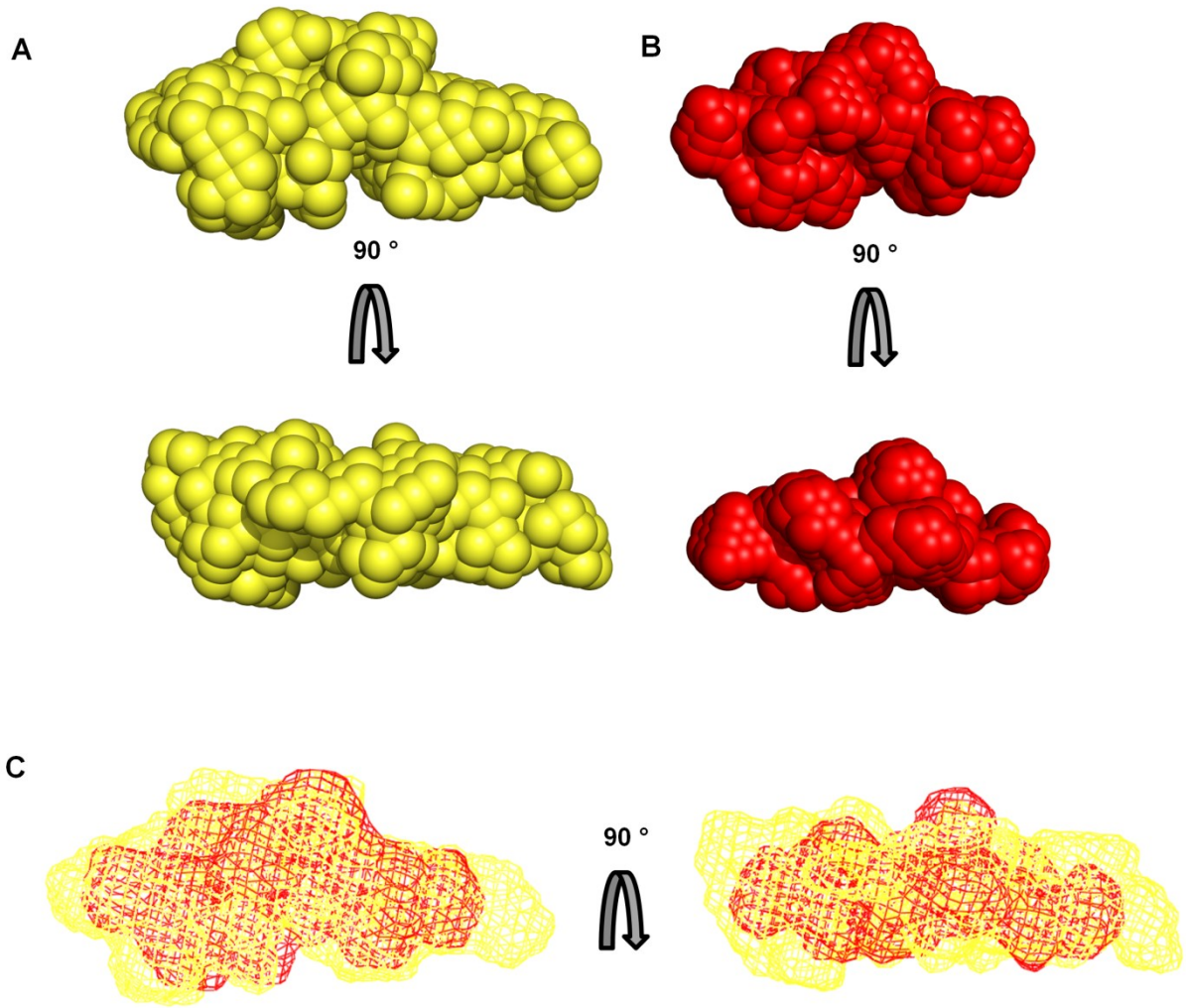


Figure 6.45 Ab-initio models of the periplakin and desmoplakin linker domains

Ab-initio bead models for periplakin (yellow) and desmoplakin (red) were created based on previously described methods and visualised using PYMOL. A: Periplakin linker domain shown in two orthogonal views. B: desmoplakin linker domain shown in two orthogonal views. C: mesh models of desmoplakin (red) manually superimposed onto periplakin (yellow). It reveals a large degree of heterogeneity between the two domains and corroborates the differences observed for the R_g and D_{max} values calculated by GNOM.

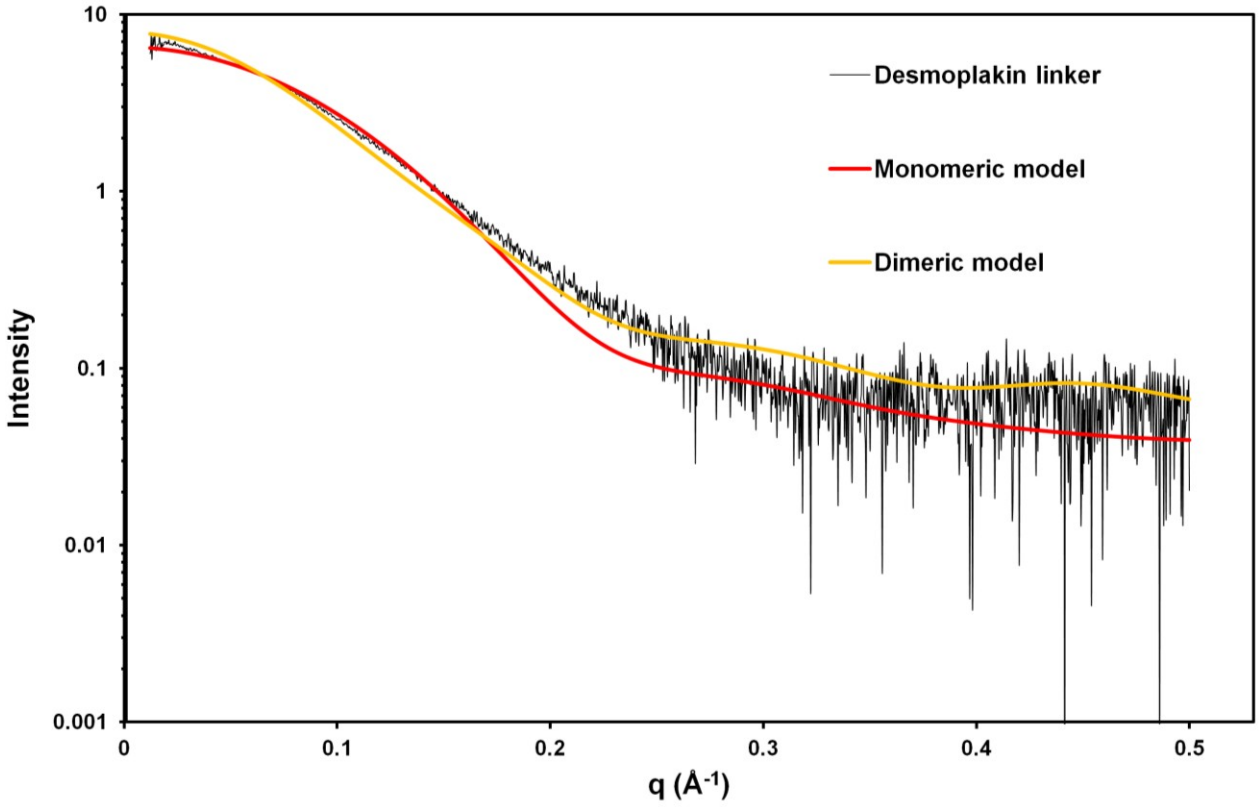


Figure 6.46 SAXS scattering curve of the desmoplakin linker domain

The SAXS scattering curve for the desmoplakin linker domain was created from scaled and merged buffer subtracted data from 10, 5 and 3.2 mg/ml samples (black) in 20 mM Hepes (pH 7.5) and 100 mM NaCl. When a monomeric sequence derived model from I-TASSER had a theoretical SAXS scattering curve superimposed over the experimental curve (red) a poor fit is observed. Similarly, when a dimeric model was created (see text) the fit is even poorer (yellow). The data analysis and purification was conducted collaboratively with Miss Penelope Rodriguez-Zamora.

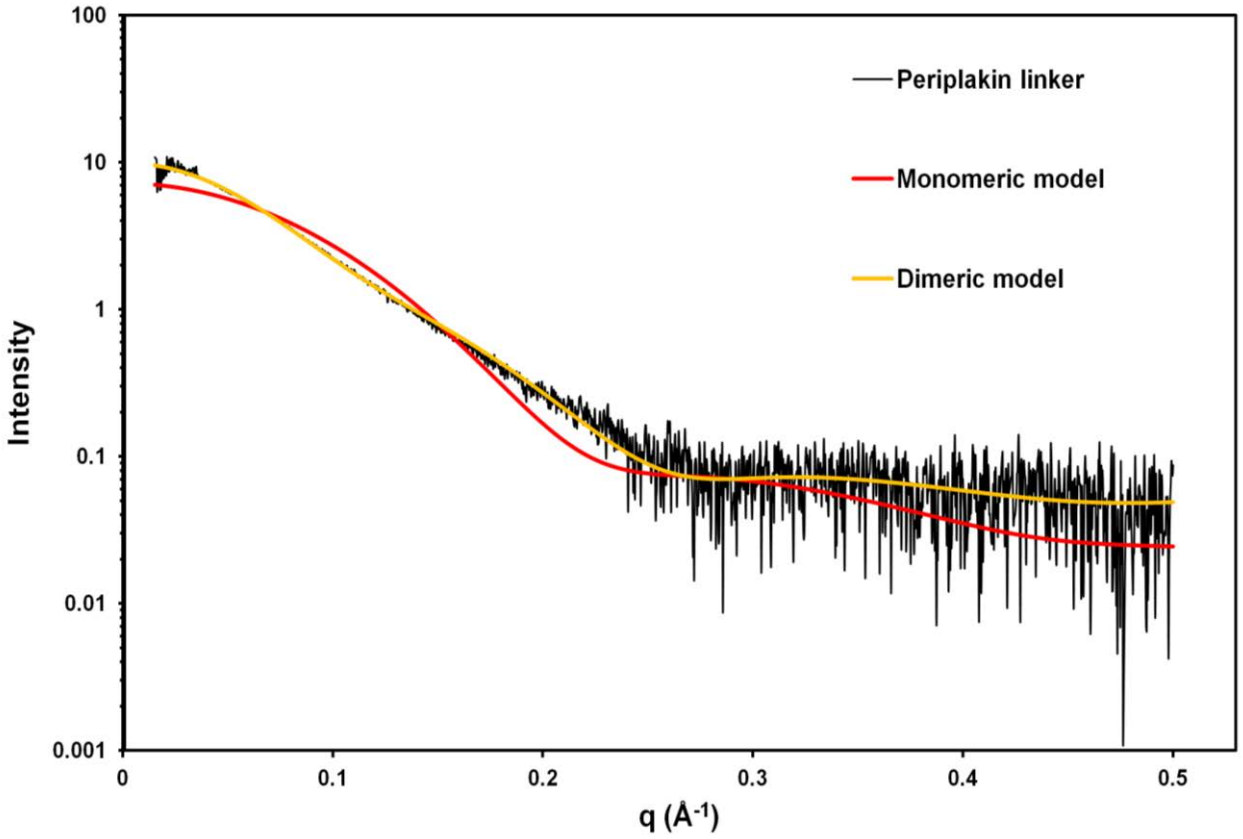


Figure 6.47 SAXS scattering curve of the periplakin linker domain

The SAXS scattering curve for the periplakin linker domain was created from scaled and merged buffer subtracted data from 10 mg/ml, 5 mg/ml and 1 mg/ml samples (black) in 20 mM Hepes (pH 7.5) and 100 mM NaCl. When a sequence derived monomeric model from I-TASSER had a theoretical SAXS scattering curve superimposed over the experimental curve (red) a poor fit is observed. However, when a dimeric model was created (see text) the fit is significantly improved. This suggests that the dimeric model is more representational of the protein in solution. Periplakin linker domain purification was conducted collaboratively with Dr Claudia Fogl.

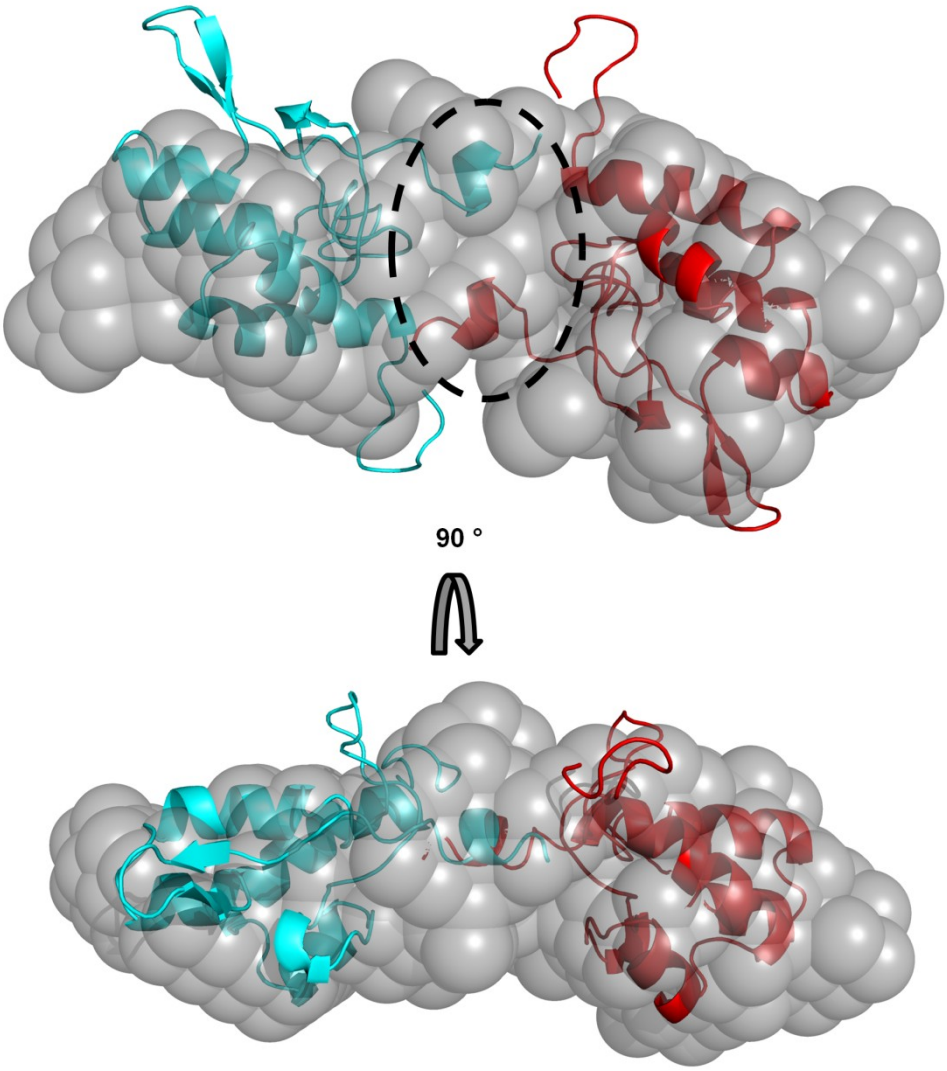


Figure 6.48 Dimeric periplakin linker domain model based on the ab-initio envelope

A dimeric model (molecule 1: cyan, molecule 2: red) for the periplakin linker domain was based on the *ab-initio* envelope created by SAXS analysis and visualised by PYMOL. The manually inserted model represents the best possible fit by visual inspection. Interestingly the N-terminal tails of the individual molecules appear to interact in the centre of the *ab-initio* model (dashed circle).

Two approaches were attempted to structurally characterize the desmoplakin linker domain in order to investigate the possibilities of dynamics, an extended structure or both. The “starting point” for each approach used the sequence derived model by I-TASSER. The first approach assumed that protein was compact (the compact model) while the other assumed the protein is extended (the extended model). Alternative rigid models and flexible ensembles utilized the programmes CRYSOL and EOM respectively. A flow diagram elaborating the two different approaches is given in figure 6.49.

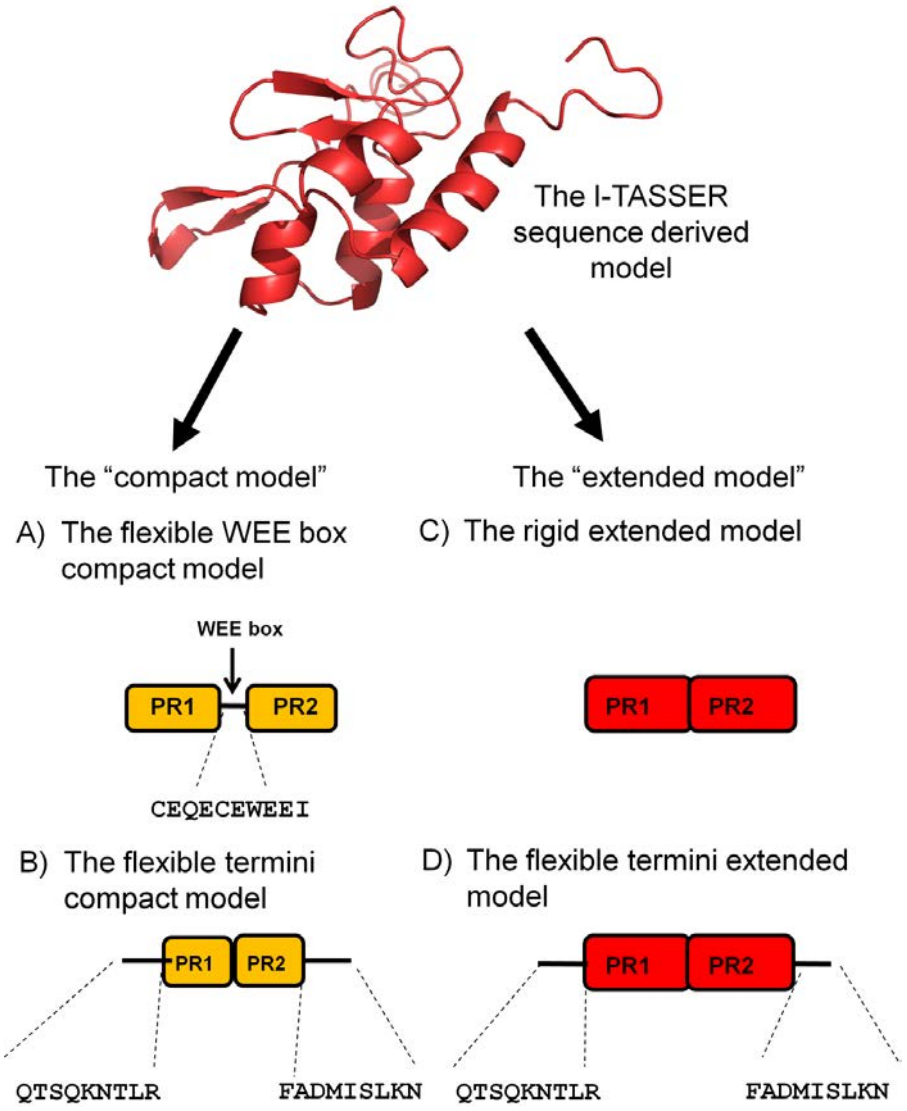


Figure 6.49 A flow diagram representing the approaches to be used for assessing the fold of the desmoplakin linker domain.

To further elucidate the fold and dynamics of the desmoplakin linker domain two approaches were used, utilizing SAXS analysis by EOM and CRYSOL. The first approach "the compact model" assumed the protein adopts a compact conformation. Two models were created using this approach in which one (A) contains a flexible WEE box and the other (B) contained flexible termini (residues indicated), while the rest of the protein was set as rigid. The other approach "the extended model" assumed the protein adopted an extended conformation which was manually extended to fit around the ab-initio model in which one model (C) was entirely rigid and the other (D) contained flexible termini (residues indicated) whereas the rest of the domain was set as rigid.

6.5 Modelling alternative approaches to ascertain the fold of the desmoplakin linker domain

6.51 Modelling dynamics of the proposed compact desmoplakin linker domain model

The first approach assumed the protein was compact based on the sequence based modelling. Plakin repeat motifs were set as structured on the basis of the observation of a mixed α-helical and β-sheet content indicated by 2D-HSQC NMR (figures 6.32-6.34), modelling results (figure 6.41) and sequence alignments to canonical motifs (figure 6.42). This left the N terminus, C terminus and the WEE box, which were of a more ambiguous topology. Two models were therefore created. In the first, the termini were designated as flexible. In the second the WEE box was set as flexible and the rest of the domain as structured. The aim was to investigate which of the two models is more representative of the SAXS scattering data and therefore indicate approximate flexible or extended regions in the linker domain sequences.

The precise residues selected to ascertain flexibility in the two desmoplakin linker domain models are indicated in figure 6.51. The flexible termini model appears to fit relatively well, and has a χ^2 value lower than that of the original compact model fit (1.5 as opposed to 1.7). This suggests that some flexibility is likely to occur within the sequence. However, the fit does not exactly match the SAXS scattering data, and is therefore unlikely to give a significant contribution to the proposed flexibility of the domain. The flexible WEE box model however, matches perfectly throughout the SAXS scattering curve and gives a χ^2 value of 0.74. This suggests that this model is more representative of the data when compared to the original compact model and a flexible termini model. An ensemble of the best 20 models calculated by RANCH and GAJOE is given in figure 6.52. Interestingly, the programme GAJOE appears to

select an ensemble in which the WEE box is predominantly extended and present within the core of the protein. This implied that extension and flexibility of the sequence derived model through the WEE box fitted the data well and was corroborated by the higher than expected R_g and D_{max} values by GNOM. This suggests that the WEE box was significantly contributing to the flexibility of the protein and suggested that the linker domain contains a mix of structured and unstructured elements. It must be stressed again, that given the low resolution of SAXS, the precise amino acid sequence contributing to the flexibility cannot be located. It does suggest however, that the flexibility and extension may be internally located within the core of the domain and that given the WEE box is of unknown structure it was a strong candidate. A second approach was also used to counter this argument by assuming that the entire protein may be extended as the *ab-initio* model suggested.

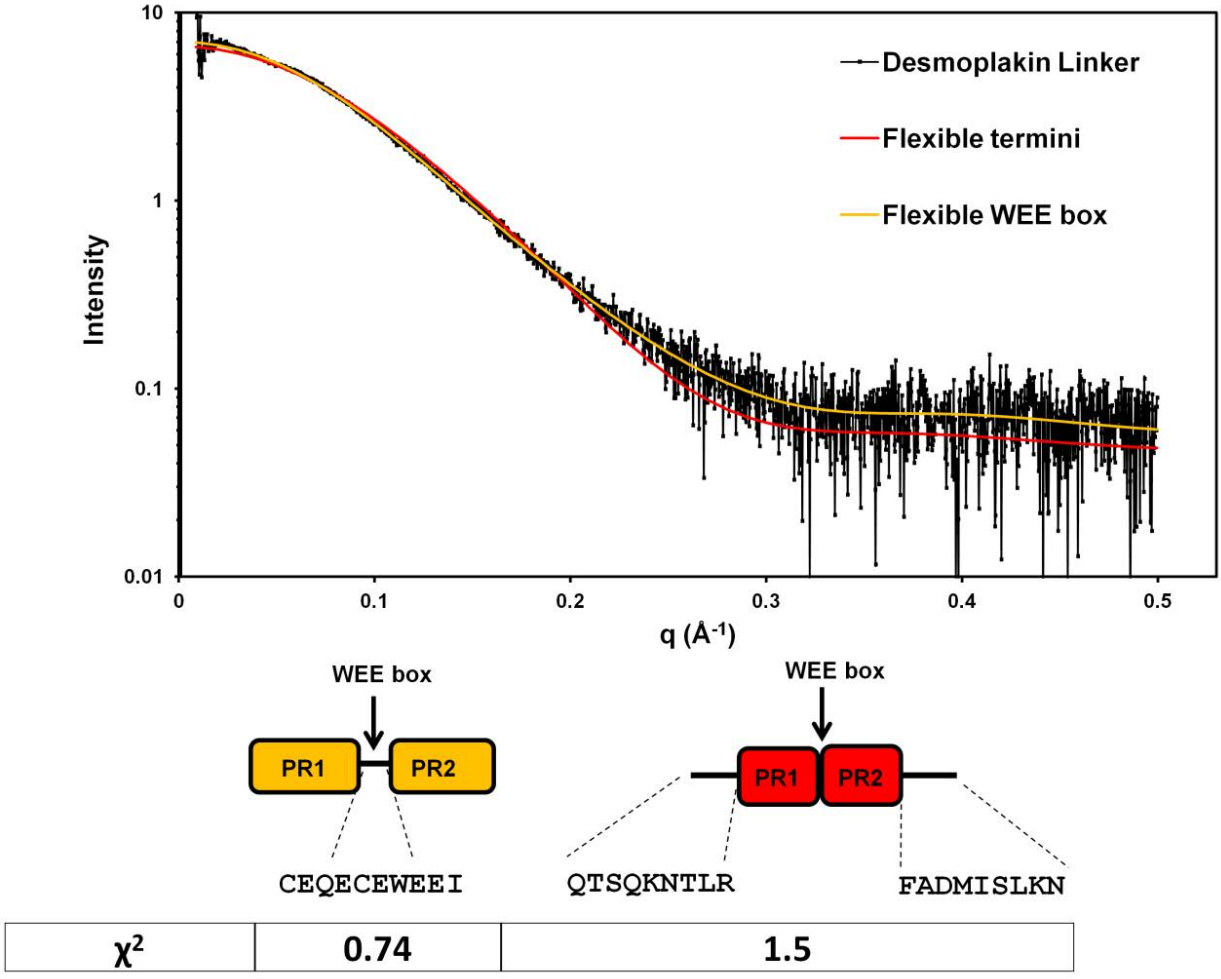


Figure 6.51 Flexibility of the compact model

The compact model was assessed for flexibility using the EOM suite discussed earlier using two different variants. One model had a flexible WEE box and rigid plakin repeat motifs and termini (yellow), while the other had a rigidified WEE box and plakin repeat motifs and flexible termini (red). Indicated are schematic diagrams of the two models, the WEE box and regions of flexibility and their sequences. 10000 random models were created by RANCH and a subset of 20 models were selected by GAJOE that best fit the SAXS scattering curve using the ATSAS online web server (Petoukhov et al., 2007). This revealed that the model containing the flexible WEE box fits the data better than the flexible termini model. This is indicated by the χ^2 values of 0.74 and 1.5 for the flexible WEE box and flexible termini models respectively.

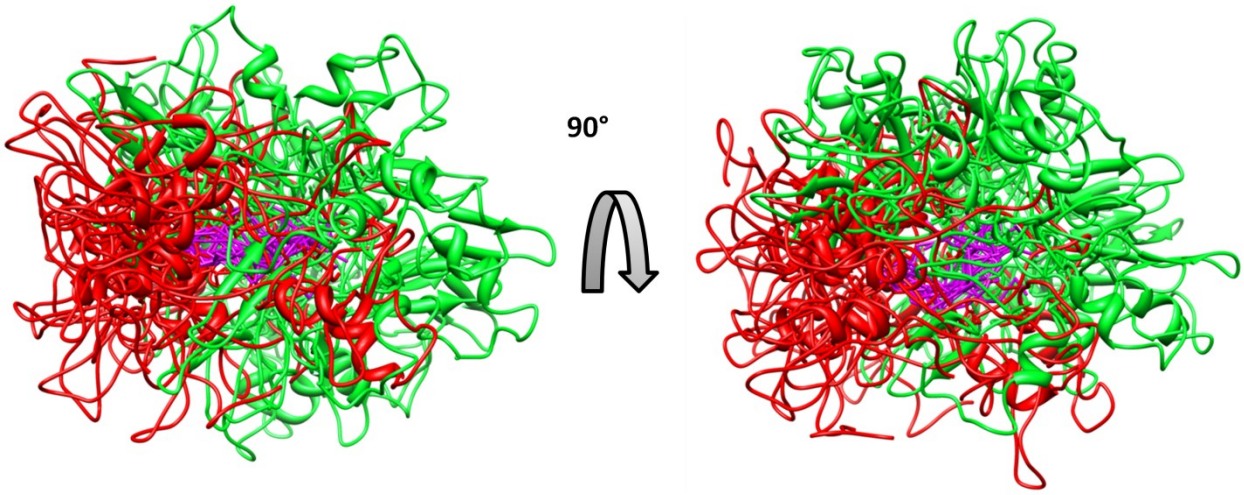


Figure 6.52 Selected ensemble of the flexible WEE box compact model

A subset of the 20 best flexible compact models representative of the SAXS scattering data was superimposed by CHIMERA. For ease of viewing the models were superimposed at the WEE box sequence (purple) with PR1 and PR2 shown in red and green respectively. It reveals extension of the protein through the centrally located WEE box but also a high degree of flexibility.

6.52 Modelling dynamics of the proposed extended desmoplakin linker domain model

To ascertain the orientation of a putative extended model the sequence derived model was used as the starting point. The sequence derived model of the desmoplakin linker domain was remodelled manually about the WEE box keeping it within the core of the protein so that the model fit the *ab-initio* envelope better (figure 6.53). To gauge whether this was representative of the SAXS data, a theoretical SAXS scattering curve of the extended desmoplakin linker domain was superimposed onto the experimental data by CRYSOL. When compared to the original sequence based model this revealed a significant improvement in the χ^2 value which dropped from 1.5 to 0.89 (figure 6.54). However, closer inspection reveals that at higher angles the fit still did not match the SAXS scattering data. This suggested that when compared to the compact model, extension around the WEE box improved the fit to the SAXS scattering curve, but further improvements were still required. Since the N and C termini were set as rigid in this model it was decided that these may by disordered based on the secondary structure predictions. A second extended model was created whereby the core region of the linker domain was set as a structured element (PR1+WEE box+PR2), and the termini were set as flexible. This revealed an excellent fit throughout the SAXS scattering curve with a χ^2 value of 0.74 and suggested this model is also representative of the protein in solution (figure 6.54). An ensemble of the best flexible extended 20 models representative of the SAXS scattering data is shown in figure 6.55. This revealed a modest degree of flexibility in the termini suggesting that the termini may possess secondary structure elements since the termini appear to “cap” the rigid domains. The same ability is retained in PRDs where the non-PR N and C termini are used to cap and correctly fold the PRD (as was the case with the envoplakin PRD). Unfortunately, until structural determination of the termini has been completed this cannot be validated.

Overall, modelling of both a flexible WEE box compact model and a flexible termini extended model both fit the data well. They appear to corroborate each other in that the flexible compact model sets the WEE box as centrally located within the core of the protein, while the flexible extended model also suggested this is required to extend the protein.

Model	χ^2	Fit to the data	Flexible residues
Sequence derived I-TASSER model (monomer)	1.73	No	None
Sequence derived I-TASSER model (dimer)	2.44	No	None
Flexible WEE box compact model	0.74	Yes	2497-2506
Flexible termini compact model	1.51	No	2454-2462 and 2557-2565
Rigid extended model	0.89	No	None
Flexible termini extended model	0.74	Yes	2454-2462 and 2557-2565

Table 6.2 Final modelling and SAXS statistics of the desmoplakin linker domain

CRYSOL (rigid models) and EOM (flexible models) based fitting of the desmoplakin linker domain to its SAXS scattering curves. Important considerations include low χ^2 values (~1) and whether the curve fit the experimental data set. Also indicated are regions selected for flexibility during the EOM calculation. Rows highlighted in grey indicate good matches to the experimental data set

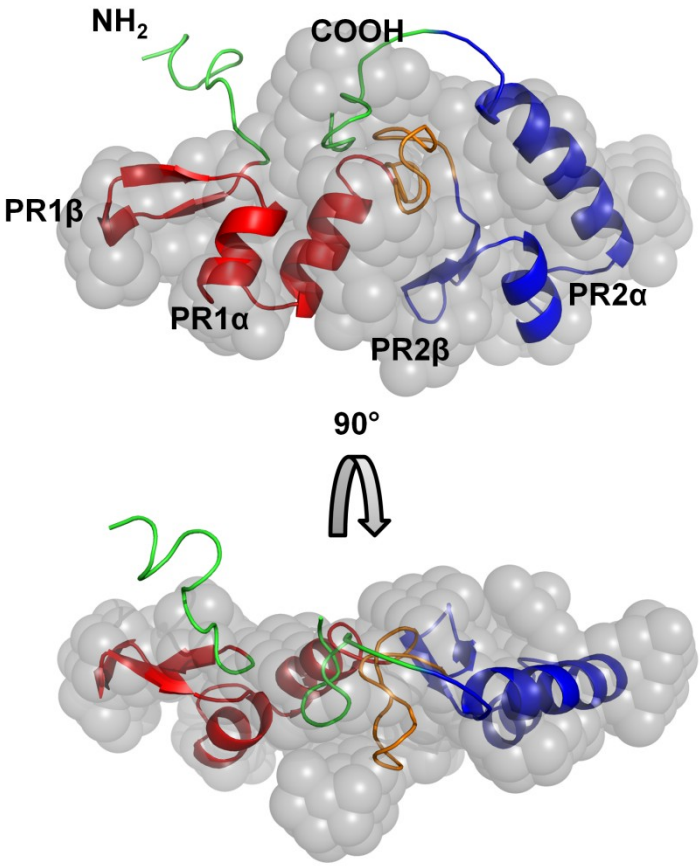


Figure 6.53 Creation of the extended desmoplakin linker domain model based on the ab-initio envelope

An extended model was manually created based on the *ab-initio* envelope created from the SAXS data, using the sequence derived compact model as the starting point in PYMOL. The WEE box was kept centrally located based on the information from the flexible model in 6.52 and also on the fact that it allowed most of the envelope to be occupied by structural regions. Indicated are plakin repeat repeat 1 β-hairpin (PR1β), plakin repeat 1 anti parallel α-helices (PR1α), both in red, the WEE box (in gold), plakin repeat repeat 2 β-hairpin (PR2β) and plakin repeat 2 anti parallel α-helices (PR2α), both in blue.

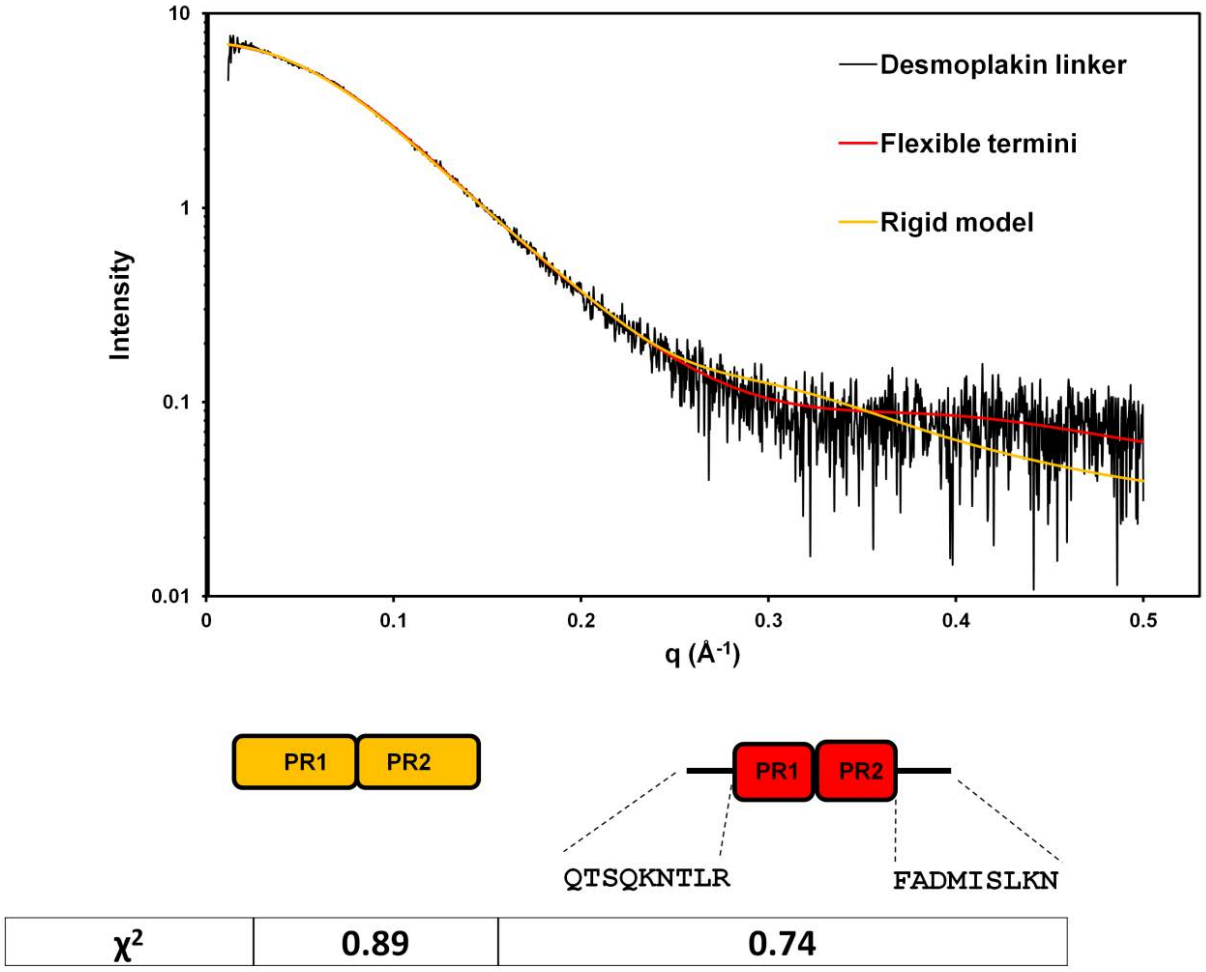


Figure 6.54 Assessment of flexibility using the extended desmoplakin linker domain model

The extended model was assessed for flexibility using the EOM suite discussed earlier using two variants. Both were based on the extended model created in figure 6.53. One was left entirely rigid (yellow), while the other was predominantly rigid with flexible termini (red). Indicated are schematic diagrams of the two models, the WEE box and regions of flexibility and their sequences. 10000 random models were created by RANCH and a subset of 20 models was selected by GAJOE that best fit the SAXS scattering curve using the ATSAS online web server (Petoukhov et al., 2007). It revealed that the model containing the flexible termini best fit the data when compared to the entirely rigid extended model. This is indicated by the χ^2 values of 0.74 and 0.89 for the flexible termini extended model and entirely rigid extended models respectively.

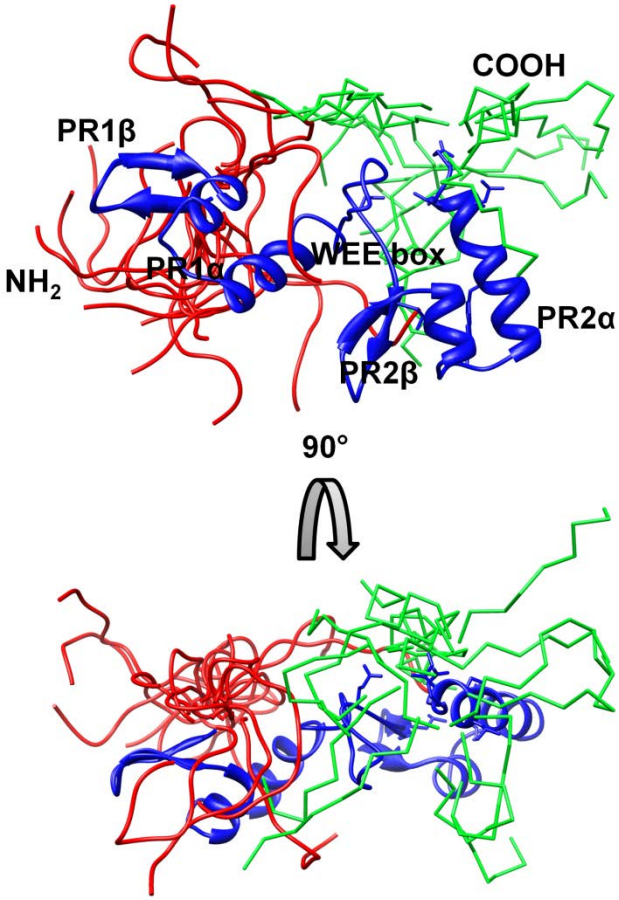


Figure 6.55 Selected ensemble of the flexible termini extended model

A subset of the 20 best flexible termini extended models representative of the SAXS scattering data was superimposed at the rigid PR1, WEE box and PR2 region (blue) by CHIMERA. The rigid PR1, WEE box and PR2 region are shown in blue. It revealed a modest degree of flexibility with the termini folding over to “cap” the rigid region. Also indicated is the N terminus (red), and C terminus (green).

To determine whether similar models were also representational of the periplakin linker domain the same approach used to estimate the conformation of the desmoplakin linker domain was also applied to periplakin. Table 6.3 reveals that the rigid dimeric conformation originally discussed in section 6.42 is still the most valid conformation based on the SAXS analysis.

Model	χ^2	Fit to the data	Flexible residues
Sequence derived I-TASSER model (monomer)	4.03	No	None
Sequence derived I-TASSER model (dimer)	0.84	Yes	None
Flexible WEE box compact model	1.95	No	1689-1698
Flexible termini compact model	4.97	No	1646-1655 and 1748-1756
Rigid extended model	3.72	No	None
Flexible termini extended model	2.02	No	1646-1655 and 1748-1756

Table 6.3 Final modelling and SAXS statistics of the periplakin linker domain

CRYSOL (rigid models) and EOM (flexible models) based fitting of the periplakin linker domain to its SAXS scattering curves. Important considerations include low χ^2 values (~1) and whether the curve fit the experimental data set. Also indicated are regions selected for flexibility during the EOM calculation. Rows highlighted in grey indicate good matches to the experimental data set

6.6 2D-HSQC NMR binding studies using vimentin and the envoplakin and periplakin linker domains

6.61 Investigating the interaction between the envoplakin and periplakin linker domains by 2D-HSQC NMR

Full length periplakin and envoplakin are hetero-dimers (Ruhrberg et al., 1997). The domain responsible for this, the rod domain, is also responsible for homo-dimerization in desmoplakin. Karashima and Watt (2002) suggested that although the rod domain was important for hetero-

dimerization the linker domains of periplakin and envoplakin could also interact and facilitate dimerization even when the rod domains were deleted. As the experiment was conducted *in vivo* it was felt that this interaction would benefit from being studied *in vitro* as it would show a direct interaction and if possible identify residues important to this interaction. A complex of both the envoplakin and periplakin linker domains was not observed by sedimentation velocity AUC (in section 6.2). To confirm that complex formation does not take place NMR spectroscopy was used. Using a 2D-HSQC, unlabelled envoplakin linker domain was titrated against ^{15}N labelled periplakin linker domain. The same experiment was conducted again except that the labelling strategy was reversed. Both titrations indicated that a direct interaction between the envoplakin linker and periplakin linker domains was unlikely to occur due to the lack of any chemical shifts or line broadening (figures 6.61 and 6.62). This was the case even when either unlabelled linker protein was present in molar excess. The results suggested that the findings of Karashima and Watt either needed further investigation or that dimerization was dependent on interaction with another ligand, most likely vimentin. This was suggested as a possibility by Karashima and Watt but was never validated. To confirm this result was valid it was important to verify the functional properties of either the periplakin or envoplakin linker domain. This is due to the possibility of non-native folding caused by *in-vitro* expression in *E.coli* which is devoid of post-translational modifications often required to fold proteins or domains in eukaryotes. To validate the lack of interaction between the periplakin linker and envoplakin linker, unlabelled vimentin was titrated against ^{15}N labelled periplakin linker. This is based on previous observations of the periplakin linker domain binding to vimentin by different labs (Karashima and Watt, 2002, Kazerounian et al., 2002).

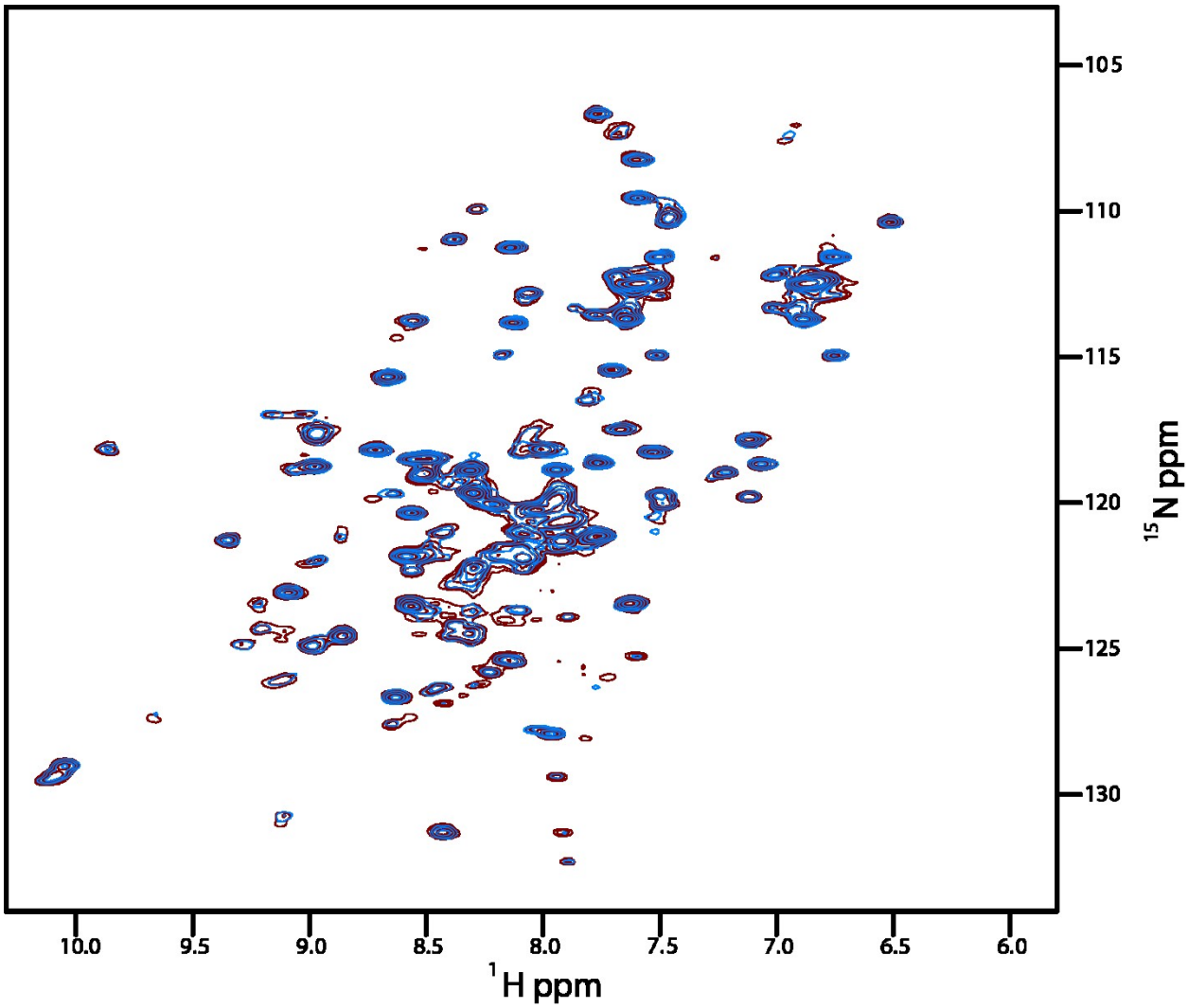


Figure 6.61 ^{15}N labelled periplakin linker titration with unlabelled envoplakin linker domain

A titration of ^{15}N labelled periplakin linker domain against unlabelled envoplakin linker domain was conducted and analyzed by 2D-HSQC NMR analysis. The titration was analyzed on an Agilent 600 MHz spectrometer in 20 mM Hepes (pH 7.2) and 100 mM NaCl at 298 K with periplakin linker at 150 µM alone (blue) and at 150 µM in the presence of 150 µM envoplakin linker domain (maroon). At a ratio of 1:1 an interaction between the two is not observed – as shown by the lack of chemical shift perturbations and signal intensity changes.

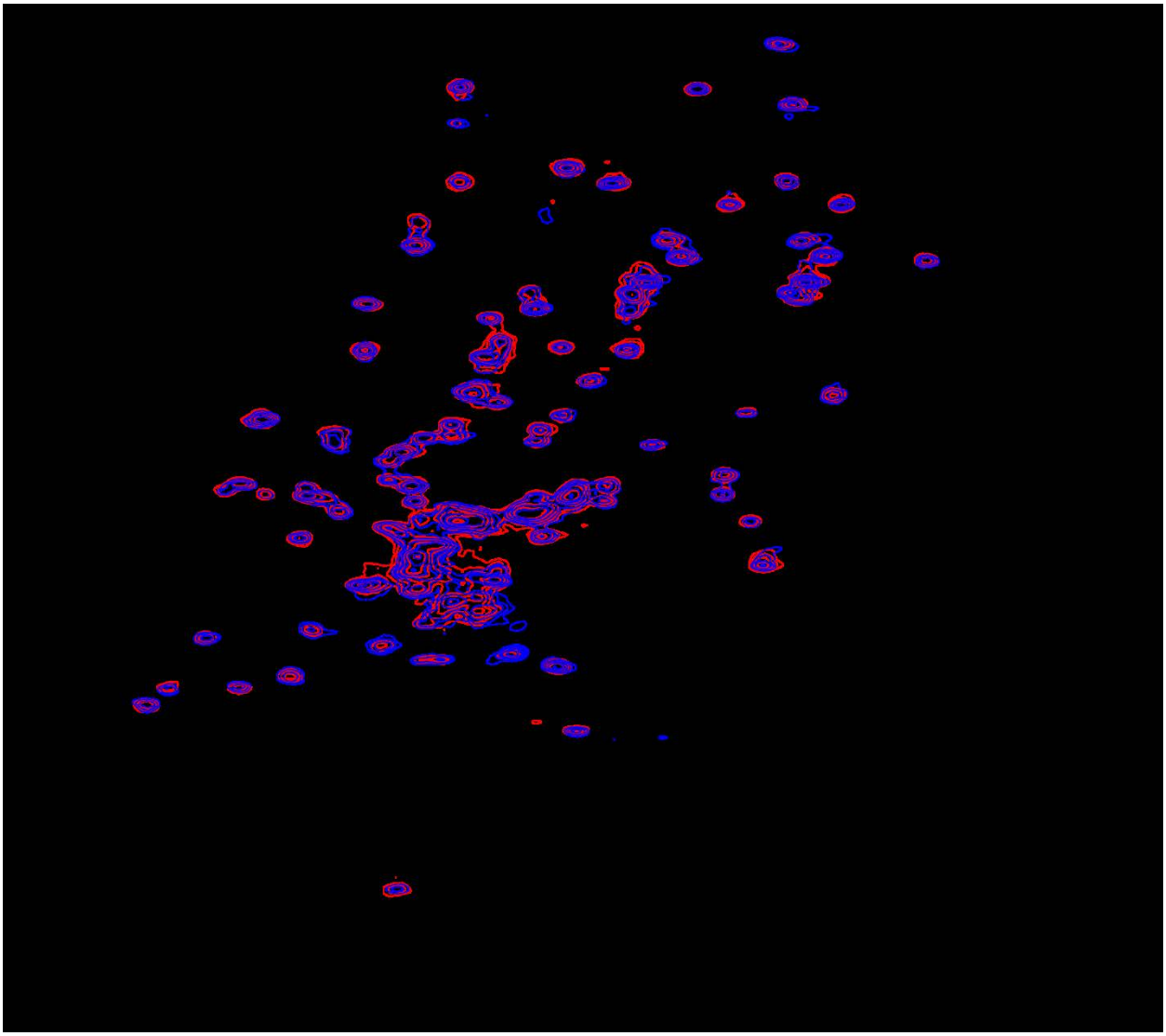


Figure 6.62 ^{15}N labelled envoplakin linker titration with unlabelled periplakin linker

A titration of ^{15}N labelled envoplakin linker domain against unlabelled periplakin linker domain was conducted and analyzed by 2D-HSQC NMR analysis. The titration was analyzed on an Agilent 600 MHz spectrometer in 20 mM Hepes (pH 7.2) and 100 mM NaCl at 298K with envoplakin linker at 250 µM alone (blue) and at 250 µM in the presence of 250 µM periplakin linker domain (red). At a ratio of 1:1 an interaction between the two is not observed – as shown by the lack of chemical shift perturbations and signal intensity changes.

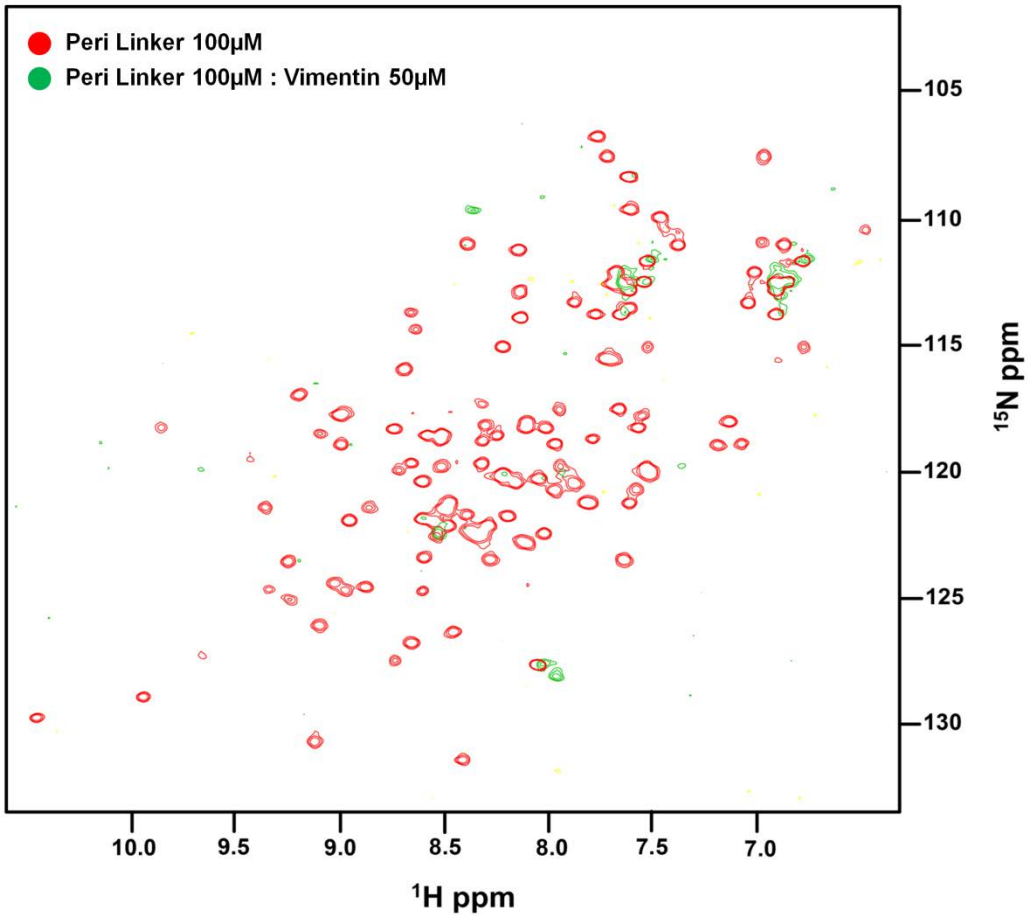
6.62 Investigating the interaction between the periplakin linker domain and vimentin by 2D-HSQC NMR

As stated in the previous chapter vimentin was purified from insoluble inclusion bodies although the folded nature of vimentin was not investigated. Once urea was completely removed vimentin had previously been shown able to independently fold and forms a mixture of monomers and tetramers *in vitro* (Quinlan et al., 1986). For simplicity the following titrations were conducted in the absence of salt (i.e. binding to monomeric/tetrameric vimentin).

Previous reports have shown that the periplakin linker domain is capable of binding to vimentin (Karashima and Watt, 2002, Kazerounian et al., 2002). As well as validating the functional capability of the *in-vitro* purified periplakin linker domain, it was important to validate experiments that had been conducted *in vivo* by an *in vitro* study as it would show a direct interaction. Using a 2D-HSQC, unlabelled vimentin was titrated against ^{15}N labelled periplakin linker in an attempt to document a direct interaction and if possible identify residues important to this interaction. Due to the lack of time and protein only one titration point was used. The first time point used periplakin linker at 100 µM alone. The second used the periplakin linker domain and vimentin at 200 µM and 100 µM respectively, in Tris pH 7.

As expected the periplakin linker domain bound to unpolymerized vimentin. This can be observed by the extreme line broadening of the periplakin linker peaks in the HSQC due to complex formation, which is unresolvable by NMR. Even when the periplakin linker domain to vimentin ratio was as low as 2:1, a significant portion of the peaks disappeared (~80%). The extreme line broadening suggested two possibilities: firstly, the interaction is extremely strong i.e. nanomolar or tighter. This is because if the interaction was weak i.e. in fast exchange, on the NMR time scale, this would be observable by seeing gradual chemical shift perturbations. In addition this it may be a reflection of the fact that interaction of the periplakin linker domain with

vimentin would lead to a complex of molecular weight of at least 65kDa (monomeric vimentin is 53kDa) and the complexed proteins would tumble too slowly for peaks to be observed in an HSQC spectrum. As significant line broadening occurred upon binding of the linker domain to vimentin it was impossible to locate peaks, and so identify residues critical to this interaction. The possibility that the periplakin linker domain binds strongly to vimentin may explain why periplakin has no plakin repeat domain at its C-terminal end. Envoplakin has both a linker domain and a plakin repeat domain (shown in the previous chapter to bind to vimentin). To test whether the ability of the periplakin linker domain to bind vimentin was unique or a common property of all linker domains, the envoplakin linker domain was ^{15}N labelled and titrated against unlabelled vimentin.

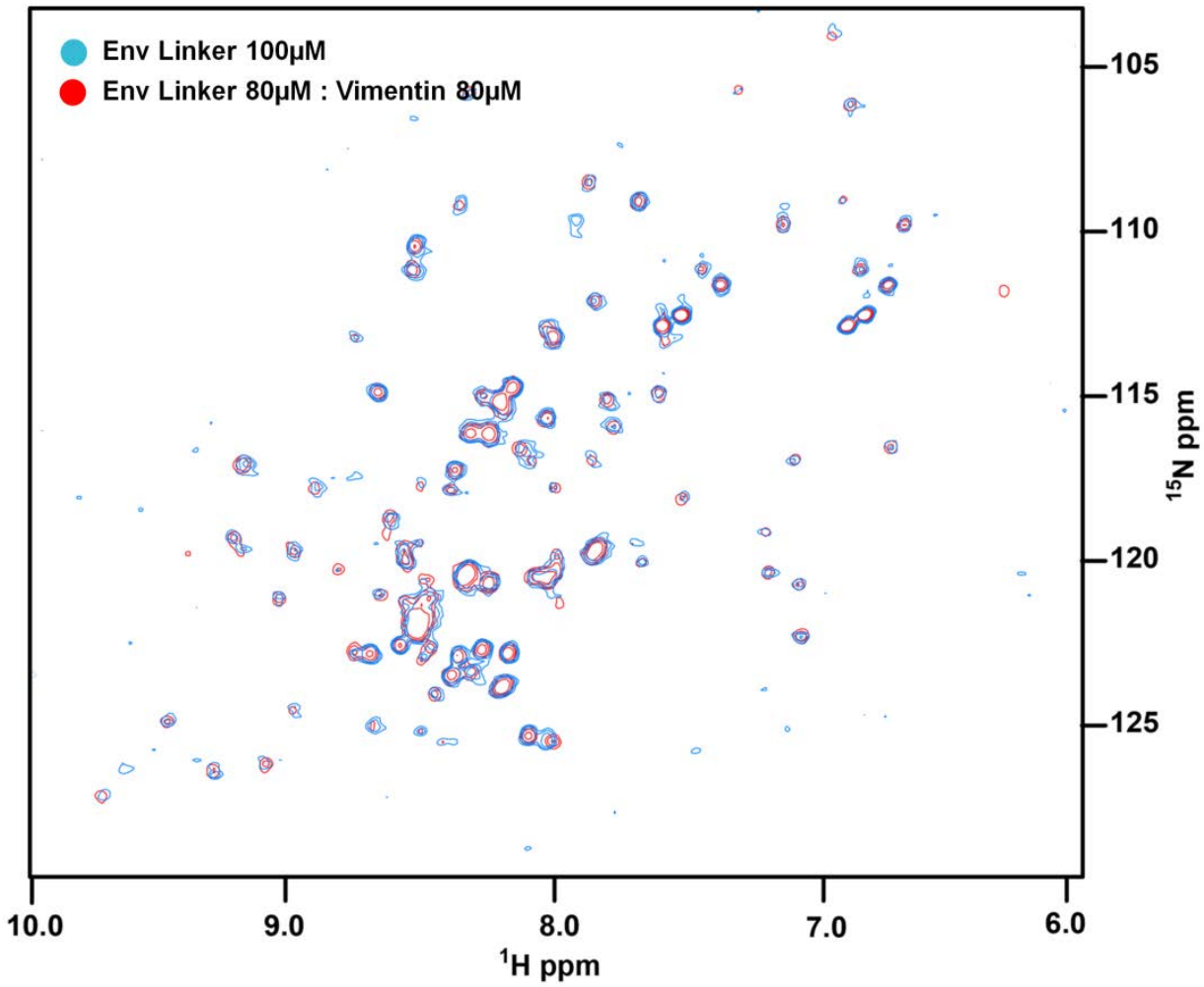


6.63 Titration of unlabelled vimentin to ^{15}N labelled periplakin linker

The proteins were analyzed by an Agilent 600 MHz spectrometer in 20 mM Tris (pH7) with periplakin linker domain alone at a concentration of 100 µM (red) and in the presence of 50 µM vimentin (green). In the presence of vimentin approximately 95% of peaks observed for the periplakin linker alone broaden out. This suggests that at a 2:1 (periplakin linker to vimentin) ratio a complex forms between the periplakin linker domain and vimentin.

6.63 Investigating the interaction between the envoplakin linker domain and vimentin by 2D-HSQC NMR analysis

Karashima and Watt (2002) suggested that in the absence of other domains, the envoplakin linker domain cannot bind to vimentin. To confirm this, unlabelled unpolymerized vimentin was titrated against ^{15}N labelled envoplakin linker domain. The 2D-HSQC showed that the envoplakin linker domain is unable to bind vimentin directly (figure 6.64). This can be shown by the lack of any significant chemical shift perturbations or broadening. Closer inspection suggested that some peaks broaden out. However, this may be an artefact of the decreased concentration of the envoplakin linker domain in the presence of vimentin relative to the envoplakin linker domain HSQC alone. Therefore it is likely that even at a ~1:1 ratio, (which had broadened out the signals of envoplakin PRD and the periplakin linker domain) the envoplakin linker domain and unpolymerized vimentin are unlikely to bind directly.



6.64 Titration of unlabelled vimentin to ^{15}N labelled envoplakin linker

A titration of ^{15}N labelled envoplakin linker domain against unlabelled vimentin was conducted and analyzed by 2D-HSQC NMR analysis. The proteins were run on an Agilent 600 MHz spectrometer in 20 mM Tris (pH 7) with envoplakin linker at 100 μM alone (blue) and at 80 μM in the presence of 80 μM vimentin (red). In the presence of vimentin, even at a ratio of 1:1, significant line broadening does not occur. Approximately 10% of peaks do broaden out or encounter chemical shift perturbations in the HSQC in the presence of vimentin. However this may be an artefact of the decreased concentration of the envoplakin linker domain. Regardless, it appears that the envoplakin linker domain does not bind significantly to unpolymerized vimentin.

6.7 Identification of a putative flexible and highly variable "spacer" sequence proceeding the linker domain

The experiments presented in this chapter suggest that the linker domains within the plakin family are folded protein domains. However, in some plakin family members a preceding amino acid sequence (of about 40 residues) separates the linker domain from the PRD C domain suggesting a specialized element (see figure 6.71). When the sequences of each of the plakin family members were aligned it becomes evident that some members do not possess this sequence and that the sequence is relatively divergent (figure 6.71). The sequence is absent in both periplakin and BPAG1. This is unsurprising for periplakin given that the linker domain is located at the extreme C-terminal end of the protein. However, it is surprising that the spacer region is absent in BPAG1 meaning that its linker domain is directly adjacent to its proceeding PRD C with no other sequence separating the two domains. Furthermore, this sequence in the remaining plakin family members desmoplakin, envoplakin and plectin are not homologous, the lengths are variable for desmoplakin (48), envoplakin (37) and plectin (34) and are predominantly made up glycines and serines. The lack of homology, variability in length and predominance of glycines and serines (amino acids which are typically found in intrinsically disordered regions (Dyson and Wright, 2005)) suggest that this region is in fact a flexible/intrinsically disordered region. When the sequences were individually input into the GLOBPLOT server (Linding et al., 2003), they were predicted to have a high propensity of disorder (data not shown). Finally, when this sequence from a variety of species from the envoplakin protein was aligned, it revealed this region for lower order eukaryotes (such as *Xenopus Laevis*) is much more divergent (figure 6.72). Again this is another indication of flexibility since regions with high conservation from lower to higher order eukaryotes are often structured and folded domains. Collectively this suggests that this region - that I would like to

tentatively call the “spacer” – (as it separates the linker domain from the PRD C domain) is a flexible element and since it is distinct from the linker domain, may have implications for IF binding.

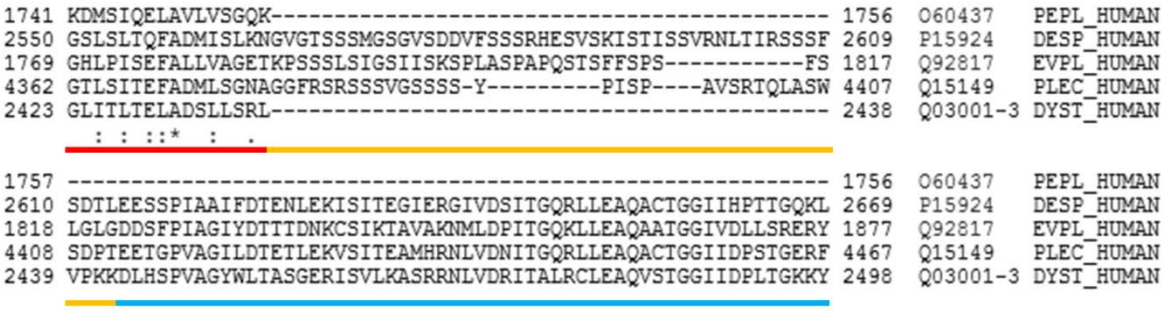


Figure 6.71 Alignment of the "spacer" region from the plakin family members

Full length human protein sequences from the plakin family members: periplakin (PEPL_HUMAN), desmoplakin (DESP_HUMAN), envoplakin (EVPL_HUMAN), plectin (PLEC_HUMAN), and BPAG1 (DYST_HUMAN) were aligned using CLUSTALW (Thompson et al., 1994). Highlighted are the C-terminal end of the linker domain (red), the spacer region (yellow) and the first 55 residues of PRD C domain (cyan). The sequence is absent in periplakin and BPAG1 and reveals a large degree of heterogeneity between the remaining plakin family members. The "spacer" region is predominantly made up of glycines and serines indicating possible flexibility.

Figure 6.72 Alignment of the "spacer" region from varying species of envoplakin

Full length protein sequences from varying species of envoplakin were aligned using CLUSTALW (Thompson et al, 1994). Highlighted are the linker domain (red), the spacer region (yellow) and the PRD C domain (cyan). The sequence is highly conserved from human (EVPL_HUMAN) to turkey (G1MZ10_MELGA) however is divergent in *Xenopus Laevis* (F6V862_XENTR) – indicated by red arrowheads. Since structured domains are often highly conserved from higher to lower order eukaryotes (and occasionally yeast), the divergence in *Xenopus Laevis* implies this region is likely to be unstructured and therefore flexible/intrinsically disordered.

6.8 Chapter 6 conclusions

The C-terminal tails of plakin family members are crucial to interacting with cytoskeletal proteins. Previous studies have highlighted the interdependent nature of the PRDs and linker domains with relation to cytoskeletal binding (Choi et al., 2002, Fontao et al., 2003, Lapouge et al., 2006, Meng et al., 1997b, Stappenbeck et al., 1993, Stappenbeck and Green, 1992). Further studies have shown that the linker domains in isolation are the most critical determinant of cytoskeletal preference (Favre et al., 2011, Karashima et al., 2012, Karashima and Watt, 2002, Kazerounian et al., 2002). To date no biophysical data is known of the linker domains, and they are simply viewed as "tethers" linking PRDs B to PRD C. However, some reviews have suggested the presence of up to 2 PR motifs (Leung et al., 2002, Jefferson et al., 2004). By analyzing the biophysical properties of linker domains from 3 plakin family members it was hoped a better understanding into how cytoskeletal preference is created would be observed. This study used a combination of solution based biophysical techniques, including NMR and SAXS, which allowed for the biophysical properties of linker domains to be probed. The novel results presented in this study indicate that unlike previous reports, the linker domains are fully folded PR motif containing domains with differential folding capabilities between desmoplakin, periplakin and envoplakin. A number of functional aspects of the linker domains were also corroborated which included cytoskeletal binding and lack of hetero-dimerization of envoplakin and periplakin linker domains, unlike previous suggestions. An additional highly variable and putatively flexible "spacer" region was also identified. Finally it must be noted that due to time constraints the PR motifs of the linker domains investigated in this thesis were not independently verified using experimental atomic resolution data. However, recent NMR backbone assignments of the desmoplakin linker domain have revealed the presence of 2 PR motifs concordant with the

suggested secondary structure in figure 6.42 (unpublished findings: Miss Penelope Rodriguez-Zamora).

Previous reports had suggested that the linker domains may possess up to 2 PR motifs. The possibility of which was initially investigated by sequence alignment and modelling which revealed the presence of two PRs based on the canonical 38 amino acid sequence (figure 6.42). The models suggested that the PR motifs retain their abilities to form intra-domain and inter-domain hydrophobic and electrostatic binding for correct fold in a manner reminiscent to those observed of the envoplakin PRD in chapter 5 (figure 6.43). The presence of a mixed α-helical β-sheet structure was also indicated in the 2D-HSQCs of each of the linker domains (figures 6.32-6.34). Furthermore the high quality dispersion and isotropic peaks suggested the domain was well folded. This confirmed that unlike previous reports the linker domains from 3 different plakin family proteins are independently folded. The presence of 2 putative PR motifs also implied that they must adopt novel mechanisms to assume their fold given the results of chapter 5 which suggested that a typical PRD must contain a minimum of 4.5 PR motifs. Unfortunately, time constraints did not allow for a full structural determination of any of the linker domains to be conducted. However, when analyzed by SAXS the linker domains of desmoplakin and periplakin revealed some novel insights which suggested that regardless of their high sequence homology they adopted different topologies.

Based on the SAXS data an unexpected extended or flexible conformation for both desmoplakin and periplakin were implied. This is in contrast to the I-TASSER sequence derived models produced for the linker domains. The extended nature was more pronounced for the periplakin linker domain relative to the desmoplakin linker domain (table 6.1). Given that the linker domains are half the size of a typical PRD (in terms of their sequence derived molecular weight) it is interesting to observe that the envoplakin PRD had similar R_g and D_{max} values. Collectively, this

implied the linker domains were extended with the extension being more prominent in periplakin. Modelling of a monomeric and dimeric structure for the desmoplakin linker domain revealed that neither fit the data, implying a more complex structure (figure 6.46). However, a dimeric model for the periplakin linker domain fit the data well which implied that it was a dimer in solution (figure 6.47). The specific molecular mechanisms underlying periplakin linker domain dimerization is beyond the resolutions permitted by SAXS. However, the N-terminal regions may play a role as implied by the dimeric model (figure 6.48). Further inspection of the N-terminal sequence of the periplakin linker domain did not clarify the situation. The N-terminal region of periplakin contains charged residues which may interact in an electrostatic manner between the two molecules of the dimer. However, the same residues are semi-conserved in the desmoplakin sequence, which did not form a dimer. Conversely, it is possible that the periplakin linker domain undertakes dimerization using its WEE box. The importance of the WEE box has been further highlighted by protein insolubility upon bacterial expression caused by the mutagenesis of the tryptophan to an alanine (Personal communication: Miss Penelope Rodriguez-Zamora). Furthermore this is corroborated by the *in vivo* binding studies of Karashima and Watt (2002), which showed the deletion of the WEE box plus the proceeding isoleucine, prevented binding of the periplakin linker to vimentin. Regardless of the mechanism that may be adopted to create a periplakin linker domain dimer its size similarity to a PRD suggested that it may in effect adopt a PRD-like topology. Unfortunately, the possibility of a dimeric periplakin linker domain raises a concern in terms of its *in vivo* behaviour. The full length periplakin protein is typically observed as a heterodimer with the envoplakin protein (Ruhrberg et al., 1997). Homo-dimerization would therefore not be possible since another periplakin protein would be required. There are reports of a periplakin homo-dimer (Kalinin et al., 2004, Ruhrberg et al., 1997) however, the functional differences between this state and the hetero-dimer are unknown. The presence of a functional dimeric periplakin linker domain therefore raises concerns about which

periplakin protein (homo/hetero-dimer) is the functionally active protein. While Ruhrberg et al (1997) implied the formation of homo/heterodimers of periplakin (and envoplakin) by sequence homology, Kalinin et al (2004) confirmed that periplakin is typically observed as a heterodimer with envoplakin. Kalinin et al also observed that envoplakin was insoluble without the presence of periplakin. Finally, they also observed the presence of higher order oligomers of the periplakin/envoplakin complex *in vivo*. If true, it may explain the homodimerization of the periplakin linker domain by stabilization of the tetrameric periplakin/envoplakin oligomers (and higher). This clearly needs more work to validate. Interestingly, the known homo-dimeric protein desmoplakin in contrast is likely to contain a monomeric linker domain.

To clarify the fold and topology of the desmoplakin linker domain two different modelling approaches were used, based on the differences observed in the SAXS scattering curve between desmoplakin and periplakin. One such approach adopted the *ab-initio* envelopes generated by DAMMIF and manually constructed an extended model whereby the core region was rigid yet extended at the same time (figure 6.53). This model revealed a very good fit when the N and C termini were set to disordered by EOM (figure 6.54). Unfortunately a compact model with a flexible WEE box model similarly gave good fits for the desmoplakin linker domain. Therefore, it is not possible to confirm one model over the other due to the low resolutions of SAXS analysis. However some assumptions can be made. The first, being that the desmoplakin linker domain is highly likely to be monomeric given the consistently lower χ^2 values observed for monomeric models (table 6.2). The second, that the desmoplakin linker domain is likely to utilize a form of inter-domain interaction divergent from that of the periplakin linker domain and that observed of typical PRDs. The cause of this structural divergence may linked to the architecture of their respective C-terminal tails. The periplakin linker domain is the only domain present at its C-terminal tail. In contrast, the C-terminal tail of desmoplakin contains a PRD A, PRD B, the

linker domain and PRD C. The architecture and the fact that the C-terminal tails of both desmoplakin and periplakin interact with different affinities to cytoskeletal proteins implies possible redundancy in the structured domains of the desmoplakin C-terminal tail. The desmoplakin linker domain is not required to bind to cytoskeletal proteins as shown by Weis and colleagues (2002) while the periplakin linker domain is required for cytoskeletal binding. Therefore these differences may be explained by the different folds they adopt as demonstrated by this study. The differences in possible functional outcomes of the linker domains were therefore further investigated.

The linker domains of periplakin and envoplakin were investigated for their ability to bind vimentin and interact with one another as previously described (Karashima and Watt, 2002). A 2D-HSQC titration series of ^{15}N labelled periplakin to unlabelled envoplakin was conducted and vice versa (figures 6.61 and 6.62). Both titrations revealed that the two linker domains cannot bind to one another even in molar excess of the titrated ligand. This is in stark contrast to the findings of Karashima and Watt (2002) which suggested that when the rod domain (the domain involved in hetero-dimerization of the two full length proteins) was deleted, binding could still take place. Closer inspection of the constructs used to test this interaction in the original study revealed that an approximately 50 amino acid sequence derived from the rod domain was present in the rod domain “deleted” full length sequence. They conceded that the presence of this sequence may have artificially allowed hetero-dimerization and based on the titration results of this study is likely to be the case. The matter is further complicated when the cytoskeletal protein vimentin is involved in binding.

The linker domains of envoplakin and periplakin were tested for their ability to bind to unpolymerized vimentin by 2D-HSQC NMR. This revealed differential binding abilities between periplakin and envoplakin which showed that periplakin was the only linker domain capable of

binding to vimentin. It showed saturation of binding at a ratio of 2:1 (periplakin to vimentin) (figure 6.63). In stark contrast, envoplakin showed no binding at a ratio of 1:1 (figure 6.64). Binding of the desmoplakin linker domain to vimentin was not tested but showed no binding to the heart cytoskeletal protein desmin in its unpolymerized state (Personal communication: Miss Penelope Rodriguez-Zamora). The lack of interaction between desmin and the desmoplakin linker domain corroborates the idea that the presence of PRDs leads to redundancy in desmoplakin C-terminal domains. However, previous reports had shown that the desmoplakin linker domain is critical for cytoskeletal preference (Lapouge et al., 2006). While this has not been explored in this study it is likely to be true as implied by the structural differences between the periplakin and desmoplakin linker domains.

The periplakin linker domain had previously been shown to bind to vimentin *in vivo* (Karashima and Watt, 2002, Kazerounian et al., 2002). Karashima and Watt (2002) proposed that the envoplakin and periplakin linker domains co-localise to vimentin with periplakin enhancing the binding propensity of envoplakin to this trimeric complex. The findings reported in this thesis corroborate the earlier results of Karashima and Watt as it is possible that the periplakin linker domain firstly binds to vimentin which allows a conformational change in either vimentin and/or periplakin that is more conducive for envoplakin linker domain binding.

The matter is further complicated by the observation of a putatively flexible and highly variable “spacer” region observed in envoplakin, desmoplakin and plectin (figure 6.71). Flexibility was proposed to be involved in increased affinity to cytoskeletal proteins by Weis and colleagues (2002). This was deduced based on a series of interactions using different constructs from the desmoplakin C-terminal tail. A construct encompassing PRD B, the linker domain, the spacer region and PRD C had the highest affinity for binding to vimentin compared to the domains in isolation (Choi et al, 2002). Given the high sequence divergence between the observed spacer

regions, it may reflect another structural mechanism for cytoskeletal preference and/or binding affinities. Indeed, intrinsically disordered regions have been reported to enhance or coordinate binding of their adjacent structured domains to ligands (Nagulapalli et al., 2012, Vuzman and Levy, 2012). Constructs encompassing PRD B to PRD C from desmoplakin and the linker domain to PRD C from envoplakin (which each contain a spacer region) were relatively unstable during purification (Personal communications: Dr Claudia Fogl and Miss Penelope Rodriguez-Zamora). This is likely to reflect spontaneous degradation or non-specific cleavage of the spacer region typical of highly disordered regions and corroborates the idea that this region is flexible. These constructs require further investigation, possibly by SAXS. It is possible that their roles in the overall structure topology and functional implications of the C-terminal tails can be investigated, since it is possible to purify these domains to homogeneity as shown by Weis and colleagues (2002).

Overall, this study has highlighted that the linker domains are fully folded domains with 2 putative PR motifs. They also show different folding capabilities which may reflect the cytoskeletal preferences observed in the literature between different plakin proteins. In addition a putative and highly variable “spacer” region may further contribute to the functional implications of C-terminal tails. This work may therefore highlight the differences in plakin protein cytoskeletal preferences and variability of C-terminal tails which may give rise to their specific roles in the cell.

Chapter 7- Biophysical implications of ARVC missense mutations in desmoplakin: the traffic light system

ARVC is a complex disease, characterized by fibrofatty replacement in the right ventricle which is often life threatening and can lead to sudden cardiac death. The symptoms of ARVC include dizziness, fatigue, breathlessness and ECG abnormalities, amongst others (Marcus et al., 2010, McKenna et al., 1994). ARVC is a genetic disease caused by mutations in proteins expressed in the cardiac desmosome (Awad et al., 2008). However, the link between these mutations and the phenotype is unclear. The matter is further complicated by the disease being difficult to diagnose due to its phenotypic similarities to other cardiomyopathies (such as long QT syndrome) which can lead to incorrect diagnosis of a patient (although great improvements have been made in the past 10 years). Furthermore, ARVC mutations are thought to have incomplete penetrance meaning that one member of a family who has a disease causing mutation may not be affected, while another with the same mutation may be affected. Diagnosis of ARVC requires the usage of the taskforce criteria where some symptoms are classed as “major” while others are classed as “minor” (McKenna et al., 1994). Furthermore symptoms which fulfil classic criteria of ARVC may only arise at a later age i.e. beyond their 20s, as highlighted by one study on 3 Dutch sisters (Otterspoor et al., 2007). A mutation in a cardiac desmosome gene has recently been increased

to a major criterion for fulfilling ARVC diagnosis further highlighting the need to understand the link between a mutation and the phenotype (Marcus et al., 2010).

At the genetic level there are a plethora of studies which suggest certain mutations are disease causing, but there are very few papers on the mutations' direct impacts on their protein products. The ARVC database which contains details of over 800 mutations is the sole primary source of information on these mutations (van der Zwaag et al., 2009). Unfortunately, the ARVC database is likely to harbour many polymorphisms as there appears to be little (if any) regulation as to what is included. Furthermore, the ARVC database classifies each mutation using the "Grantham score" a scoring system which calculates the proposed difference in amino acid structure caused by a mutation. The problem with the Grantham score is that it disregards the amino acid's position in the tertiary structure of the protein (Grantham, 1974). The lack of a link between the genotype and clinical phenotype therefore appeared to require further attention. To date only a few studies have investigated the functional and/or structural consequences of mutations at the molecular level.

Yang et al investigated the direct functional consequences of 2 ARVC mutations at the extreme N-terminus of desmoplakin (V30M and Q90R) (Yang et al., 2006). The study found that plakophilin binding was adversely affected but plakoglobin binding was unaffected (Yang et al, 2006). This was the first study to directly show the adverse functional consequence of an ARVC mutation at the molecular level. Another very recent study investigating the biophysical effects of ARVC mutations on the ARM domain of PKP2a (where the majority of desmosomal ARVC mutations are reported) has recently been published (Kirchner et al., 2012). The authors suggested that ARVC mutations had varying degrees of biophysical effects on the ARM domain of PKP2a. The authors crystallized a WT fragment from the ARM domain of PKP2a which showed homology to the previously crystallized ARM domain of PKP1 (Kirchner et al, 2012).

This allowed them to investigate the effects of the mutations on the entire PKP2a ARM domain by modelling. They showed that some ARM domain constructs with disease causing ARVC mutations (I487S and V453I) could be purified in similar amounts to the WT protein. However, ARM domain constructs incorporating the disease causing ARVC mutations C752R, S571F and C649fsX697, were unstable, resulting in degradation of these proteins. They concluded that the altered amino acids I487S and V453I which resulted in stable ARM domain constructs were not involved in structural stabilization/folding events. However, the mutations that resulted in protein degradation (C752R, S571F and C649fsX697) were essential for inter and intra domain stabilization/folding effects. The study showed for the first time that some ARVC mutations were more deleterious than others. Therefore, it appears that ARVC mutations can cause functional as well as structural consequences at the molecular level which may rationalise their pathogenecity. The recent study on PKP2a also highlights the possible presence of polymorphisms which may have previously been incorrectly suggested to “cause” ARVC.

The two studies discussed above focused on specific regions of their target proteins; the desmoplakin extreme N-terminal and the PKP2a ARM domain. The main focus of both studies was on specific domains as opposed to investigating different mutations present in varying structural domains which adopt different folds. The experiments described in this chapter investigated mutations in desmoplakin across a variety of domains and therefore it was hoped that by gauging their biophysical implications it would answer the following questions:

1) What is the structural consequence of ARVC mutations in desmoplakin?
2) Are certain mutations more deleterious to protein structure than others?

The ultimate aim of the experiments described in this chapter was to gauge whether some information could be used to aid cardiologists in facilitating diagnosis based on genetic

screening results of the patient. Part of the findings presented in this chapter was published in the Journal of Molecular Biology (see appendix A1).

7.1 Mutation selection criteria

Over 800 mutations are reported to date in the ARVC database (van der Zwaag et al., 2009) and the choice of which mutations to investigate needed to be narrowed down. It appeared more pertinent to investigate those more susceptible to deleterious effects physiologically and therefore only mutations classed as “pathogenic” were selected. So that direct comparisons could be made between the WT and mutant variants only missense mutations were selected. Next, since the focus of this study was the plakin family and only mutations in desmoplakin (as opposed to envoplakin or periplakin) have been reported to cause ARVC, this was the selected as the target protein. Finally the mutations were selected based on the constructs to hand. Using this approach allowed for a wide variety of different domains and amino acid substitutions to be compared. The final mutations selected for biophysical analysis were: K470E, R808C, G2375R, R2541K and R2639Q. A modified table based on the ARVC database with the selected mutations is given in table 7.1. A schematic diagram representing the positions of these proteins on the desmoplakin full length sequence is given in figure 7.1.

Mutation	DNA change	Domain location	Patients affected	Controls (chromosomes)	Notes	Original reporting article
K470E	1408 A>G	SH3	4*	0/200	Fulfilled international taskforce criteria	Basso et al (2006)
R808C+	2422 C>T	SR8	1	0/1200	Presyncope and VT	Al-Jassar et al (2011)
G2375R	7123 G>C	PRD B	8	0/180	VT, SCD, Skin fragility wooly hair syndrome	Alcalai et al (2003)
R2541K	7622 G>A	Linker domain	4$	0/500	SCD	Bauce et al (2010)
R2639Q	7916 G>A	PRD C	1	0/200	VT	Yu et al (2008)

Table 7.1 Selected desmoplakin ARVC mutations for biophysical analysis

The table highlights the alterations and physiological consequences of each mutation. * and $ represent incorrectly reported number of patients affected with ARVC from the ARVC database, however on closer inspection of the reporting papers more likely to be 1. VT = ventricular tachycardia, SCD = sudden cardiac death. + A R808H mutation has also been reported (Personal communication: Dr Hennie Bikker, University of Amsterdam, Netherlands).

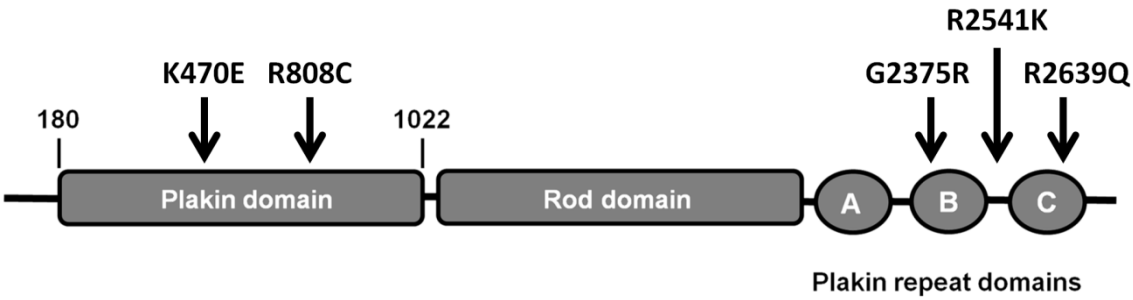


Figure 7.1 Schematic diagram showing the positions of the mutations studied in the full length desmoplakin sequence

Mutations K470E and R808C originate within the plakin domain and were analysed in the context of full length plakin domain and SR subdomain proteins. Mutations G2375R and R2639Q originate in the PRD B and PRD C subdomains respectively and were analysed in the context of their respective PRD proteins. R2541K originated in the desmoplakin linker domain.

7.2 Purification of ARVC mutant proteins

All amino acid substitutions were carried out by site directed mutagenesis using the Lightning quick change kit (Stratagene). Amino acid substitutions of K470E and R808C were carried out on the full length plakin domain construct and their component structural domains (SR56 and SR78 respectively). G2375R was mutated in the PRD B construct of desmoplakin. R2541K was mutated in the desmoplakin linker domain. R2639Q was mutated in the PRD C construct of desmoplakin. Purification was carried out in the same fashion as their WT counterparts. With the exception of the PRD mutant constructs, expression and purification of the mutants gave no significant differences in yield, purity or final stability at 4°C.

Interestingly proteins with mutations in the two PRDs were entirely insoluble when expressed in bacteria (figure 7.2). Even when a variety of different expression conditions were altered (temperature, expression time, IPTG concentration) the proteins remained insoluble when compared to their WT counterparts which remained soluble in nearly all conditions (data not shown). This suggested that a single amino acid substitution in both constructs was significant enough to alter its folding and/or stability and therefore its expression profile.

The following sections will therefore concentrate on plakin domain and linker domain mutants. The final section will attempt to model the possible structural differences caused by each mutant including the PRDs.

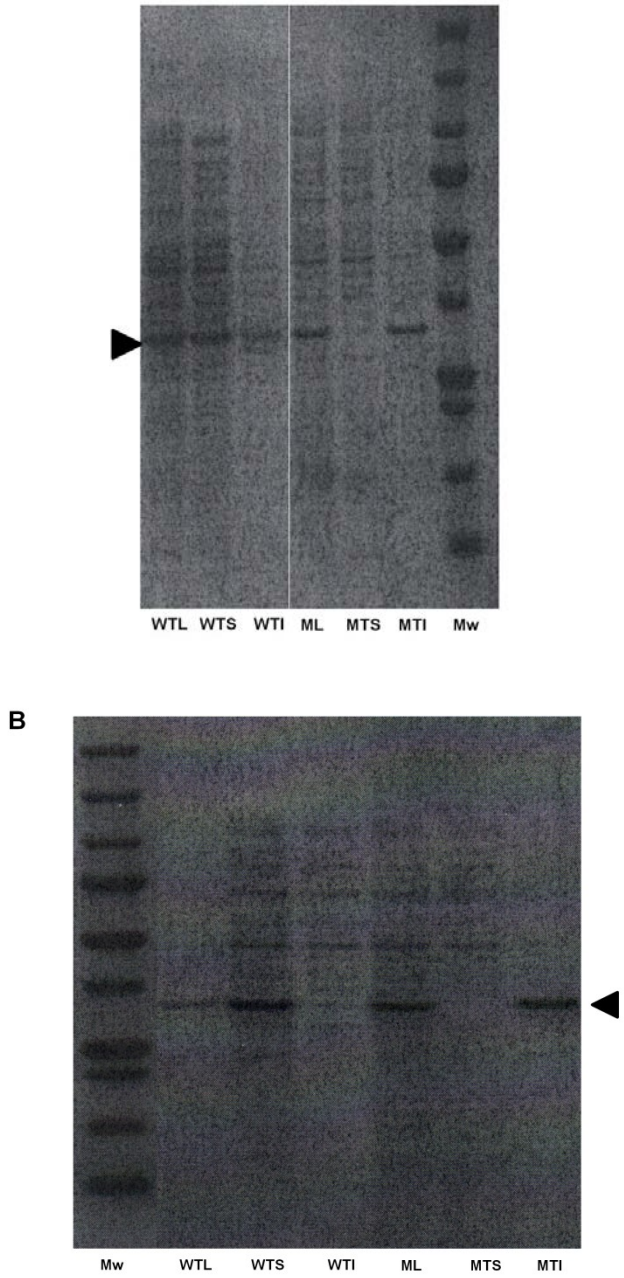


Figure 7.2 Comparison of expression profiles of wt and mutant PRD proteins

A: SDS-PAGE analysis of PRD B and the mutant variant G2375R. B: SDS-PAGE analysis of PRD C and the mutant variant R2639Q. WT and mutant variants of the PRDs B and C were expressed, lysed and analysed for solubility. Both mutant proteins are insoluble. WTL: wt lysate, WTS: wt soluble fraction, WTI: wt insoluble fraction, ML: mutant lysate, MTS: mutant soluble fraction, MTI: mutant insoluble fraction. Arrowheads indicate the respective proteins.

7.3 Comparison of gel filtration elution profiles of SR56 and SR78 subdomains

During the purification of the plakin domain subdomain mutants a Superdex S75 size exclusion column was used to separate them from contaminants of different molecular weights. This method can also be used to determine the molecular weight and oligomerization state when compared to known standards. This was not conducted for this thesis as AUC was used instead (e.g. see 7.4). During size exclusion purification the molecular weight determination is heavily influenced by the hydrodynamic radius of the protein. Therefore, if the hydrodynamic properties had altered due to an amino acid substitution it would be observed as a significant shift in the elution profile of the mutant. When wild type and mutant SR56 and SR7-8 proteins were subjected to size exclusion chromatography a single peak was observed in the elution profile. When superimposed, the peaks of SR56 and its mutant variant SR56-K470E matched (figure 7.3). The same applied to the wt and mutant variants of SR78 (figure 7.3). This suggested that the wt and mutant variants match in terms of the hydrodynamic radius and therefore their molecular weight. Therefore, it appears the amino acid substitutions have had no effect on the hydrodynamic radius of the proteins. Although the hydrodynamic radius had not been altered it wasl possible that the oligomerization state of the protein had been altered by the mutation. This was assessed using sedimentation velocity AUC.

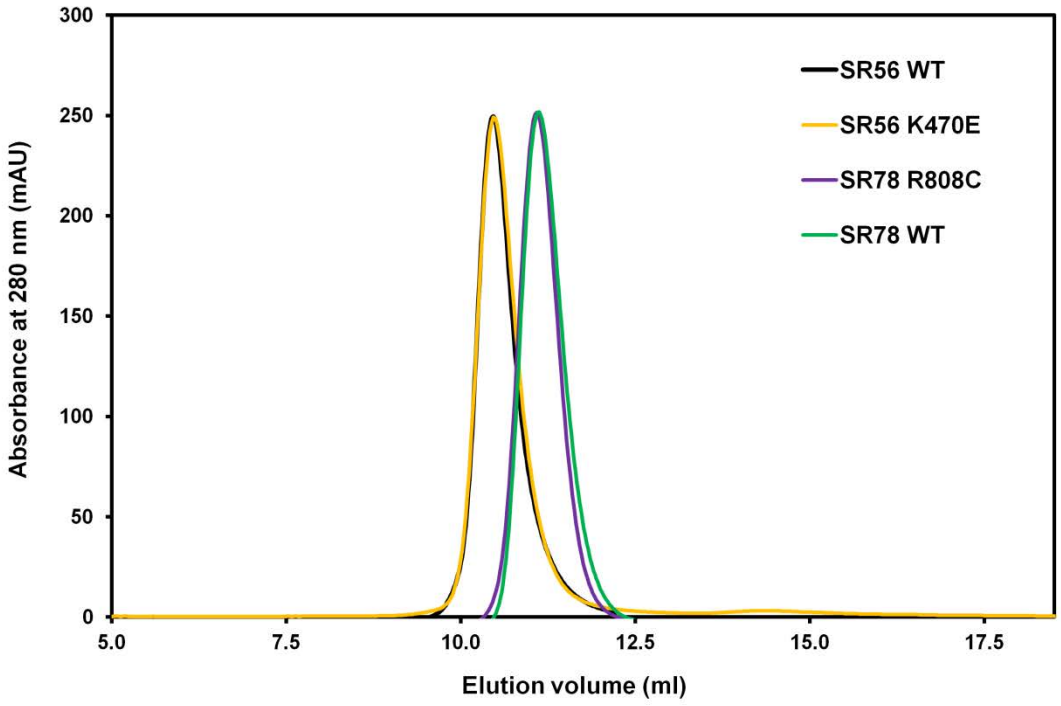


Figure 7.3 Gel filtration elution profiles of wt and mutant subdomain proteins

WT and mutant variants of SR56 and SR78 were subjected to size exclusion chromatography in 20 mM sodium phosphate (pH 7.2) 100 mM NaCl. The elution profiles were scaled to 250 mAU and superimposed. SR56 wt (black) and SR56-K470E (yellow) match closely indicating that they are the same size and have a comparable hydrodynamic radius. The same applies to SR78 wt (green) and SR78-R808C (purple).

7.4 Sedimentation velocity AUC profiles

To gauge the oligomerization state further AUC was conducted on wt and mutant SR56 and SR78 proteins. Size exclusion chromatography showed that K470E and R808C ARVC mutations had no affect on hydrodynamic radius so for simplicity only oligomerization state was analysed by AUC. SR56 wt and SR56-K470E AUC profiles were acquired using SEDFIT (Brown and Schuck, 2006). A sedimentation profile with a single peak estimated to be 28.0 kDa was obtained for SR56-K470E (figure 7.41). This is in good agreement with SR56 wt profile which gave an estimated size of 31.6 kDa. This suggested that the protein remained monomeric in the presence of a predicted pathogenic ARVC causing mutation. The slight variations in size and distribution may represent a slight difference in hydrodynamic properties or an artefact created by slightly different processing. An alternative explanation may be that the SR56-K470E has decreased heterogeneity in the sample by increasing intra-molecular stability. SR78 wt and SR78-R808C were also analysed by sedimentation velocity AUC. Both gave matching sedimentation velocity profiles when scaled with single peaks estimated to 25.1 and 25.1 kDa for SR78 wt and SR78-R808C respectively (figure 7.42). Hence it appears that neither the K470E nor the R808C mutations significantly altered the oligomerization state of their respective subdomains.

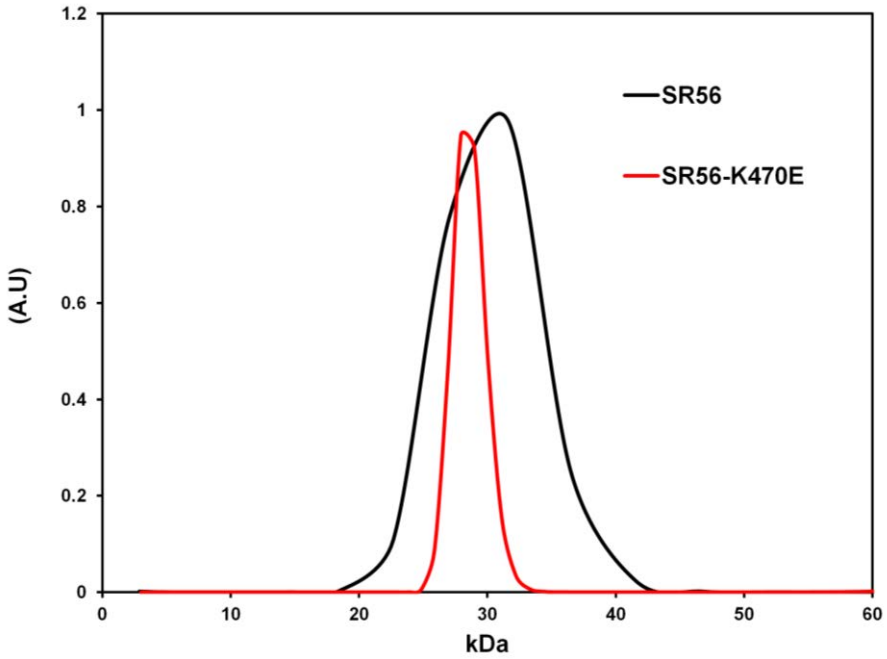


Figure 7.41 Sedimentation velocity profiles for SR56 and SR56-K470E

SR56 wt (black) and SR56-K470E (red) were subjected to sedimentation velocity AUC at 20000 rpm on a Beckman XL-1 in 20mM sodium phosphate (pH 7.2) and 100 mM NaCl and analyzed by SEDFIT (Brown and Schuck, 2006) then scaled. Both revealed a single peak around approximately 30 kDa which matches closely to the predicted sequence derived molecular weight of 31.6 kDa.

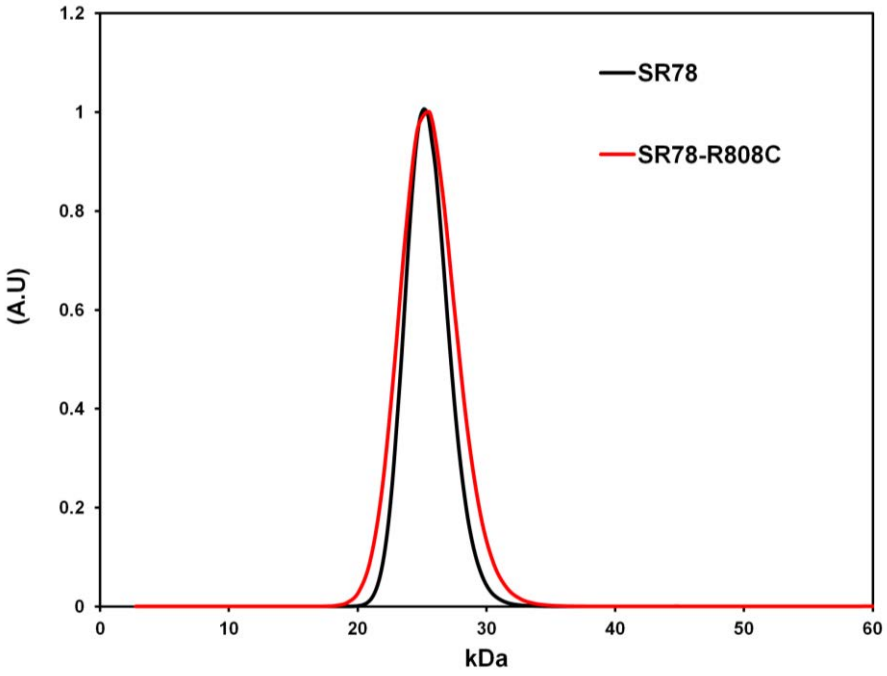


Figure 7.42 Sedimentation velocity profiles for SR78 and SR78-R808C

SR78 wt (black) and SR78-R808C (red) were subjected to sedimentation velocity AUC at 20000 rpm on a Beckman XL-1 in 20mM sodium phosphate (pH 7.2) and 100 mM NaCl and analyzed by SEDFIT (Brown and Schuck, 2006) and scaled. Both revealed a single peak around approximately 25 kDa which matches the predicted sequence derived molecular weight of SR78.

7.5 Comparison of far-UV CD profiles

CD is a method used to analyse the secondary structure content of proteins and was used to analyse the SR56 and SR78 subdomains. As observed in figure 7.51 and 7.52, the ARVC mutations do not appear to change the CD profile shown by the superimposition of the wt and mutant spectra. My conclusions were independently validated by Professor Alison Rodger (Warwick University, 2010), Dr Matthew Hicks (Warwick University, 2010) and Dr Raul Pinchero-Gomez (University of Birmingham, 2010) all experts in the use of CD. This suggested that both K470E and R808C do not alter the secondary structure content of the protein when compared to their wt counterparts. Since subdomain secondary structure was not altered, full length plakin domain mutants were not compared to the wt counterparts by far-UV CD, although were compared by SAXS (see 7.7).

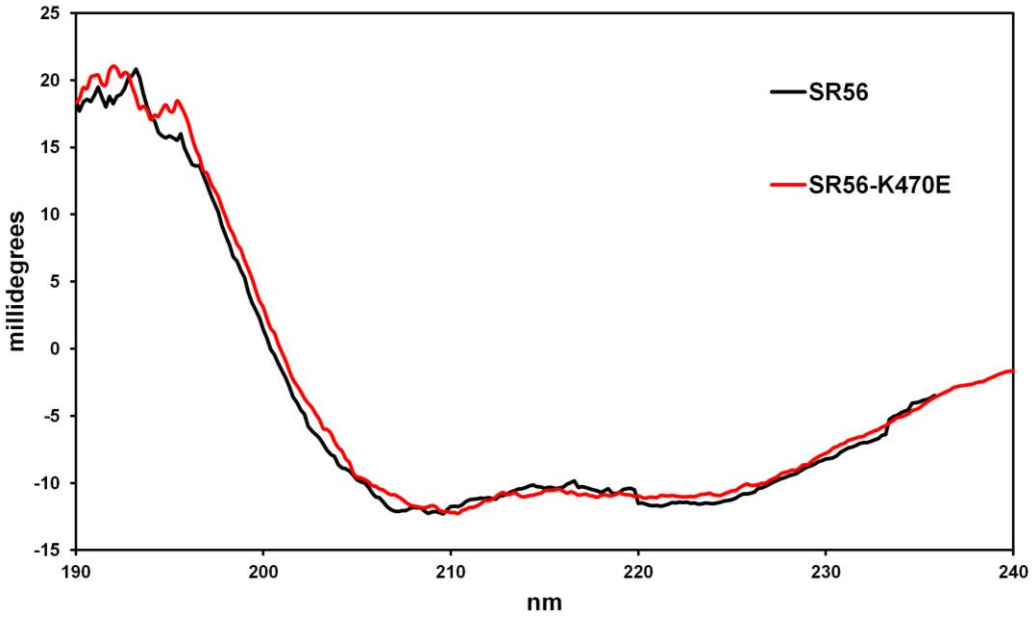


Figure 7.51 Typical far-UV CD profiles of SR56 WT versus SR56-K470E

SR56 wt (black) was compared to the K470E variant (red). Comparisons were run multiple times at the University of Birmingham and the University of Warwick at 0.1 mg/ml in 20mM sodium phosphate (pH 7.2). The K470E mutation did not cause any significant alterations to the far-UV CD profiles and secondary structure.

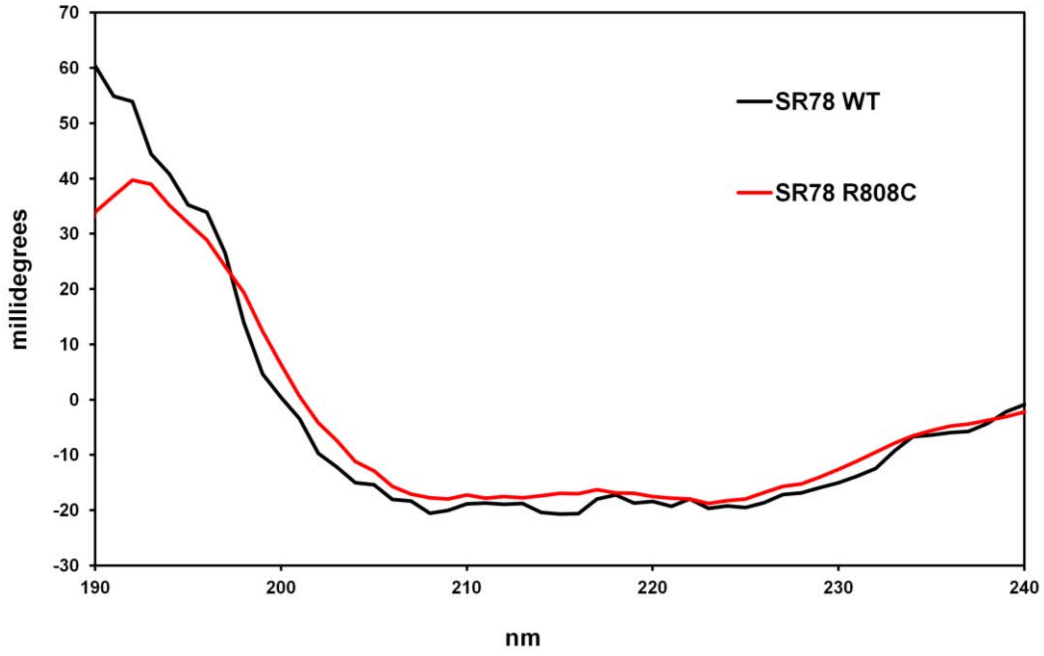


Figure 7.52 Typical far-UV CD profiles of SR78 WT versus SR78-R808C

SR78 wt (black) was compared against the R808C variant (red). Comparisons were run multiple times at the University of Birmingham and the University of Warwick at 0.1 mg/ml in 20mM sodium phosphate (pH 7.2). The R808C mutation did not cause any significant alterations to the far-UV CD profiles and secondary structure.

7.6 Thermofluor melting temperature comparisons between wt and mutants

Although secondary structure content was not altered by either of the mutants it is possible that intra and inter molecular stabilization events (i.e. amino acid interactions) may have been altered as observed by Kirchner et al for PKP2a (2012). This could lead to a less stable structure which can be analyzed by thermal stability assays as the melting curves correlate with the general stability of the protein. Circular dichroism is the preferred choice of analyzing melting temperatures by structural biologists. Typically, a sharp transition from ordered to disordered is correlated with temperature and a melting curve is created. Unfortunately, CD analysis for ascertaining melting curves of each subdomain and respective mutants were attempted at the University of Warwick but both wt and mutant variants of SR56 and SR78 aggregated at temperatures above 60°C. An accurate melting curve could not be created as a result. It was felt that although the proteins may not be aggregating themselves the proteins could have aggregated on the glass present in the cuvette (Personal communication: Dr Matthew Hicks, University of Warwick, 2010). Indeed when a water only solution was added to the (thoroughly washed) cuvette immediately after analysis of SR56, an anomalous CD signal was observed owing to the presumed aberrant aggregation of SR56 on the glass of the cuvette. As an alternative, wt and mutant variants were subjected to thermofluor analysis where a plastic 96 well plate is used instead and a different signal acquired by means of fluorescence.

Wild type and mutant variants were analysed at three different pH value ranges. Different pHs were used due to the fact that both the original wt amino acids (K470 and R808) were charged and may have required a certain chemical environment to stabilize an intramolecular interaction. The program XLFit was chosen for thermal shift analysis as it is the gold standard for thermofluor analysis and gives accurate melting temperatures when dealing with multiple data sets (to 1°C) (Ericsson et al., 2006). These were replicated 6 times to gain a more accurate

melting temperature. SR56 and the mutant variant revealed superimposable melting curves at all pHs attempted (data not shown). This suggested that the mutation had no effect at all on the general stability of this subdomain. SR78 and SR78-R808C revealed differences in melting temperature at all pHs attempted (table 7.2). When analysed by XLFit, a decrease of 2 °C in melting temperature was observed for the mutant protein at both pH 6.5 and pH 7.1. A larger decrease of 4 °C was observed at pH 8.2 (figure 7.6). This suggested that at a range of different pHs thermal stability was consistently lower in the R808C variant when compared to the wt.

pH	**SR78WT** Melting temperature (^{0}C)	**SR78-R808C** Melting temperature (^{0}C)	**ΔT_m** (^{0}C)
6.5	57	55	-2
7.2	59	57	-2
8.2	60	56	-4

Table 7.2 Thermofluor melting temperatures as ascertained by XLFit.

Maximum and minimum points were ascertained by XLFit and the midpoints were calculated as the melting temperature with the caveat of an accuracy of no more than 1 °C. In each case a negative shift is observed. The most pronounced affect is observed at pH 8.2 where the negative shift is 4 °C.

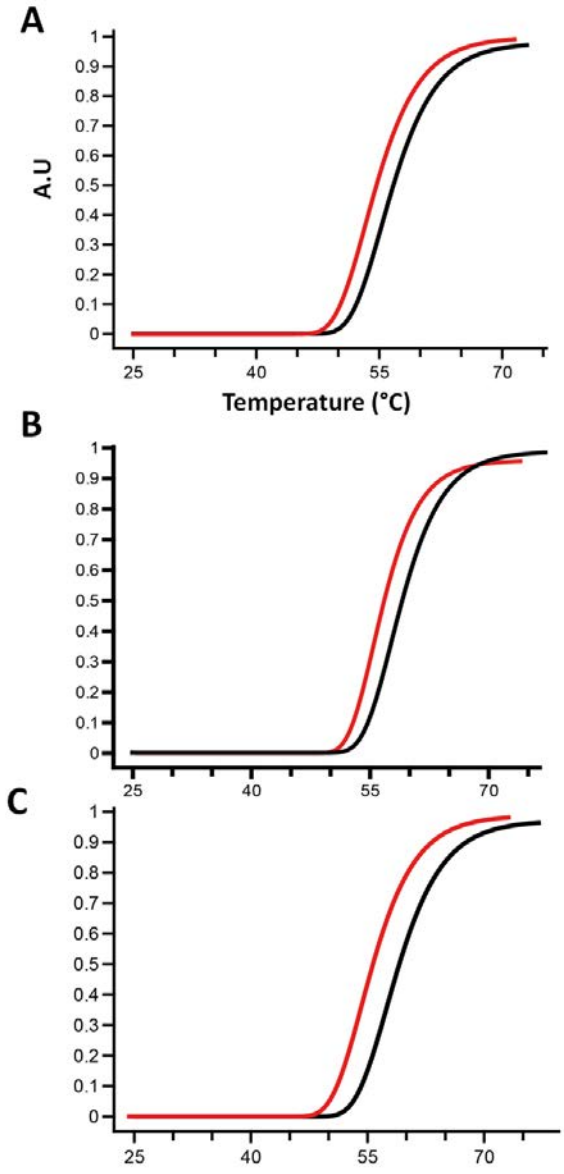


Figure 7.6 Thermofluor analysis of SR78 wt and SR78-R808C

SR78 (black) and SR78-R808C (red) at 4 µM were analysed by thermofluor in A) 20 mM Bis Tris (pH 6.5), 100 mM NaCl; B) 20mM sodium phosphate (pH 7.1), 100 mM NaCl; C) 20 mM Tris (pH 8.2), 100mM NaCl. Each experiment was repeated 6 times. For each condition replicates superimposed very well. This was then input into XLFit and a sigmoidal fit was applied in line with previously published methods (Ericsson et al., 2006). In each curve a negative shift is observed for the mutant when compared to the wt which suggested a decrease in thermal stability with the most pronounced effect observed at pH 8.2. See table 7.2 for detailed differences as recorded by XLFit.

7.7 SAXS analysis of wt and mutant proteins

Although secondary structure content alteration was not observed as a direct consequence of an amino acid substitution, a global change may still have occurred. For instance, since both K470E (large positively charged residue to a negatively charged residue) and R808C (large positively charged residue to a small sulphur containing residue) mutations are dramatic changes it was hypothesized that they could have an effect on the tertiary fold. Wt and mutant full length (plakin domain) and subdomain variants were compared by SAXS.

SR56 wt and the K470E mutant variants were analysed by SAXS analysis. Using PRIMUS their merged data sets (using similar concentrations for signal to noise ratio comparison) were superimposed to ascertain global structural differences. As observed in figure 7.71 their scattering curves superimpose very well. This is true for the entire scattering curve range (figure 7.71A) and importantly at low angles (figure 7.71B) where the size and shape is predominantly calculated from the Guinier region. Thus the K470E mutation has had no observable effect on the globular shape of the SR56 subdomain at the low resolutions permitted by SAXS analysis.

SR78 WT and R808C mutant variants were similarly analysed by SAXS using the same protocol as described above. The scattering curves of the SR78 wt and mutant proteins superimposed very well (figure 7.72). This was the case across the entire scattering curve range (figure 7.52A) and at lower angles (figure 7.72B). Thus it appears that the R808C mutation has had no significant effect on the globular shape of the SR78 subdomain.

To independently verify that these mutations did not significantly alter globular fold, the raw data sets were sent to Dr Dmitri Svergun (DESY, Hamburg, 2010) for analysis. He confirmed that these data sets do indeed superimpose and described any differences as “marginal” and that the curves in general “practically coincide”. The independent analysis of the wt versus mutant SAXS

scattering curves corroborated the previously suggested lack of globular change within the subdomains.

WT and mutant counterparts were also compared using the full length plakin domain proteins. Since the SH3 domain was known to be folded in the recently crystallized SR3-6 fragment (Choi and Weis, 2011) it was hypothesized that the K470E mutation could disrupt fold and/or the interaction with SR4. Similarly for R808C it was possible that the mutation had an effect on the entire plakin domain. Wild type plakin domain was compared to its mutant counterparts by SAXS analysis. As shown in figure 7.73 neither the K470E or R808C mutants cause a significant alteration in the globular structure of the plakin domain. SAXS is also sensitive to differences in degree of flexibility and changes in flexibility can also be ruled out (Personal communication: Dr Pau Bernadó, Montpellier University, 2012).

Overall it appears that the global structures of the subdomain and full length plakin domain proteins are not altered by the ARVC mutations investigated. Although global structure appears not to have been altered within the low resolutions permitted by SAXS, it is entirely possible that local conformational changes could have occurred especially as each of the amino acid substitutions are dramatic. 2D-HSQC NMR is sensitive to local conformational changes since the position of each peak within a HSQC is determined by its local chemical environment. Due to the size limitations of NMR (<30 kDa) comparisons of wt and mutant proteins of the full length plakin domain were not possible. Instead the subdomain wt and mutant variants were compared by 2D-HSQC NMR.

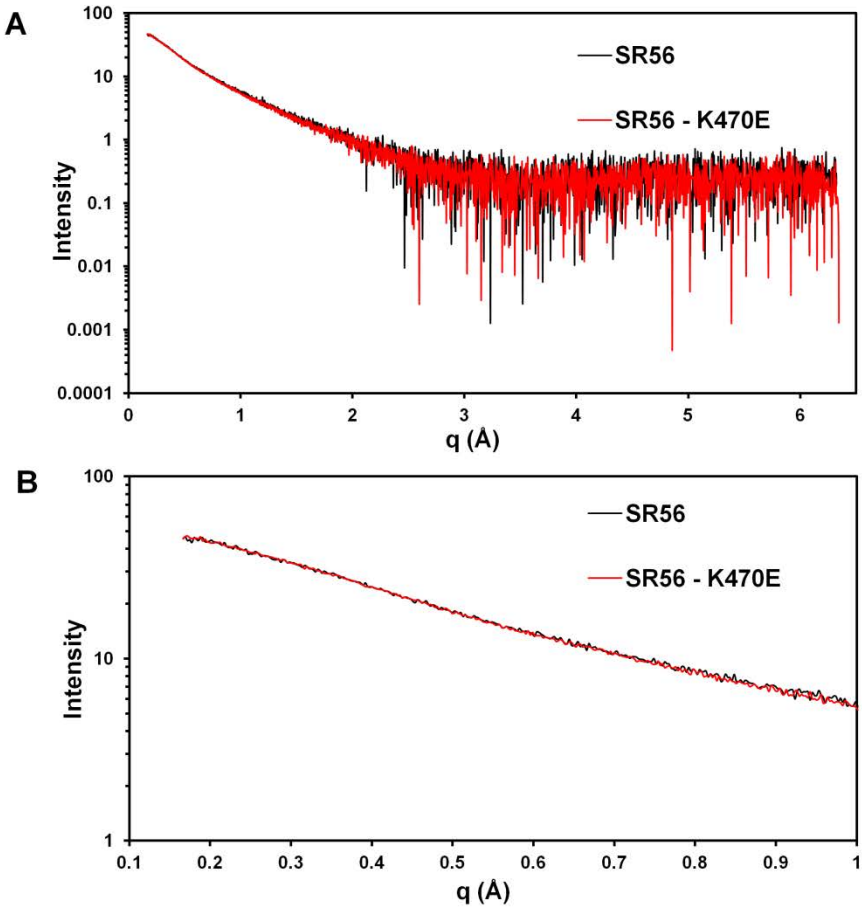


Figure 7.71 SAXS analysis of SR56 subdomain wt and mutant counterpart

SR56 and its mutant variant were subjected SAXS analysis in 20mM sodium phosphate (pH 7.2) and 100 mM NaCl at 1.0, 2.5 and 3.8 mg/ml for SR56 WT and 1.0, 2.5 and 4.2 mg/ml for SR56-K470E. Each scattering profile was scaled and superimposed on to one another. Shown are the full scattering range (A) and the Guinier region (B) for each protein. The scattering curves of the wt and mutant proteins were similar suggesting that the mutation does not affect overall structure at the low resolutions permitted by SAXS. This was also the case in the Guinier region in (B) which is the predominant determinant of overall protein dimensions.

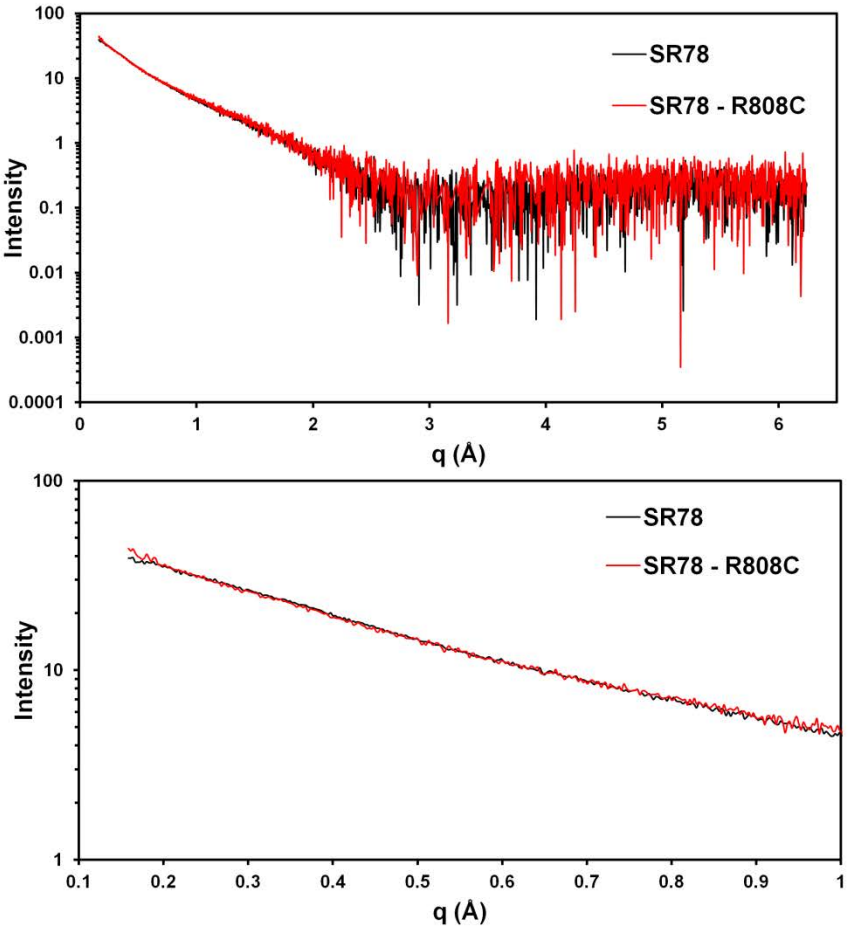


Figure 7.72 SAXS analysis of SR78 subdomain wt and mutant counterpart

SR78 and its mutant variant were subjected to SAXS analysis in 20mM sodium phosphate (pH 7.2) and 100 mM NaCl at 1.0, 2.5 and 3.7 mg/ml for SR78 WT and 0.8, 1.2 and 2.4 mg/ml for SR78-R808C. Each were then scaled and superimposed on to one another. Shown are the full scattering range (A) and the guinier region (B) for each protein. The scattering curves of the wt and mutant proteins were similar suggesting that the mutation does not affect overall structure at the low resolutions permitted by SAXS. This was also the case in the Guinier region in (B) which is the predominant determinant of overall protein dimensions.

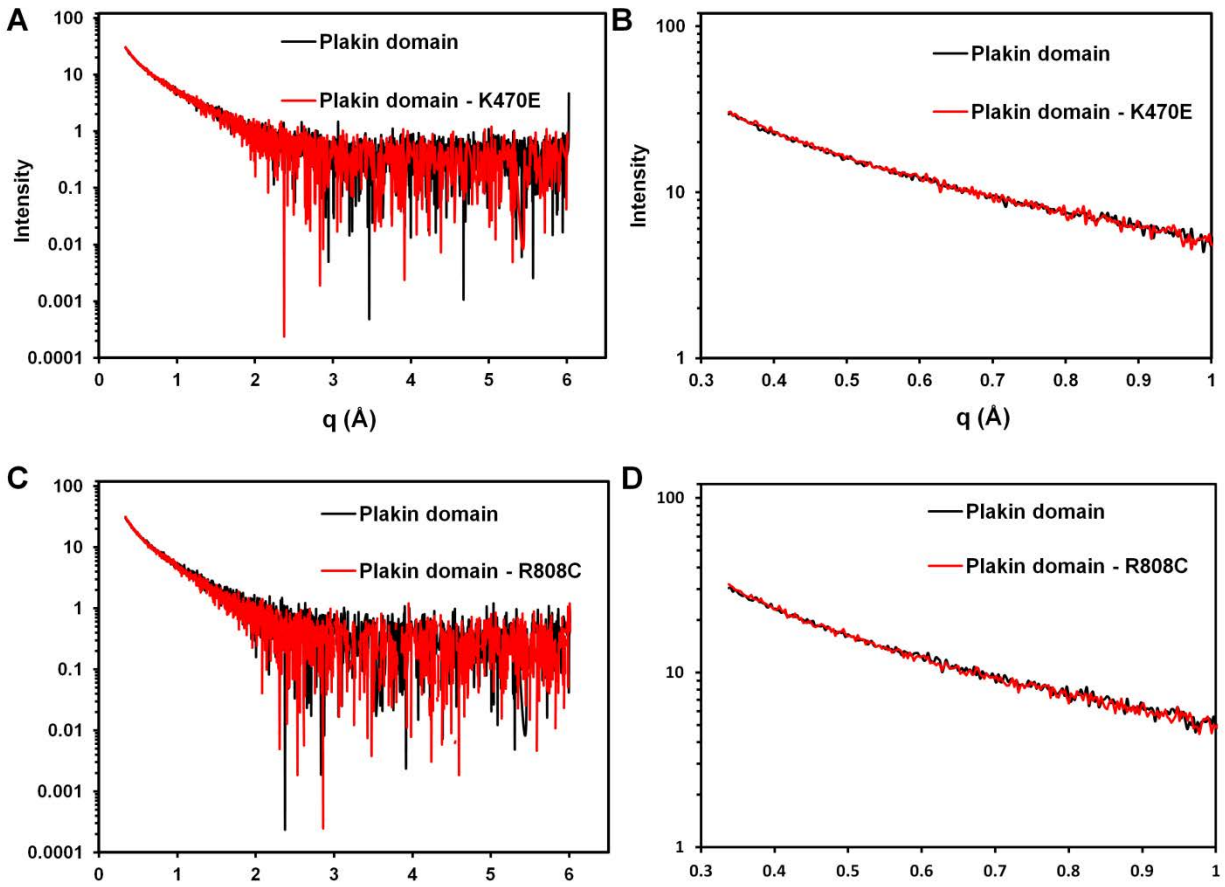


Figure 7.73 SAXS analysis of the plakin domain wt and mutant counterparts

The plakin domain and both its mutant variants were subjected SAXS analysis in 20mM sodium phosphate (pH 7.2) and 100 mM NaCl. Shown are the full scattering range (A) and (C) and the Guinier region (B) and (D). Merged SAXS scattering curves of approximately the same signal to noise ratio from each of the wt and mutant variants were superimposed. This is also the case in the Guinier regions (B) and (D) which is the predominant determinant of overall protein dimensions. This suggested that the K470E and R808C mutations did not alter the globular shape of the plakin domain within the low resolutions permitted by SAXS.

7.81 2D-HSQC NMR analysis of the plakin domain mutants

It was shown in chapter 3 that both SR56 and SR78 contained tertiary fold by 2D-HSQC NMR, with SR78 being the more dispersed of the two subdomains. Mutant variants were analysed by 2D-HSQC NMR under the same conditions (spectrometer, buffer conditions, spectral widths) as their wt counterparts. The 2D-HSQC of SR56-K470E was superimposed onto the SR56 wt variant. It revealed a similar degree of dispersion when compared to the wt indicating that the mutation neither increased nor decreased tertiary structure (figure 7.81), consistent with the SAXS data. On closer inspection however, approximately 25% of the observable peaks of the mutant exhibited chemical shift perturbations. This indicated a local conformational or electrochemical change as a direct result of the amino acid substitution. Unfortunately, due to the lack of assignment data, the precise way in which the local conformation had been altered could not be identified. The same process was then applied to the SR78-R808C mutation. SR78-R808C revealed a similar degree of dispersion and chemical shift perturbations (~30% of observable peaks) (figure 7.82). Overall it appears that mutations cause local conformational changes since each peak reports on their local physiochemical environments.

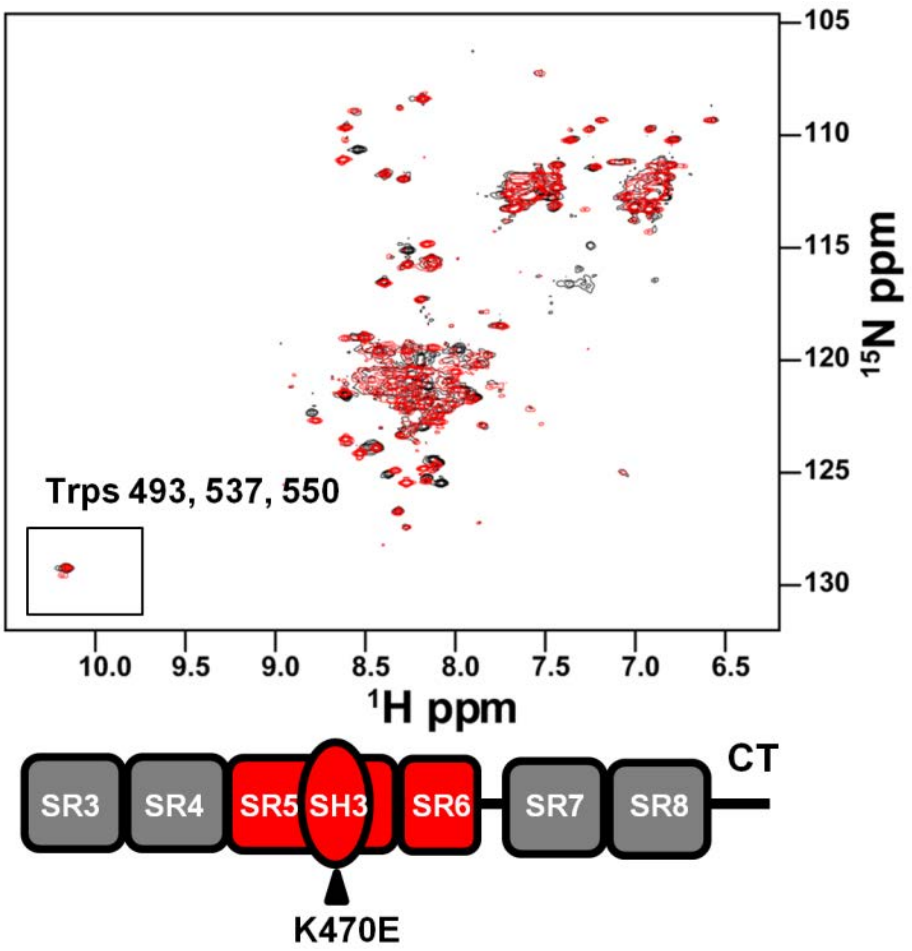


Figure 7.81 2D-HSQC NMR analysis of SR56 and its mutant counterpart

Both the wt (red) and mutant (black) variant of SR56 were subjected to 2D-HSQC NMR analysis at 100 µM in 20 mM sodium phosphate (pH 7.2) and 100 mM NaCl using a Agilent 600 MHz spectrometer. The HSQC shows that overall dispersion remains similar although chemical shift perturbations occur due the movement of some peaks (~25%). The box represents the expected positions of the $N^{\varepsilon}H$ resonance of tryptophans 493, 537 and 550 with only one being observable. A schematic diagram of the full length plakin domain with SR56 highlighted in red is shown. The position of the K470E ARVC mutation is indicated.

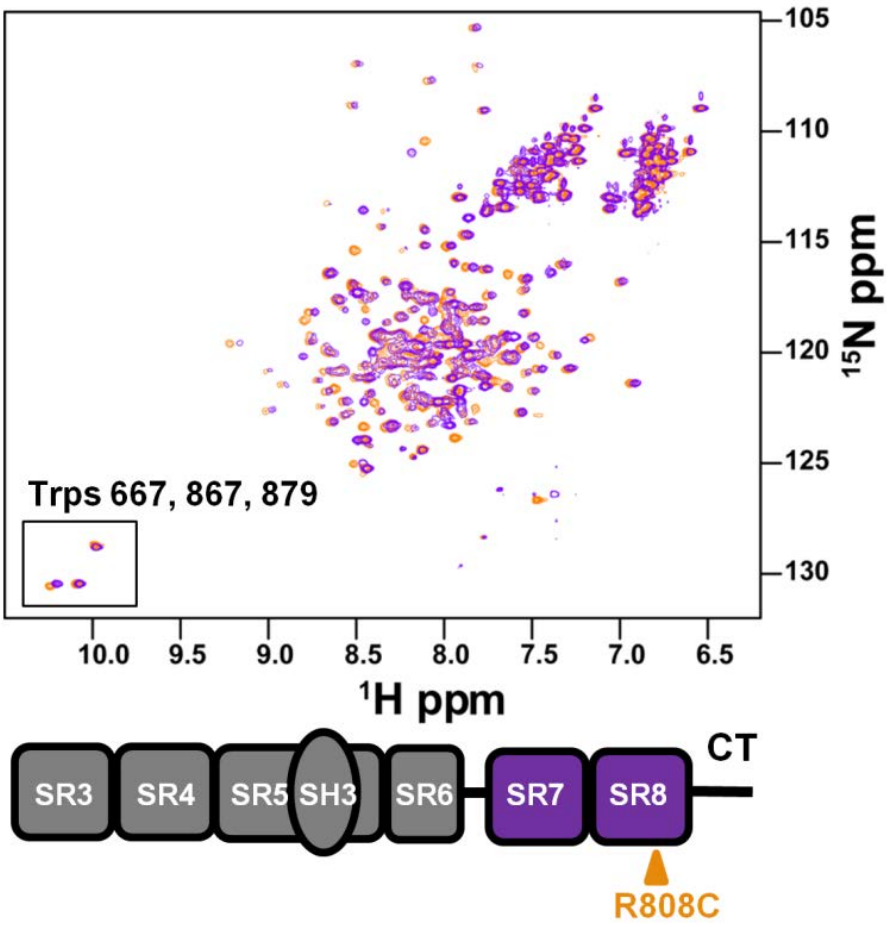


Figure 7.82 2D-HSQC NMR analysis of SR78 and its mutant counterpart

Both the wt (purple) and mutant (orange) variants of SR78 were subjected to 2D-HSQC NMR analysis at 100 µM in 20 mM sodium phosphate (pH 7.2) and 100 mM NaCl on an Agilent 800 MHz spectrometer. The HSQC reveals that overall dispersion remains similar however, chemical shift perturbations occur due the movement of some peaks (~30%). The box represents the expected positions of the $N^{\varepsilon}H$ resonance of tryptophans 667, 867, 879. A schematic diagram of the full length plakin domain with SR78 highlighted in purple and the mutant in orange is highlighted below the spectrum.

7.82 2D-HSQC NMR analysis of the desmoplakin linker domain mutant

It was shown in chapter 6 that the desmoplakin linker domain was folded and well dispersed by 2D-HSQC NMR analysis. Mutant variant expression and purification of the ^{15}N labelled desmoplakin linker domain yielded no significant differences in yield or purity when compared to those grown in LB.

The two spectra were superimposed to analyse chemical shift perturbations as a direct result of the mutation. The degree of dispersion appeared to be similar between the wt and mutant counterparts indicating the mutation has had little/no effect on the overall fold of the protein (figure 7.83). Closer inspection revealed that a few peaks (~10%) have chemical shift perturbations. This suggested that some local chemical environments or conformations had been altered although the majority remain unchanged. Due to the low number of chemical shift perturbations relative to the changes observed in SR56 and SR78 it suggested that this mutation is less deleterious in its effects on the conformation of the protein. The precise way in which the local chemical environments or conformations have been altered will require a high resolution structure of the linker domain.

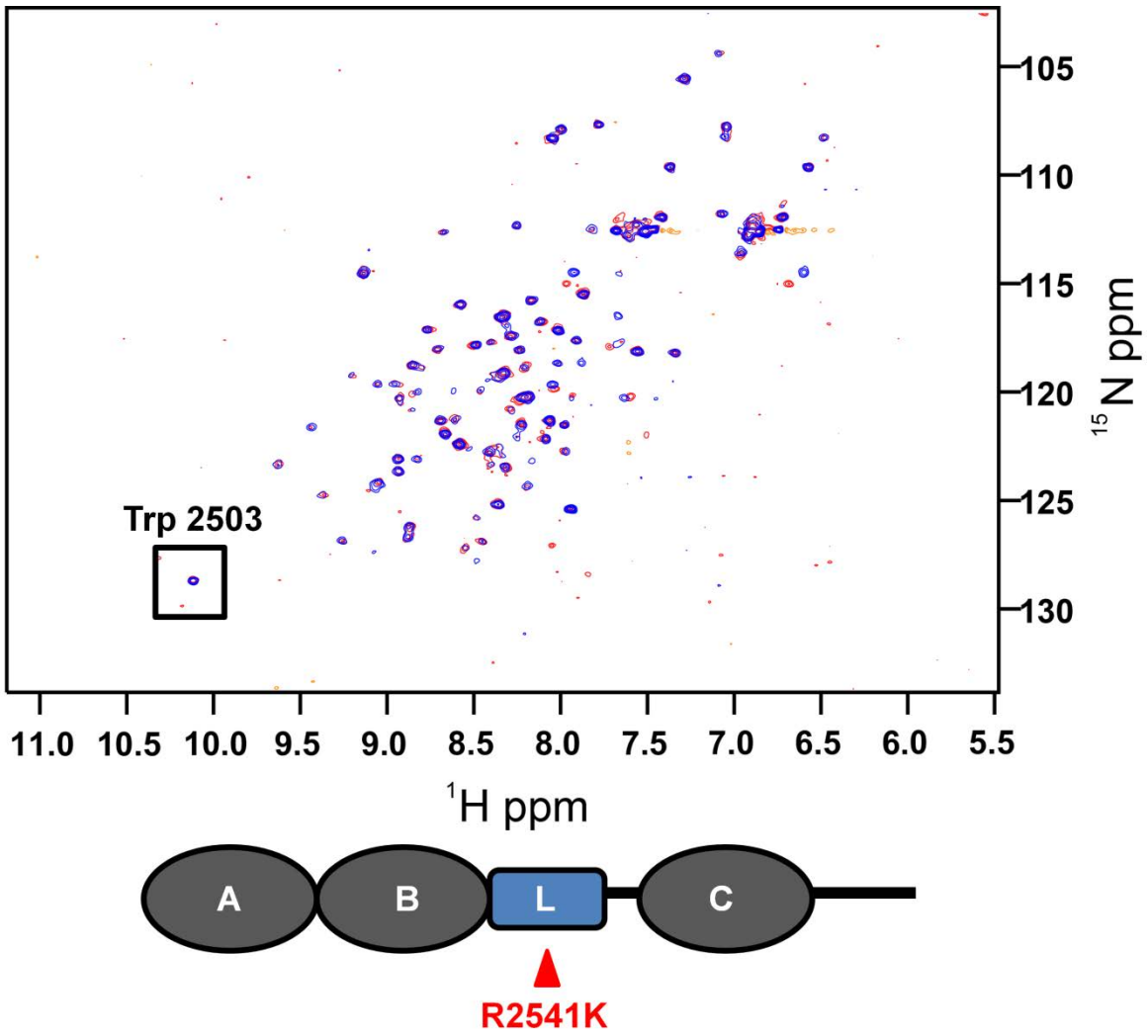


Figure 7.83 2D-HSQC NMR analysis of the desmoplakin linker domain and its mutant counterpart

Both the wt (blue) and mutant (red) variant of the desmoplakin linker were subjected to 2D-HSQC NMR analysis at 100 µM in 20 mM HEPES (pH 7) and 100 mM NaCl on an Agilent 600 MHz spectrometer. The HSQC shows that overall dispersion remains similar although a few chemical shift perturbations occur due the movement of some peaks (~10%). The box represents the expected position of the $N^{\varepsilon}H$ resonance of tryptophan 2503. A schematic diagram of the desmoplakin C-terminal tail with the linker domain highlighted in blue is given. The position of the R2541K mutation is indicated. Purification and NMR analysis conducted in collaboration with Miss Penelope Rodriguez-Zamora.

7.91 Modelling of ARVC mutants

To further evaluate the effects of each mutant modelling was performed. Three of the mutations analyzed in this chapter already have their wt structures elucidated. One of the mutations R808C within SR78 can be modelled using human erythroid spectrin, repeats 8 and 9 (pdb code: 1S35) as a template as it was shown that SR78 is a canonical spectrin repeat in chapter 3. A high quality model for the desmoplakin linker domain cannot be created due to the lack of structural information and therefore will not be discussed.

K470E in the desmoplakin plakin domain resides in an exposed position on the RT loop of the SH3 domain. It is not directly involved in the SH3-SR4 binding interface that has been shown to rigidify the SR3-6 fragment (Choi and Weis, 2011). The SH3 domain does not have a known ligand. Chemical shift perturbations in the 2D-HSQC of SR56-K460E compared to the wt suggested flexibility based on the non-isotropic dispersion. This may be due to the fact that it requires SR4 for stabilization. When a model for the SR3-6-K470E mutant was created superimposition to the wt fragment suggested few (if any) changes (figure 7.91). The mutant does however lie next to Y469 which is directly involved in the SH3-SR4 binding interface with W360. Since local conformational / electrochemical changes were observed in the HSQCs of SR56 WT and SR56-K470E it is possible that the mutation has altered the position of Y469 and therefore altered the binding propensity to W360 previously suggested (Choi and Weis, 2011). The mutant model however suggested this was not the case. Structural determination of the K470E mutant will validate this, with the use of SR3-6 as the preferred construct of choice since SR56 and the full length plakin domain were refractory to crystallization.

SR78 has not been structurally determined and therefore a PHYRE model based on erythroid spectrin, repeats 8 and 9 (pdb code: 1S35) was used. It revealed that the side chain of R808 had a potential electrostatic interaction with a spatially adjacent glutamic acid (E882) (figure

7.92). This was present in all 5 models (based on 5 different spectrin repeat templates from the PDB) created by I-TASSER which revealed an average distance of 2.7 Å between the carbonyl and amide groups. This is well within maximum distance required to form such an interaction. Electrostatic interactions throughout spectrin repeats are a typical feature known to stabilize their overall fold (Broderick and Winder, 2005, Djinovic-Carugo et al., 2002, Grum et al., 1999, Jefferson et al., 2007, Yan et al., 1993, Ylänne et al., 2001). The alteration to a cysteine (a small uncharged amino acid) amino acid would therefore abolish this proposed interaction. This idea is supported by the decreased thermal stability observed at a variety of pHs. Although unpublished, I received further information from a cardiologist specializing in cardiomyopathies identifying a desmoplakin R808H (histidine is a bulky residue) mutant as the sole determinant of ARVC in a patient (Personal communication: Dr Hennie Bikker, University of Amsterdam, Netherlands). This mutation would also abolish this interaction and indicates this residue is critical for stabilization of the spectrin repeat.

The importance of the G2375R mutation to the PRD B fold is more “clear cut”. G2375 in PR3 is a highly conserved residue across all plakin family members at the end of PR1 and PR3. Its importance is highlighted by the fact that it creates a sharp bend which precedes the beginning of PR2 and PR4 (figure 7.93). Glycine is the only amino acid capable of a wide variety of backbone conformations. Amino acid substitutions would prevent this bend from occurring, which was indicated by the insolubility / instability of the protein upon overexpression in bacteria (figure 7.2).

Furthermore the R2639Q mutation in PRD C of desmoplakin is similarly more “clear cut” than those of the plakin domain. In all plakin repeats an interaction is typically observed between the positively charged residue at position 19 and an aspartic acid at position 4 to set the position of the β-hairpin. R2639Q lies at position 19 of PR1 of PRD C, which interacts with D2624 which

lies at position 4 of PR1. As shown in figure 7.94 the alteration to a glutamine (a polar uncharged amino acid) could abolish this interaction. All together this suggests that this amino acid substitution is not tolerable in maintaining the overall fold of the domain as shown by insolubility of the protein upon overexpression in bacteria.

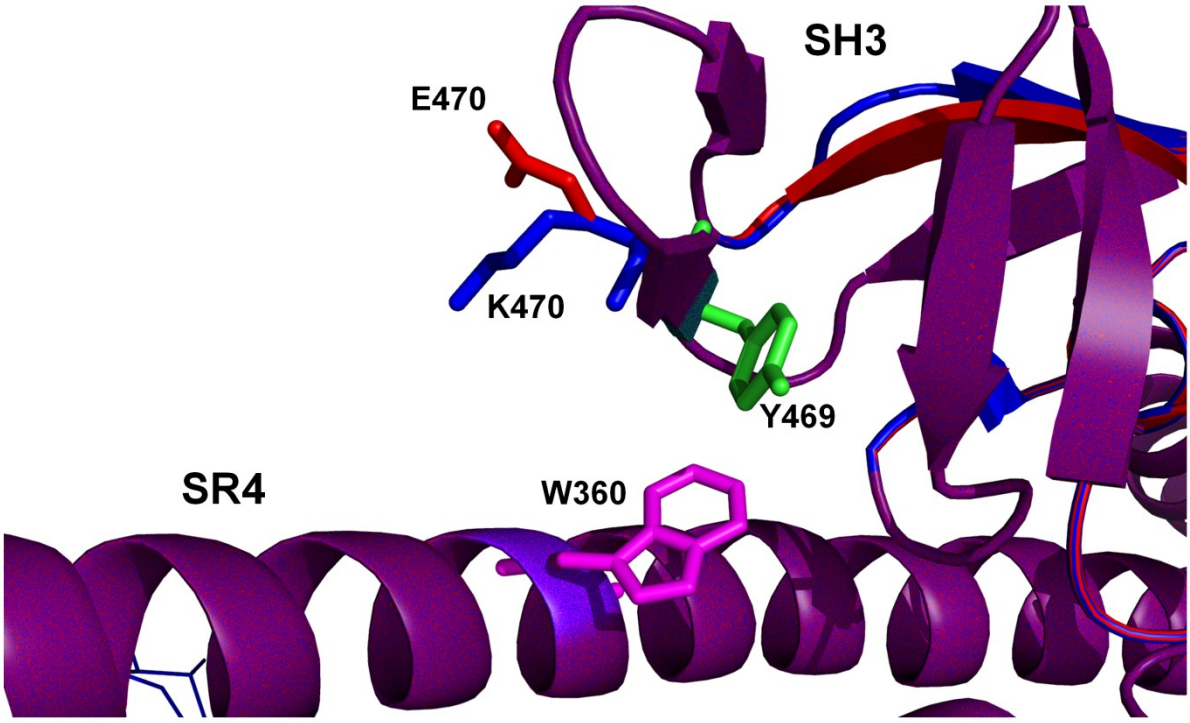


Figure 7.91 Modelling of the K470E mutant versus the wt

The desmoplakin plakin domain (pdb code: 3R6N) and a model of the K470E mutant protein were superimposed. Residue K470 is shown in blue and E470 in red. SR4 and the SH3 domain are shown in purple. Also shown are residues Y469 (green) and W360 (magenta) that are thought to be involved in the SH3-SR4 binding interface. The orientation of these residues is not altered in the mutant model.

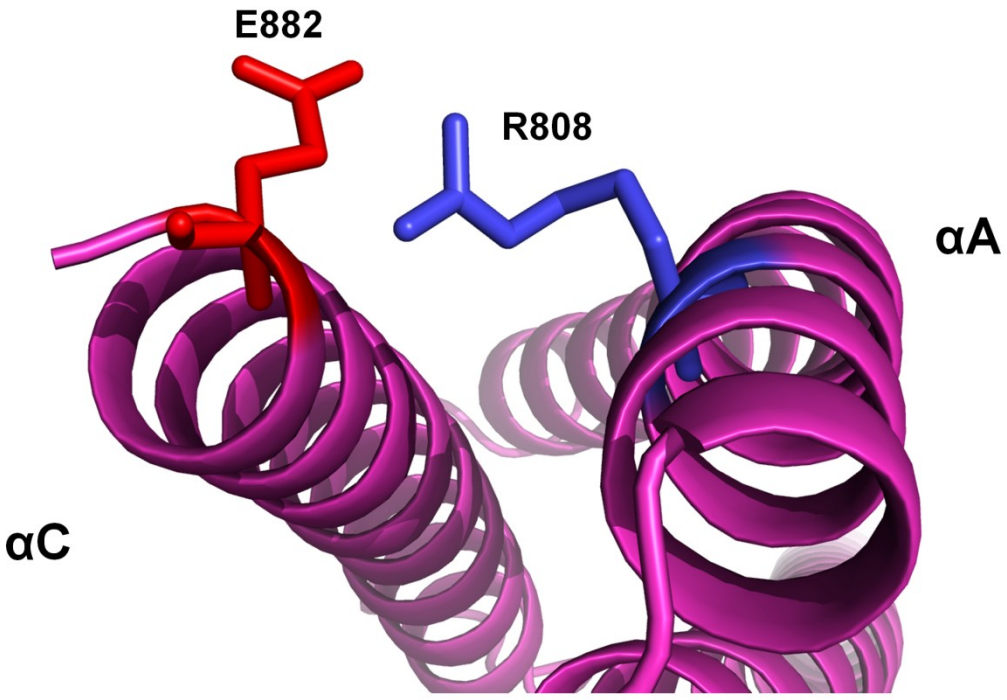


Figure 7.92 Modelling the R808 residue in a SR78 model

SR78 was modelled by the PHYRE server using erythroid spectrin, repeats 8 and 9 (pdb code: 1S35) as a template shown in magenta. The positively charged side chain of R808 is shown in blue, which is present on the “A” alpha helix of SR8. The negatively charged side chain of E882 is shown in red, which is present on the “C” alpha helix of SR8. This suggests that an electrostatic interaction between the two is required for stabilization of the domain.

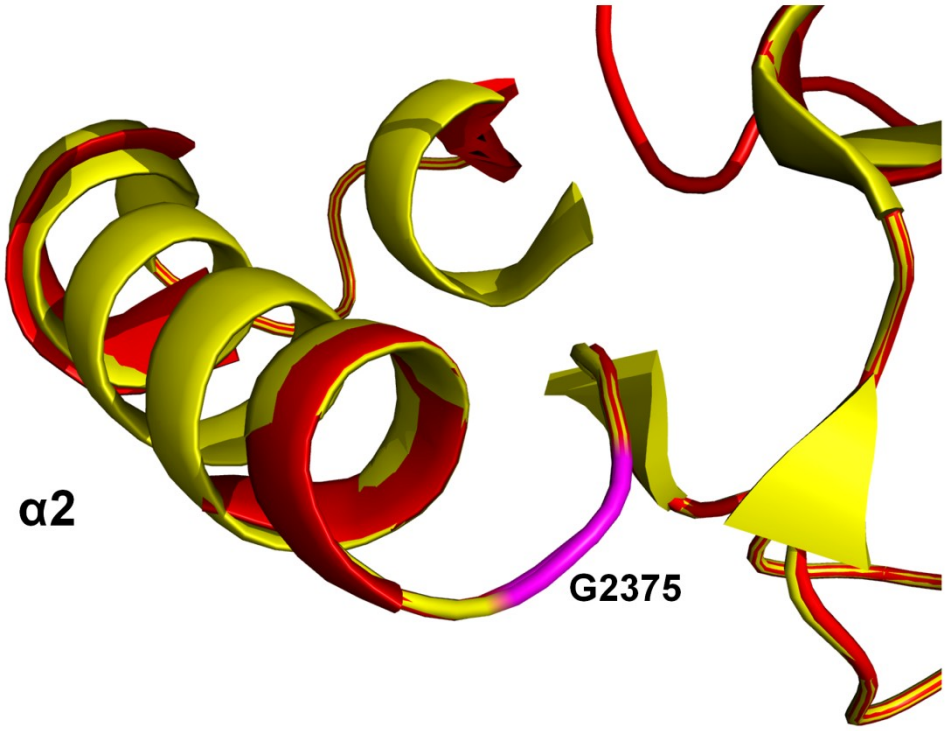


Figure 7.93 Modelling of the G2375R mutation

The crystal structure of the PRD B of desmoplakin (yellow) and a model harbouring a G2375R mutant created by I-TASSER (red) were superimposed. The second alpha helix of PR3 is highlighted as “α2” with the backbone of the critical bend created by G2375 depicted in magenta. Although both models superimpose well, (including a proposed bend created by the mutant R2375 amino acid) the bend in reality must be altered as R2375 cannot adopt as many backbone conformations as a glycine.

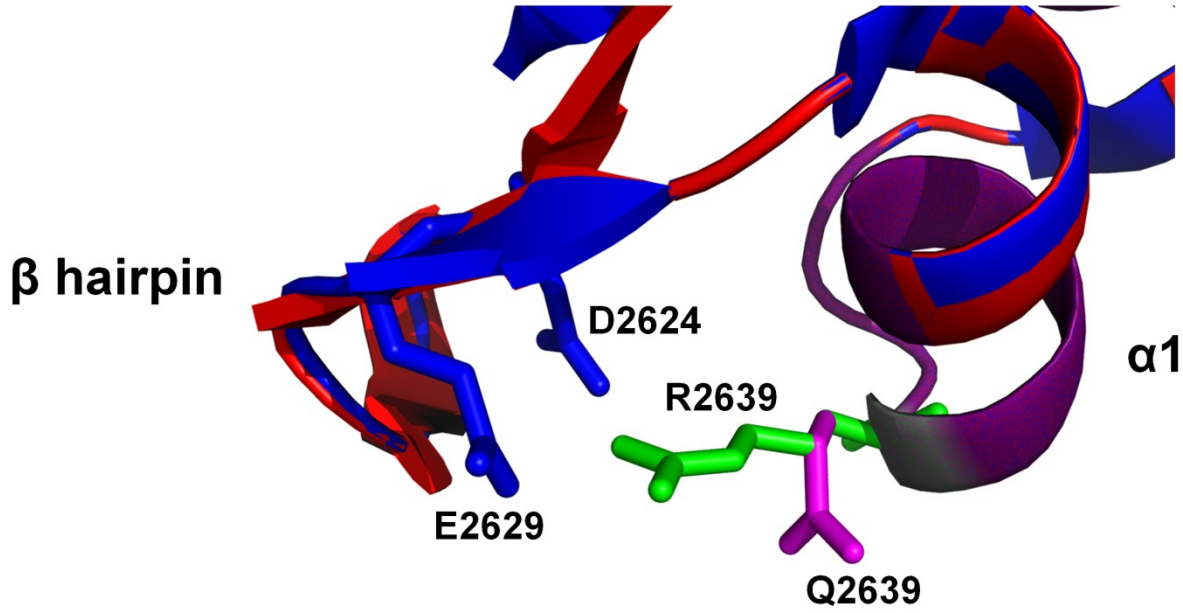


Figure 7.94 Modelling of the R2639Q mutant

WT (blue) and mutant (red) variants of the PRD C domain of desmoplakin were superimposed. The β-hairpin and first α-helix ("α1") of PR1 where R2639 resides is shown. It clearly shows the importance of the interaction between the R2639 side chain (green) and the D2624 side chain (blue) which has been previously shown to set the position of the β-hairpin. Also shown is another residue E2629 (blue) which may also allow the arginine to set the β-hairpin. Depicted in magenta is the side chain of the R2639Q mutant which would abolish the interaction of R2639Q to D2624.

7.92 Chapter 7 Conclusions: The traffic light system?

The experiments described in this chapter have for the first time highlighted the wide variety of differences ARVC mutations cause on the biophysical properties of desmoplakin domains. Since the function of proteins is intimately linked with their structure they show how a wide range of different types of mutations have a varying effect on the structure. It therefore appears critical to view each mutation individually for its proposed deleterious effect. The results indicate a wide variety of deleterious effects each mutation may have at least at the biophysical level. Therefore since the ARVC field is devoid of mutant-physiological effect correlations it may be beneficial to "class" certain mutations as highly deleterious (red), moderately deleterious (amber), and mildly deleterious (green) which the Overduin lab and I would like to call the "ARVC traffic light system". G2375R and R2639Q would be classed as red mutants, K470E and R808C as amber mutants, and R2541K as a green mutant.

7.93 Overall chapter 7 discussion

Given the recent taskforce criteria upgrade of a desmosomal mutation as a major criterion (Marcus et al., 2010), it was important that the mutation-phenotype correlation was explored further. The experiments of this chapter and the creation of the ARVC traffic light system hopefully show that mutations could be considered in terms of their ability to affect the overall fold of the protein.

Mutations from both N and C termini of the desmoplakin protein were investigated. K470E and R808C from the N-terminal plakin domain, G2375R from the desmoplakin PRD B domain, R2541K from the desmoplakin linker domain and R2639Q from the desmoplakin PRD C domain were analyzed in their respective subdomains. Both mutations from the plakin domain (K470E and R808C) were soluble and comparable to their WT counterparts in terms of their purification. While SAXS revealed no global alterations their structures, 2D-HSQC NMR revealed probably

localized chemical shift perturbations indicative of local electrochemical changes. SR56-K470E gave no differences in terms of its thermal stability profile compared to its WT counterpart. SR78-R808C however, did reveal a difference in its thermal stability profile compared to its WT counterpart, at a variety of different pHs. The difference in their behaviour may underlie the type of construct used. SR56 consists of a tandem spectrin repeat and a SH3 domain which appeared to contain some internal flexibility or disorder based on the CD and 2D-HSQC NMR data (chapter 3). In contrast SR78 is solely a tandem spectrin repeat construct, contained the highest thermal stability of all plakin domain constructs and was a well folded as shown by CD and 2D-HSQC NMR. With the aid of modelling, how the mutants affected the structure may complement the experimental observations.

The SR78 models clearly indicated that the side chain of R808 on alpha helix A was in close proximity to that of E882 on alpha helix C (figure 7.92). The proximity between the two charged side chains implied an electrostatic interaction between the two is feasible and is within the distances required to form. Electrostatic interactions between the α-helices of individual SRs are critical throughout all structures known to date for stabilization (Pascual et al., 1996, Yan et al., 1993). As a result the interaction is likely to exist and as such the mutation from R808 to a cysteine would disrupt this. Given that thermal stability is lower in the mutant variant it suggested that a destabilization event has occurred and corroborated the suggestion that the R808C mutation had disrupted the R808-E882 interaction. How this correlates to the clinical manifestation of ARVC is less clear. It was originally suggested in section 3.9 that SR78 may be acting as a “shock absorber” given its relatively high thermal stability. The destabilization may therefore result in an aberrant ability to resist mechanical force and possibly result in inadequate desmosomal formation or mechanotransduction. Confirmation of this hypothesis would require structural characterization of the SR78 domain.

The other plakin domain mutant analyzed (K470E) appeared to have less significant alterations. The most pronounced affect was observed when the K470E mutant was compared to its WT counterpart by 2D-HSQC NMR. This revealed a shift perturbation in the spectrum. While this suggests an alteration in the local chemical environment of the residue affected, the lack of assignment data and the slightly poorer dispersion (relative to that of SR78) suggests caution in interpreting the cause of the K470E mutation on the structure. However, the recent crystal structure revealed the position of the mutation to within the SH3 domain and may explain some of the observations. The K470 residue resides in an exposed position on the RT loop and is not directly involved in the SH3-SR4 interface implicated in rigidifying the SR3-6 domain (Choi and Weis, 2011, Ortega et al., 2011). It is difficult to interpret its biophysical or functional implication of the K470E mutation given its innocuous position on the RT loop. However, K470 resides next to Y469 which is directly implicated in the SH3-SR4 interface (Choi and Weis, 2011). Given this residue is critical to the interface, the K470E mutation may alter the backbone orientation of Y469 and prevent interaction with W360. This may explain the chemical shift perturbations in the 2D-HSQC spectrum. However the mutation is not sufficient enough to unfold the protein given the lack of an increased number of broadend signals. At present no known ligands are known for the plakin SH3 family. Although the typical binding region for SH3 domains is occluded in desmoplakin and plectin (Choi and Weis, 2011; Ortega et al, 2011) alternative modes of binding have been shown for other SH3 domains (Mongiovı et al., 1999). However, with the lack of ligand binding data only the structural implications can be assessed. Given that SR56 did not behave as well (biochemically) as the larger SR3-6 construct, a more pertinent way of addressing the ambiguity in the observations would be to crystallise SR3-6 in the presence of the K470E mutation. Regardless, the observations for the K470E mutation may also suggest that it is a polymorphism and others may be present throughout the ARVC database. Clearly other explanations may be involved in leading to ARVC from the K470E mutation (such as

preventing of binding to an unknown ligand) but the evidence from this study suggests otherwise.

The G2375R mutation was assessed in the PRD B of desmoplakin. It revealed that in stark contrast to the WT variant it turned the protein entirely insoluble. In context, the WT variant expresses in large amounts and is entirely soluble. Furthermore, the wt behaviour in solution makes it a relatively easy protein to purify. Therefore, this mutation is clearly causing a significant alteration to the overall fold or stability of the protein. With the aid of a crystal structure and bioinformatic analysis the biophysical effect on the fold of the protein by this mutation can be explained. G2375 resides at the end of PR3 of PRD B. Glycines are critical in this position to create a sharp bend for PR1 and PR3. There is an absence of a glycine at the end of PR2 and PR4 which results in an alternate conformation for the proceeding PR motif. Glycine is the only amino acid capable of creating such a sharp bend as the ones observed at the ends of PR1 and PR3. Mutagenesis of this residue to an arginine would therefore prevent such a sharp bend occurring and more than likely alter the fold. Since the protein was predominantly located in the insoluble fraction the altered conformation of the proceeding PR motif is likely to have had an effect on the overall conformation of the protein and destabilized the overall fold. The other PRD subdomains A and C were reported to bind vimentin with the same affinity as PRD B and the entire C-terminal tail encompassing all 3 (Choi et al, 2002). It is possible that misfolding of PRD B is unlikely to affect the fold of its neighbouring PRDs given that they can each be individually expressed and fold correctly. This would need validation with the full length desmoplakin C-terminal tail construct. However, what still remains to be answered is whether all 3 are required for desmin (the predominant heart cytoskeletal protein) binding and whether (for example in this particular instance) PRDs A and C are sufficient to substitute the destabilized fold in PRD B. Clearly this mutation has a profound effect on the patient as a reported 8 members of the same

family were affected and resulted in sudden cardiac death at an early age (Alcalai et al., 2003). Since the main function of the PRD is to bind to cytoskeletal proteins it is highly likely that PRD B will not be able to do so. Another possibility is that the expression of the entire full length desmoplakin protein may result in partial aggregation rendering the full length protein less able to conduct its role to its full potential. Indeed altered gene expression of plakoglobin was shown to lead to ARVC like symptoms in endurance-trained mice (Kirchhof et al., 2006). Interestingly a study looking at a G2056R mutation which similarly is involved in bend formation at the end of PR1 of PRD A resulted in less profound effects (Christensen et al., 2010a). The authors reported that the patient had mild skin defects such as skin blistering with no other cardiac associated problems. Clearly then for a mutation to be deleterious to life it must destabilize the overall fold but the domain it is located in may be more functionally important than others. This implies that while the G2375 residue is critical, the G2056 residue may not be, although causes skin problems which are tolerable.

The R2639Q mutation in PRD C of desmoplakin appears to have a similar affect on the subdomain fold in a similar fashion to that of G2375R. Mutation of R2639 to a glutamine results in the protein being expressed into inclusion bodies. As explained in the previous paragraph expression into inclusion bodies implies a destabilization event which is directly linked to the mutation. Again, similarly to the G2375R with the aid of the crystal structure and bioinformatics, the critical nature of the R2639Q mutation can be explained. R2639 is located in PR1 of PRD C at position 19. Residue 19 of each PR motif is usually occupied by a positively charged residue (typically arginines or lysines), although exceptions are observed. The positively charged residue is required for interaction with a negatively charged amino acid at position 4 of the motif. In this case the R2639 residue interacts with D2624 in an electrostatic manner. These interactions are required so that the β-hairpin is “locked” in position. Disruption of this interaction by

mutagenizing the arginine to a glutamine is therefore presumably deleterious to the conformation of the protein so that it results in a destabilized fold and causes the protein to be sequestered in inclusion bodies. As in the case of the G2375R mutation it is unlikely to have a direct effect on the folds of the other neighbouring PRDs. However, it cannot be discounted that the destabilized fold of the PRD C may result in aggregation of the full length desmoplakin protein *in vivo*. Furthermore, binding to cytoskeletal proteins may be disrupted although it should be noted that the G2375R and R2639Q mutations do not reside in the proposed vimentin binding groove. Furthermore to date no mutations have been reported in any of the PRD binding grooves. Instead it appears that the mutations studied in the PRDs of desmoplakin destabilize the fold resulting in ARVC.

The R2541K mutation was analyzed within the desmoplakin linker domain. Although a full biochemical analysis was not conducted due to time constraints, the relative differences observed were far less significant than those of the other mutations studied. The mutant variant of the protein was stable, soluble and no significant alterations were observed in its final yield or purity. Furthermore, the 2D-HSQC of the desmoplakin linker mutant variant superimposed onto the WT variant revealed very few chemical shift perturbations relative to those observed for the plakin domains. Although approximately 10% of the peaks experience chemical shift perturbations this is likely to be a reflection of an NMR artefact whereby minor backbone alterations reorientate the amides slightly. In contrast to the 2D-HSQCs observed for the plakin domain mutants, the slight reorientations are likely to be highly localized. As a result, although 10% of peaks had chemical shift perturbations the actual number of amino acids experiencing backbone conformational change is likely to be lower. Therefore, altogether this suggests that the R2541K mutation is having a minor effect on the biophysical aspects of the desmoplakin

linker domain. Unfortunately, to date the atomic structure of the linker domain is unknown and therefore the effect of the mutation on the fold remains to be seen.

Another possibility is that since so few chemical shift perturbations were observed R2541 may not be involved in folding but may in fact be surface exposed. This in turn would reduce the number of other amides the mutation would have an effect on. It was previously shown that the desmoplakin linker domain can bind to desmin (Lapouge et al., 2006) and therefore if the R2541K mutation was surface exposed could prevent binding to cytoskeletal proteins. The test whether the mutation affected desmin binding titrations of WT and mutant desmoplakin linker against unpolymerized desmin were conducted (unpublished results, Miss Penelope Rodriguez-Zamora, 2012). No binding was observed for either the WT or mutant variant which suggested that R2541K does not significantly affect fold, nor is the desmoplakin linker domain involved in binding to unpolymerized desmin. All together the biophysical data imply that this mutation may in fact be a polymorphism. Given it was previously shown to bind to polymerized desmin, it may have affected binding to only the polymerized variant of desmin and therefore needs investigating. Interestingly, the lead author who published the original R2541K mutation said the descendents of the deceased with whom also had the mutation were currently middle-aged and reported no signs of ARVC (Personal communication: Dr Alessandra Rampazzo, University of Padova, 2012). This was despite routine checkups with their doctors and highlights the need for caution for over interpreting the importance of mutations. Dr Alessandra Rampazzo conceded that the R2541K may be a mutation in which incomplete penatrance may be at play.

Recently it has become apparent that those with ARVC may in fact harbour multiple mutations across different genes in the desmosome (Xu et al., 2010). Xu et al highlighted the need for all cardiac expressed genes of the desmosome to be sequenced. Many previous studies have investigated only up to two genes and would therefore suggest that many mutations in the ARVC

database may not represent the “causative” mutation. Xu et al also suggested that currently it would be difficult to assess which of the mutations may/may not be contributing to the pathogenesis of ARVC. They concluded however, that it is highly likely that they presence of two or more mutations would significantly increase the chances of harbouring ARVC. Furthermore the prevalence of possible polymorphisms in the ARVC database was highlighted by two independent studies which suggest that some mutations reported as pathogenic are prevalent in the general population (Posch et al., 2008, Christensen et al., 2010b). Both authors conclude that while mutations may be disease modifying, they cannot be clearly extrapolated to being disease causing. To date the DSG2 mutations originally reported as pathogenic have been altered in light of this data in the ARVC database while, unfortunately, the same cannot be said for those reported as unknown pathogenecity in PKP2. This therefore presents a new problem in which scientists are overzealous in their reporting of novel “disease causing mutations” in desmosomal genes. Regardless, the presence of an ARVC database to discuss these sorts of phenomena is likely to be beneficial to the field as it allows healthy debate to further understand the ambiguities involved in the diagnosis of ARVC and the correlation between the genotype and phenotype.

The ARVC mutations investigated in this study represent a wide variety of types of mutants, in a diverse range of domains. It suggested that some mutations are more deleterious than others. Furthermore, and most importantly, it highlights the need for caution in terms of interpreting the importance of a desmosomal mutation which recently has been upgraded as a “major” criterion for the revised task force criteria (Marcus et al., 2010). It is highly likely that each mutation at the biophysical level will have a highly specific effect on the structure of the protein (like that observed for the mutations studied in PKP2a by Kirscher et al (2012)). Unfortunately, the speed at which new mutations are reported and the level of work required to interpret these proteins at

the biochemical level will never balance. However, this study on ARVC mutations lays a foundation that suggests that some mutations can be classified in a certain way. It was earlier suggested that a traffic light system could be used in which some mutations (G2375R and R2639Q) are highly deleterious or red, others are moderately deleterious or amber (K470E and R808C) and some are mildly deleterious or green (R2541K). Furthermore the grantham score (Grantham, 1974), a common tool used to quantify the deleterious effect on the protein in the ARVC database appears to have no correlation with the suggested traffic light system (figure 7.95). This highlights the need for rather looking at the amino acid substitution solely, its position in the three-dimensional fold of the domain to be taken as the predominant determinant of its effect. While the traffic light system is clearly an over-simplification it can set some ground rules for how mutations can be classed in the future based on their amino acid position, which ultimately would be helpful for diagnosis and prognosis. For example if a residue known to be critical to fold was mutagenized (i.e. W360 involved in the SH3-SR4 interface) it would likely cause the SH3 domain to detach from SR4 and become unfolded based on the evidence presented in chapter 3. A patient with a W360 mutation would be a mutation of top priority for cardiologists to maintain careful observation of. Unfortunately, very few mutations within the known atomic structures of desmoplakin have such mutations. Indeed 50% of the classified missense “pathogenic” desmoplakin mutations in the entire ARVC database reside in the plakin domain, yet the few examined in the C-terminal tail within this study resulted in the most deleterious effects. This therefore implies that some mutations are more tolerable than others, with those who have these mutants reaching an age where they can be passed on. The G2375R mutant is particularly fatal with most of the patients not reaching even 25 (Alcalai et al., 2003). This therefore implies that through natural selection the most deleterious of ARVC mutations are eliminated. However, it is clear that even ARVC mutations that are “tolerable” present a huge challenge to cardiologists treating those with ARVC. Clearly there are many questions that

remain unanswered and these are likely to come from understanding the biology of the desmosome further. These include desmosomal assembly and binding domains of one desmosomal protein to another.

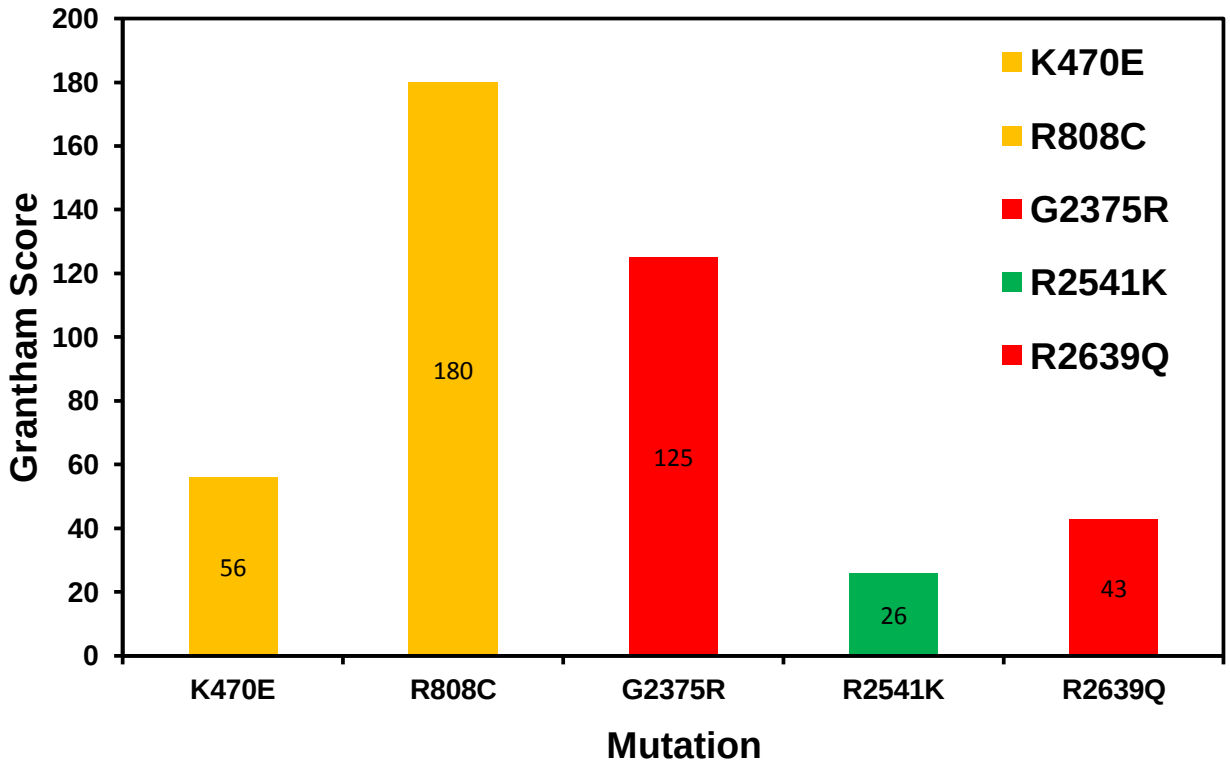


Figure 7.95 The Grantham score does not correlate with the traffic light system

Grantham scores were calculated for mutations investigated. Bars are coloured according to the traffic light system. The figure indicates a lack of correlation between the Grantham score for each mutation and the experimentally derived classification within the traffic light system.

Chapter 8 – Overall conclusions

This study has shed light on the biophysical properties of the functional N and C-termini "arms" of the desmosomally located plakin proteins desmoplakin, envoplakin and periplakin. Using a variety of biophysical techniques such as NMR, X-ray crystallography and SAXS (amongst others) it has provided a plethora of information which no individual technique could have given. Using SAXS and NMR the N-terminal plakin domain was shown to have a modular organization with rigid SR3-6 and SR78-CT regions separated by a flexible hinge. This was the case in all plakin domains examined. X-ray crystallography allowed for the first time atomic resolution comparisons of a non desmoplakin PRD from envoplakin, which SAXS confirmed was a monomer in physiological solution. Furthermore SAXS and NMR shed light on the plakin linker domains (which were previously thought of as semi-structured/disordered tethers). The data showed they were actually fully folded domains with PR motifs capable of binding IFs with differential abilities. Finally, using all the above mentioned major techniques, this study revealed the highly specific biophysical impacts of desmoplakin ARVC missense mutations. Overall it reveals a highly complex architecture for the plakin proteins, which is likely to impart protein specific functions such as cytoskeletal binding preferences.

The N-terminal head region of plakin proteins had previously been shown to be essential for desmosome assembly (Bornslaeger et al., 2001, Kowalczyk et al., 1997). The N-terminal region has also been shown to bind to plakophilin1, plakoglobin and weakly to desmoglein. This highlights its critical role in desmosome formation. Unfortunately, the binding site for these proteins has not been further narrowed down within the N-terminal region with the exception of plakophilin1. In a study published with my colleagues, we showed that plakophilin1 was unable to bind to SR56 and SR78-CT, and therefore deduced by inference that the plakophilin 1 binding site is located within SR34 (Al-Jassar et al, 2011). Unfortunately, this could not be directly

validated since the SR34 construct was insoluble upon expression in *E.coli* (chapter 3). Despite this the question remained – how does the N-terminal head region of desmoplakin bind to multiple binding partners and what is the role of the plakin domain? To answer this, the plakin domain architecture was assessed. It revealed that SR3-6 is a protease resistant, rigid entity (Choi and Weis, 2011; Al-Jassar, et al, 2011) which is followed by a flexible hinge region and a thermally stable SR78 domain. Indeed, SAXS analysis of each plakin domain revealed flexibility and excluded the possibility that plakin domains are linear rigid elements. Given that SRs are often exposed to mechanical stress and the N-terminal head region binds to multiple binding partners it opens up the possibility that the plakin domain could bind to certain partners depending on the mechanical stresses imposed on the cell. In a beautiful study under the umbrella of a new field termed "intracellular mechanotransduction" conducted by Michael Sheetz's lab, they investigated the two proteins vinculin and talin. Together they are functionally similar to plakin proteins in that they tether anchoring junctions (in this case focal adhesions) to the cytoskeletal network (Ziegler et al., 2008). The study showed, using a variety of single molecule techniques (such as magnetic tweezers, total internal reflection, fluorescence, and atomic force microscopy), that upon mechanical stimulation (i.e. stretching) of talin, vinculin was able to bind to cryptic binding sites on talin (del Rio et al., 2009). Although no protein has been directly implicated in mechanotransduction in the desmosomal field, the plakin protein plectin has been shown to have an influence on the mechanical stability of cells (Na et al., 2009). Furthermore it is well established that desmosomes experience mechanical stress and are expressed in tissues known to encounter mechanical stress (such as the epidermis and myocytes). Sheetz's findings have implications for the observations presented in this study. Firstly, all plakin domains were found to be non-linear *in-vitro*. Secondly unpublished *in-vitro* findings by Dr Martyn Chidgey showed that plakophilin1 was able to bind to desmoplakin but a number of different plakoglobin constructs could not. This is contrary to previous *in-vivo*

experiments which showed that both plakophilin1 and plakoglobin could bind to desmoplakin (Al-Jassar et al, 2011). Collectively, it could be argued that the N-terminal head region is acting as a "mechanosensor" to the cell's mechanical situation. The flexible plakin domain could potentially linearize (as opposed to being U or L-shaped in chapter 4) and the SRs could withstand the mechanical stress to compensate for the altered mechanical state of the cell. This could then initiate a change in conformation for desmoplakin where cryptic sites could be exposed for PG. PG is known to be important for stability in desmosomes (Kirchhof et al., 2006, Palka and Green, 1997). Finally, plakophilin1 could be released upon aberrant mechanical stresses and translocate to the nucleus to initiate downstream effects, something it is known to do (Mertens et al., 2001, Sobolik-Delmaire et al., 2010). These potential mechanisms may underpin the dynamic assembly of the desmosome in which it is proposed to disassemble under high mechanical stresses (Green and Gaudry, 2000). Exploring these possibilities would need a lab competent in single molecule techniques.

Given the possibility of a dynamic role in desmosome assembly at the N-terminal region, a similar process may occur at the C-terminal tail in terms of plakin proteins' cytoskeletal binding preferences and aiding IF assembly. The crystal structure of the envoplakin PRD and SAXS structures of the desmoplakin and periplakin linker domains were deduced in this study. In addition to this NMR titrations were conducted using unlabelled unpolymerized full length vimentin which revealed differential abilities to bind vimentin. The envoplakin PRD retained a conserved structure similar to that of its crystallized homologue, the desmoplakin PRD C. The putative positively charged vimentin binding groove was also preserved in the structure. Subtle differences were observed in the envoplakin PRD's ability to fold when compared to the desmoplakin counterpart. These included the presence of a random coil feature in PR4 which is centrally located in the putative vimentin binding groove and a negatively charged patch in a

different position found in desmoplakin which may also play a role in vimentin binding. The envoplakin PRD was also shown to be able to bind to polymerized vimentin with the same binding affinity as the desmoplakin PRD C. Although few structural differences were observed, the biophysical data obtained for the plakin linker domains may explain why differences are observed in cytoskeletal binding of full length plakin proteins as previously reported.

The linker domains of desmoplakin, envoplakin and periplakin were analyzed by NMR and SAXS. They revealed they are fully folded domains with a likely novel fold encompassing 2 PR motifs separated by a highly conserved WEE box. Unfortunately, the envoplakin linker domain was not analyzed by SAXS. However, the two remaining linker domains from desmoplakin and periplakin were successfully analyzed by SAXS and the data revealed significant differences. The desmoplakin linker domain was shown to be a flexible monomeric protein (figure 6.55). In contrast the periplakin linker domain was shown to be a rigid dimeric protein (figure 6.48). When unpolymerized vimentin was titrated against periplakin it saturated at a ratio of 1 to 2 (vimentin : linker domain) ratio. Binding between the envoplakin linker domain and vimentin was not observed under similar conditions. In addition to this unpolymerized desmin was unable to bind to the desmoplakin linker domain (unpublished results: Miss Penelope Rodriguez-Zamora). Previous reports have shown that linker domains from a variety of plakin proteins are the sole (or at least in combination with a PRD) determinants of cytoskeletal binding (Favre et al., 2011, Fontao et al., 2003, Kazerounian et al., 2002, Lapouge et al., 2006, Steinböck et al., 2000). Therefore, although binding was not observed for desmoplakin and envoplakin, linker domains are likely to have a significant effect on binding to polymerized cytoskeletal proteins. Furthermore, a heterogeneous sequence of a likely flexible “spacer” region was also identified which was present in desmosomal located proteins desmoplakin, plectin and envoplakin but not the hemidesmosomal protein BPAG1. Again this may reflect a determinant of cytoskeletal

preference. Collectively, given the large heterogeneity of their structural and binding capabilities it suggests that the linker domains are in fact the domains that dictate the cytoskeletal binding preferences of plakin proteins at their C-terminal tails. The heterogeneity between the PRDs and linker domain binding capabilities could allow the C-terminal tails of plakin proteins to be viewed as “hands”. The similar binding capabilities of the desmoplakin and envoplakin PRDs suggest they may act as a “palm”, while the linker domains act as the “fingers” to specify the plakin protein’s cytoskeletal preference. The binding of both unpolymerized and polymerized vimentin also implies they have a role in IF assembly too, although would need to be validated.

The greater understanding of the highly specialized structural domains from the N and C-termini of plakin proteins may therefore reflect the ambiguity surrounding the genotype-phenotype link in ARVC, especially in terms of missense mutations. Before this study was conducted no data was available regarding the biophysical impact of ARVC missense mutations on desmosomal proteins. The data presented in this study suggests that the precise location (as opposed to the type of substitution) of the ARVC mutation is highly likely to reflect how damaging it would be to the protein. For instance the G2375R and R2639Q had predicted to have the most destabilizing effect on the fold of the PRDs B and C of desmoplakin respectively. In contrast, R2541K and to a lesser extent K470E had the least significant effects. As a result, the Michael Overduin lab and I have proposed the “traffic light system”. The system proposes that the likelihood of a missense mutation's deleterious effects can be inferred from the mutation’s position in the fold of the domain. For instance if the amino acid is surface exposed, not conserved and non-critical to fold it would be classed as a “green mutation”. While an amino acid critical to fold, highly conserved or is involved in a binding pocket would be classed as a “red mutation”. This is particularly important now that a mutation within a desmosomal protein is classed as a major criterion given the recent revision of the taskforce criteria (Marcus et al, 2010). The lack of any significant

effects on some domains by some mutations raises the question of potential polymorphisms in the ARVC database, which although may be disease related, are not disease causing. The possibility of polymorphisms in the ARVC database has been raised by other labs (Posch et al., 2008, Christensen et al., 2010b). This is also corroborated by a very recent study from Kirchner et al (2012) which showed that some ARVC mutants in plakophilin2a caused destabilization of protein fold whereas others produced no observable effects. However, clearly some mutations may not have a significant impact on the structure or biochemistry of the domain yet alterations of binding grooves to as yet unknown ligands may occur. This therefore highlights the importance of combining biochemical, structural and medical studies in future ARVC research. There are however some limitations to the proposed traffic light system. For instance the mutagenesis of a glycine at the end of PR1 or PR3 in any PRD would destabilize the fold of most PRDs *in vitro* (as seen for the G2375R mutant in PRD B of desmoplakin). However, a G2056R mutant in desmoplakin PRD A at the end of PR1 in a middle aged patient caused only minor skin abnormalities. Using the system it would suggest it would be a severe problem to the patient at the molecular level, as for the patients who harboured the G2375R mutant could not reach beyond the age of 25 (Alcalai et al., 2003). However, clearly the patient with a similar mutation (G2056R) suggests that severity of the disease could not be extrapolated using the traffic light system. Thus caution should be applied when using the traffic light system. Interestingly, the G2056R data raises another question – why do some desmoplakin mutants cause skin defects while others cause skin and/or heart defects? Unfortunately, the answer to this is not clear based on the evidence in the literature or this study. The answer may lie in the two desmoplakin isoforms, DPI and DPII. DPI and DPII are expressed at similar levels; however, DPII is expressed at lower levels in the heart. It would be interesting to investigate whether expression patterns and/or protein function between the two desmoplakin isoforms differ with or without the presence of skin/heart causing mutants. Again this proposed investigation would

allow cardiologists to make more informed decisions about how to go about maintaining care for a patient and whether other relatives should similarly be screened. The presence of over 800 “disease causing” mutants in the ARVC database suggests that rather than focus on finding new mutants, the paradigm should be shifted towards new ways of dealing with the medical and genetic information to hand in order to help those afflicted with ARVC. Hopefully the traffic light system, in addition to biochemical studies on ARVC mutants, will take the first few steps in translating the vast amount of medical and genetic information to channelling research efforts to help maintain and treat those with ARVC.

Overall this study has shed light into the biophysical properties of the functional N and C-termini “arms” of plakin proteins. It suggests a complex interplay in the way the N-terminal region is organized as well as potential determinants of protein specific functions at the C-terminal tail. This study hopefully sets a foundation of work which can be built upon to further elaborate the mechanisms of plakin protein functions of tethering the desmosome to the IF cytoskeletal system.

Limitations

A number of limitations were present in this study and will be discussed.

Firstly since the work was conducted *in vitro*, it is possible that previously documented interactions *in vivo* that were not observed in this study are unlikely to reflect a real lack of an interaction. Rather an alternative view would be to interpret lack of interaction (as was the case with the envoplakin linker domain and vimentin) down to a requirement of an intermediary or synergistic interaction with another protein.

Secondly, although attempts were made to correlate structural/biophysical information of ARVC mutations with patient outcome in chapter 7 the picture is likely to be far more complex. Evidence which suggests this is the case include: the likely presence of polymorphisms in the ARVC database, incomplete genetic profiling of ARVC patients (i.e. multiple rather than single mutations in cardiac desmosomal genes), and as yet unknown binding sites for cardiac desmosomal proteins to one another. Other complexities include the possibility of misdiagnosis of ARVC due to its inherent difficult nature to diagnose.

Thirdly was the lack of AUC data on the effect of high concentrations on potential protein oligomerization. Throughout the study proteins were tested at low concentrations (typically 0.1 – 1 mg/ml) by AUC. However, as was the case with the periplakin linker domain it was evident that at low concentrations the protein was monomeric but this model did not fit very well when compared to a dimeric model when analyzed by SAXS at higher concentrations.

Finally another limitation was the lack of data confirming that the recombinant vimentin purified for this thesis was independently folded. Although the case, the divergent functional ability to bind between the different c-terminal structural domains of envoplakin, periplakin and desmoplakin implied it was indeed folded although not implicitly verified.

Future work

Clearly much work needs to be conducted to elaborate the findings presented in this study. The study aimed to investigate the biophysical mechanisms of plakin families that allowed desmosomal assembly, tethering to IFs as well as identifying how cell specific function was created. Since PKP and PG binding to the N-terminal of desmoplakin is imperative to desmosomal integrity, how these proteins interact should be investigated further. Whether there are cryptic sites for PG or whether both PG and PKP are able to bind together to desmoplakin should be looked into. This could be conducted by pull down assays as well as NMR titrations such as those used in this study.

The identification of the plakin linker domains as structured elements is an important discovery. Clearly their structural elucidation is of top priority and has almost been completed by other members of the Chidgey/Overduin lab as this thesis is being finalised. Importantly, why and how there are structural differences between the linker domains will also need to be investigated, as is whether this causes plakin protein cytoskeletal preferences.

Finally, how the PRDs and linker domains bind to IFs will be an important priority as currently this is unknown, although some mechanisms were proposed in chapter 5.

Bibliography

Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., Mccoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D Biological Crystallography,* 66**,** 213-221.

Aho, S., Li, K., Ryoo, Y., Mcgee, C., Ishida-Yamamoto, A., Uitto, J. & Klement, J. F. 2004. Periplakin gene targeting reveals a constituent of the cornified cell envelope dispensable for normal mouse development. *Molecular and cellular biology,* 24**,** 6410-8.

Al-Jassar, C., Knowles, T., Jeeves, M., Kami, K., Behr, E., Bikker, H., Overduin, M. & Chidgey, M. 2011. The Nonlinear Structure of the Desmoplakin Plakin Domain and the Effects of Cardiomyopathy-Linked Mutations. *Journal of molecular biology,* 411**,** 1049-1061.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. & Walter, P. 2002. *Molecular Biology of the Cell, 4th Edition,* New York, Garland Science.

Alcalai, R., Metzger, S., Rosenheck, S., Meiner, V. & Chajek-Shaul, T. 2003. A recessive mutation in desmoplakin causes arrhythmogenic right ventricular dysplasia, skin disorder, and woolly hair. *Journal of the American College of Cardiology,* 42**,** 319-327.

Amagai, M., Klaus-Kovtun, V. & Stanley, J. R. 1991. Autoantibodies against a novel epithelial cadherin in pemphigus vulgaris, a disease of cell adhesion. *Cell,* 67**,** 869-877.

Awad, M. M., Calkins, H. & Judge, D. P. 2008. Mechanisms of disease: molecular genetics of arrhythmogenic right ventricular dysplasia/cardiomyopathy. *Nature Clinical Practice Cardiovascular Medicine,* 5**,** 258-267.

Banks, G. B., Judge, L. M., Allen, J. M. & Chamberlain, J. S. 2010. The polyproline site in hinge 2 influences the functional capacity of truncated dystrophins. *PLoS genetics,* 6**,** e1000958.

Basso, C., Czarnowska, E., Della Barbera, M., Bauce, B., Beffagna, G., Wlodarska, E. K., Pilichou, K., Ramondo, A., Lorenzon, A., Wozniek, O., Corrado, D., Daliento, L., Danieli, G. A., Valente, M., Nava, A., Thiene, G. & Rampazzo, A. 2006. Ultrastructural evidence of intercalated disc remodelling in arrhythmogenic right ventricular cardiomyopathy: an electron microscopy investigation on endomyocardial biopsies. *European Heart Journal,* 27**,** 1847-1854.

Bauce, B., Nava, A., Beffagna, G., Basso, C., Lorenzon, A., Smaniotto, G., De Bortoli, M., Rigato, I., Mazzotti, E., Steriotis, A., Marra, M. P., Towbin, J. A., Thiene, G., Danieli, G. A. & Rampazzo, A. 2010. Multiple mutations in desmosomal proteins encoding genes in arrhythmogenic right ventricular cardiomyopathy/dysplasia. *Heart Rhythm,* 7**,** 22-29.

Bernadó, P., Mylonas, E., Petoukhov, M. V., Blackledge, M. & Svergun, D. I. 2007. Structural characterization of flexible proteins using small-angle X-ray scattering. *Journal of the American Chemical Society,* 129**,** 5656-64.

Bhasin, N., Law, R., Liao, G., Safer, D., Ellmer, J., Discher, B. M., Sweeney, H. L. & Discher, D. E. 2005. Molecular extensibility of mini-dystrophins and a dystrophin rod construct. *Journal of molecular biology,* 352**,** 795-806.

Bierkamp, C., Mclaughlin, K. J., Schwarz, H., Huber, O. & Kemler, R. 1996. Embryonic heart and skin defects in mice lacking plakoglobin. *Developmental biology,* 180**,** 780-5.

Bishr Omary, M. & Coulombe, P. A. 2004. *Intermediate Filament Cytoskeleton,* San Diego, USA, Elsevier Inc.

Bizzozero, G. 1864. Delle cellule cigliate, del reticolo Malpighiano dell'epedermide. *Annal Univ Med,* 190**,** 110-118.

Blikstad, I. 1983. Vimentin filaments are assembled from a soluble precursor in avian erythroid cells. *The Journal of Cell Biology,* 96**,** 1803-1808.

Boczonadi, V., Mcinroy, L. & Määttä, A. 2007. Cytolinker cross-talk: Periplakin N-terminus interacts with plectin to regulate keratin organisation and epithelial migration. *Experimental cell research,* 313**,** 3579-3591.

Bolling, M. C. & Jonkman, M. F. 2009. Skin and heart: une liaison dangereuse. *Experimental dermatology,* 18**,** 658-668.

Bonné, S., Gilbert, B., Hatzfeld, M., Chen, X., Green, K. J. & Van Roy, F. 2003. Defining desmosomal plakophilin-3 interactions. *The Journal of cell biology,* 161**,** 403-16.

Bornslaeger, E. A., Godsel, L. M., Corcoran, C. M., Park, J. K., Hatzfeld, M., Kowalczyk, A. P. & Green, K. J. 2001. Plakophilin 1 interferes with plakoglobin binding to desmoplakin, yet together with plakoglobin promotes clustering of desmosomal plaque complexes at cell-cell borders. *Journal of cell science,* 114**,** 727-38.

Broderick, M. J. F. & Winder, S. J. 2005. Spectrin, alpha-actinin, and dystrophin. *Advances in protein chemistry,* 70**,** 203-46.

Brown, P. H. & Schuck, P. 2006. Macromolecular size-and-shape distributions by sedimentation velocity analytical ultracentrifugation. *Biophysical journal,* 90**,** 4651-4661.

Chen, C. S. 2008. Mechanotransduction - a field pulling together? *Journal of Cell Science,* 121**,** 3285-3292.

Chen, X., Bonne, S., Hatzfeld, M., Van Roy, F. & Green, K. J. 2002. Protein binding and functional characterization of plakophilin 2. Evidence for its diverse roles in desmosomes and beta -catenin signaling. *The Journal of biological chemistry,* 277**,** 10512-22.

Chernyatina, A. A., Nicolet, S., Aebi, U., Herrmann, H. & Strelkov, S. V. 2012. Atomic structure of the vimentin central α-helical domain and its implications for intermediate filament assembly. *Proceedings of the National Academy of Sciences*.

Chidgey, M. 2011. Human Disease and the Desmosome. *Encyclopedia of Life Sciences.* John Wiley and Sons.

Chitaev, N., Averbakh, A., Troyanovsky, R. & Troyanovsky, S. 1998. Molecular organization of the desmoglein-plakoglobin complex. *Journal of Cell Science,* 111**,** 1941-1949.

Choi, H.-J., Gross, J. C., Pokutta, S. & Weis, W. I. 2009. Interactions of plakoglobin and beta-catenin with desmosomal cadherins: basis of selective exclusion of alpha- and beta-catenin from desmosomes. *The Journal of biological chemistry,* 284**,** 31776-88.

Choi, H.-J. & Weis, W. I. 2005. Structure of the armadillo repeat domain of plakophilin 1. *Journal of molecular biology,* 346**,** 367-76.

Choi, H.-J. & Weis, W. I. 2011. Crystal structure of a rigid four-spectrin-repeat fragment of the human desmoplakin plakin domain. *Journal of molecular biology,* 409**,** 800-12.

Choi, H. J., Park-Snyder, S., Pascoe, L. T., Green, K. J. & Weis, W. I. 2002. Structures of two intermediate filament-binding fragments of desmoplakin reveal a unique repeat motif structure. *Nature Structural and Molecular Biology,* 9**,** 612-620.

Christensen, A. H., Benn, M., Bundgaard, H., Tybjaerg-Hansen, A., Haunso, S. & Svendsen, J. H. 2010a. Wide spectrum of desmosomal mutations in Danish patients with arrhythmogenic right ventricular cardiomyopathy. *Journal of medical genetics,* 47**,** 736-44.

Christensen, A. H., Benn, M., Tybjaerg-Hansen, A., Haunso, S. & Svendsen, J. H. 2010b. Missense variants in plakophilin-2 in arrhythmogenic right ventricular cardiomyopathy patients--disease-causing or innocent bystanders? *Cardiology,* 115**,** 148-54.

Dauter, Z., Dauter, M. & Rajashankar, K. R. 2000. Novel approach to phasing proteins: derivatization by short cryo-soaking with halides. *Acta Crystallographica Section D Biological Crystallography,* 56**,** 232-237.

Del Rio, A., Perez-Jimenez, R., Liu, R., Roca-Cusachs, P., Fernandez, J. M. & Sheetz, M. P. 2009. Stretching single talin rod molecules activates vinculin binding. *Science,* 323**,** 638-41.

Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. 1995. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR,* 6**,** 277-293.

Dicolandrea, T., Karashima, T., Määttä, A. & Watt, F. M. 2000. Subcellular Distribution of Envoplakin and Periplakin. *The Journal of Cell Biology,* 151**,** 573-586.

Djinovic-Carugo, K., Gautel, M., Ylänne, J. & Young, P. 2002. The spectrin repeat: a structural platform for cytoskeletal protein assemblies. *FEBS letters,* 513**,** 119-23.

Dolinsky, T. J., Nielsen, J. E., Mccammon, J. A. & Baker, N. A. 2004. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Research,* 32**,** W665-W667.

Dominguez, C., Boelens, R. & Bonvin, A. M. J. J. 2003. HADDOCK: A Protein−Protein Docking Approach Based on Biochemical or Biophysical Information. *Journal of the American Chemical Society,* 125**,** 1731-1737.

Dyson, H. J. & Wright, P. E. 2005. Intrinsically unstructured proteins and their functions. *Nature reviews. Molecular cell biology,* 6**,** 197-208.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. 2010. Features and development of Coot. *Acta Crystallographica Section D Biological Crystallography,* 66**,** 486-501.

Ericsson, U. B., Hallberg, B. M., Detitta, G. T., Dekker, N. & Nordlund, P. 2006. Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Analytical Biochemistry,* 357**,** 289-298.

Favre, B., Schneider, Y., Lingasamy, P., Bouameur, J.-E., Begré, N., Gontier, Y., Steiner-Champliaud, M.-F., Frias, M. A., Borradori, L. & Fontao, L. 2011. Plectin interacts with

the rod domain of type III intermediate filament proteins desmin and vimentin. *European journal of cell biology,* 90**,** 390-400.

Foisner, R. 2001. Intermediate Filaments. *eLS.* John Wiley & Sons, Ltd.

Foisner, R. & Wiche, G. 1987. Structure and hydrodynamic properties of plectin molecules. *Journal of molecular biology,* 198**,** 515-531.

Fontao, L., Favre, B., Riou, S., Geerts, D., Jaunin, F., Saurat, J.-H., Green, K. J., Sonnenberg, A. & Borradori, L. 2003. Interaction of the Bullous Pemphigoid Antigen 1 (BP230) and Desmoplakin with Intermediate Filaments Is Mediated by Distinct Sequences within Their COOH Terminus. *Molecular Biology of the Cell,* 14**,** 1978-1992.

Franke, D. & Svergun, D. I. 2009. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *Journal of Applied Crystallography,* 42**,** 342-346.

Gallicano, G. I., Kouklis, P., Bauer, C., Yin, M., Vasioukhin, V., Degenstein, L. & Fuchs, E. 1998. Desmoplakin is required early in development for assembly of desmosomes and cytoskeletal linkage. *The Journal of cell biology,* 143**,** 2009-22.

Garboczi, D. N., Hung, D. T. & Wiley, D. O. N. C. 1992. HLA-A2-peptide complexes : Refolding and crystallization of molecules expressed in Eschericlhia coli and complexed with single antigenic peptides. 89**,** 3429-3433.

Garcia-Gras, E., Lombardi, R., Giocondo, M. J., Willerson, J. T., Schneider, M. D., Khoury, D. S. & Marian, A. J. 2006. Suppression of canonical Wnt/beta-catenin signaling by nuclear plakoglobin recapitulates phenotype of arrhythmogenic right ventricular cardiomyopathy. *The Journal of clinical investigation,* 116**,** 2012-21.

Garrod, D. & Chidgey, M. 2008. Desmosome structure, composition and function. *Biochimica et Biophysica Acta,* 1778**,** 572-587.

Garrod, D. R., Berika, M. Y., Bardsley, W. F., Holmes, D. & Tabernero, L. 2005. Hyper-adhesion in desmosomes: its regulation in wound healing and possible relationship to cadherin crystal structure. *Journal of cell science,* 118**,** 5743-54.

Geourjon, C. & Deléage, G. 1995. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Computer applications in the biosciences : CABIOS,* 11**,** 681-4.

Gomez, G. A., Mclachlan, R. W. & Yap, A. S. 2011. Productive tension: force-sensing and homeostasis of cell–cell junctions. *Trends in Cell Biology,* 21**,** 499-505.

Grantham, R. 1974. Amino Acid Difference Formula to Help Explain Protein Evolution. *Science,* 185**,** 862-864.

Green, K. J. & Gaudry, C. A. 2000. Are desmosomes more than tethers for intermediate filaments? *Nature reviews Molecular cell biology,* 1**,** 208-216.

Green, K. J., Parry, D. A., Steinert, P. M., Virata, M. L., Wagner, R. M., Angst, B. D. & Nilles, L. A. 1990. Structure of the human desmoplakins. Implications for function in the desmosomal plaque. *The Journal of biological chemistry,* 265**,** 11406-11407.

Groot, K. R., Sevilla, L. M., Nishi, K., Dicolandrea, T. & Watt, F. M. 2004. Kazrin, a novel periplakin-interacting protein associated with desmosomes and the keratinocyte plasma membrane. *The Journal of cell biology,* 166**,** 653-9.

Grum, V. L., Li, D., Macdonald, R. I. & Mondragón, A. 1999. Structures of two repeats of spectrin suggest models of flexibility. *Cell,* 98**,** 523-35.

Guinier, A. 1939. La diffraction des rayons X aux tres petits angles; application a l'etude de phenomenes ultramicroscopiques. *Annales de Physique (Paris),* 12**,** 161-237.

Guinier, A. F., G 1955. *Small Angle Scattering of X-rays,* New York, Wiley.

Haegel, H., Larue, L., Ohsugi, M., Fedorov, L., Herrenknecht, K. & Kemler, R. 1995. Lack of beta-catenin affects mouse development at gastrulation. *Development,* 121**,** 3529-3537.

Hatzfeld, M. & Weber, K. 1990. Tailless keratins assemble into regular intermediate filaments in vitro. *Journal of Cell Science,* 97**,** 317-324.

Haynes, C., Oldfield, C. J., Ji, F., Klitgord, N., Cusick, M. E., Radivojac, P., Uversky, V. N., Vidal, M. & Iakoucheva, L. M. 2006. Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. *PLoS computational biology,* 2**,** e100.

Heer, K. R., Heer, K., Bear, C., Kitty, S., Bear, P., Heer, C. a. S. & Bear, Z. 2012. Investigations into the interactions amongst teddy bears in an urban environment. *Teddy Bear Quarterly,* 4**,** 45-67.

Herrmann, H. & Aebi, U. 2004. INTERMEDIATE FILAMENTS: Molecular Structure, Assembly Mechanism, and Integration Into Functionally Distinct Intracellular Scaffolds. *Annual Review of Biochemistry,* 73**,** 749-789.

Herrmann, H., Häner, M., Brettel, M., Müller, S. A., Goldie, K. N., Fedtke, B., Lustig, A., Franke, W. W. & Aebi, U. 1996. Structure and assembly properties of the intermediate filament protein vimentin: the role of its head, rod and tail domains. *Journal of molecular biology,* 264**,** 933-53.

Hesse, M., Magin, T. M. & Weber, K. 2001. Genes for intermediate filament proteins and the draft sequence of the human genome. *Journal of Cell Science,* 114**,** 2569-2575.

Heuvel, A. P. J. V. D., Vries-Smits, A. M. M. D., Weeren, P. C. V., Dijkers, P. F., Bruyn, K. M. T. D., Riedl, J. A. & Burgering, B. M. T. 2002. Binding of protein kinase B to the plakin family member periplakin. *Journal of Cell Science,* 9.

Huber, A. H., Nelson, W. J. & Weis, W. I. 1997. Three-Dimensional Structure of the Armadillo Repeat Region of β-Catenin. *Cell,* 90**,** 871-882.

Insall, R. & Machesky, L. 2001. Cytoskeleton. *eLS.* John Wiley & Sons, Ltd.

Jacques, D. A. & Trewhella, J. 2010. Small-angle scattering for structural biology--expanding the frontier while avoiding the pitfalls. *Protein science : a publication of the Protein Society,* 19**,** 642-57.

Janda, L., Damborský, J., Rezniczek, G. A. & Wiche, G. 2001. Plectin repeats and modules: strategic cysteines and their presumed impact on cytolinker functions. *BioEssays : news and reviews in molecular, cellular and developmental biology,* 23**,** 1064-9.

Jefferson, J. J., Ciatto, C., Shapiro, L. & Liem, R. K. 2007. Structural analysis of the plakin domain of bullous pemphigoid antigen1 (BPAG1) suggests that plakins are members of the spectrin superfamily. *Journal of molecular biology,* 366**,** 244-257.

Jefferson, J. J., Leung, C. L. & Liem, R. K. H. 2004. Plakins: goliaths that link cell junctions and the cytoskeleton. *Nature reviews. Molecular cell biology,* 5**,** 542-53.

Jk3rd, W., Nieset, J. E., Sacco-Bubulya, P. A., Sadler, T. M., Johnson, K. R. & Wheelock, M. J. 2000. The amino- and carboxyl-terminal tails of (beta)-catenin reduce its affinity for desmoglein 2. *Journal of cell science,* 113 (Pt 1**,** 1737-45.

Kabsch, W. 2010. Xds. *Acta Crystallographica Section D Biological Crystallography,* 66**,** 125-132.

Kalinin, A. E., Idler, W. W., Marekov, L. N., Mcphie, P., Bowers, B., Steinert, P. M. & Steven, A. C. 2004. Co-assembly of Envoplakin and Periplakin into Oligomers and Ca2+-dependent Vesicle Binding. *The journal of Biological Chemistry,* 279**,** 22773-22780.

Kami, K., Chidgey, M., Dafforn, T. & Overduin, M. 2009. The desmoglein-specific cytoplasmic region is intrinsically disordered in solution and interacts with multiple desmosomal protein partners. *Journal of molecular biology,* 386**,** 531-543.

Kaneko, T., Li, L. & Li, S. S. 2008. The SH3 domain--a family of versatile peptide- and protein-recognition module. *Frontiers in bioscience : a journal and virtual library,* 13**,** 4938-4952.

Karashima, T., Tsuruta, D., Hamada, T., Ishii, N., Ono, F., Hashikawa, K., Ohyama, B., Natsuaki, Y., Fukuda, S., Koga, H., Sogame, R., Nakama, T., Dainichi, T. & Hashimoto, T. 2012. Interaction of plectin and intermediate filaments. *Journal of Dermatological Science,* 66**,** 44-50.

Karashima, T. & Watt, F. 2002. Interaction of periplakin and envoplakin with intermediate filaments. *Journal of Cell Science,* 115**,** 5027-5037.

Kazerounian, S., Uitto, J. & Aho, S. 2002. Unique role for the periplakin tail in intermediate filament association: specific binding to keratin 8 and vimentin. *Experimental dermatology,* 11**,** 428-38.

Kelley, L. A. & Sternberg, M. J. 2009. Protein structure prediction on the Web: a case study using the Phyre server. *Nature Protocols,* 4**,** 363-371.

Kirchhof, P., Fabritz, L., Zwiener, M., Witt, H., Schäfers, M., Zellerhoff, S., Paul, M., Athai, T., Hiller, K.-H., Baba, H. A., Breithardt, G., Ruiz, P., Wichter, T. & Levkau, B. 2006. Age- and Training-Dependent Development of Arrhythmogenic Right Ventricular Cardiomyopathy in Heterozygous Plakoglobin-Deficient Mice. *Circulation,* 114**,** 1799-1806.

Kirchner, F., Schuetz, A., Boldt, L.-H., Martens, K., Dittmar, G., Haverkamp, W., Thierfelder, L., Heinemann, U. & Gerull, B. 2012. Molecular Insights into Arrhythmogenic Right Ventricular Cardiomyopathy Caused by Plakophilin-2 Missense Mutations. *Circulation. Cardiovascular genetics***,** CIRCGENETICS.111.961854-.

Kiyokawa, C., Ruhrberg, C., Nie, Z., Karashima, T., Mori, O., Nishikawa, T., Green, K. J., Anhalt, G. J., Dicolandrea, T., Watt, F. M. & Hashimoto, T. 1998. Envoplakin and Periplakin are Components of the Paraneoplastic Pemphigus Antigen Complex. 111**,** 1236-1238.

Koenig, M. & Kunkel, L. M. 1990. Detailed analysis of the repeat domain of dystrophin reveals four potential hinge segments that may confer flexibility. *The Journal of biological chemistry,* 265**,** 4560-6.

Konarev, P. V., Petoukhov, M. V., Volkov, V. V. & Svergun, D. I. 2006. ATSAS 2.1, a program package for small-angle scattering data analysis. *Journal of Applied Crystallography,* 39**,** 277-286.

Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. 2003. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *Journal of Applied Crystallography,* 36**,** 1277-1282.

Kouklis, P. D., Hutton, E. & Fuchs, E. 1994. Making a connection: direct binding between keratin intermediate filaments and desmosomal proteins. *The Journal of cell biology,* 127**,** 1049-60.

Kowalczyk, A. P., Bornslaeger, E. A., Borgwardt, J. E., Palka, H. L., Dhaliwal, A. S., Corcoran, C. M., Denning, M. F. & Green, K. J. 1997. The amino-terminal domain of desmoplakin binds to plakoglobin and clusters desmosomal cadherin-plakoglobin complexes. *The Journal of cell biology,* 139**,** 773-84.

Kozin, M. B. & Svergun, D. I. 2001. Automated matching of high- and low-resolution structural models. *Journal of Applied Crystallography,* 34**,** 33-41.

Krissinel, E. & Henrick, K. 2007. Inference of Macromolecular Assemblies from Crystalline State. *Journal of molecular biology,* 372**,** 774-797.

Lahtinen, A. M., Lehtonen, E., Marjamaa, A., Kaartinen, M., Heliö, T., Porthan, K., Oikarinen, L., Toivonen, L., Swan, H., Jula, A., Peltonen, L., Palotie, A., Salomaa, V. & Kontula, K. 2011. Population-prevalent desmosomal mutations predisposing to arrhythmogenic right ventricular cardiomyopathy. *Heart Rhythm,* 8**,** 1214-1221.

Lapouge, K., Fontao, L., Champliaud, M.-F., Jaunin, F., Frias, M. A., Favre, B., Paulin, D., Green, K. J. & Borradori, L. 2006. New insights into the molecular basis of desmoplakinand desmin-related cardiomyopathies. *Journal of Cell Science,* 119**,** 4974-4985.

Laskowski, R. A. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography,* 26**,** 283-291.

Law, R., Liao, G., Harper, S., Yang, G., Speicher, D. W. & Discher, D. E. 2003. Pathway shifts and thermal softening in temperature-coupled forced unfolding of spectrin domains. *Biophysical journal,* 85**,** 3286-3293.

Lee, C.-H., Kim, M.-S., Chung, B. M., Leahy, D. J. & Coulombe, P. A. 2012. Structural basis for heteromeric assembly and perinuclear organization of keratin filaments. *Nature Structural and Molecular Biology,* 19**,** 707-715.

Legardinier, S., Legrand, B., Raguenes-Nicol, C., Bondon, A., Hardy, S., Tascon, C., Le Rumeur, E. & Hubert, J. F. 2009. A Two-amino Acid Mutation Encountered in Duchenne Muscular Dystrophy Decreases Stability of the Rod Domain 23 (R23) Spectrin-like Repeat of Dystrophin. *The Journal of biological chemistry,* 284**,** 8822-8832.

Leung, C. L., Green, K. J. & Liem, R. K. H. 2002. Plakins: a family of versatile cytolinker proteins. *Trends in cell biology,* 12**,** 37-45.

Li, J., Mahajan, A. & Tsai, M.-D. 2006. Ankyrin repeat: a unique motif mediating protein-protein interactions. *Biochemistry,* 45**,** 15168-78.

Liew, C. K., Gamsjaeger, R., Mansfield, R. E. & Mackay, J. P. 2008. NMR spectroscopy as a tool for the rapid assessment of the conformation of GST-fusion proteins. *Protein science : a publication of the Protein Society,* 17**,** 1630-5.

Linding, R., Russell, R. B., Neduva, V. & Gibson, T. J. 2003. GlobPlot: exploring protein sequences for globularity and disorder. *Nucleic Acids Research,* 31**,** 3701-3708.

Lobley, A., Whitmore, L. & Wallace, B. A. 2002. DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra. *Bioinformatics,* 18**,** 211-212.

Määttä, A., Dicolandrea, T., Groot, K. & Watt, F. M. 2001. Gene targeting of envoplakin, a cytoskeletal linker protein and precursor of the epidermal cornified envelope. *Molecular and cellular biology,* 21**,** 7047-53.

Mahoney, G., Aho, S., Uitto, J. & Stanley, J. R. 1998. The Members of the Plakin Family of Proteins Recognized by Paraneoplastic Pemphigus Antibodies Include Periplakin. *Society***,** 308-313.

Marblestone, J. G., Edavettal, S. C., Lim, Y., Lim, P., Zuo, X. & Butt, T. R. 2006. Comparison of SUMO fusion technology with traditional gene fusion systems: enhanced expression and solubility with SUMO. *Protein science : a publication of the Protein Society,* 15**,** 182-9.

Marcus, F. I., Fontaine, G. H., Guiraudon, G., Frank, R., Laurenceau, J. L., Malergue, C. & Grosgogeat, Y. 1982. Right ventricular dysplasia: a report of 24 adult cases. *Circulation,* 65**,** 384-98.

Marcus, F. I., Mckenna, W. J., Sherrill, D., Basso, C., Bauce, B., Bluemke, D. A., Calkins, H., Corrado, D., Cox, M. G. P. J., Daubert, J. P., Fontaine, G., Gear, K., Hauer, R., Nava, A., Picard, M. H., Protonotarios, N., Saffitz, J. E., Sanborn, D. M. Y., Steinberg, J. S., Tandri, H., Thiene, G., Towbin, J. A., Tsatsopoulou, A., Wichter, T. & Zareba, W. 2010. Diagnosis of arrhythmogenic right ventricular cardiomyopathy/dysplasia. *European Heart Journal,* 31**,** 806-814.

Mccoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. 2007. Phaser crystallographic software. *Journal of Applied Crystallography,* 40**,** 658-674.

Mckenna, W. J., Thiene, G., Nava, A., Fontaliran, F., Blomstrom-Lundqvist, C., Fontaine, G. & Camerini, F. 1994. Diagnosis of arrhythmogenic right ventricular

dysplasia/cardiomyopathy. Task Force of the Working Group Myocardial and Pericardial Disease of the European Society of Cardiology and of the Scientific Council on Cardiomyopathies of the International Society. *British heart journal,* 71**,** 215-8.

Meng, J.-J., Bornslaeger, E. A., Green, K. J., Steinert, P. M. & Ip, W. 1997a. Two-hybrid Analysis Reveals Fundamental Differences in Direct Interactions between Desmoplakin and Cell Type-specific Intermediate Filaments. *The journal of Biological Chemistry,* 272**,** 21495-21503.

Meng, J. J., Bornslaeger, E. A., Green, K. J., Steinert, P. M. & Ip, W. 1997b. Two-hybrid analysis reveals fundamental differences in direct interactions between desmoplakin and cell type-specific intermediate filaments. *The Journal of biological chemistry,* 272**,** 21495-503.

Mertens, C., Hofmann, I., Wang, Z., Teichmann, M., Chong, S. S., Schnölzer, M. & Franke, W. W. 2001. Nuclear particles containing RNA polymerase III complexes associated with the junctional plaque protein plakophilin 2. *Proceedings of the National Academy of Sciences,* 98**,** 7795-7800.

Mirza, A., Sagathevan, M., Sahni, N., Choi, L. & Menhart, N. 2010. A biophysical map of the dystrophin rod. *Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics,* 1804**,** 1796-1809.

Mongiovı, A. M., Romano, P. R., Panni, S., Mendoza, M., Wong, W. T., Musacchio, A., Cesareni, G., Paolo, P. & Fiore, D. 1999. A novel peptide – SH3 interaction. *Embo Journal,* 18**,** 5300-5309.

Musacchio, A., Noble, M., Pauptit, R., Wierenga, R. & Saraste, M. 1992. Crystal structure of a Src-homology 3 (SH3) domain. *Nature,* 359**,** 851-855.

Musacchio, A., Saraste, M. & Wilmanns, M. 1994. High-resolution crystal structures of tyrosine kinase SH3 domains complexed with proline-rich peptides. *Nature Structural and Molecular Biology,* 1**,** 546-551.

Na, S., Chowdhury, F., Tay, B., Ouyang, M., Gregor, M., Wang, Y., Wiche, G. & Wang, N. 2009. Plectin contributes to mechanical properties of living cells. *American journal of physiology. Cell physiology,* 296**,** C868-77.

Nagulapalli, M., Parigi, G., Yuan, J., Gsponer, J., Deraos, G., Bamm, Vladimir v., Harauz, G., Matsoukas, J., De planque, Maurits r. R., Gerothanassis, Ioannis p., Babu, M. M., Luchinat, C. & Tzakos, Andreas g. 2012. Recognition Pliability Is Coupled to Structural Heterogeneity: A Calmodulin Intrinsically Disordered Binding Region Complex. *Structure (London, England : 1993),* 20**,** 522-533.

Nie, Z., Merritt, A., Rouhi-Parkouhi, M., Tabernero, L. & Garrod, D. 2011. Membrane-impermeable Cross-linking Provides Evidence for Homophilic, Isoform-specific Binding of Desmosomal Cadherins in Epithelial Cells. *The Journal of Biological Chemistry,* 286**,** 2143-2154.

Nikolic, B., Mac Nulty, E., Mir, B. & Wiche, G. 1996. Basic amino acid residue cluster within nuclear targeting sequence motif is essential for cytoplasmic plectin-vimentin network junctions. *The Journal of cell biology,* 134**,** 1455-67.

North, A., Bardsley, W., Hyam, J., Bornslaeger, E., Cordingley, H., Trinnaman, B., Hatzfeld, M., Green, K., Magee, A. & Garrod, D. 1999. Molecular map of the desmosomal plaque. *Journal of Cell Science,* 112**,** 4325-4336.

Ortega, E., Buey, R. M., Sonnenberg, A. & De Pereda, J. M. 2011. The structure of the plakin domain of plectin reveals a non-canonical SH3 domain interacting with its fourth spectrin repeat. *The Journal of biological chemistry*.

Otterspoor, L. C., Reichert, C. L. A., Cramer, M. J. M., Bhuiyan, Z. A., Wilde, A. a. M. & Hauer, R. N. W. 2007. Arrhythmogenic right ventricular cardiomyopathy: asymptomatic to life threatening as illustrated by the cases of two sisters. *Netherlands heart journal : monthly*

journal of the Netherlands Society of Cardiology and the Netherlands Heart Foundation, 15**,** 348-53.

Palka, H. L. & Green, K. J. 1997. Roles of plakoglobin end domains in desmosome assembly. *Journal of cell science,* 110 (Pt 1**,** 2359-71.

Papadakis, M., Sharma, S., Cox, S., Sheppard, M. N., Panoulas, V. F. & Behr, E. R. 2009. The magnitude of sudden cardiac death in the young: a death certificate-based review in England and Wales. *Europace,* 11**,** 1353-1358.

Park, S., Caffrey, M. S., Johnson, M. E. & Fung, L. W.-M. 2003. Solution structural studies on human erythrocyte alpha-spectrin tetramerization site. *The Journal of biological chemistry,* 278**,** 21837-44.

Parry, D. A., Crewther, W. G., Fraser, R. D. & Macrae, T. P. 1977. Structure of alpha-keratin: structural implication of the amino acid sequences of the type I and type II chain segments. *Journal of molecular biology,* 113**,** 449-54.

Pascual, J., Pfuhl, M., Rivas, G., Pastore, A. & Saraste, M. 1996. The spectrin repeat folds into a three-helix bundle in solution. *FEBS Letters,* 383**,** 201-207.

Pascual, J., Pfuhl, M., Walther, D., Saraste, M. & Nilges, M. 1997. Solution structure of the spectrin repeat: a left-handed antiparallel triple-helical coiled-coil. *Journal of molecular biology,* 273**,** 740-51.

Petoukhov, M. V., Konarev, P. V., Kikhney, A. G. & Svergun, D. I. 2007. ATSAS 2.1 - towards automated and web-supported small-angle scattering data analysis. *Journal of Applied Crystallography,* 40**,** s223-s228.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of Computational Chemistry,* 25**,** 1605-1612.

Posch, M. G., Posch, M. J., Perrot, A., Dietz, R. & Ozcelik, C. 2008. Variations in DSG2: V56M, V158G and V920G are not pathogenic for arrhythmogenic right ventricular dysplasia/cardiomyopathy. *Nature clinical practice. Cardiovascular medicine,* 5**,** E1.

Protonotarios, N., Anastasakis, A., Antoniades, L., Chlouverakis, G., Syrris, P., Basso, C., Asimaki, A., Theopistou, A., Stefanadis, C., Thiene, G., Mckenna, W. J. & Tsatsopoulou, A. 2011. Arrhythmogenic right ventricular cardiomyopathy/dysplasia on the basis of the revised diagnostic criteria in affected families with desmosomal mutations. *European Heart Journal,* 32**,** 1097-104.

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

Quinlan, R. A., Hatzfeld, M., Franke, W. W., Lustig, A., Schulthess, T. & Engel, J. 1986. Characterization of dimer subunits of intermediate filament proteins. *Journal of molecular biology,* 192**,** 337-49.

Ren, J., Wen, L., Gao, X., Jin, C., Xue, Y. & Yao, X. 2009. DOG 1.0: illustrator of protein domain structures. *Cell Res,* 19**,** 271-273.

Roy, A., Kucukural, A. & Zhang, Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols,* 5**,** 725-738.

Ruhrberg, C., Hajibagheri, M. A., Parry, D. A. & Watt, F. M. 1997. Periplakin, a novel component of cornified envelopes and desmosomes that belongs to the plakin family and forms complexes with envoplakin. *The Journal of cell biology,* 139**,** 1835-49.

Ruiz, P. 1996. Targeted mutation of plakoglobin in mice reveals essential functions of desmosomes in the embryonic heart. *The Journal of Cell Biology,* 135**,** 215-225.

Russell, D., Andrews, P. D., James, J. & Lane, E. B. 2004. Mechanical stress induces profound remodelling of keratin filaments and cell junctions in epidermolysis bullosa simplex keratinocytes. *Journal of cell science,* 117**,** 5233-43.

Schaffer, J. 1920. *Vorlesungen ber Histologie und Histogenese* Leipzig, Germany, W Engelmann.

Schrodinger, Llc 2010. The PyMOL Molecular Graphics System, Version 1.3r1.

Semenyuk, A. V. & Svergun, D. I. 1991. GNOM - a program package for small-angle scattering data processing. *Journal of Applied Crystallography,* 24**,** 537-540.

Sevilla, L. M., Nachat, R., Groot, K. R., Klement, J. F., Uitto, J., Djian, P., Määttä, A. & Watt, F. M. 2007. Mice deficient in involucrin, envoplakin, and periplakin have a defective epidermal barrier. *The Journal of cell biology,* 179**,** 1599-612.

Skerrow, C. & Matoltsy, A. 1974. Chemical characterization of isolated epidermal desmosomes. *The Journal of Cell Biology,* 63**,** 524-530.

Sobolik-Delmaire, T., Reddy, R., Pashaj, A., Roberts, B. J. & Wahl, J. K. 2010. Plakophilin-1 localizes to the nucleus and interacts with single-stranded DNA. *The Journal of investigative dermatology,* 130**,** 2638-46.

Sonnenberg, A. & Liem, R. K. 2007a. Plakins in development and disease. *Experimental cell research,* 313**,** 2189-203.

Sonnenberg, A. & Liem, R. K. H. 2007b. Plakins in development and disease. *Experimental cell research,* 313**,** 2189-203.

Sonnenberg, A., Rojas, A. M. & De Pereda, J. M. 2007. The structure of a tandem pair of spectrin repeats of plectin reveals a modular organization of the plakin domain. *Journal of molecular biology,* 368**,** 1379-1391.

South, A. P. 2004. Plakophilin 1: an important stabilizer of desmosomes. *Clinical and Experimental Dermatology,* 29**,** 161-167.

Spurny, R. 2008. *Plectin-vimentin interaction: intermediate filament network formation, dynamics, and nitrosylation-induced collapse.* Vienna University.

Stappenbeck, T. S., Bornslaeger, E. A., Corcoran, C. M., Luu, H. H., Virata, M. L. & Green, K. J. 1993. Functional analysis of desmoplakin domains: specification of the interaction with keratin versus vimentin intermediate filament networks. *The Journal of cell biology,* 123**,** 691-705.

Stappenbeck, T. S. & Green, K. J. 1992. The desmoplakin carboxyl terminus coaligns with and specifically disrupts intermediate filament networks when expressed in cultured cells. *The Journal of cell biology,* 116**,** 1197-209.

Stappenbeck, T. S., Lamb, J. A., Corcoran, C. M. & Green, K. J. 1994. Phosphorylation of the desmoplakin COOH terminus negatively regulates its interaction with keratin intermediate filament networks. *The Journal of biological chemistry,* 269**,** 29351-4.

Steinböck, F. A., Nikolic, B., Coulombe, P. A., Fuchs, E., Traub, P. & Wiche, G. 2000. Dose-dependent linkage, assembly inhibition and disassembly of vimentin and cytokeratin 5/14 filaments through plectin's intermediate filament-binding domain. *Journal of cell science,* 113 (Pt 3**,** 483-91.

Strelkov, S. V., Herrmann, H., Geisler, N., Wedig, T., Zimbelmann, R., Aebi, U. & Burkhard, P. 2002. Conserved segments 1A and 2B of the intermediate filament dimer: their atomic structures and role in filament assembly. *Embo Journal,* 21**,** 1255-1266.

Svergun, D., Barberato, C. & Koch, M. H. J. 1995. CRYSOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *Journal of Applied Crystallography,* 28**,** 768-773.

Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L.-W., Read, R. J. & Adams, P. D. 2007. Iterative model building, structure refinement

and density modification with the PHENIX AutoBuild wizard. *Acta Crystallographica Section D Biological Crystallography,* 64**,** 61-69.

Thomason, H. A., Scothern, A., Mcharg, S. & Garrod, D. R. 2010. Desmosomes: adhesive strength and signalling in health and disease. *The Biochemical journal,* 429**,** 419-33.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research,* 22**,** 4673-4680.

Troyanovsky, R., Chitaev, N. & Troyanovsky, S. 1996. Cadherin binding sites of plakoglobin: localization, specificity and role in targeting to adhering junctions. *Journal of Cell Science,* 109**,** 3069-3078.

Van Der Zwaag, P. A., Jongbloed, J. D., Van Den Berg, M. P., Van Der Smagt, J. J., Jongbloed, R., Bikker, H., Hofstra, R. M. & Van Tintelen, J. P. 2009. A genetic variants database for arrhythmogenic right ventricular dysplasia/cardiomyopathy. *Human Mutation,* 30**,** 1278-1283.

Van Tintelen, J. P., Entius, M. M., Bhuiyan, Z. A., Jongbloed, R., Wiesfeld, A. C. P., Wilde, A. a. M., Van Der Smagt, J., Boven, L. G., Mannens, M. M. a. M., Van Langen, I. M., Hofstra, R. M. W., Otterspoor, L. C., Doevendans, P. a. F. M., Rodriguez, L.-M., Van Gelder, I. C. & Hauer, R. N. W. 2006. Plakophilin-2 Mutations Are the Major Determinant of Familial Arrhythmogenic Right Ventricular Dysplasia/Cardiomyopathy. *Circulation,* 113**,** 1650-1658.

Vasioukhin, V., Bowers, E., Bauer, C., Degenstein, L. & Fuchs, E. 2001. Desmoplakin is essential in epidermal sheet formation. *Nature Cell Biology,* 3**,** 1076-1085.

Virata, M. 1992. Molecular Structure of the Human Desmoplakin I and II Amino Terminus. *Proceedings of the National Academy of Sciences,* 89**,** 544-548.

Volkov, V. V. & Svergun, D. I. 2003. Uniqueness of ab initio shape determination in small-angle scattering. *Journal of Applied Crystallography,* 36**,** 860-864.

Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J. & Laue, E. D. 2005. The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins,* 59**,** 687-696.

Vuzman, D. & Levy, Y. 2012. Intrinsically disordered regions as affinity tuners in protein-DNA interactions. *Molecular bioSystems,* 8**,** 47-57.

Wegener, J. 2001. Cell Junctions. *eLS.* John Wiley & Sons, Ltd.

Whitmore, L. & Wallace, B. A. 2004. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Research,* 32**,** W668-73.

Whittock, N. V., Wan, H., Morley, S. M., Garzon, M. C., Kristal, L., Hyde, P., Mclean, W. H. I., Pulkkinen, L., Uitto, J., Christiano, A. M., Eady, R. a. J. & Mcgrath, J. A. 2002. Compound Heterozygosity for Non-Sense and Mis-Sense Mutations in Desmoplakin Underlies Skin Fragility//Woolly Hair Syndrome. 118**,** 232-238.

Wilkinson, D. L. & Harrison, R. G. 1991. Predicting the solubility of recombinant proteins in Escherichia coli. *Bio/technology (Nature Publishing Company),* 9**,** 443-8.

Wriggers, W. & Chacon, P. 2001. Using Situs for the registration of protein structures with low-resolution bead models from X-ray solution scattering. *Journal of Applied Crystallography,* 34**,** 773-776.

Xu, T., Yang, Z., Vatta, M., Rampazzo, A., Beffagna, G., Pillichou, K., Scherer, S. E., Saffitz, J., Kravitz, J., Zareba, W., Danieli, G. A., Lorenzon, A., Nava, A., Bauce, B., Thiene, G., Basso, C., Calkins, H., Gear, K., Marcus, F. & Towbin, J. A. 2010. Compound and Digenic Heterozygosity Contributes to Arrhythmogenic Right Ventricular Cardiomyopathy. *Journal of the American College of Cardiology,* 55**,** 587-597.

Yan, Y., Winograd, E., Viel, A., Cronin, T., Harrison, S. C. & Branton, D. 1993. Crystal structure of the repetitive segments of spectrin. *Science,* 262**,** 2027-2030.

Yang, Z., Bowles, N. E., Scherer, S. E., Taylor, M. D., Kearney, D. L., Ge, S., Nadvoretskiy, V. V., Defreitas, G., Carabello, B., Brandon, L. I., Godsel, L. M., Green, K. J., Saffitz, J. E., Li, H., Danieli, G. A., Calkins, H., Marcus, F. & Towbin, J. A. 2006. Desmosomal dysfunction due to mutations in desmoplakin causes arrhythmogenic right ventricular dysplasia/cardiomyopathy. *Circulation Research,* 99**,** 646-655.

Ylänne, J., Scheffzek, K., Young, P. & Saraste, M. 2001. Crystal Structure of the α-Actinin Rod Reveals an Extensive Torsional Twist. *Structure,* 9**,** 597-604.

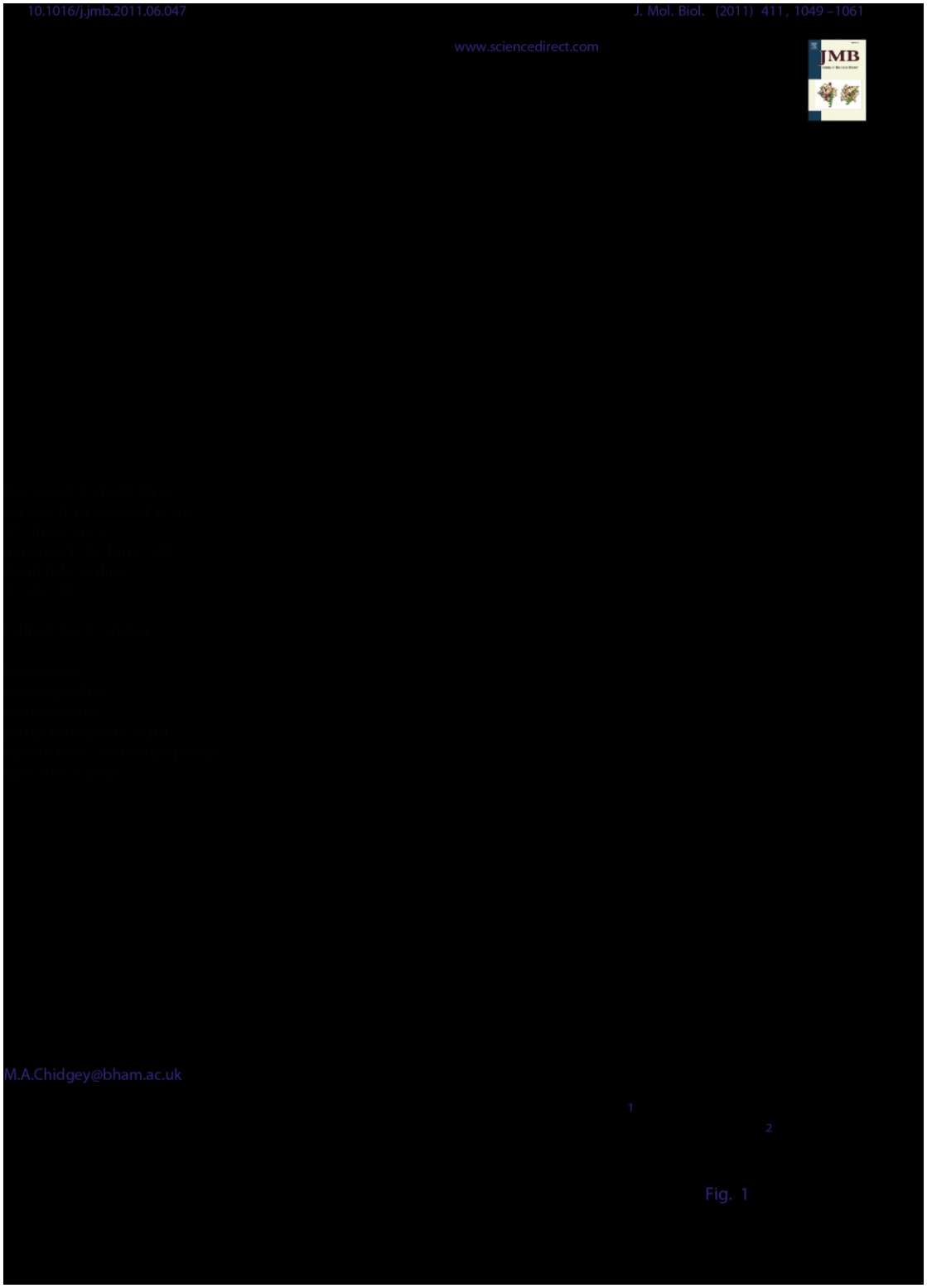
Yu, C.-C., Yu, C.-H., Hsueh, C.-H., Yang, C.-T., Juang, J.-M., Hwang, J.-J., Lin, J.-L. & Lai, L.-P. 2008. Arrhythmogenic Right Ventricular Dysplasia: Clinical Characteristics and Identification of Novel Desmosome Gene Mutations. *Journal of the Formosan Medical Association,* 107**,** 548-558.

Yu, H., Chen, J. K., Feng, S., Dalgarno, D. C., Brauer, A. W. & Schrelber, S. L. 1994. Structural basis for the binding of proline-rich peptides to SH3 domains. *Cell,* 76**,** 933-945.

Zhurinsky, J., Shtutman, M. & Ben-Ze'ev, A. 2000. Plakoglobin and beta-catenin: protein interactions, regulation and biological roles. *Journal of Cell Science,* 113**,** 3127-3139.

Ziegler, W. H., Gingras, A. R., Critchley, D. R. & Emsley, J. 2008. Integrin connections to the cytoskeleton through talin and vinculin. *Biochemical Society transactions,* 36**,** 235-9.

Appendix - A1


10.1016/j.jmb.2011.06.047
J. Mol. Biol. (2011) 411, 1049–1061
www.sciencedirect.com
JMB
M.A.Chidgey@bham.ac.uk
1
2
Fig. 1

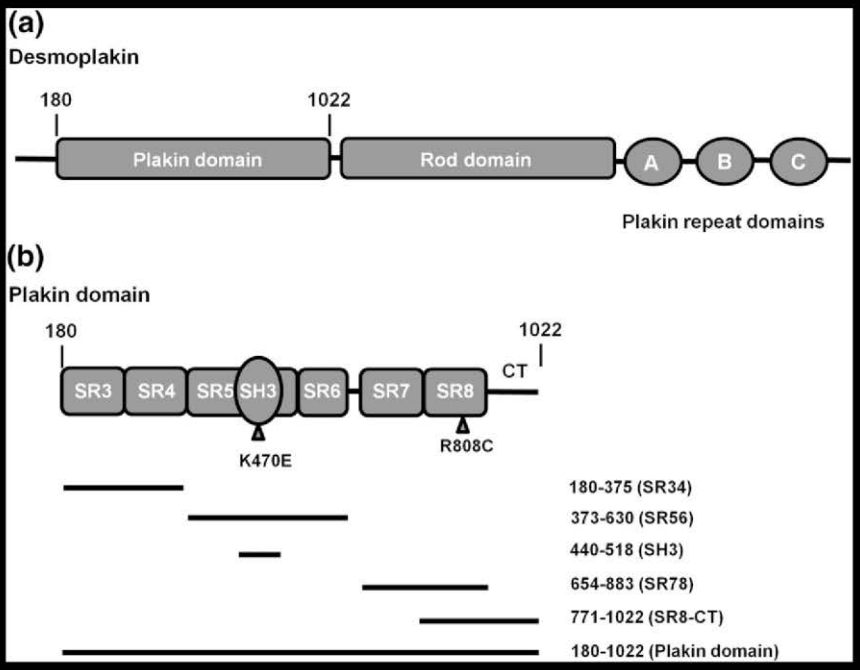

(a)
Desmoplakin
180
1022
Plakin domain
Rod domain
A
B
C
Plakin repeat domains
(b)
Plakin domain
180
1022
CT
SR3
SR4
SR5
SH3
SR6
SR7
SR8
K470E
R808C
180-375 (SR34)
373-630 (SR56)
440-518 (SH3)
654-883 (SR78)
771-1022 (SR8-CT)
180-1022 (Plakin domain)

Fig. 1

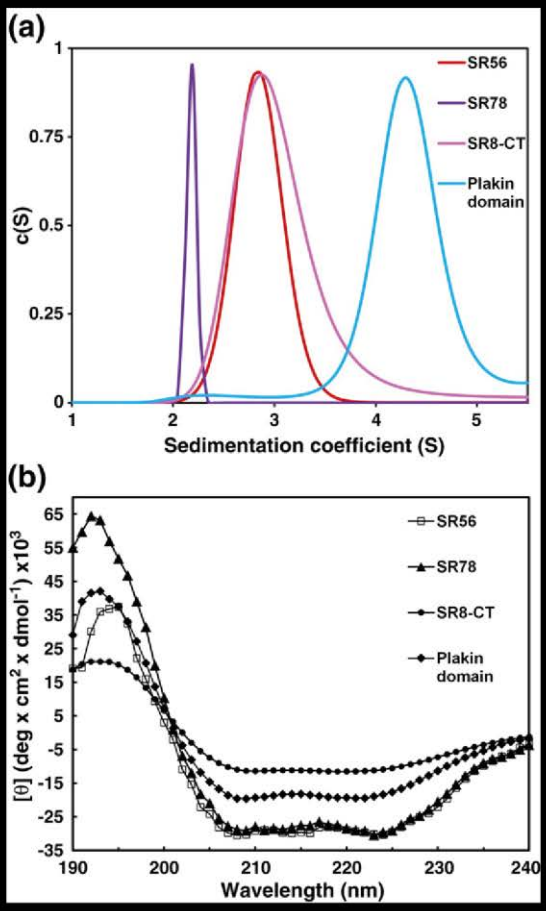

(a)
SR56
SR78
SR8-CT
Plakin domain
c(S)
Sedimentation coefficient (S)
(b)
SR56
SR78
SR8-CT
Plakin domain
[θ] (deg x cm² x dmol⁻¹) x10³
Wavelength (nm)

Fig. 1

Fig. 1

Fig. 2

Fig. 2

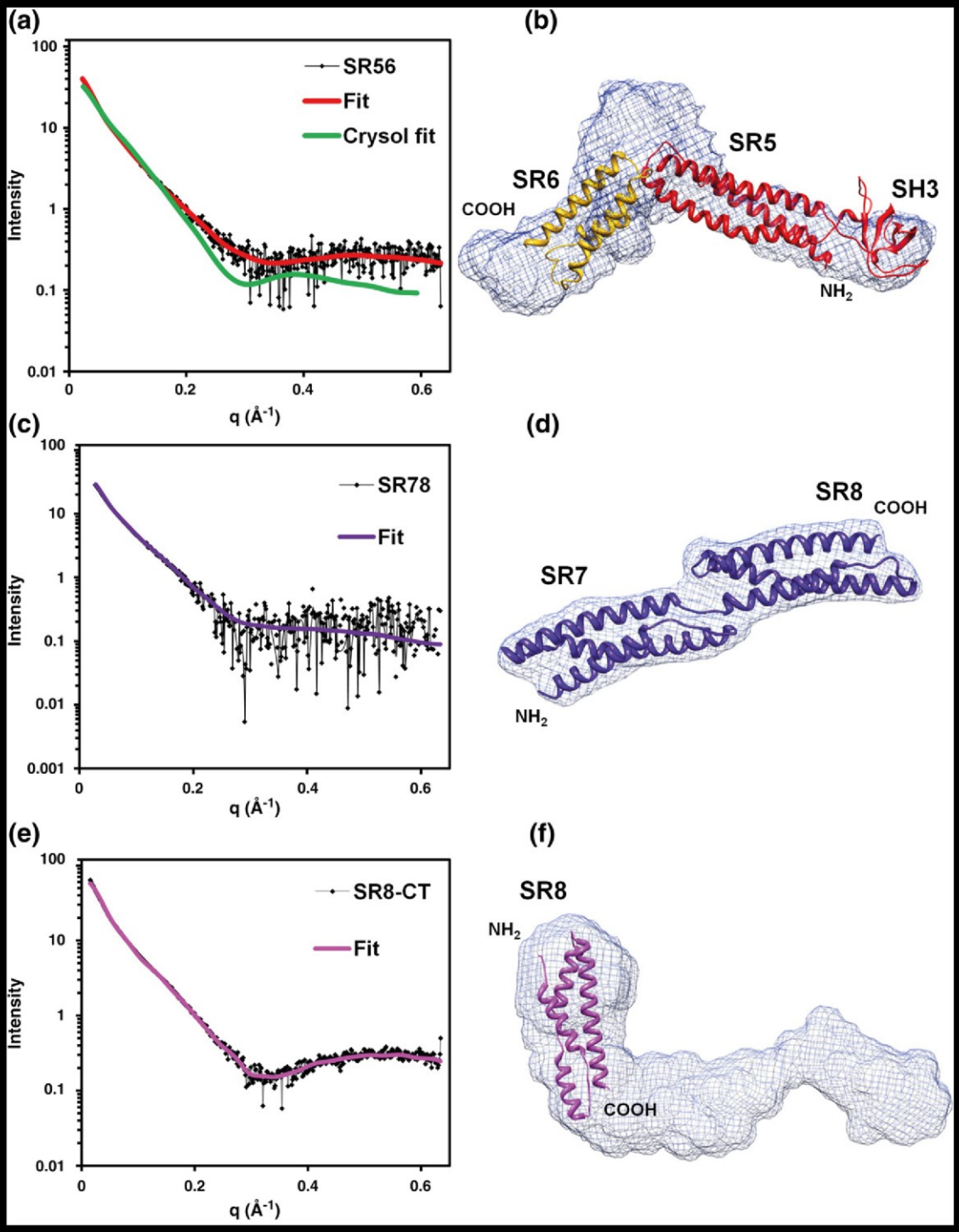

(a)
SR56
Fit
Crysol fit
Intensity
q (Å^{-1})
(b)
SR5
SR6
SH3
COOH
NH$_2$
(c)
SR78
Fit
Intensity
q (Å^{-1})
(d)
SR8
COOH
SR7
NH$_2$
(e)
SR8-CT
Fit
Intensity
q (Å^{-1})
(f)
SR8
NH$_2$
COOH

Fig. 3a and b

Fig. 3c and d

Fig. 2

Fig. 3e and f

16

17

Fig. 2

Table 1

Fig. 3

18

Fig. 3

Table 1

Fig. 2

19

Fig. 3

17

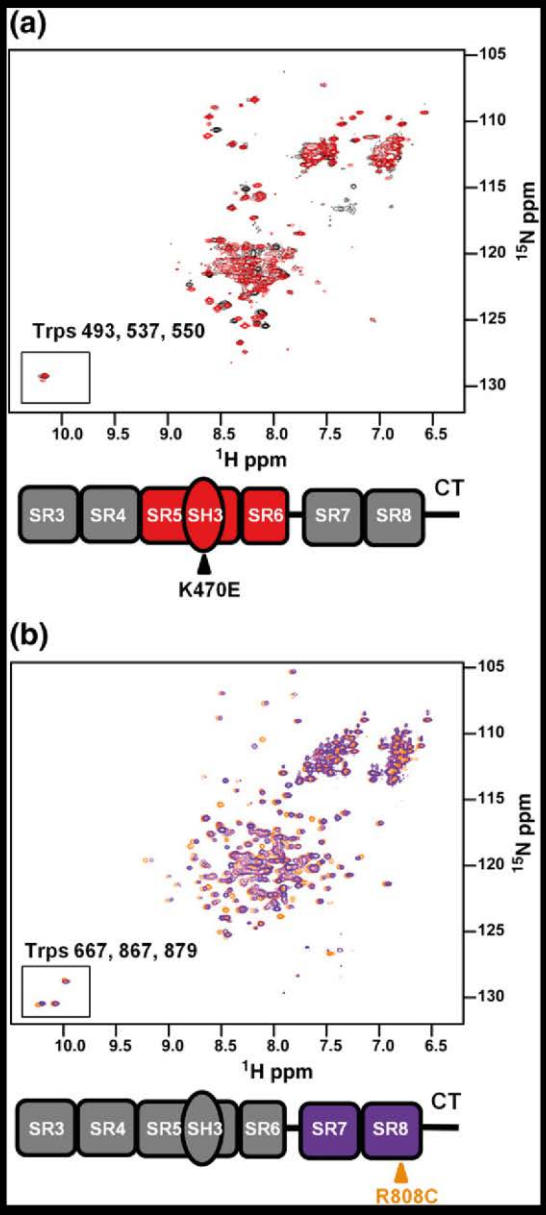


Fig. 3

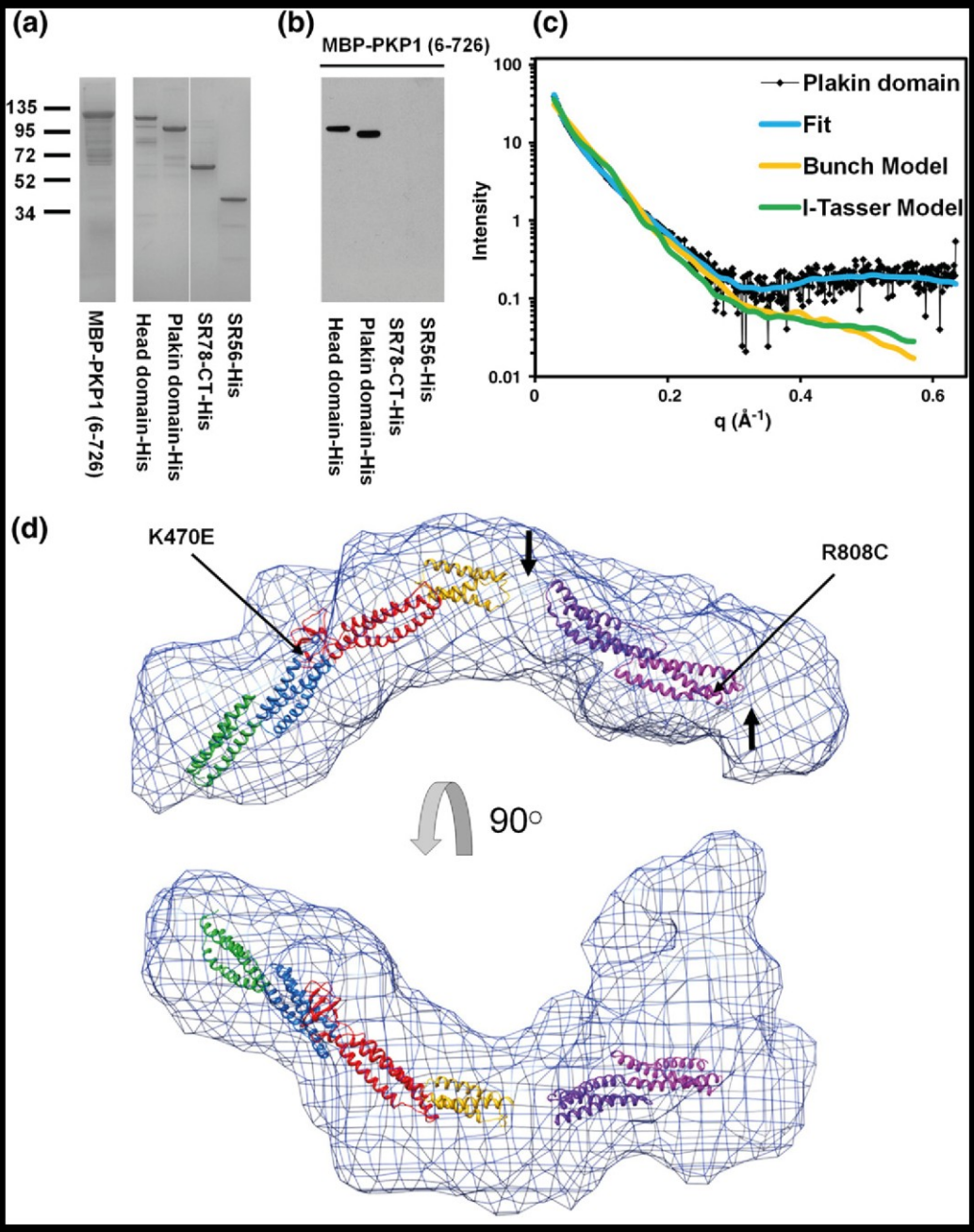
Fig. 2

20

Table 2

Fig. 4

21


(a)
135
95
72
52
34
MBP-PKP1 (6-726)
Head domain-His
Plakin domain-His
SR78-CT-His
SR56-His
(b)
MBP-PKP1 (6-726)
Head domain-His
Plakin domain-His
SR78-CT-His
SR56-His
(c)
Plakin domain
Fit
Bunch Model
I-Tasser Model
Intensity
100
10
1
0.1
0.01
0
0.2
0.4
0.6
q (Å-1)
(d)
K470E
R808C
90°

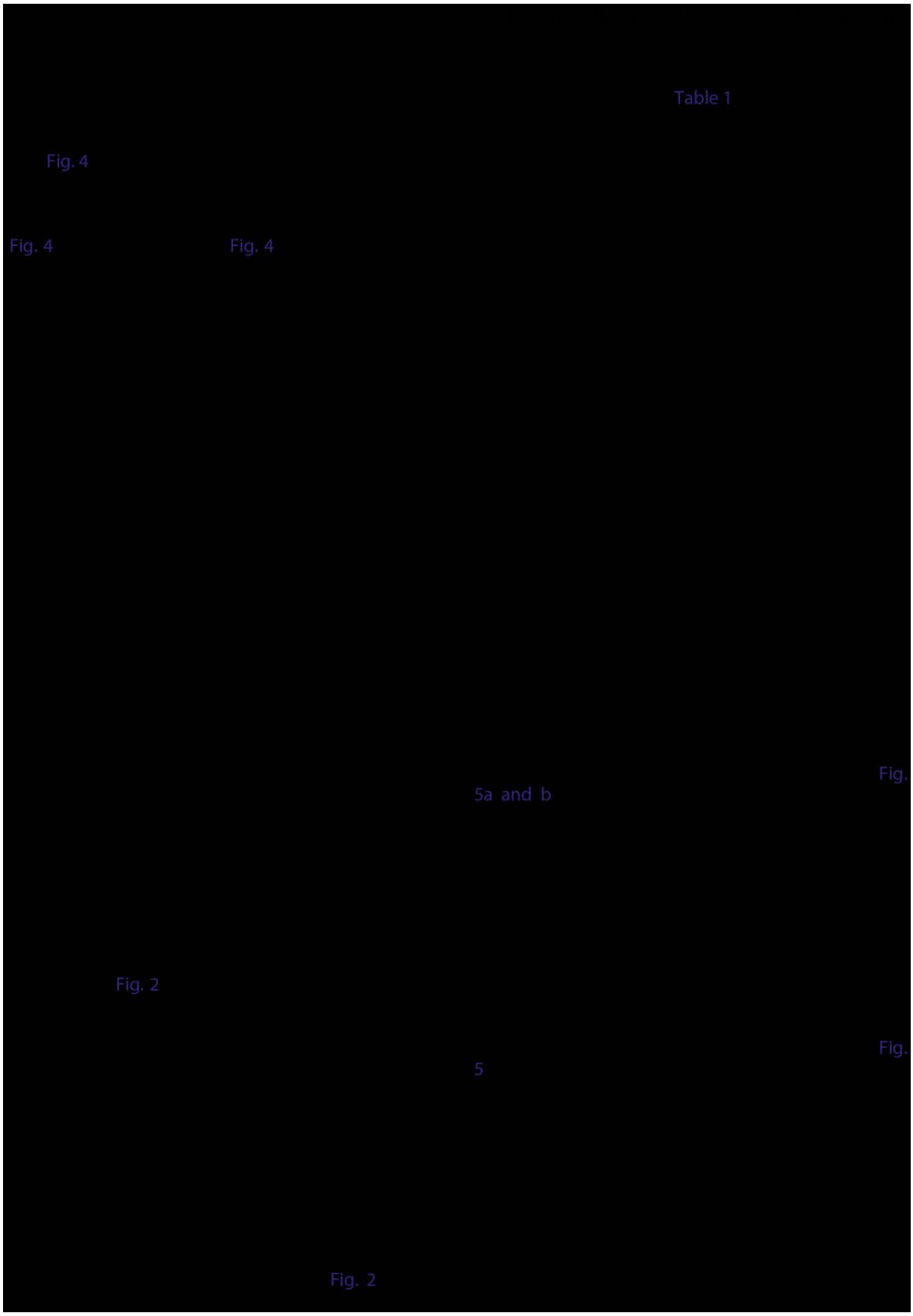
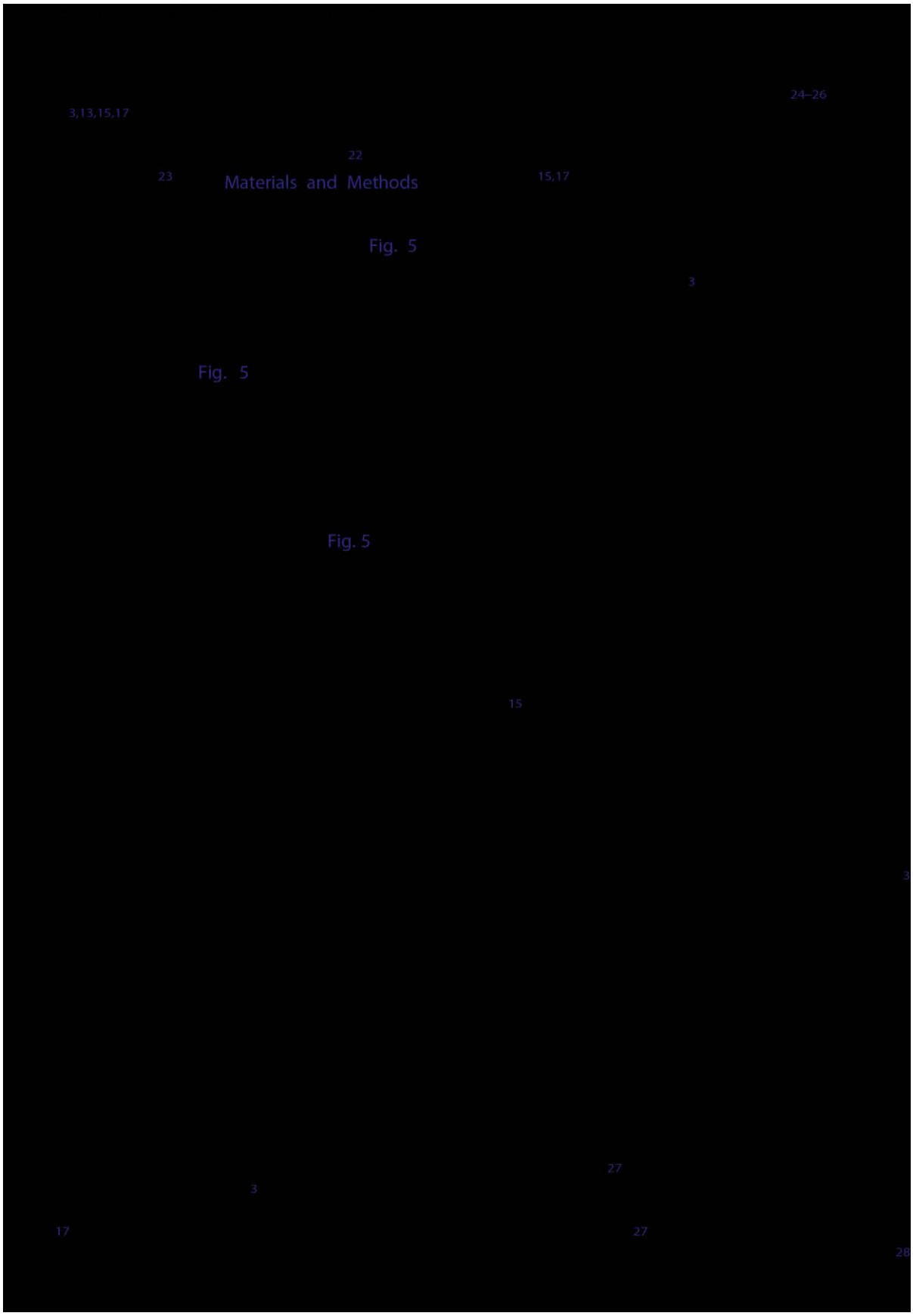


Table 1
Fig. 4
Fig. 4
Fig. 4
Fig.
5a and b
Fig. 2
Fig.
5
Fig. 2


24–26
3,13,15,17
22
23
Materials and Methods
15,17
Fig. 5
3
Fig. 5
Fig. 5
15
3
27
3
17
27
28

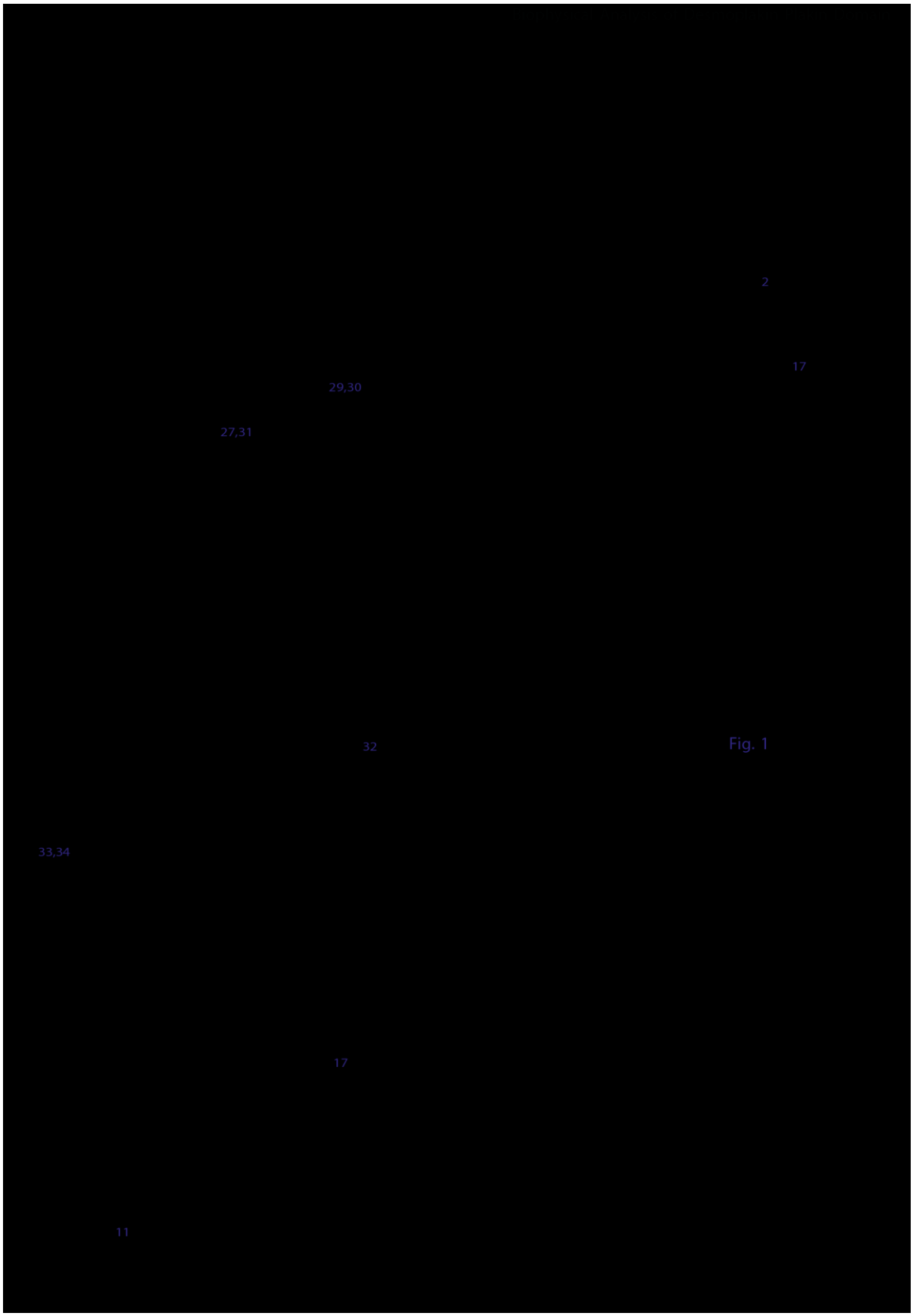
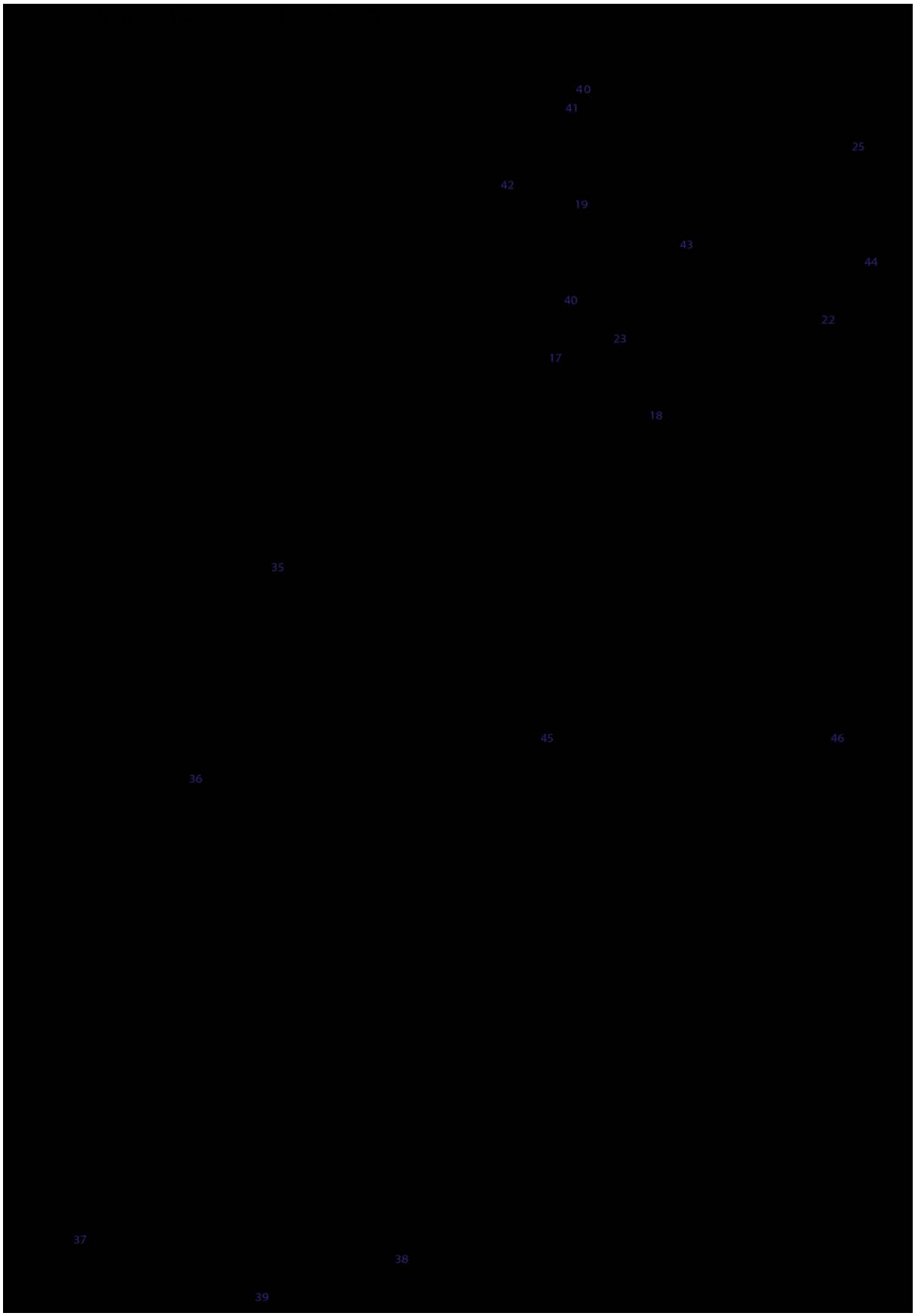
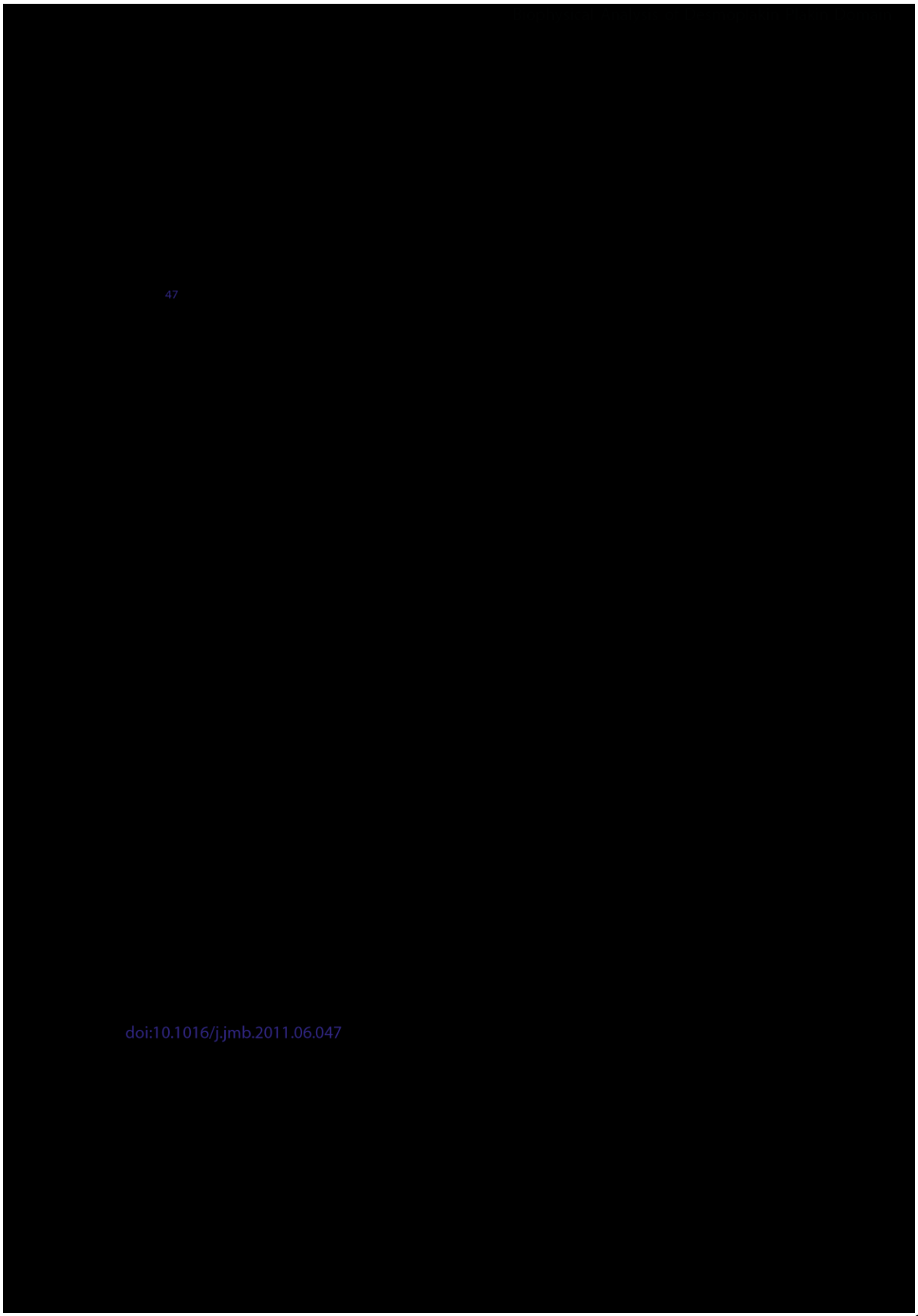
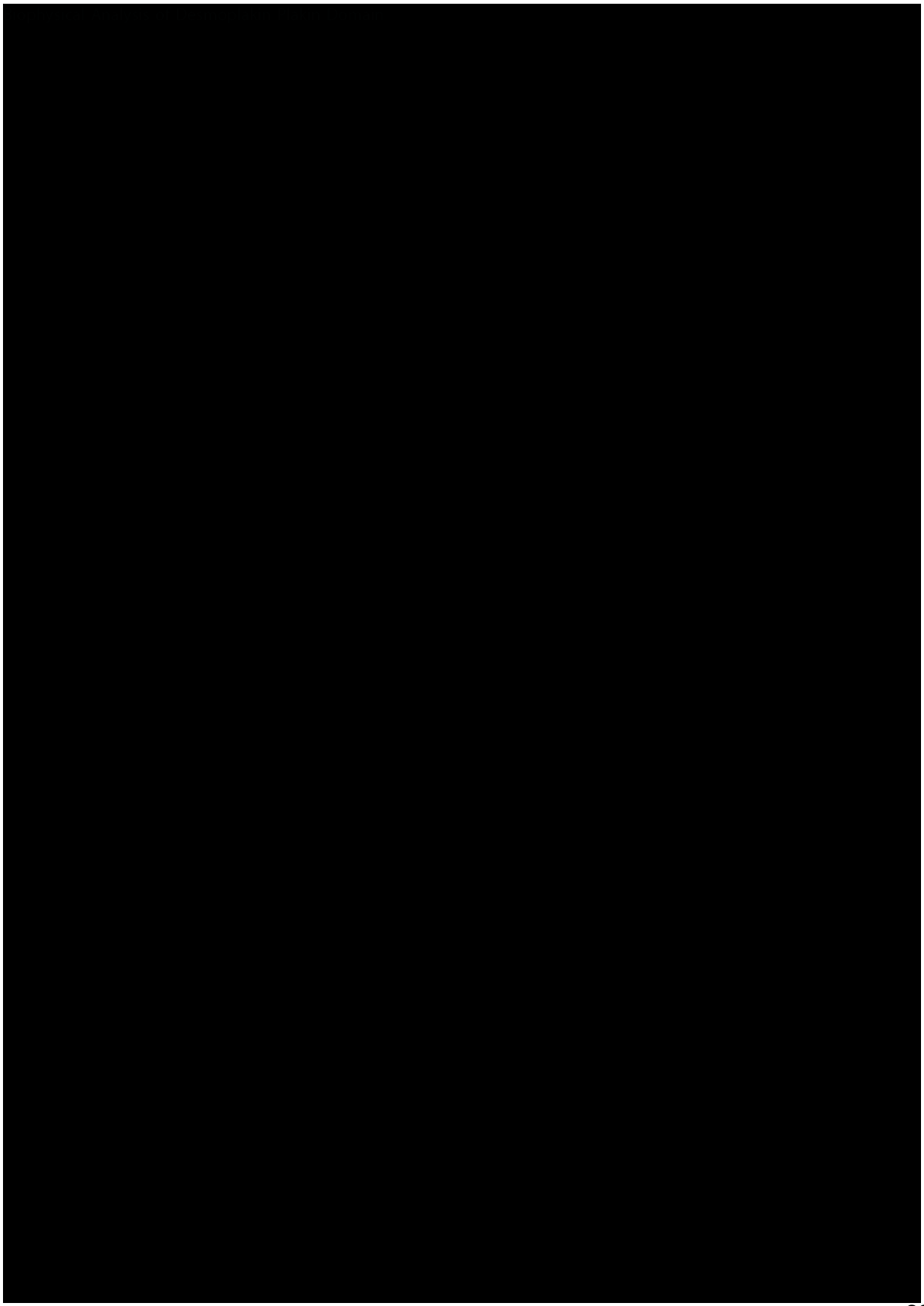


Fig. 1


40
41
25
42
19
43
44
40
22
23
17
18
35
45
46
36
37
38
39



doi:10.1016/j.jmb.2011.06.047



Appendix – A2

Buffer list

Recipes for GST purification buffers:

0.5M-sodium phosphate (pH7.2)	5ml
1M-NaCl	10ml

	100ml

(25mM-phosphate (pH7.2)/100mM-NaCl)

Elution buffer

Glutathione	80mg
1M-Tris.Cl (pH8.0)	2ml
1M-NaCl	2.5ml
Water	5.5ml

	10ml

(30mM glutathione/200mM-Tris.Cl (pH8.0)/250mM-NaCl)

His-purification buffers:

His-trap binding buffer:
20mM Hepes, 500mM NaCl, 10mM Imidazole (pH 7.5)

His-trap wash buffer:
20mM Hepes, 500mM NaCl, 30mM Imidazole (pH 7.5)

His-trap Elution Buffer:
20mM Hepes, 500mM NaCl, 250mM Imidazole (pH 7.5)

Ion-exchange Q sepharose buffers

Binding buffer:
20mM Hepes pH 7.5

Elution buffer
20mM Hepes pH 7.5 and 500mM NaCl

Recipes for vimentin refold buffers

Triton Wash buffer:

(for 1L), 5ml Triton X-100 (100%), 50ml 4M NaCl, 20 ml 0.5 M EDTA (pH8), 10 ml 10 % Na Azide, 50 ml 1M Tris (pH 8)

Non-Triton wash buffer:

(for 1L), 20 ml 0.5 M EDTA (pH 8), 10 ml 10 % Na Azide, 50 ml 1M Tris (pH 8)

Solubilization buffer:

20ml of 10mM Tris-HCl pH 8, 0.2mM EGTA/EDTA (Optional), 5mM DTT (optional), and 8M Urea

NMR expression media:

M9 media salts (1 l)

6 g	Na_2HPO_4
3g	KH_2PO_4
0.5	NaCl

Nutrient Mix

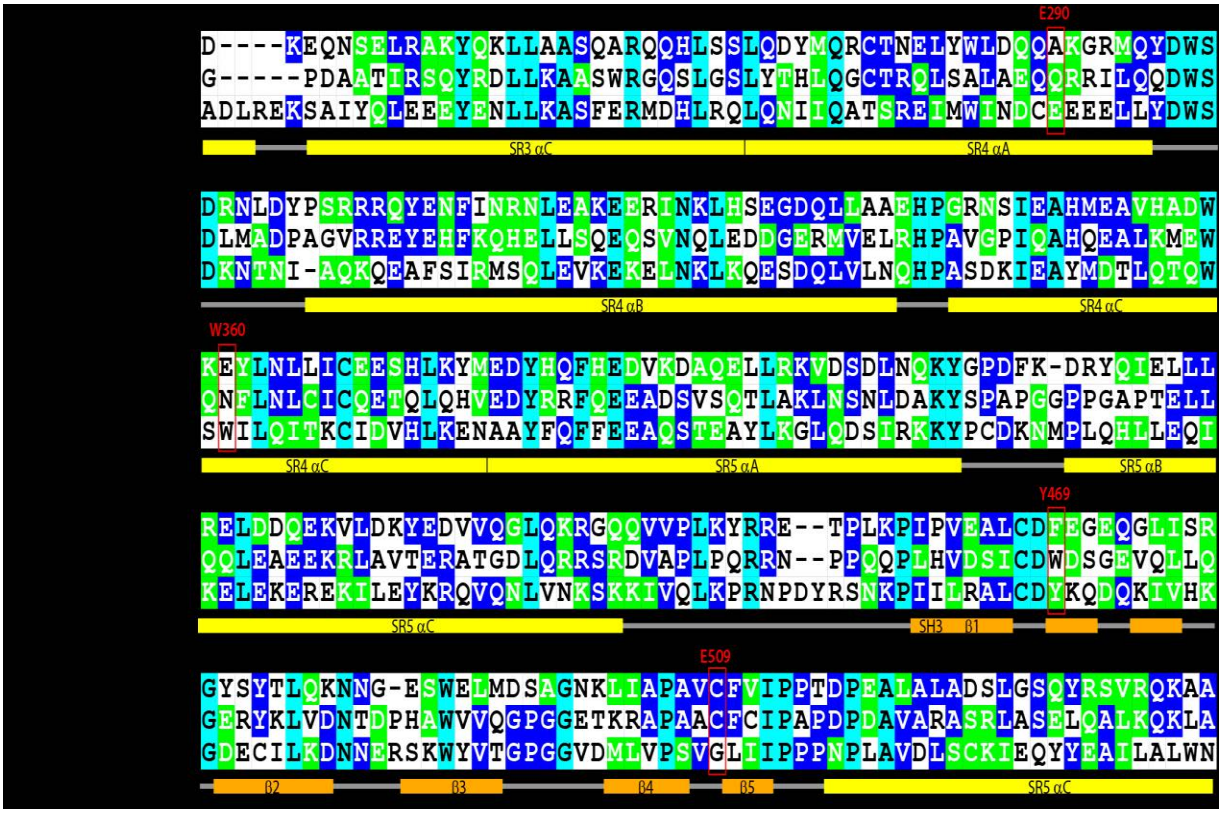
10 µl	1 M $CaCl_2$
2 ml	1M $MgSO_4$
400 µl	1M $FeCl_3$
1 ml	20 mg/ml Thiamine
1 g	$^{15}NH_4Cl$
0.5 ml	“Metal Mix”

Metal mix (100 ml)

0.115 g	$ZnSO_4$
0.0169 g	$MnSO_4$
0.029 g	$H3BO_3$
0.0175 g	$CuSO_4$

* In the exception of cleavage experiments, most purification buffers contained a complete protease inhibitor.

Appendix B


E290
SR3 αC
SR4 αA
SR4 αB
SR4 αC
W360
SR4 αC
SR5 αA
SR5 αB
Y469
SR5 αC
SH3
β1
E509
β2
β3
β4
β5
SR5 αC

Appendix - C

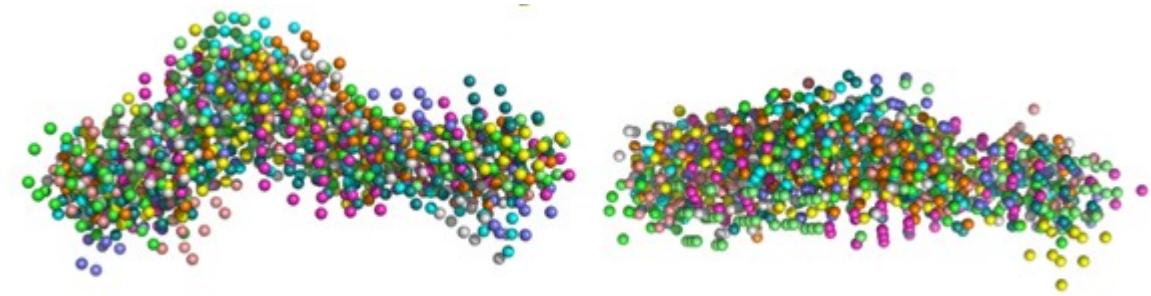
```
          character*3  wed
          character*5  xmodel

          dimension g1mat(11000,3)
          dimension v654mat(11000,3)
          dimension F8767mat(11000,3)
          dimension Q447mat(11000,3)

          dimension mat_d1(1000)
          dimension mat_d2(1000)

          dimension mat_d1sel(1000)
          dimension mat_d2sel(1000)

          dimension d1_vec(11000)
          dimension d2_vec(11000)

          OPEN(1,FILE='P1.txt',STATUS='OLD')
          OPEN(2,FILE='G523.txt',STATUS='OLD')
C         OPEN(3,FILE='F8767_coord.txt',STATUS='OLD')
C         OPEN(4,FILE='Q447_coord.txt',STATUS='OLD')

          OPEN(5,FILE='best_curve001.txt',STATUS='OLD')

C         Definition of the histrogram

          start=20.0   !holaaaaa
          fin=400.0
          width=2.0

          nbins=(fin-start)/width

          do i=1,nbins
           mat_d1(i)=0
           mat_d2(i)=0
          enddo

100       format(i5,7x,f8.3,f8.3,f8.3)
101       format(i5,7x,f8.3,f8.3,f8.3)
          do k=1,10000
           read(1,100) ista,xc,yc,zc
           g1mat(ista,1)=xc
           g1mat(ista,2)=yc
           g1mat(ista,3)=zc

           read(2,101) istb,xc,yc,zc
           V654mat(istb,1)=xc
```

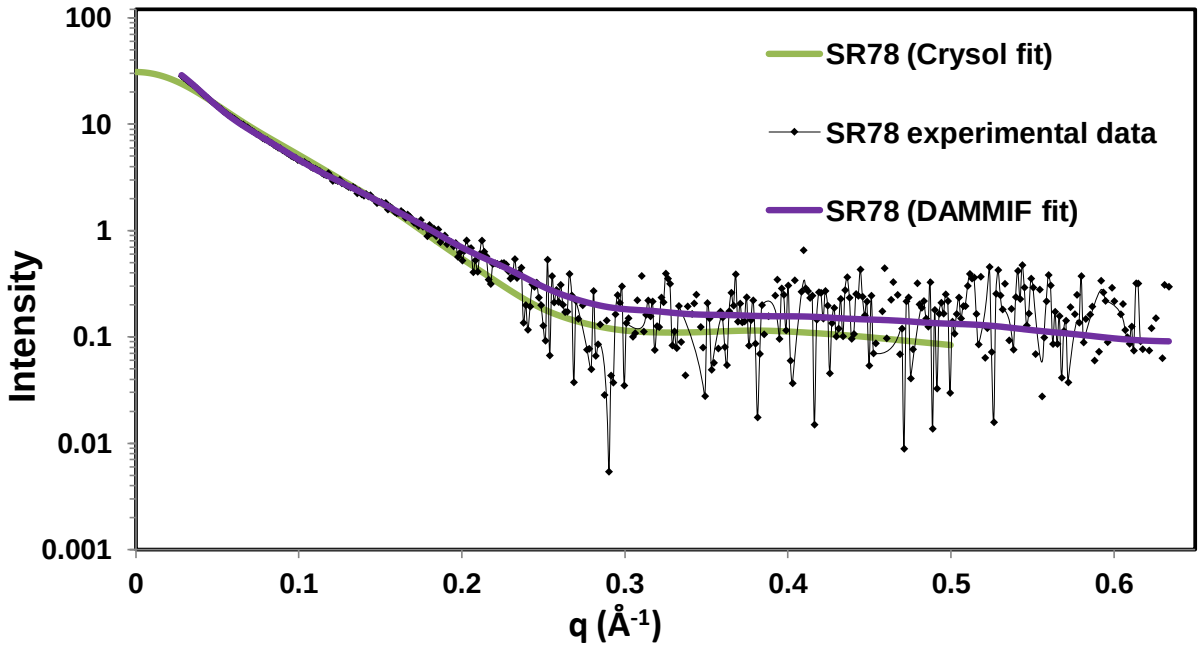
```
           V654mat(istb,2)=yc
           V654mat(istb,3)=zc

C           read(3,100) istc,xc,yc,zc
C           F8767mat(istc,1)=xc
C           F8767mat(istc,2)=yc
C           F8767mat(istc,3)=zc

C           read(4,100) istd,xc,yc,zc
C           Q447mat(istd,1)=xc
C           Q447mat(istd,2)=yc
C           Q447mat(istd,3)=zc

          enddo

          do kk=1,10000
C  Distance between G1 and V654

           xx=g1mat(kk,1)-V654mat(kk,1)
           xx2=xx*xx
           yy=g1mat(kk,2)-V654mat(kk,2)
           yy2=yy*yy
           zz=g1mat(kk,3)-V654mat(kk,3)
           zz2=zz*zz

           d1=sqrt(xx2+yy2+zz2)

C           print *, d1,kk

           d1_vec(kk)=d1

C  Let's put d1 into a histogram

           do j=1,nbins
            xmin=start+width*(j-1)
            xmax=start+width*j
            if(d1.gt.xmin .and. d1.le.xmax) then
               mat_d1(j)=mat_d1(j)+1
            endif
           enddo

C  Distance between G1 and F8767

C           xx=g1mat(kk,1)-F8767mat(kk,1)
C          xx2=xx*xx
C           yy=g1mat(kk,2)-F8767mat(kk,2)
C           yy2=yy*yy
C           zz=g1mat(kk,3)-F8767mat(kk,3)
C           zz2=zz*zz
```

```
C          d2=sqrt(xx2+yy2+zz2)

C          d2_vec(kk)=d2

C  Let's put d2 in a histogram

C          do j=1,nbins
C           xmin=start+width*(j-1)
C           xmax=start+width*j
C           if(d2.gt.xmin .and. d2.le.xmax) then
C              mat_d2(j)=mat_d2(j)+1
C           endif
C          enddo

          enddo

           read(5,*) wed
           read(5,*) xmodel

           do while (.true.)
            read(5,*,end=999) nstr,ifreq
C            print *, nstr,ifreq

            do j=1,nbins
            xmin=start+width*(j-1)
            xmax=start+width*j
            if(d1_vec(nstr).gt.xmin .and. d1_vec(nstr).le.xmax) then
               mat_d1sel(j)=mat_d1sel(j)+ifreq
            endif
           enddo

           do j=1,nbins
            xmin=start+width*(j-1)
            xmax=start+width*j
            if(d2_vec(nstr).gt.xmin .and. d2_vec(nstr).le.xmax) then
               mat_d2sel(j)=mat_d2sel(j)+ifreq
            endif
           enddo

           enddo

102       format(2X,f5.1,5x,4(i4,3x))
999       do j=1,nbins
            print 102, start+width*(j-1)+width/2.0,mat_d1(j),
     &        mat_d1sel(j),mat_d2(j),mat_d2sel(j)
          enddo
```

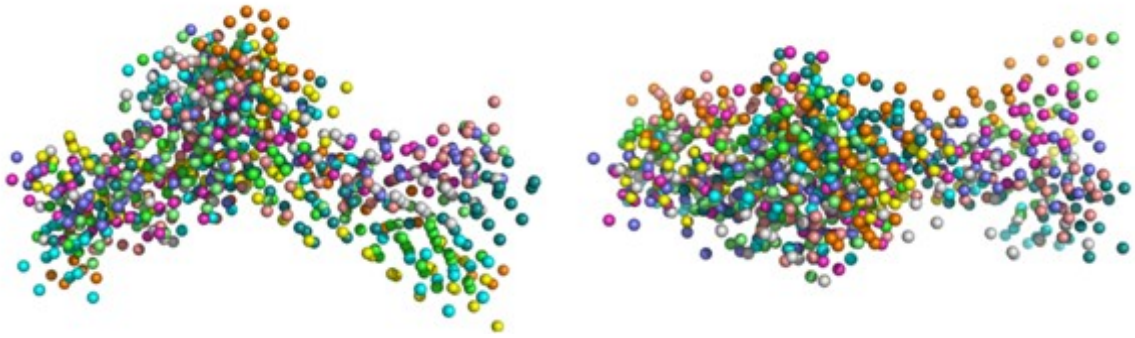
```
stop
end
```

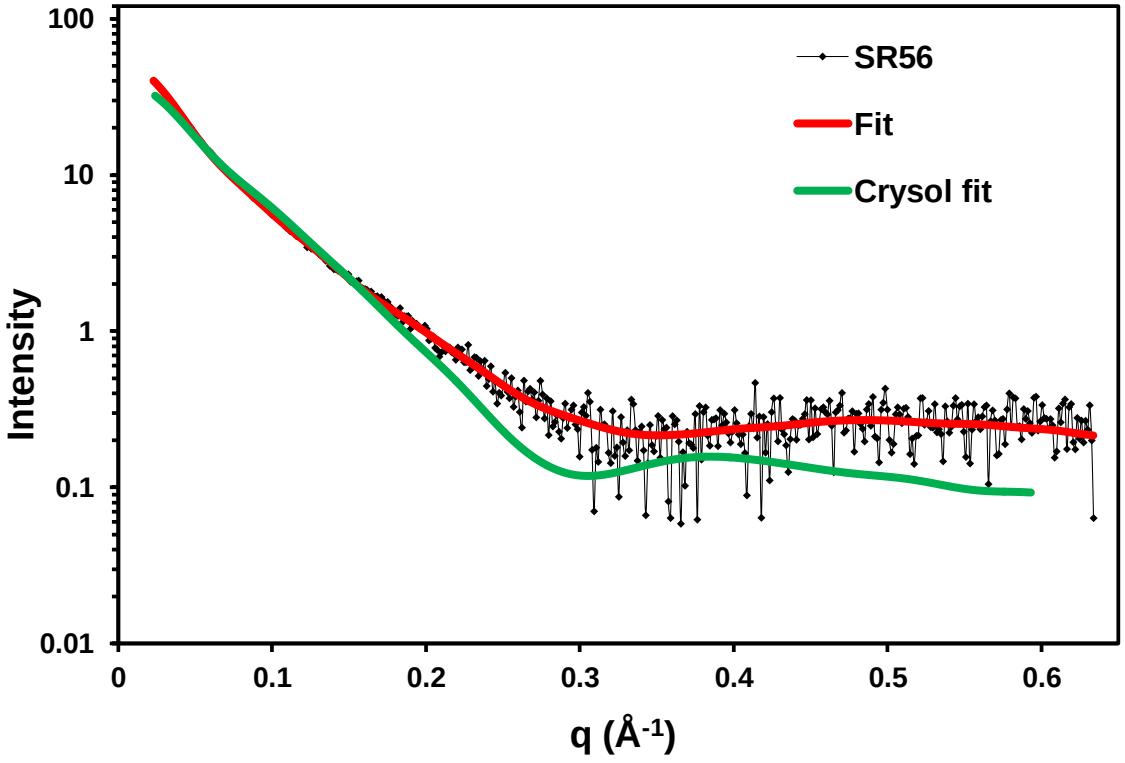
Appendix - D



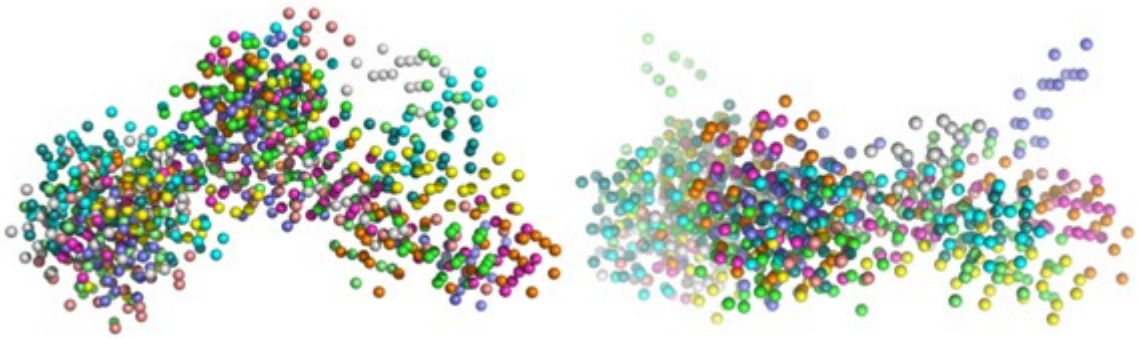

100
10
1
0.1
0.01
0.001
Intensity
0
0.1
0.2
0.3
0.4
0.5
0.6
q (Å-1)
SR78 (Crysol fit)
SR78 experimental data
SR78 (DAMMIF fit)

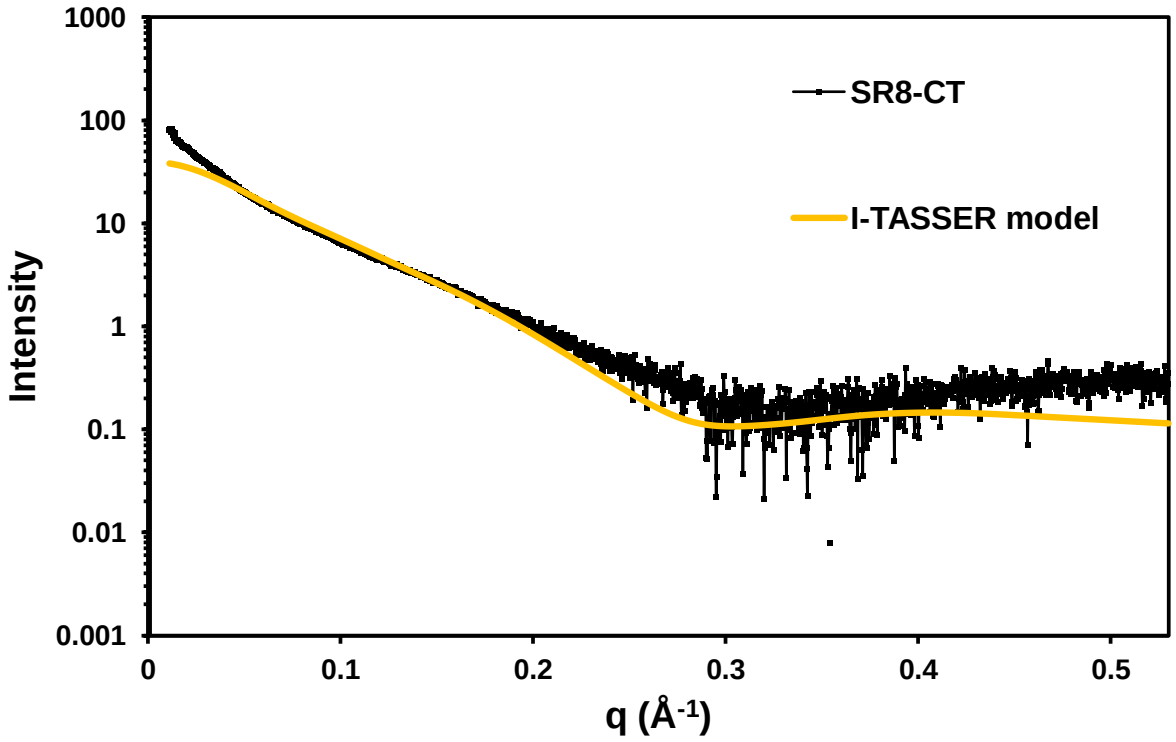
Appendix - E



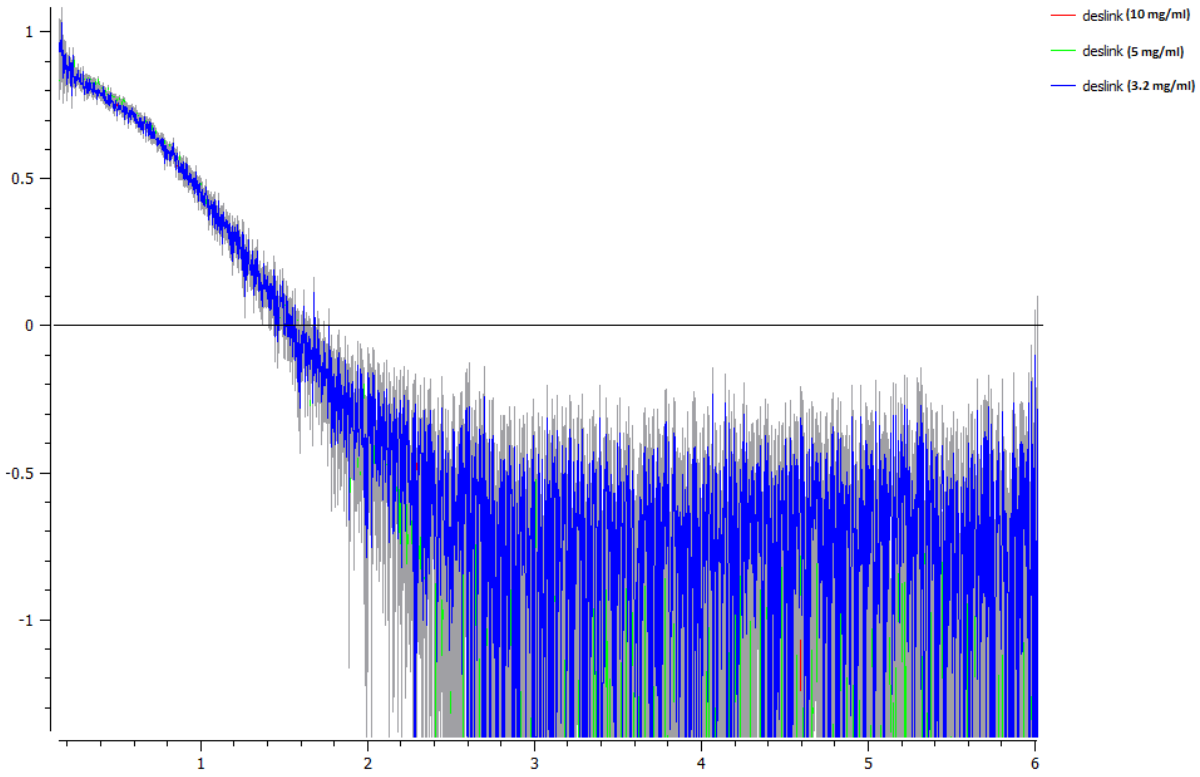

100
10
1
0.1
0.01
Intensity
0
0.1
0.2
0.3
0.4
0.5
0.6
q (Å-1)
SR56
Fit
Crysol fit

Appendix - F

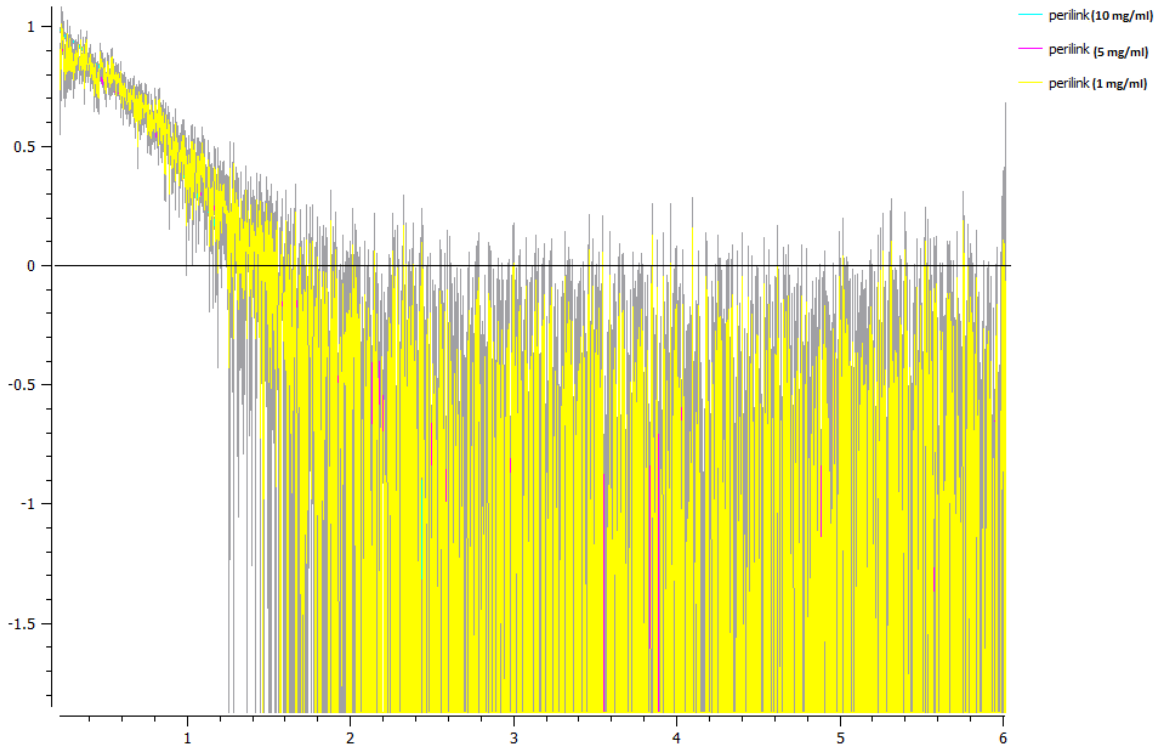



1000
100
10
1
0.1
0.01
0.001
Intensity
0
0.1
0.2
0.3
0.4
0.5
q (Å-1)
SR8-CT
I-TASSER model

Appendix - G


deslink (10 mg/ml)
deslink (5 mg/ml)
deslink (3.2 mg/ml)
1
0.5
0
-0.5
-1
1
2
3
4
5
6

Appendix - H


perilink (10 mg/ml)
perilink (5 mg/ml)
perilink (1 mg/ml)
1
0.5
0
-0.5
-1
-1.5
1
2
3
4
5
6