Structural and Biochemical Characterisation of Utrophin and Dystrophin Spectrin Repeat Domains

by

Muralidharan Muthu

Submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

in

Biochemistry

Massey University,

Palmerston North, NZ

2012

Abstract

Duchenne and Becker muscular dystrophies are muscle-wasting disorders caused by mutations in the X-linked dystrophin gene. Dystrophin is a large cytoskeletal protein belonging to the spectrin superfamily, that links intracellular F-actin to the extracellular matrix *via* a membrane-associated glyco-protein complex thus maintaining structural rigidity and flexibility. Utrophin is a widely expressed protein that has been shown to functionally compensate for dystrophin in cultured muscle cells as well as in the muscular dystrophy mice model. Both utrophin and dystrophin share a similar domain architecture with N-terminal actin-binding domains and C-terminal variable domains separated by 22 or 24 spectrin-like repeats respectively. Therapeutic strategies to replace individuals having defective dystrophin with utrophin require full characterisation of these proteins.

In this thesis, high-resolution structures of the N-terminal first spectrin repeat domains from utrophin and dystrophin have been determined by x-ray crystallography to 1.95 and 2.3 Å. Despite multiple structures of spectrin repeats in the Protein Data Bank from α-actinin, spectrin and plectin these are the first structures determined for any of the spectrin repeats from utrophin and dystrophin. These structures are similar to one another and display a three-helix bundle spectrin repeat fold. The repeat domain structure reveals the relationship between the canonical spectrin repeat domain sequence motif and the structural domain for utrophin and dystrophin, showing the N-terminal first spectrin repeat to be extended at the C-terminal end. Earlier biochemical studies revealed that the extension at the C-terminus was required for the protein's stability. Studies have also shown that spectrin repeats of utrophin are required for a higher affinity interaction of the actin-binding domain with F-actin. However, it was unclear whether the N-terminal repeat domain has an intrinsic affinity for F-actin. In the present study the actin-binding properties of these spectrin repeats are elucidated.

Previous experiments using molecular dynamic simulations and atomic force microscopy of tandem spectrin repeat domains from erythroid spectrin and α-actinin suggested that flexibility of multiple repeats depends upon the linker region. However

under higher extension, the triple helical domain further undergoes unfolding and refolding and thus functions as an elastic element within the cell. Studies using steered molecular dynamics suggested that the force required for unfolding the N-terminal first spectrin repeat domain from utrophin is higher in comparison to that of the N-terminal first spectrin repeat from dystrophin.

Keywords

Dystrophin, Utrophin, F-actin, Spectrin repeats.

Acknowledgements

First and foremost, I would like to thank my supervisor, Dr. Andrew J Sutherland-Smith for his support and patience during my PhD. I feel incredibly lucky and honoured to have been able to work for him. I have learned so much while working in his laboratory. He gave me the opportunity to do amazing research and also supported me for various conferences throughout my PhD and for that, I am truly thankful.

I want to thank Prof Geoff B. Jameson, who taught me an "Advanced Chemistry" course and the course during the NZ Synchrotron workshop, these courses provided me with a much broader understanding of x-ray crystallography and synchrotron radiation. I would like to thank Dr Steven Pascal and his group for their help to understand concepts of NMR spectroscopy and the incredible things that one can do with it. Thanks to Dr Jen Hsin at Stanford University for conducting simulation experiments and scientific advice on molecular dynamics.

I am truly thankful to Prof. Barry Scott and Assoc. Prof. Kathryn Stowell, for their support and advice during my PhD and providing a pleasant research environment in IMBS. I would like to thank Dr Gill Norris and Dr Mark Patchett for letting me use some of their chemicals and instruments when required. I would also like to thank Dr Greg Sawyer for his assistance with the cloning, protein purification and assays. I would like to thank Simon Oakley and Bryan Anderson for their help to understand crystallography software and visualisation tools. I would like to thank past and present members in structural biology for their support and providing me with a great atmosphere especially to Matt, Trevor, Alice, Judith, Jana, Jan and Komala. I would also like to express my gratitude to Massey University for a Doctoral Scholarship. I would like to thank my friends Srinu, Keerti, Sridhar, Pradeep and Harinath for their incredible support for many years.

Lastly but not least, I would like to thank my parents Mr. Muthu and Mrs. Bhuvana, grandparents Mr. Bangaru and Mrs. Kasthuri, and my wife Kavitha for their love and support throughout my life.

Contents

List of Figures

List of Tables

List of Abbreviations

ABA	actin binding assay
ABD	actin binding domain
ABS	actin binding sequence
AFM	atomic force microscopy
Amp	ampicillin
APS	ammonium persulphate
ATP	adenosine triphosphate
AWCGS	Allan Wilson Centre Genome Service
BMD	Becker muscular dystrophy
BSA	bovine serum albumin
CBB	Coomassie brilliant blue
CCP4	Collaborative Computational Project No. 4
CCP4MG	Collaborative Computational Project No. 4 Molecular Graphics
CD	circular dichroism
CH	calponin homology
COOT	Crystallographic Object-Oriented Toolkit
CV-SMD	constant velocity-steered molecular dynamics
DAP	dystrophin-associated protein
DAPC	dystrophin-associated protein complex
DIW	deionised water
dko	double knockout
DMD	Duchenne muscular dystrophy
DSR1	dystrophin N-terminal first spectrin repeat (residues 338 to 456)
DSR12	dystrophin N-terminal first two spectrin repeat (residues 338 to 567)
DTT	dithiothreitol
Dys	dystrophin
EM	electron microscopy
EDTA	ethylenediaminetetaacetic acid
F-actin	filamentous actin

HCl	hydrochloric acid
IDT	Integrated DNA Technologies
IPTG	isopropyl-β-D-thiogalactopyranoside
MCS	multiple cloning site
MD	molecular dynamics
MPD	2-methyl-2, 4-pentanediol
MR	molecular replacement
MTJ	myotendinous junction
NaCl	sodium chloride
NAMD	Not (just) Another Molecular Dynamics program
Ni-NTA	nickel nitrotriacetic acid
NMJ	neuromuscular junction
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PDB	protein data bank
PEG	polyethylene glycol
Pfam	database of Protein families
PMSF	phenylmethanesulfonyl fluoride
pN	pico Newton
rTEV	tobacco etch virus, recombinant
s/n	supernatant
SMART	Simple Modular Architecture Research Tool
SR	spectrin repeat
TCEP	tris [2–carboxyethyl] phosphine
TEMED	N,N,N',N'-tetramethylethylenediamine
Tris	tris(hydroxymethyl)aminomethane
USR1	utrophin N-terminal first spectrin repeat (residues 308 to 425)
USR12	utrophin N-terminal first two spectrin repeats (residues 308 to 537)
Utr	utrophin
Ve	elution volume
Vo	void volume
VMD	visual molecular display

Xp21	short arm(p) of the X-chromosome at position 21
Å	Ångström or Angstrom
λ	wavelength
θ	angle of reflection
a, b, c	axial lengths of a unit cell along x, y and z coordinates
α, β, γ	interaxial angles between b & c, c & a, and a & b respectively

Related Publications

Part of the work presented in this thesis is in the following publication. I am very grateful to all co-authors of these papers.

Muthu M, Richardson KA, Sutherland-Smith AJ, (2012). The crystal structures of dystrophin and utrophin spectrin repeats: Implications for domain boundaries. *PLoS ONE 7(7): e40066. Doi:10.1371/journal.pone.0040066.*

PDB Coordinates: 3UUL, 3UUM, 3UUN

1. Introduction

1.1 Muscular dystrophy

Muscular dystrophy refers to a group of inherited disorders that result in progressive muscle weakness, degeneration of the skeletal muscles that control movement, increased proteolytic activity and abnormal calcium homeostasis [Imbert *et al.,* 1995; Alderton & Steinhart, 2000; Armstrong & Shivell, 2001; Culligan & Ohlendrick, 2002].

There are various types of muscular dystrophies known; in the present research work we focused on the muscular dystrophies associated with the absence or abnormal expression of the dystrophin (Dys) protein, that leads to the development of severe and incurable forms of either Duchenne or Becker muscular dystrophies, both of which are a devastating form of neuromuscular disorder caused by mutations in the *DMD* gene located at Xp21 locus. [Emery, 1993; Ferlini *et al.,* 1999; Biggar *et al.,* 2002].

1.1.1 Duchenne muscular dystrophy (DMD)

Duchenne muscular dystrophy (DMD) is an X-linked recessive neuromuscular disorder caused by mutation in the 2.4 Mb dystrophin gene (*DMD* gene), that encodes the formation of dystrophin protein [Monaco *et al.,* 1986; Koenig *et al.,* 1987; Blankinship *et al.,* 2006], an important component of muscle tissue. Dys is localised at the cytoplasmic surface of the sarcolemma of the skeletal muscle plasma membrane [Zubrzycka-Gaarn *et al.,* 1988; Blake *et al.,* 1996; Pichavant *et al.,* 2011]. DMD is characterised by decreasing muscle mass and a progressive loss of muscle function affecting 1 in every 3,500 male children [Emery*,* 1991; Emery*,* 1993]. Patients affected with DMD are usually bound to a wheelchair in their early teens and die when they reach their early twenties due to respiratory and cardiac failure. Females can be carriers but generally do not experience the symptoms of this condition [Tinsley *et al.,* 1994].

1.1.2 Becker muscular dystrophy (BMD)

Becker muscular dystrophy is frequently described as a more benign X-linked muscular disorder, because of its late onset and slow progressive course compared to DMD [Koenig *et al.,* 1987]. Earlier studies have stated that 85% of BMD patients have dystrophin of abnormal molecular weight and 15% have a reduced amount of normal size protein resulting in a longer life expectancy compared to DMD [Beggs *et al.,* 1991; Pozzoli *et al.,* 2002].

1.2 Dystrophin and Utrophin Expression and Regulation

1.2.1 Dystrophin

Dystrophin is a large multidomain cytoskeletal protein of 3,685 amino acids with a molecular size of 427 kDa, which is required for maintaining muscle cell membrane structural integrity during muscle contraction [Menke *&* Jockusch, 1991; Stedman *et al.,* 1991].

Dys can be divided into four distinct domains; its N-terminal domains, also known as CH-domains (Calponin Homology), bind to the F-actin of the subsarcolemmal cytoskeleton and its C-terminal domains to transmembrane proteins, such as syntrophin, dystrobrevin and dystroglycans. Together they form the dystrophin associated protein complex (DAG or DAPC) as shown in Figure 1-1 [Blake *et al.,* 1995; Blake *et al.,* 2002; Ervasti, 2007]. Numerous studies have also demonstrated that the loss or deletion of Dys is accompanied by the reduction of DAG complex formation [Ervasti*,* 1990; Matsumura *et al.,* 1992; Campbell*,* 1995; Worton, 1995; Cohn & Campbell, 2000; Rando, 2001].

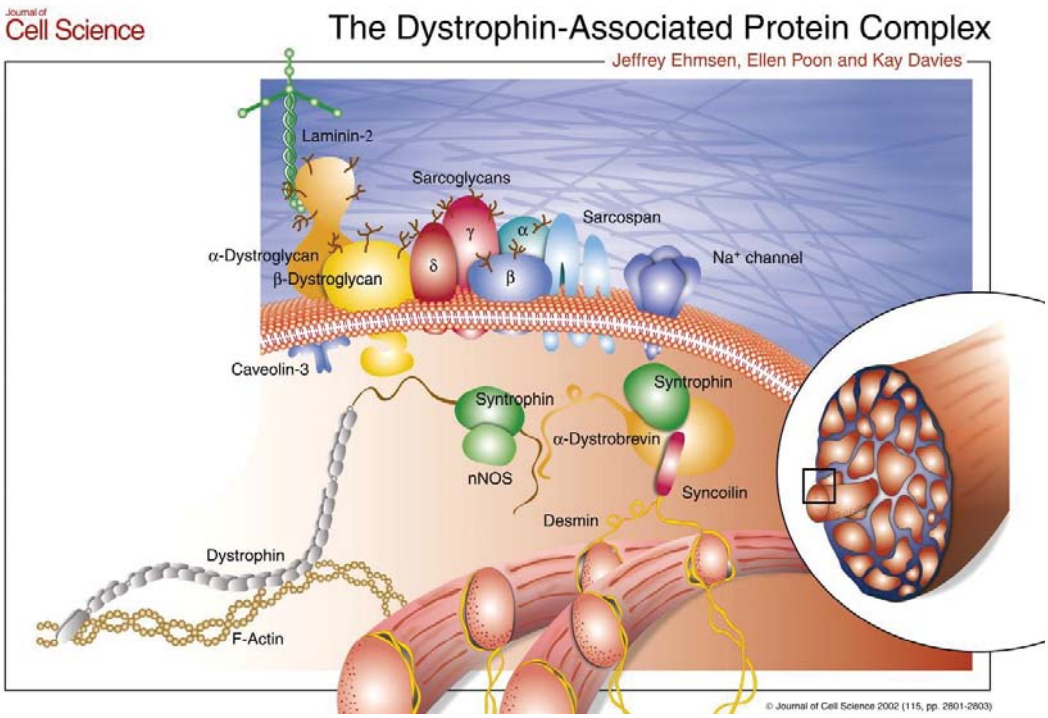


Figure 1-1. Schematic representation of dystrophin *(in gray)* cross-linking F-actin cytoskeleton (N-terminus) and extracellular complex (C-terminus).

Ehmsen *et al.*, 2002

Studies have shown that the *mdx* (dystrophin deficient) mouse that has a nonsense mutation in exon 23 of the Dys gene [Sicinski *et al.,* 1989] and lacks functional Dys protein is able to survive and live an almost normal life span in contrast to human DMD patients [Mattson, 2001]. The survival was due to the presence of Utr (dystrophin homologue) in *mdx* mouse skeletal and cardiac muscle that compensates for the lack of Dys in the assembly of the DAG complex. It was also reported that the over expression of Utr in transgenic *mdx* mice was found to inhibit muscular dystrophy [Tinsley *et al.,* 1998]. Furthermore, double knockout (dko) mice deficient of both Dys and Utr producing genes exhibit symptoms similar to that seen in DMD patients [Deconinck *et al.,* 1997a].

1.2.2 Utrophin

Utr is located on chromosome 6q24 and its translation generates a large multidomain cytoskeletal protein with a molecular size of 395 kDa [Love *et al.,* 1989; Bewick *et al.,* 1992; *Law*, 1994; Pons *et al.,* 1994]. The Utr protein shares a considerable sequence homology [Tinsley *et al.,* 1996], and similar domain architecture, to Dys [Broderick *&* Winder, 2002], indicating that the Utr gene has evolved by considerable duplication of the Dys gene [Pearce *et al.,* 1993].

Utr is expressed ubiquitously in humans, being found in brain, kidney, liver, spleen, lungs, stomach, neurons and vascular tissues [Man *et al.,* 1992a; Blake *&* Tinsley, 1994]. Utr is found along the inner plasma membrane of developing fetal muscle [Clerk *et al.,* 1993; Man *et al.,* 1992b; Gramolini *&* Jasmin, 1999; Weir *et al.,* 2004], but is restricted to the myotendinous and neuromuscular junctions (MTJ and NMJ) in normal adult skeletal muscle fibers [Bewick *et al.,* 1992; Law, 1994; Pons *et al.,* 1991]. However, high levels of Utr are present around the inner plasma membrane of regenerating myofibers in *mdx* mouse muscle [Deconinck *et al.,* 1997b; Matsumura *et al.,* 1993] and in DMD patients [Mizumo *et al.,* 1993].

Earlier studies have shown that *mdx* mice have the distribution pattern of Utr in neurons of the cerebral cortex, in several sensor and motor brainstem nuclei as well as in blood vessels. However, mice lacking full-length Dys have shown no change in the expression pattern of Utr [Deconinck *et al.,* 1997b]. Moreover, studies have also shown that there were no mutations found in the *Utr* gene in human patients suffering with DMD. If upregulation of the *Utr* gene is enhanced, the production of Utr protein increases and symptoms disappear. Eventually this may lead to a potential therapeutic replacement [Tinsley *et al.,* 1996; Blake *et al.,* 2002; Rybakova *et al.,* 2002; Miura *&* Jasmin, 2006].

1.3 Spectrin repeat family

Spectrin was the first protein identified in the erythrocyte membrane cytoskeleton [Speicher & Marchesi, 1984]. Based on protein sequence analysis, each spectrin repeat consists of ~106 amino acids and is predicted to form a characteristic triple-helical coiled-coil bundle [Parry *et al.,* 1992].

DNA sequencing of erythroid spectrin was completed first [Sahr *et al.,* 1990] followed by the DNA sequences of α-actinin [Baron *et al.,* 1987] and Dys [Koenig *et al.,* 1988]. The entire DNA sequences revealed the similarities between the repeating regions of spectrin [Davison *et al.,* 1988] to the repeating regions of α-actinin and dystrophin and hence the spectrin family of proteins was born. Gradually Utr and plakin proteins were added into the family.

1.3.1 Members of spectrin family

The discrete group of cytoskeletal proteins that comprises spectrin, α-actinin, Dys and Utr, function to maintain cell stability by connecting cytosolic F-actin to the extracellular matrix. All together this constitutes the spectrin family of proteins [Dubreuil, 1991]. In this thesis, the focus was on Dys and Utr proteins from the spectrin superfamily.

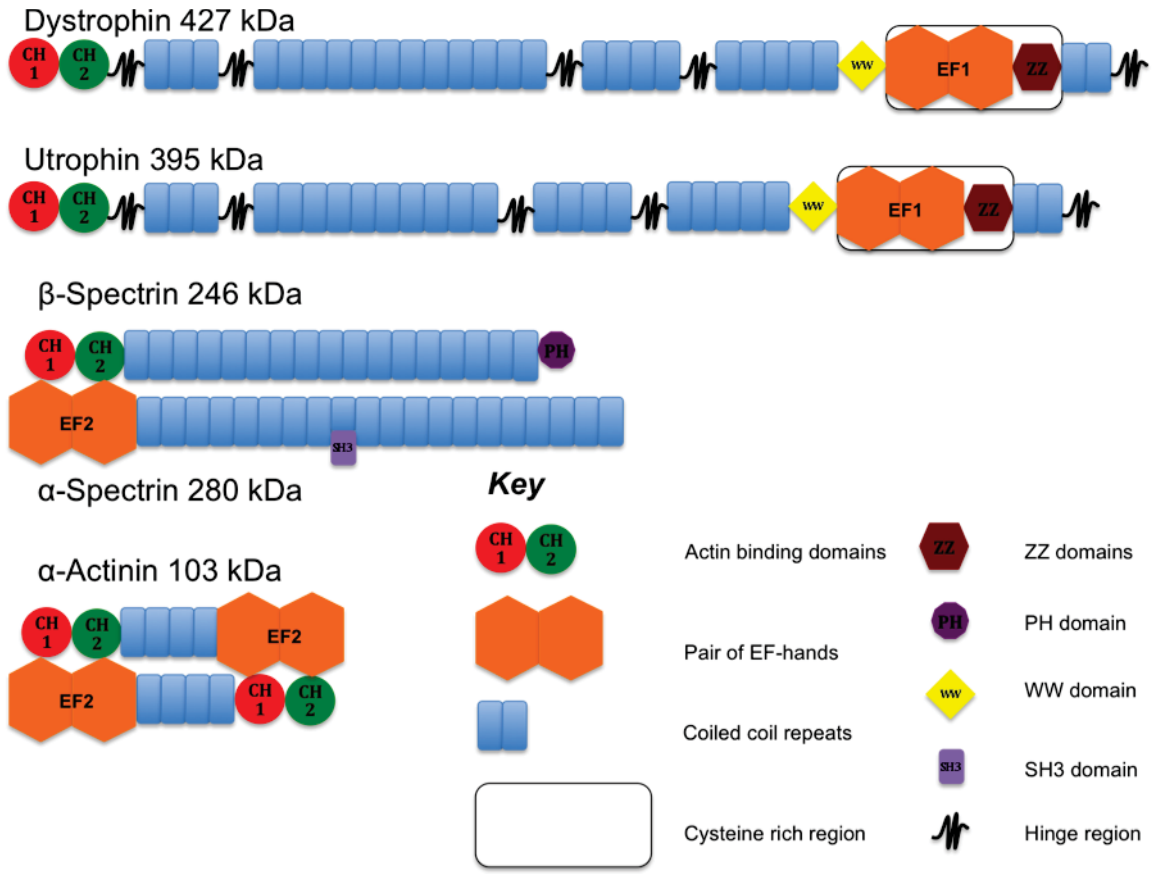


Figure 1-2. Schematic representation of spectrin family of proteins.

Broderick & Winder, 2002

Dys and Utr are large cytoskeletal proteins. Utr shares 69% sequence identity [Galkin *et al.,* 2002] and equal domain identity [Broderick & Winder, 2002] to Dys. All contain an N-terminal acting binding domains (ABD) consisting of two Calponin homology domains that bind F-actin [Way *et al.,* 1992] and keratin [Stone *et al.,* 2005], followed by the central rod region consisting of multiple spectrin-like repeats and the C-terminal region containing a WW-domain, EF hands, a cysteine-rich ZZ domain, and a coiled-coiled motif [Koenig & Kunkel, 1990].

1.3.1.1 Actin-binding domain (ABD)

The actin-binding domain of Dys and Utr consists of approximately 240 residues that comprise two functionally distinct but structurally equivalent calponin homology (CH) domains, CH1 and CH2 [Castresana & Saraste, 1995; Gimona & Mital, 1998] based on the sequence similarities to the muscle regulatory protein calponin [Winder & Walsh, 1990]. Unlike calponin, which has only one CH-domain, spectrin family proteins require

two such domains to provide an interaction with F-actin of µM affinity [Gimona *et al.,* 2002].

1.3.1.1.1 Structure

Due to the flexibility and larger size, it is difficult to obtain either the crystal or NMR structure of full-length Dys and Utr [Rybakova *et al.,* 2002]. However the atomic models of small lengths and/or fragments of Dys and Utr have been determined.

The atomic structure of the actin-binding domains of Utr was solved by x-ray crystallography [Keep *et al.,* 1999] and followed soon after by the Dys actin-binding domains [Norwood *et al.,* 2000] shown in Figure 1-3.

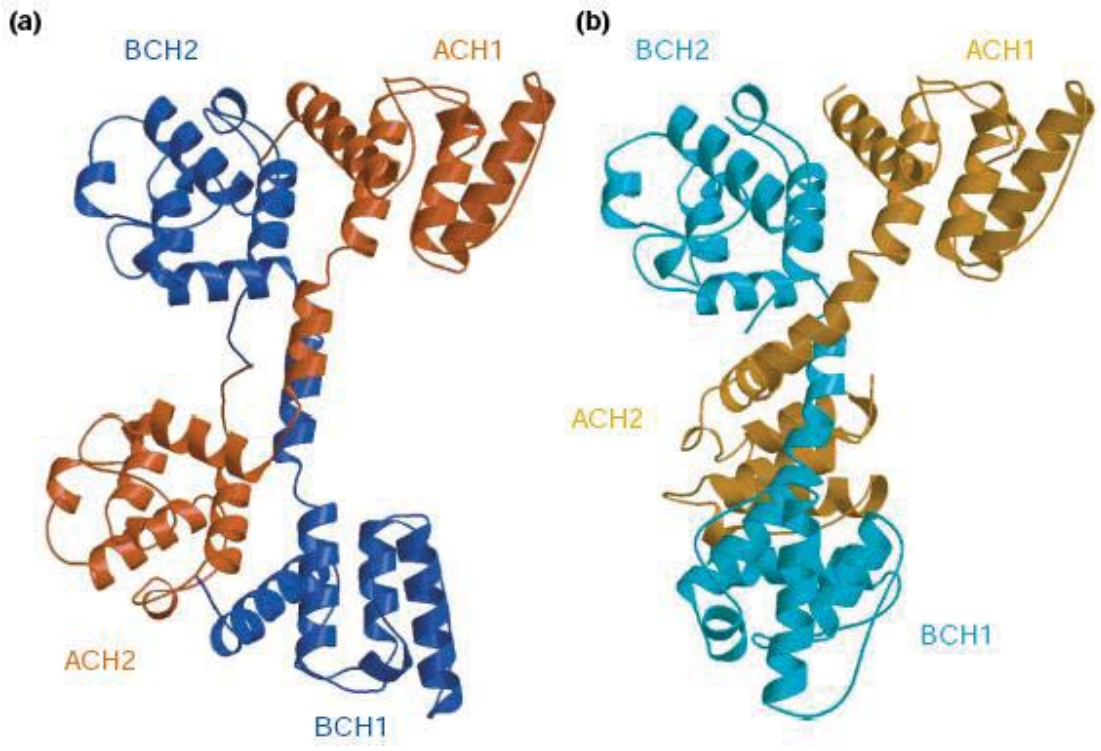


Figure 1-3. Ribbon representation of actin-binding domains of, **a)** dystrophin and **b)** utrophin.

Norwood *et al.,* 2000

The crystal structures of individual CH1 and CH2-domains of Dys and Utr ABDs comprised four α-helices, three of these parallel to each other and the fourth is approximately perpendicular to other three [Banuelos *&* Saraste*,* 1998, Keep *et al.,* 1999].

The CH-domains of both Dys and Utr form an antiparallel dimer with the CH1 domain of one monomer forming a close association with the C-terminal CH2 domain of the other, adopting a dumbbell conformation separated by a long helix [Keep *et al.,* 1999; Norwood *et al.,* 2000]. The relative orientation of CH1 and CH2 within the crystal structure is different for both Dys and Utr as shown in Figure 1-3.

An electron microcrospy (EM) reconstruction of the Utr CH-domains and first spectrin repeat (Utr416) protein bound to F-actin showed that both CH1 and CH2 domains of Utr bound to the F-actin. The elongated EM density was bound laterally to F-actin at the end of the C-terminus of the actin-binding domain, which clearly suggests that the first spectrin repeat of Utr may bind to F-actin laterally along with the CH-domains [Sutherland-Smith *et al.,* 2003] (Figure 1-4).

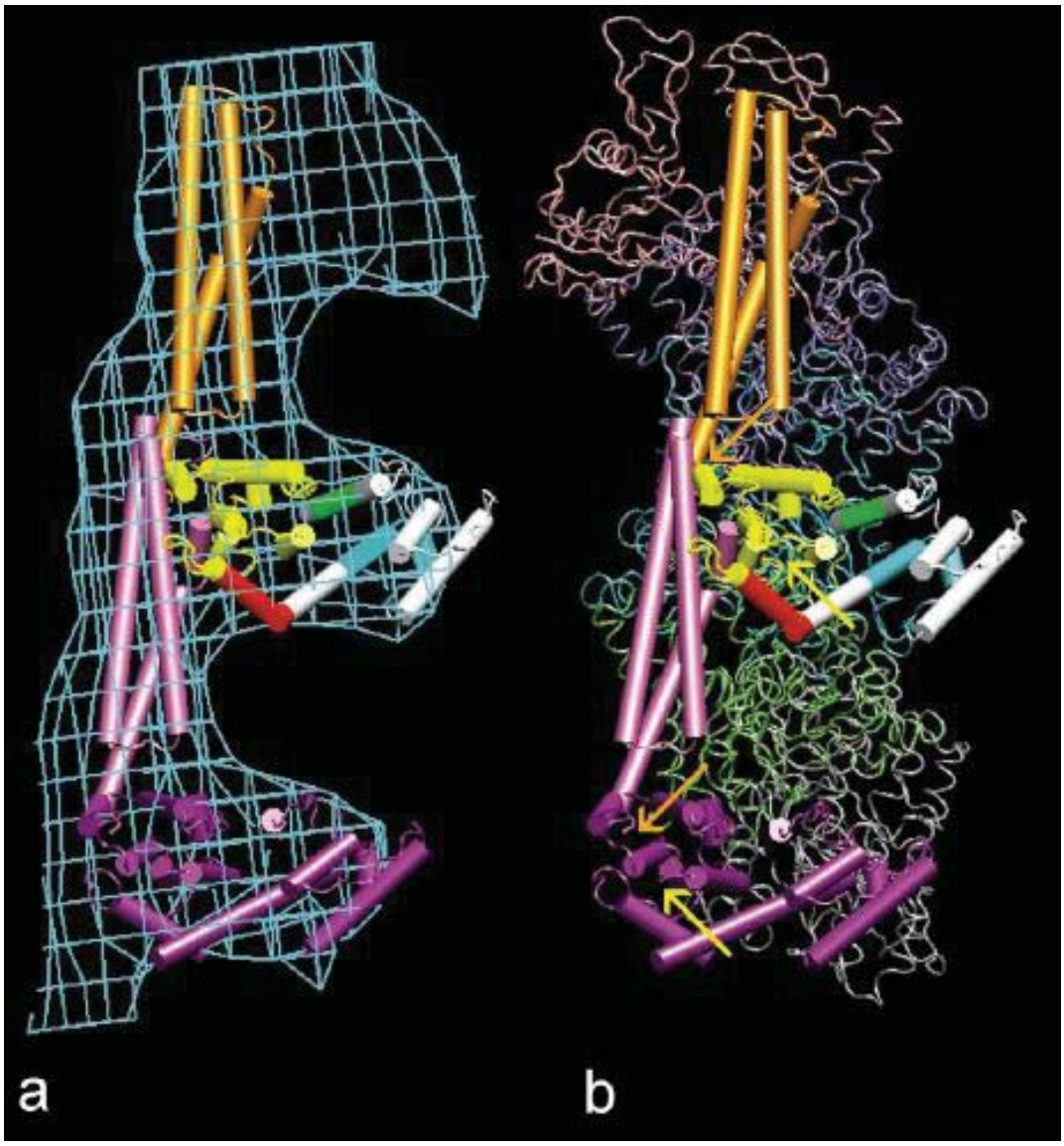


Figure 1-4. Electron microscopy reconstruction of Utr416 (Utr actin-binding domain with first spectrin repeat) bound to F-actin.

Sutherland-Smith *et al.,* 2003

1.3.1.1.2 Function

The calponin homology domain has been identified in many proteins having various functions. However, its presence usually signifies an interaction with F-actin [Winder, 1997, Morres *et al.,* 2000, Galkin *et al.,* 2002; Clark *et al.,* 2009; Sawyer *et al.,* 2009]. Studies have also shown that the Utr actin-binding domain alone has low affinity with F-actin (K_d of 98.8±40 μM) when compared to the full-length Utr [Moores *et al.,* 2000]. Full-length Dys binds actin at a K_d of 0.5μM affinity, whereas full-length Utr has an affinity of 0.15μM and ABDs of Utr shows higher affinity with actin from both muscle

and non-muscle cells when compared to the Dys ABD. Neither the Utr nor the Dys ABDs have any affinity for monomeric G-actin [Way *et al.,* 1992; Winder *et al.,* 1995].

The actin-binding domain of spectrin family proteins has three actin-binding sequences (ABS) represented by ABS1, ABS2 and ABS3 as shown in Figure 1-5. ABS1 and ABS3 have been localized in the first α-helices of CH1 and CH2. ABS2 has been localized in the CH1 domain helices 5 and 6. Initially ABS1 and ABS3 were identified using synthetic peptides derived from dystrophin [Levine *et al.,* 1990]. ABS2 was later identified in α-actinin using *in vitro* actin-binding studies with glutathione S-transferase (GST) fusion proteins [Kuhlman *et al.,* 1992].

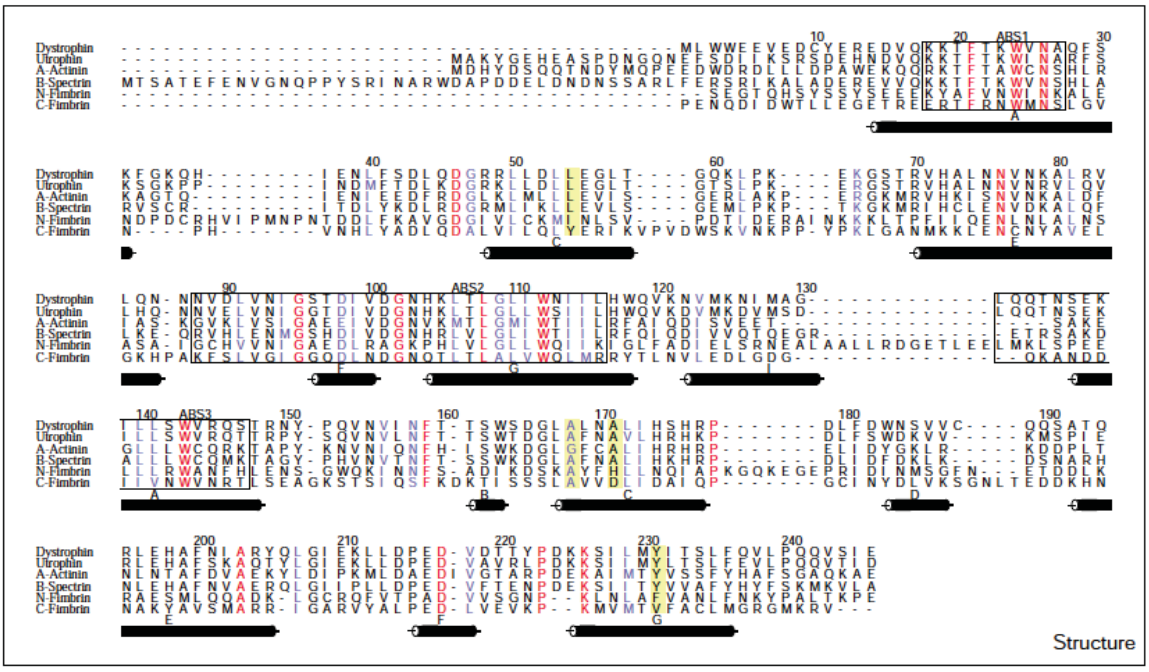


Figure 1-5. Sequence alignment of ABDs from the spectrin superfamily. The actin-binding sequences (ABS) are indicated by boxed regions. The positions (residues 54, 168, 171 and 231) where pathogenic dystrophin mutations occur are shaded yellow.

Norwood *et al.*, 2000

1.3.1.1.3 Mechanism of actin-binding by CH-domains

Both CH domains contribute to actin binding. Although CH1 of the Utr actin-binding domain binds to F-actin without CH2 present, its affinity is 10-fold lower than the complete actin-binding region containing both CH domains [Gimona & Mital, 1998]. Even though CH2 has a low intrinsic actin-binding activity by itself, it helps CH1 to bind F-actin more effectively, than the CH1-domain alone [Way *et al.,* 1992, Winder *et al.,* 1995].

Utr ABDs exists as a monomer in solution but when it was crystallised, it showed two extended molecules forming a dimer in the asymmetric unit [Keep *et al.,* 1999]. However, there exists an equilibrium between the dimeric and monomeric forms for Dys [Norwood *et al.,* 2000].

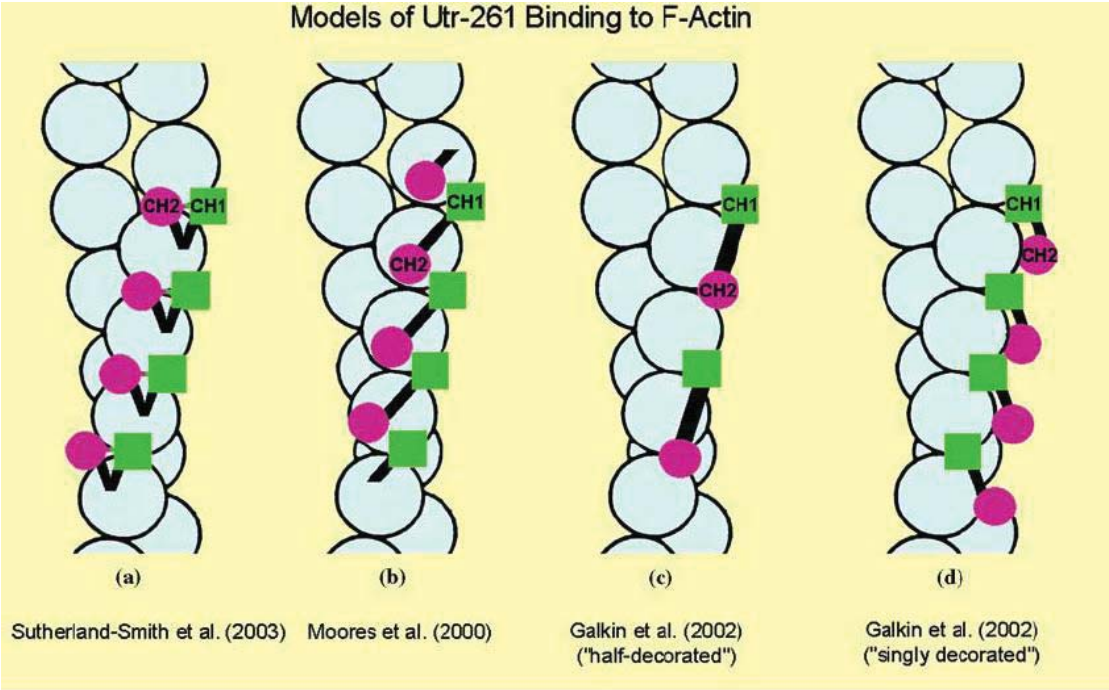


Figure 1-6. Schematic representations of the modes of interaction of Utr ABD on F-actin. Utr ABD is coloured green (CH1) and purple (CH2). Actin is light blue.

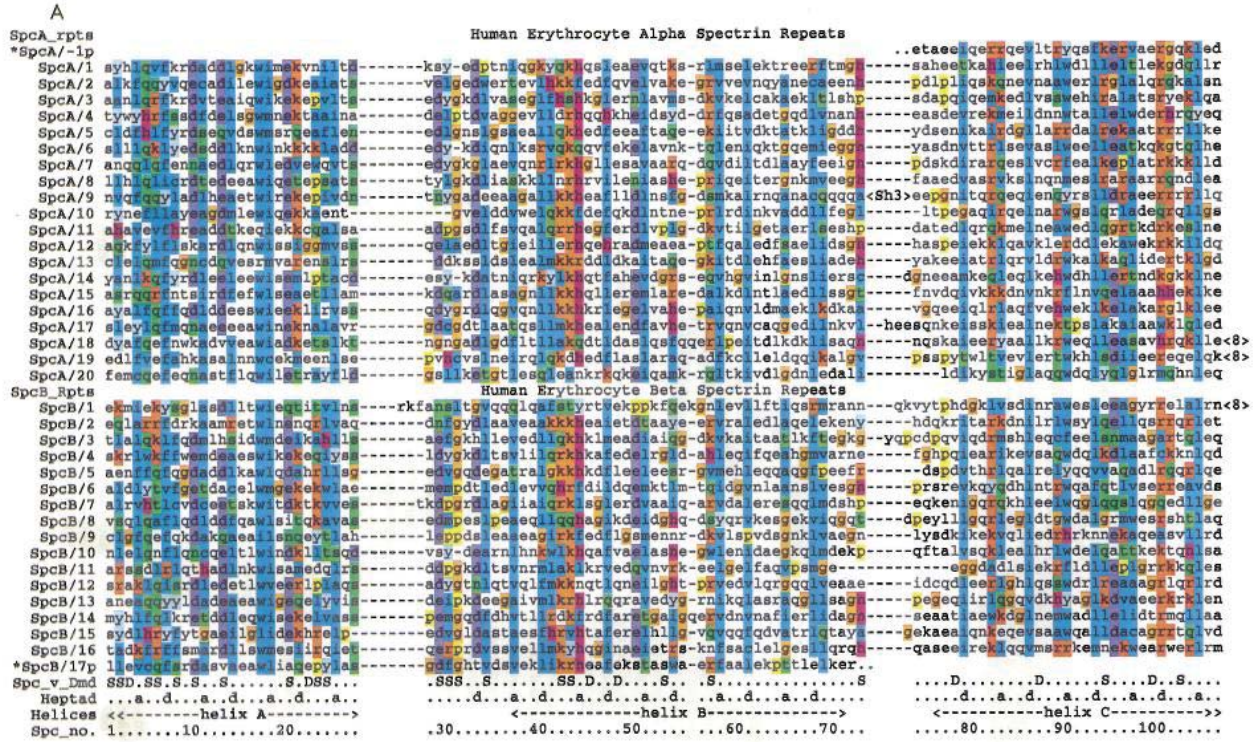
Sutherland-Smith *et al.,* 2003; Lehman *et al.,* 2004

Various F-actin binding mechanisms have been proposed, showing different modes of binding by the CH-domains to F-actin (Figure 1-6). A closed compact conformation of CH-domains bound with F-actin was proposed by Sutherland-Smith *et al.,* (2003), and open extended conformations were proposed by Moores *et al.,* (2000) and Galkin *et al.,* (2002). The crystal structures of the ABDs of Dys and Utr both exhibit 'open' or 'extended' confirmations [Keep *et al.,* 1999; Norwood *et al.,* 2000].

1.3.1.2 Spectrin Repeat Domain

The central rod region of Dys and Utr consists of weakly repeating domain units of ~109 amino acid residues. Sequence analysis of Dys revealed 24 repeat domains separated by 4 proline-rich hinge regions [Koenig & Kunkel, 1990; Cross *et al.,* 1990; Parry *et al.,* 1992]. Similarly, Utr is thought to contain 22 repeats with 2-hinge regions. Furthermore the arrangement of the nested repeats and hinge regions, suggests the elastic nature of the proteins, and thus Dys and Utr are hypothesised to act as molecular shock absorbers within cells [Broderick & Winder, 2002; Djinovic-Carugo *et al.,* 2002].

Sequence alignment of the Dys and Utr spectrin repeats shows less similarities at the conserved residues when compared with α- and β- spectrin (Figure 1-7A and 1-7B). The lengths of the three helices ('A', 'B' & 'C') are more consistent in the spectrins with fewer insertions and deletions when compared to the repeats observed in Dys and Utr. Furthermore, the central 'B' helices in Dys repeat 10 and 14, and Utr repeat 14 are less regular and are truncated. The large insertions between the helix 'B' and 'C', in repeat 23 of Dys and repeat 21 of Utr were predicted by sequence alignment [Winder *et al.,* 1995]. Insertions within helix 'A' in repeats 3, 15 and 19 of Dys and 3, 13 and 17 of Utr were observed and are unfavourable for anti-parallel dimer formation unlike α-actin and spectrin.


A
Human Erythrocyte Alpha Spectrin Repeats
Human Erythrocyte Beta Spectrin Repeats
SpcA_rpts
SpcB_Rpts
Spc_v_Dmd
Heptad
Helices
Spc_no.

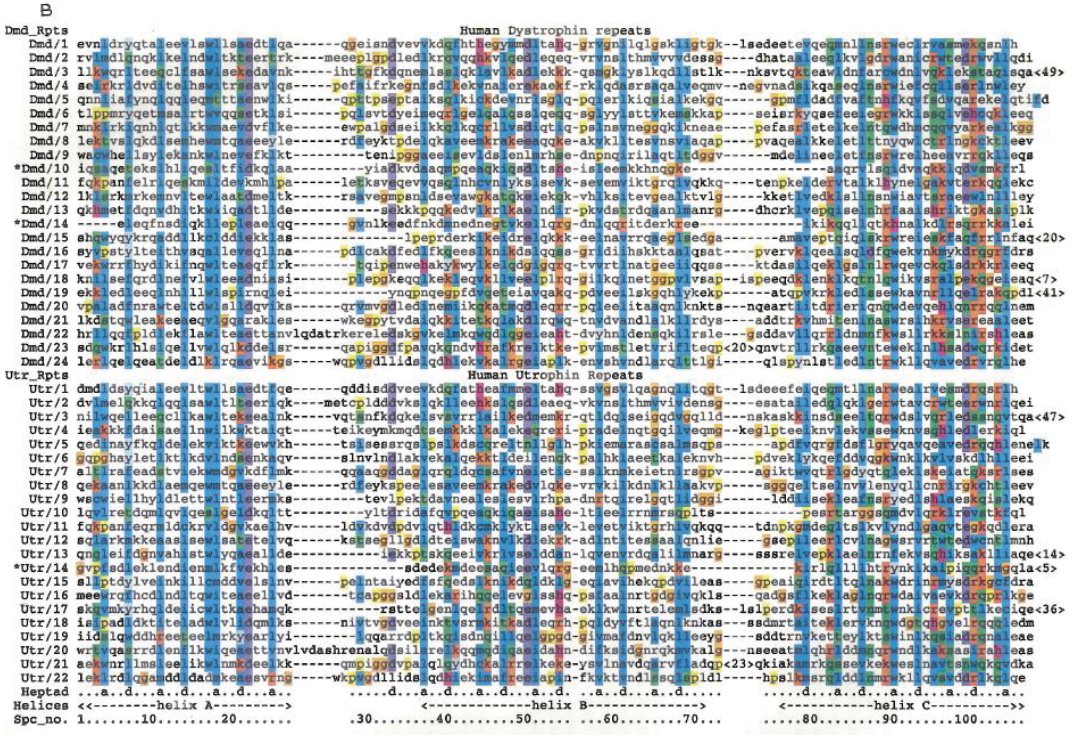


Figure 1-7. Sequence alignment of triple helical spectrin repeats with the heptad sequence repeat labelled from 'a' to 'g' of, A) α- (SpcA) and β- (SpcB) Spectrin, B) dystrophin (Dmd) and utrophin (Utr).

Winder *et al.*, 1995

1.3.1.2.1 Structure

The repeating nature of the spectrin repeat coiled-coil was first observed in erythroid spectrin, based on evidence of partial protein sequence data [Speicher & Marchesi, 1984]. The number of repeats typically ranges from 4 (for α-actinin) to 24 (for dystrophin) consecutive spectrin repeats in the spectrin family of proteins. The tandem arrangements of domain pairs constitutes to an elongated shape for these proteins [Speicher & Marchesi, 1984; Parry *et al.,* 1992]. The internal sequence identity between each repeat is as low as ~20-30% for spectrin [Speicher & Marchesi, 1984] and ~10-25 %) for Dys [Koenig & Kunkel, 1990], however it was thought the repeats retained a similar three-dimensional structure [Parry *et al.,* 1992]. Each spectrin repeat comprised three α-helices, predicted to be wrapped around each other to form a coiled-coil structure containing approximately 106 amino acids [Speicher & Marchesi, 1984]. Further analysis of the protein sequences revealed a heptad pattern with sequence alignment showing the periodicity of conserved hydrophobic and/or charged residues [Koenig *et al.,* 1988; Parry *et al.,* 1992], which suggested that the folding of the three helices of a single repeat into an antiparallel coiled-coil structure. Furthermore, the repeat length varies between spectrin (106 amino acids), α-actinin (122) and Dys (109), however the main difference lies in the number of residues in the loop regions. The atomic structure of a single spectrin repeat has been solved, e.g. those of drosophila α-spectrin repeat 14 by x-ray crystallography [Yan *et al.,* 1993] and repeat 16 of chicken brain α-spectrin by NMR [Pascual *et al.,* 1997]. Both single repeat structures formed a three helix anti-parallel left-handed coiled-coil, consistent with the predicted model [Parry *et al.,* 1992]. Each three helix coiled-coil or spectrin repeat domain is also referred to as three-helix bundle [Lovejoy *et al.,* 1993]. The three-helix bundle structure was mediated by the presence of hydrophobic interactions at 'a' and 'd' heptad repeat sequence positions, and ionic interaction between the 'e' and 'g' positions of the heptad, a repeating pattern of seven amino acids from 'a' to 'g' $(\text{-a-b-c-d-e-f-g-})_n$ [Koenig *et al.,* 1988; Parry *et al.,* 1992].

Sequence analysis suggested that a short region termed the linker, joins the repeat domains. This linker region has been predicted to be either disordered and flexible [Speicher & Marchesi, 1984; Bloch & Pumplin, 1992] or ordered and inflexible [Lux & Palek, 1995; Nicolas *et al.,* 1998]. Furthermore, CD analysis has shown that a helical linker exists between two repeats [Dubreuil *et al.,* 1989; Cross *et al.,* 1990; Parry *et al.,* 1992].

An initial structure of single repeats from x-ray crystallography and NMR [Yan *et al.,* 1993; Pascual *et al.,* 1997] has not revealed any atomic details of the linker region between the two repeat units. It was not until multiple repeats of α-actinin and spectrin were solved by x-ray crystallography that the structure of the linker region was revealed [Djinovic-Carugo *et al.,* 1999; Grum *et al.,* 1999; Ylanne *et al.,* 2001] and found to be α-helical. The linker region consists of 5 amino acids in the long CA' helix (Figure 1-8). Studies have shown that this linker region is necessary to maintain the stability of the repeats by making contacts with the loop AB of repeat 16 and loop B'C' of repeat 17 [Grum *et al.,* 1999].

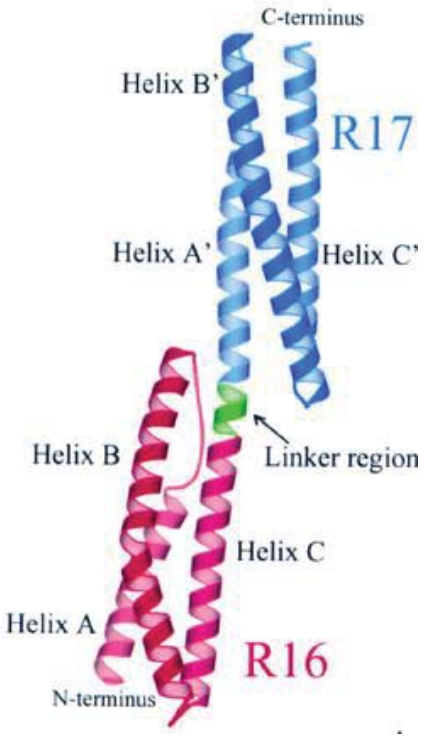


Figure 1-8. Crystal structure of two repeat domains from chicken brain α-spectrin. Repeat 16 (R16) is in red, repeat 17 (R17) is in blue and the linker region is in green.

Grum *et al.,* 1999

1.3.1.2.2 Stability and flexibility

Previous studies using CD spectroscopy show that Dys spectrin repeat domains two and three on their own have reduced molar residue ellipticity values of -14400 and -13700 ($deg^2.cm^2.dmol^{-1}$) observed at 222nm compared to a double repeat domain polypeptide containing both repeat 2 and repeat 3 of -23800 ($deg^2.cm^2.dmol^{-1}$) [Kahana & Marsh, 1994; Calvert *et al.,* 1996]. CD melting temperatures studies of multiple spectrin repeat domains from dystrophin (repeat 14+15, repeat 16+17 and repeat 14 to 17) showied *Tm* values of >55 °C [Menhart, 2006]. It was also shown that the minimum stable folding length of an individual spectrin repeat domain from Dys is 117 amino acids long, a conclusion suggested from proteolysis, tryptophan fluorescence spectra, CD melting temperature and urea denaturation spectra [Kahana & Gratzer, 1995]. Furthermore, individual spectrin domains (repeat 1, repeat 2 and repeat 3) from erythroid spectrin have shown lower melting temperatures of 40, 50 and 48 °C, compared to the multiple repeat domain proteins (repeat 1+2, repeat 2+3, repeat 1 to 3) of 50, 56 and 54 °C [Menhart *et al.,* 1996]. Additionally, it has also been shown that a multiple spectrin repeat construct is thermodynamically more stable than a single repeat [MacDonald & Pozharski, 2001]. Although, structural data of spectrin repeats for spectrin and α-actinin are available [Djinovic-Carugo *et al.,* 1999; Grum *et al.,* 1999; Ylanne *et al.,* 2001], due to the low sequence identity in the rod domains from Utr and Dys the domain boundaries of spectrin repeats were still unclear.

There are approximately 500 homologous spectrin repeat motifs found in the human genome [Altmann *et al.,* 2002] with the consensus that the rod domain is found in proteins that constantly undergo mechanical stress. Spectrin repeat proteins potentially act as molecular shock-absorbers by withstanding external force and limiting damage to the plasma membrane, with the spectrin repeat domains acting as spring units by undergoing force-induced unfolding [Bhasin *et al.,* 2005]. Initial studies on force induced unfolding of spectrin repeats using atomic force microscopy suggested that the α-helical spectrin repeat can be unfolded one domain at a time [Rief *et al.,* 1999; Lenne *et al.,* 2000; Altmann *et al.* 2002] as shown for spectrin repeats from spectrin and α-actinin (Figure 1-9). It was also suggested that multiple repeat

domains do not unfold one at a time. Rather, they unfold in a cooperative manner [Law *et al.,* 2003]. However further investigation in force induced unfolding of spectrin repeat is required.

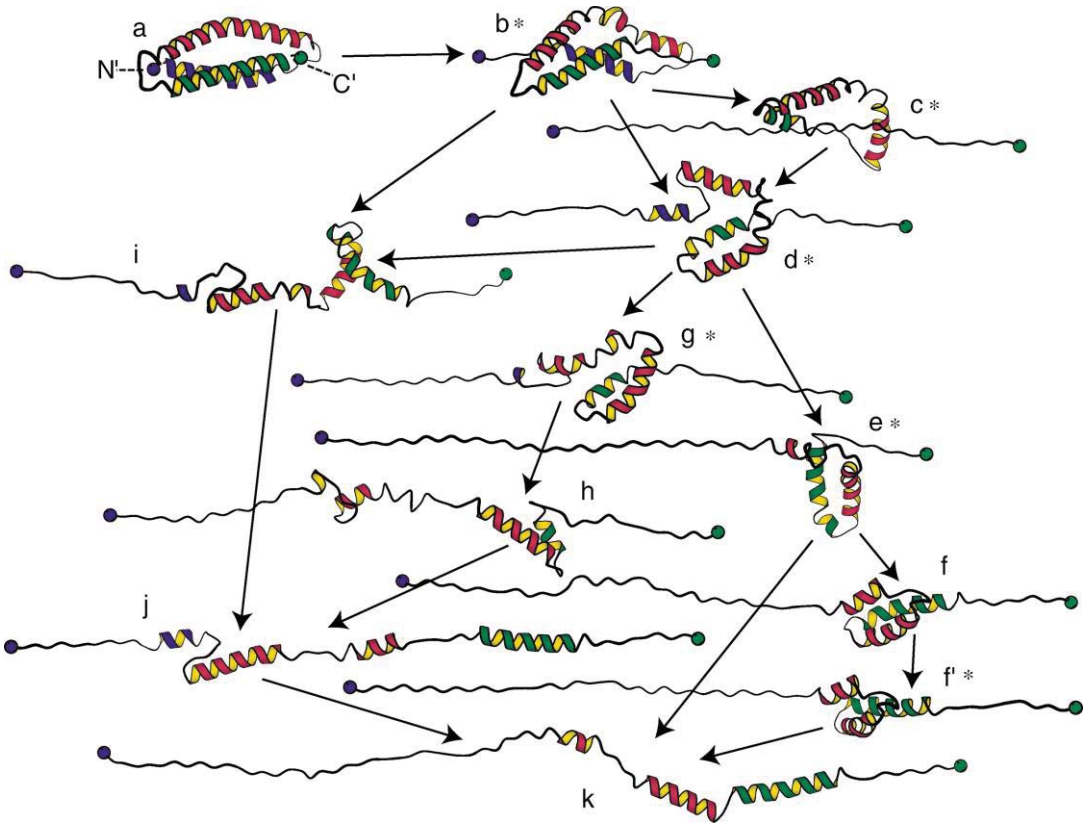


Figure 1-9. Molecular Dynamics simulation experiments using solution structure of repeat 16 from chicken brain α-spectrin showing characteristic unfolding topologies. The distances between N- and C- terminus for each topology are (given in Å) a-42, b-71, c-135, d-126, e-173, f'-205, g-158, h-175, i-143, j-192, k-241.

Altmann *et al.*, 2002

1.3.1.2.3 Function

The number of repeating units differs between members of the spectrin family of cytoskeletal proteins. Spectrin repeats are triple helical coiled-coil structures initially thought to act as a spacer between the functional N- and C- terminal domains [Winder, 1997]. However, this is not the only function of spectrin repeat. Biochemical characterisation revealed that repeats 11 – 17 from Dys bind F-actin but the corresponding repeats from Utr don't [Amann *et al.,* 1998; Amann *et al.,* 1999]. The affinity of the Utr actin-binding domain alone for F-actin is less when compared with

Utr-ABD plus the first spectrin repeat, suggesting that the rod domain also contributes in the actin-binding mechanism [Rybakova *et al.,* 2002; Sutherland-Smith *et al.,* 2003].

The group of basic repeats (repeats 11-14 with a pI of 8.04) in the Dys central rod domain bound F-actin with a Kd of 14.2±5 μM and the binding persists until the ionic strength was increased to 400 mM whereas the group of Dys acidic repeats (repeat 7 to 10, pI of 5.05) showed no intrinsic actin-binding activity [Amann *et al.,* 1998], suggesting the binding is through electrostatic interactions to acidic F-actin. In contrast the Utr repeat domains 11-16 (pI of 4.86) showed no actin-binding activity [Amann *et al.,* 1999]. However in contrast Utr repeats 1 to 10 in combination with the ABD showed a lateral interaction with F-actin; this suggested that both Utr and Dys have distinct modes of domain interaction with F-actin [Rybakova *&* Ervasti, 2005]. Other binding partners of Dys and Utr were identified using a blot-overlay assay, revealing that the intermediate filament protein synemin binds to repeat 4 of Dys and weakly to repeat 3 of Dys and Utr [Bhosle *et al.,* 2006]. Furthermore, repeats 8 and 9 of Utr interact with β-dystroglycan through a stable polarity regulatory kinase, PAR-1b [Masuda-Hirata *et al.,* 2009; Yamashita *et al.,* 2010]. Transfection of cDNA containing repeats 16 and 17 from Dys shows that they interact with sarcolemma-associated neuronal nitric-oxide-synthase (nNOS) [Lai *et al.,* 2009]. However, the mechanism of these interactions was unknown. In muscle physiology, nNOS plays a key role in muscle contraction, muscle regeneration, glucose uptake and in blood perfusion [Stamler & Meissner, 2001].

Studies have also shown that the spectrin repeats bind to membrane phospholipids [Bonnet *&* Begard*,* 1984; DeWolf *et al.,* 1997; Diakowski *et al.,* 1999]. It was shown that Dys repeat 2 binds specifically to anionic phospholipids [Rumeur *et al.,* 2007] and Dys repeats 4 to 19 also bind strongly to lipids through electrostatic and hydrophobic forces [Legardinier *et al.,* 2009a]. The relative importance of spectrin repeats was initially thought to be low as it was observed that the in-frame deletion of repeats 4 to 19 (constitutes 46%) in the Dys gene resulted in the milder BMD phenotype [England *et al.,* 1990]. However, this assumption has been increasingly challenged after it was observed that the in-frame deletion of two amino acids that destabilises repeat-23 is

associated with a severe DMD phenotype [Legardinier *et al.,* 2009b]. This shows that the spectrin repeats play a distinct role in muscle cells apart from acting like a shock-absorber.

1.3.1.3 C-terminal domains

The C-terminal domains of Dys and Utr consist of a WW domain, a pair of EF-hands, ZZ-domains and a coiled-coil region (shown in Figure 1-2). Both Dys and Utr contain copies of these domains. The presence of these domains is important for protein-protein or protein-membrane interaction [Broderick & Winder, 2002].

1.3.1.3.1 WW domain

The WW domain is a protein-protein interaction module that binds with proline-rich sequences [Kay *et al.,* 2000]. The name WW-domain is derived from the presence of two tryptophan residues that are spaced 20-22 amino acids apart [Bork & Sudol, 1994]. WW-domains are short, having approximately 40 amino acids. The structure of the dystrophin WW-domain was solved as a part of the larger structure including the EF-hand in complex with the fragment from β-dystroglycan [Huang *et al.,* 2000] (Figure 1-10). The WW-domain of Dys and Utr binds to the extracellular matrix receptor dystroglycan [Winder, 2001]. It was shown that the WW-domain mediates interaction between Utr and β-dystroglycan by recognising the PPPY peptide at the carboxy terminus of β-dystroglycan [Macias *et al.,* 2002].

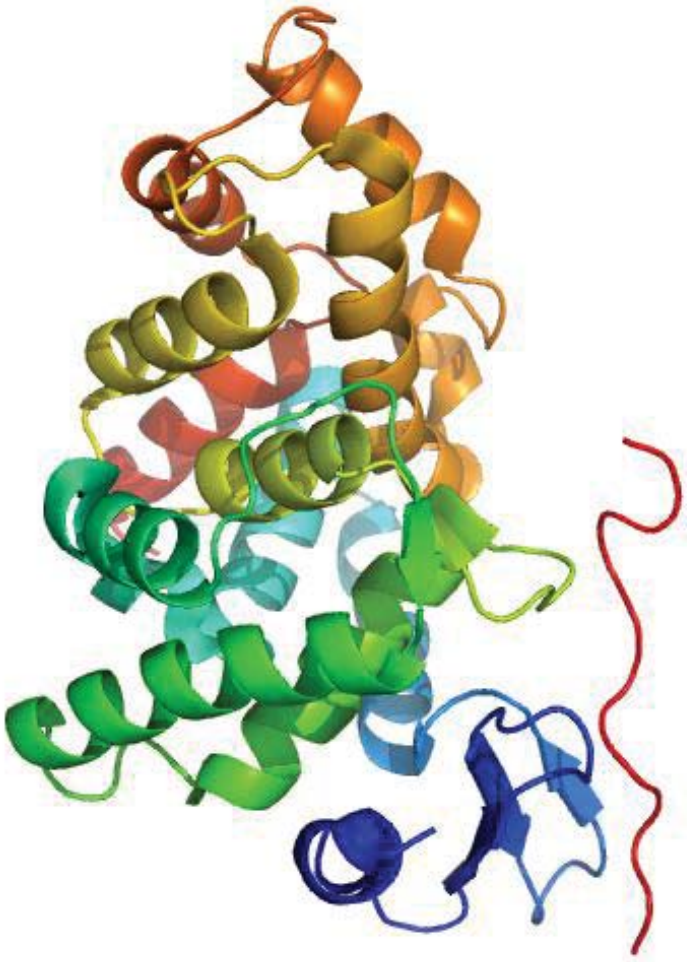


Figure 1-10. Crystal structure of the WW-domain (cartoon representation) from dystrophin in complex with the fragment from β-dystroglycan (Red, Loop representation).

Huang *et al.,* 2000

1.3.1.3.2 EF-hands

The EF-hand motif contains the helix-loop-helix topology found in a large family of calcium-binding proteins. It consists of a two α-helices, roughly perpendicular to one another joined by a loop containing approximately 12 amino acids. These motifs usually bind to calcium, but occasionally do bind magnesium [Tufty & Kretsinger, 1975]. Calcium binding at the loop region triggers the conformational change from 'closed' to 'open' exposing a hydrophobic surface that binds target proteins [Nakayama & Kretsinger, 1994; Huang *et al.,* 2000].

1.3.1.3.3 ZZ-domain

Dys and Utr contain a zinc finger motif (ZZ-domain) at their C-terminus. These motifs contain 5 cysteine residues that participate in binding two zinc ions [Ponting *et al.,* 1996; Winder*,* 1997; Hnia *et al.,* 2007]. ZZ-domains strengthen the interaction between Dys/Utr WW-EF regions and β-dystroglycan [Rentschler *et al.,* 1999; Nuttall *et al.,* 2000].

1.4 Aim of the Study:

The aim of this study is the structural and biochemical characterisation of the N-terminal first spectrin repeat and N-terminal first two spectrin repeats from utrophin (USR1 and USR12) and dystrophin (DSR1 and DSR12), with the hypothesis being that the first N-terminal utrophin spectrin repeat binds to F-actin [Sutherland-Smith *et al.,* 2003]. Furthermore, I aim to compare the structural stability using circular-dichroism spectroscopy (CD) and protein unfolding experiments using molecular dynamics simulation.

Therefore, the key objectives of the experimental work described in this dissertation were as follows,

1) To clone USR1, USR12, DSR1 and DSR12 in the pPROEx HTb vector and transform into *E.coli* BL21 DE3 chemically competent cells.

2) To establish purification protocols for the recombinant proteins, which includes affinity purification followed by gel filtration chromatography.

3) To setup crystal plates, optimising the crystallisation condition, data collection and structural determination of purified recombinant proteins.

4) To determine whether Utr and Dys N-terminal spectrin repeats have any intrinsic actin-binding activity in the absence of the actin-binding domain.

5) To predict secondary structure using circular-dichrosium spectroscopy of recombinant proteins.

6) To investigate the influence of urea concentration and temperatures on the stability of spectrin repeats.

7) To perform equilibrium and steered molecular dynamic simulations of solved structures using NAMD and VMD program.

8) To study the protein surface of both Utr and Dys N-terminal spectrin repeat.

1.5 Significance

This detailed study of utrophin and dystrophin in regards to structural, biochemical and biophysical properties will contribute to an understanding of how the expression of utrophin in muscle cells may replace the function of missing dystrophin in DMD and BMD patients.

2. Experimental Procedures

2.1 Materials

2.1.1 Water

All water used for this work was milliQ, purified over two ionic-exchange filters and two organic filters in a Barnstead NANOpure II system (Thermoscientific) and autoclaved when necessary.

2.1.2 Buffers and solutions

All buffers and solutions were made using milli-Q water and filtered using 0.2 and 0.4 μm filters (Satorius Stedim biotech).

2.1.3 Commercial suppliers

Fine chemicals used in this work were purchased in analytical grade from Merck, Ajax Finechem, SIGMA, BIO-RAD, Fluka, Carl Roth GmbH, Hampton Research and Molecular Dimensions. Enzymes for molecular biology were purchased from Roche and Invitrogen. Protease inhibitors were purchased from Roche. Antibiotics were purchased from Roche and Invitrogen. Crystallisation screens and crystallisation plates were purchased from Hampton Research and Molecular Dimensions. Protein concentrators were purchased from GE Healthcare and Sartorius.

2.1.4 Bacterial strains

TOP10 and XL1-Blue® (Stratagene) *E. coli* cells were used for subcloning and vector amplification and *E. coli* BL21 (DE3) (Invitrogen) was used for protein expression studies.

2.1.5 Media for *E. coli* growth

The Luria Broth (LB) media (Invitrogen) was made up in milli-Q water at 25 g/L and autoclaved at 121 °C for 15 minutes before use. To make LB agar, 1% agar was added to the broth before autoclaving. 2YT media was prepared by mixing 16 g of Bacto Tryptone, 5 g of Yeast extract and 5 g of NaCl made up in 1 L volume of milli-Q water and autoclaved at 121 °C and stored at room temperature until use.

2.1.6 Antibiotics

Both ampicillin (Amp) and kanamycin (Kan) was used in this study, when required. Stock solutions were prepared according to the manufacturer's instructions (Roche hand book) and stored at -20 °C for further use.

2.1.7 cDNA Clones for PCR subcloning

The pRSV plasmid containing the full length human dystrophin (NCBI Nucleotide Data Base Code – Human NM_004006| *Homo sapiens* dystrophin), was kindly donated by Prof. George Dickson (University of London) and the pQE vector containing a fragment of rat utrophin (NCBI Nucleotide Data Base Code - Rat NM_013070| *Rattus norvegicus* utrophin) open reading frame (UT11: residues encoding from 2 to 594) was obtained from Prof Marcus C. Schaub (University of Zurich) [Zuellig *et al.,* 2000].

2.2 Methods

2.2.1 Polymerase chain reaction

The polymerase chain reaction (PCR) is a most popular technique developed in 1983 by Kary Mullis [Barlett *&* Stirling, 2003], used for rapid and specific amplification of target DNA-sequences.

2.2.1.1 PCR amplification

The oligonucleotide primers used for PCR amplification were synthesized by IDT (Integrated DNA Technologies, USA). The amplification procedure comprised denaturation at 92 °C for 2 min; this was followed by 35 cycles including denaturation for 1 min at 92 °C, annealing for 1 min at 55 °C, and polymerization for 1 to 5 min at 68 or 72 °C. The reactions were performed in a final volume of 50 µl with 2.5 U of pwo DNA polymerase (Roche). The PCR products were analysed by agarose gel electrophoresis, then purified and eluted using a PCR product purification kit (Roche). DNA fragments were digested and were subcloned into plasmids pPROEx HTb (invitrogen) as required [Sambrook *et al.,* 1989].

2.2.1.2 PCR Screening

This method was used to screen transformed bacterial colonies to determine if they contained a vector with the inserted cDNA of interest. Single transformed bacterial colonies were selected from the LB-agar-plate and transferred into a PCR-tube containing 20 µL of the pre-pipetted PCR-reaction mixture. PCR was performed immediately and the presence of a product confirmed by agarose gel electrophoresis.

Table 2-1. *Reaction mixture for whole cell PCR.*

PCR Master mix	**Volume in µl**
10X Taq PCR Buffer	2.5 µl
$MgCl_2$	1.0 µl (50 mM Stock)
Primer 1 (5' end of vector or gene specific)	0.1 µl from 100 µM stock
Primer 2 (3' end of vector or gene specific)	0.1 µl from 100 µM stock
dNTP mix	0.2 µl from 25 mM Stock
dH_2O	15.6 µl
Taq Polymerase	0.5 µl

2.2.2 Cloning and transformation

2.2.2.1 Digestion with Restriction Endonuclease

The vector pPROEx HTb and PCR products were digested using restriction endonucleases purchased from Roche/Invitrogen by incubation for 1 hour at 37 °C. The digestion mixture contained 1.0 µl of 10X enzyme buffer (appropriate buffer system selected from Roche/Invitrogen manual), 6.0 µl dH_2O, 2 µl DNA and the corresponding restriction endonucleases, each of 0.5 µl volume. Digestion was confirmed only after running the vectors and PCR products in a 1% agarose gel electrophoresis.

2.2.2.2 Ligation

The digested vectors and PCR products were mixed in 1:3 molar ratio and the resultant reaction mixture was incubated at room temperature for 2 hours with 5U (1.0 µl) T4 DNA ligase (Invitrogen) before transformation.

2.2.2.3 PCR product purification

A high pure PCR product purification kit (Roche) was used to purify and concentrate the amplified DNA.

2.2.2.4 Plasmid isolation

Plasmids were isolated from the 5 mL cell cultures that were grown over night at 37 °C with shaking in LB-broth using a high pure plasmid isolation kit (Roche).

2.2.2.5 Preparation of competent *E. coli* cells for transformation

5.0 µL of frozen competent *E. coli* cells were added to a flask containing 50 µL of LB-medium with out any antibiotics. The flask was incubated at 37 °C with constant shaking at 200 rpm. When the OD_{600} of the bacterial culture reached approximately 0.5, the flask was removed from the shaker and cooled on ice for 10 minutes and the

cells were harvested by centrifugation at 4400 rpm, 4 °C for 20 minutes. The cell pellets were re-suspended in 20 ml of ice cold TFB1-buffer (30 mM KAc, 100 mM $RbCl_2$, 10 mM $CaCl_2$, 50 mM $MnCl_2$, 15% (v/v) glycerol, pH 5.8) and incubated on ice for 5 minutes and then the cells were harvested by centrifugation at 4400 rpm, 4 °C for 20 minutes. The cell pellets were re-suspended in 2 ml ice cold TFB2-buffer (10 mM MOPS, 75 mM $CaCl_2$, 10 mM $RbCl_2$, 15% (v/v) glycerol, pH 6.5). 50 µl aliquots of cell suspension were dispensed into tubes and stored at -80 °C until further use.

2.2.2.6 Transformation into *E. coli*

Competent *E. coli* cells (50 µl) were thawed on ice and 2 µl of plasmid DNA was added and mixed gently. The cells and DNA mixture were incubated on ice for 15 minutes before placing them in a 42 °C water bath for 60 sec and then cooled on ice for 15 minutes. After transformation, the cells were plated on LB-agar plates containing the appropriate antibiotic(s) and incubated overnight at 37 °C.

2.2.2.7 Isolation of plasmid DNA from *E. coli* cells

A microcentrifuge tube was filled with an overnight bacterial culture grown in LB + antibiotic and centrifuged to a 10,000*g* for 1 minute. The supernatant was discarded and cell pellets were used for extracting plasmid DNA using a high pure plasmid isolation kit (Roche) according to the manufacturer's instructions.

2.2.2.8 Determination of DNA concentration

DNA concentration was determined using a nano drop spectrophotometer (NanoDrop® ND-1000). 2 µL DNA solution was sufficient to quantitate the amount of DNA present by absorbance at 260 nM wavelength.

2.2.2.9 DNA sequencing

Expression vectors used in this study containing the cDNA of interest were sequenced using the Applied Bioscience (ABI) 3730 DNA Analyser from the Allan Wilson Centre genome service (AWCGS). The sequences and chromatograms were analysed using the software 4Peaks (http://www.mekentosj.com/science/4peaks) and were found to be error free.

2.2.3 Gel electrophoresis

Electrophoresis is one of the important techniques for separating molecules based on charge, shape and size of the molecules. There are several forms of electrophoresis technique used depending on the requirements. In this research I used both agarose gel electrophoresis (AGE) and sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE).

2.2.3.1 Agarose gel electrophoresis (AGE)

AGE is the most common technique for separating and analysing DNA. 1% Agarose (w/v) (Roche) was made in 1X TAE buffer (40 mM Tris HCl, 20 mM acetic acid, 2 mM EDTA, pH 8.0), by heating the suspension till the agarose was completely dissolved. 4 µl of DNA sample was mixed with 4 µl of loading dye (0.2 % w/v bromophenol blue, 50 % v/v glycerol). Gels were submerged in a 1X TAE buffer and samples were added onto the wells before running it at 100 V in an Easy-cast electrophoresis system (Owl scientific) for 40 minutes. A 1 Kb Plus DNA ladder (Invitrogen) was run on all the gels. The DNA was visualised by staining the gels with ethidium bromide for 15 minutes, followed by destaining in water for another 15 minutes. The images were captured under ultraviolet light on a Geldoc (BioRad), using QuantityOne software (BioRad).

2.2.3.2 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

The purpose of SDS-PAGE is to determine protein sample purity by separating a protein mixture according to their sizes. Both 15% and 12% polyacrylamide separating

gels were used depending on the size of the proteins analysed in combination with a 4% stacking gel poured on the top of the separating gel. Samples were mixed 1:1 with 5X loading buffer (10% w/v SDS, 10 mM DTT, 20% glycerol, 0.2 M Tris-HCl pH 6.8, 0.05% bromophenol blue) and boiled for 5 minutes, then loaded on the gel. Precision plus protein molecular weight marker (BioRad) was used for all the gels. SDS-PAGE gels were clipped onto Mighty small II mini vertical gel electrophoresis unit (Amersham) and were run in running buffer (25 mM Tris, 0.2 M glycine, 0.1% SDS). Gels were run at 200V until the dye reached to the bottom end of the gel, then the gel was removed carefully and stained (0.125% w/v coomassie brilliant blue R250, 40% v/v ethanol, 10% acetic acid) for 1 hour and then destained for 3 hours in a destain solution (20% ethanol, 10% acetic acid).

2.2.4 Protein expression, purification and analysis

2.2.4.1 Expression of recombinant protein in *E. coli*

The recombinant plasmid was transformed into chemically competent *E. coli* BL21 (DE3) cells using a heat shock method. These cells were grown in 1.5L LB broth with 0.1 mg/ml ampicillin (Sigma) at 37 °C with shaking at 200 rpm, until reaching an optical density (A_{600}) of ~ 0.5 – 0.7. The flasks were then removed from 37 °C and were placed at 25 °C, with 200 rpm shaking. After 1 hour, isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 0.1 mM. The cells were incubated with shaking over night at 25 °C and were then harvested by centrifugation at 5,000 x *g* for 20 min at 4 °C. The cell pellets were washed in 100 ml of PBS (150 mM NaCl, 20 mM Na_2HPO_4/NaH_2PO_4 pH 8.0) and kept frozen at -20 °C until use.

2.2.4.2 *E. coli* cell lysis

Cell pellets were resuspended in 40 ml of lysis buffer (20 mM Tris, 1 mM β-mercaptoethanol, 5 mM imidazole and one EDTA free protease inhibitor tablet (Roche) (dissolved) and were subjected to French press (twice at 4,000 psi), followed by sonication using a Misonix ultrasonic cell disrupter twice for 20 sec on a setting of 3, to shear the nucleic acid. Samples were then centrifuged 17,000 x *g* for 20 minutes using

a SORVALL SS34 fixed angle rotor at 4 °C. The supernatant was then filtered using a 0.20 µm filter (Sartorius stedim biotech).

2.2.4.3 Sonication

Cells and cell lysate were sonicated by an Ultrasonic Processor (Misonix) probe sonicator, at a power level 3 on ice for 20 sec, done twice with 1-minute intervals to cool the samples from the heat generated by sonication.

2.2.4.4 Ni^{2+} Affinity chromatography

Affinity chromatography using Ni^{2+}-NTA Sepharose was the first step in the chromatographic purification of recombinant proteins. It was done to remove most cellular proteins from the recombinantly expressed $(His)_6$ tagged proteins. In these experiments 5 ml prepacked HisTrap HP column (GE Healthcare) were used. These columns were washed with milli-Q water and then equilibrated with lysis buffer before the filtered lysate was passed through at a flow rate of 0.5 ml per minute at 4 °C. After proteins were eluted, the columns was washed with 10 column volumes (CV) of dH_2O and stored in 20% EtOH (Ethanol) for further use.

2.2.4.5 Dialysis

Dialysis is a separation technique driven by a concentration gradient; this technique facilitates the removal of small and unwanted molecules from the macromolecules in solution by selective and passive diffusion through a semi-permeable membrane. The membrane itself is made of cellulose acetate and is chemically inert and will not bind with the protein samples [McPhie, 1971].

Before the protein solution was transferred to the dialysis membrane (Spectrum Laboratory, Inc.), the membrane was rinsed with deionised water for 10 to 20 minutes. The dialysis reservoir was filled with 2 L of dialysis buffer (20 mM Tris-HCl pH 8.0, 50 mM NaCl) and the dialysis was continued overnight at 4 °C, to remove any imidazole present. Finally the dialyzed protein solution was concentrated using Vivaspin 20

columns (GE Healthcare).

2.2.4.6 rTEV protease production, purification and storage

The pET 19b vector containing rTEV protease cDNA in *E. coli* BL21 (DE3+pARG) was kindly donated by Trevor Loo. The cells were grown in 1.5L 2YT containing both 0.1 mg/ml ampicillin and kanamycin (SIGMA) at 37 °C. After reaching an OD_{600} of ~0.7, the flasks were then shifted to 25 °C. After 1 hour, IPTG was added to a final concentration of 0.1 mM and the flasks were kept shaken at 25 °C overnight. The cells were harvested by centrifugation at 5,000 x *g* for 20 min at 4 °C. The cell pellets were subjected to French press (twice at 4,000 psi) in 40 ml of lysis buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 10 mM β-mercaptoethanol, 150 mM PMSF) containing 10 mM imidazole, followed by sonication using a Misonix ultrasonic cell disrupter twice for 20 sec on a setting of 3, to shear the nucleic acid. The lysate was centrifuged about 17,000 x *g* for 30 min using a SORVALL SS34 fixed angle rotor at 4 °C. The supernatant was then filtered using 0.20 μm filter (Sartorius stedim biotech) and passed through a 5 ml prepacked Ni^{2+}-NTA resin column (nickel nitrotriacetic acid, GE Healthcare) with a flow rate of 1.0 ml per minute at 4 °C. The column was washed with 5X column volume (CV) of lysis buffer contain 20 mM imidazole and followed by 5X column volume (CV) of lysis buffer containing 50 mM imidazole. Furthermore, the protein was eluted using 5X column volume of elution buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 250 mM imidazole, 10 mM β-mercaptoethanol, 150 mM PMSF) and the purification was confirmed by 12% SDS-PAGE. The fractions containing 30 kDa rTEV protease were pooled and dialysed overnight at 4 °C in a storage buffer (20% glycerol, 50 mM Tris-HCl pH 8.0, 1 mM EDTA, 1 mM DTT) concentrated to 1 mg/ml using VivaSpin20 (MWCO 10 kDa), flash frozen and stored in 20 μl aliquots at -80 °C until use.

2.2.4.7 Cleaving N-terminal $(His)_6$ tagged fusion proteins

The Hexa histidine tag $(His)_6$ was cleaved by adding rTEV protease to the target protein in a ratio (w/w) of 1:50 with the reaction incubated overnight at 4 °C. The His tag removal was confirmed by SDS PAGE. The reaction mixture was then run through a

HisTrap column at a flow rate of 0.5 mL/min, the flow through was collected and concentrated using VivaSpin columns (5 kDa MWCO).

2.2.4.8 Size Exclusion Chromatography (SEC) / Gel Filtration Chromatography

The Superdex 75 10/300 GL column (Amersham Biosciences) size exclusion column was pre equilibrated with running buffer (20 mM Tris HCl pH 8.0, 50 mM NaCl). Concentrated protein samples (>5 mg/mL) were then loaded onto the column using a superloop with 250 µl sample injected on to the column during each run at a flow rate of 0.4 mL/min using an ÄKTA Explorer system (Amersham Biosciences). 0.5 mL fractions were collected during each run. These fractions were then analysed using SDS PAGE and were found to be over 95% pure. The collected fractions were concentrated using VivaSpin columns (5 kDa MWCO) for further studies.

2.2.4.9 Determination of protein concentration

The concentration of purified protein solutions were determined spectrophotometrically by absorbance at 280 nm [Perkins & Sim, 1986]. The absorbance was converted to concentration (mg/mL) using Beer's Law:

= (A_{280} X Molecular weight in Daltons) / (molar extinction coefficient at 280 nm X pathlength in mm).

The molecular weight and molar extinction coefficient were calculated by submitting the amino acid sequence to the ProtParam tool [Gasteiger *et al.,* 2005] in ExPASy server (http://web.expasy.org/protparam/).

2.2.5 Protein Crystallisation

2.2.5.1 Principles of crystallisation

The crystallisation of proteins is still in general, a complex procedure. Successful X-ray crystallography requires a single crystal that comprises homogenous protein in a well-ordered lattice [Bergfors, 1999]. Crystallisation of macromolecules is still considered a

time consuming task that consists of testing several hundreds of potential screening conditions and optimizing the condition with a subtle changes. Two of the most commonly used methods for protein crystallisation are hanging and sitting-drop vapour diffusion. Both methods feature a droplet of purified protein, buffer and precipitant equilibrating with a larger reservoir containing similar buffers and precipitants in higher concentrations. When the equilibration process occurs, the water vapor diffuses from the drop to the reservoir until the precipitant concentration reaches a level suitable for crystallisation.

2.2.5.2 Initial screening

Initial crystallisation experiments were setup with commercially available crystal screens from various manufacturers as shown in *Table 2-2* utilizing both the sitting and hanging-drop vapour diffusion method. Equal volumes (1:1) of purified protein and precipitant solution were mixed using a Mosquito Nanodrop Crystallisation Robot (TTP Labtech). After mixing protein and buffer, the plates were sealed and stored at 4 °C or 21 °C. Drops were first observed daily and later monthly with a light microscope (Nikon microsystems).

Table 2-2. Commercially available crystal screens used.

Name of the screen	**Type of screen**	**Distributor**
Crystal screen I	Sparse matrix	Hampton research
Crystal screen II	Sparse matrix	Hampton research
Structural screen I	Sparse matrix	Molecular dimensions
Structural screen II	Sparse matrix	Molecular dimensions
PEG/Ion screen	Anion, cation and pH matrix screen	Hampton Research
JCSG^{+}	Sparse matrix	Molecular dimensions
PACT *premier*^{TM}	Anion, cation and pH matrix screens	Molecular dimensions

2.2.5.3 Optimisation of crystallisation

Optimisation of crystallisation was done to enlarge the size of the crystals and also to enhance the occurrence of single crystals. This can be achieved by varying the concentration of precipitant, pH of the buffer, protein concentration, temperature and additives.

2.2.5.4 Data collection and processing

The protein crystal was mounted onto the goniometer head using a cryo-loop (Hampton Research). The cryo-jet (Oxford Instruments) as part of the X-ray diffractometer was used to freeze the crystal. The cryo-jet used liquid nitrogen cooled air to achieve the temperature desired for cryo-temperature diffraction. The temperature was set to 120 K. The X-rays of wavelength of 1.5418 Å were generated using a copper-*K*α rotating anode generator (Rigaku MicroMax 007). The mounted and centered crystals were then exposed to X-rays with various exposure times and the diffraction pattern was observed on an image plate detector (R-Axis IV^{++}). Data collection was carried out using CrystalClear 1.3.2 (Rigaku) software with the data collected from images optimally having at least 95% completeness and 4 times the redundancy. The data processing was done with the CCP4 suite (Collaborative Computational Project, 1994).

Table 2-3. Crystallographic methods for structure determination.

Program	**Name/Software and Reference**
Data processing	MOSFLM [Batty *et al.,* 1990; Leslie *et al.,* 1990], XDS [Kabsch, 1993]
Scaling and averaging the intensities	SCALA [Evans, 2006]
Phasing	PHASER [McCoy *et al.,* 2007]
Model building	COOT [Emsley & Cowtan, 2004]
Model refinement	REFMAC [Murshudov *et al.,* 1997], PHENIX [Adams *et al.,* 2002]
Structure evaluation and validation	PROCHECK [Laskowski *et al.,* 1993], MolProbity [Davis *et al.,* 2007]
Visualisation	PyMOL [DeLano Scientific], CCP4mg [Collaborative Computational Project, 1994]

2.2.6 Circular dichroism (CD) spectroscopy

CD is regarded as one of the most suitiable and popular techniques available to determine the secondary structure of macromolecules. It refers to the differential absorption of left and right circularly polarised light by chromophores [Johnson *et al.,* 1990], which either possess intrinsic chirality or are placed in chiral environments [Kelly & Price, 2000; Kelly *et al.,* 2005].

The CD measurements were taken on a Chirascan CD spectrophotometer (Applied Photophysics) using quartz cells (Hellma) with a path length of 0.1 mm or 1.0 mm. For the thermal denaturation profiles, the temperature was varied between 5 and 90 °C in 5 °C intervals and the temperature ramping at 1 °C per minute. The urea denaturation profile was varied between 0.5 and 8.0 M urea concentration in steps of 0.5 M.

2.2.7 Actin-binding assay (ABA)

An actin-binding assay was carried out based on the protocol of Moores *et al.,* 2000. 15.3 µL of Monomeric actin (4 mg/ml suspended in water) obtained from Cytoskeleton Inc, was mixed with 1.7 µL of 10x F-buffer (20 mM Tris-HCl pH 8.0, 20 mM $MgCl_2$, 10 mM ATP, 500 mM KCl) with the mixture kept at room temperature for 1 hour to enable a viscous filamentous actin (F-actin) suspension to form. The protein of interest was transferred into ligand buffer (20 mM Tris-HCl pH 8.0, 6.25 mM DTT, 120 mM NaCl) and was centrifuged at 75,000 x *g* at 25 °C to remove any precipitate or insoluble proteins. 20 µl of the supernatant was mixed with 5 µl F-actin and 5 µl of binding buffer (120 mM Tris-HCl pH 8.0, 150 mM NaCl, 20 mM $MgCl_2$, 10 mM ATP, 1 mM $CaCl_2$) and the assay was incubated at 25 °C for 30 min followed by ultracentrifugation at 75,000 x *g.* The protein that binds to F-actin precipitates (pellet fractions) and the protein that doesn't bind to F-actin remains in the supernatant (soluble fractions). Both soluble and pellet fraction were loaded onto an SDS-PAGE gel, and the density of protein bands in the soluble and pellet fractions was determined.

2.2.8 Bioinformatics

2.2.8.1 Sequence alignment and structural analysis

Protein sequences were aligned using ClustalX [Larkin *et al.,* 2007]. Protein domain boundaries were predicted using the pFAM [Finn *et al.,* 2008] and SMART [Schultz *et al.,* 1998] databases. Molecular weight, isoelectric point (pI) and extension coefficients were calculated using the ProtParam tool from the ExPASY Server. Protein secondary structure was predicted using PSIPRED [Bryson *et al.,* 2005] and 3D structure was predicted using the I-TASSER Server [Zhang *et al.,* 2008]. The PISA program (Protein Interfaces, Surfaces and Assemblies) was used to study the protein subunit interfaces and surface [Krissinel*,* 2009] properties. DaliLite Ver3 server was used to compare the solved 3D structure of proteins against those in the protein databank (PDB). Vector sequences and vector maps were constructed using Serial Cloner 2.1 software.

2.2.9 Molecular dynamics simulation

Molecular dynamics (MD) relaxation and simulation were carried out using the NAMD [Nelson *et al.,* 1996] and XPLOR [Brunger, 1992] sets of programs employing the CHARMM22 force field parameters [MacKerell *et al.,* 1998]. Steered molecular dynamics at constant velocity (SMD-CV) were carried out with the solvated (neutralising by adding water and ions) and equilibrated structure, by fixing the Cα atom of the C-terminus of the domain and applying a force to the Cα atom of the N-terminus in the other direction. The force experienced by the pulled atom was given by the following equation, $F = K (vt – x)$. Where x is the displacement of the pulled atom from its original position, v is the specified velocity used for pulling the atom and K is a spring constant. From the above equation the force extension profile was obtained and the analysis of trajectories was performed using Visual Molecular Display (VMD) [Humphrey *et al.,* 1996].

3. Cloning, expression in *E. coli* and purification of utrophin (Utr) and dystrophin (Dys) spectrin repeats

3.1 Introduction

Utr constructs consisted of either the N-terminal first spectrin repeat (USR1: residues from 308 to 426) alone or the N-terminal first two spectrin repeats (USR12: residues from 308 to 537). Dys constructs consisted of either the N-terminal first spectrin repeat alone (DSR1: residues from 338 to 456) or the N-terminal first two spectrin repeat (DSR12: residues from 338 to 567). cDNA encoding these proteins was subcloned in the multiple cloning site (MCS) of the pPROEx HTb vector (invitrogen) and expressed in *E. coli* BL21 (DE3) cells as described in chapter 2.2.2. Protein expression was induced by isopropyl-thiogalactoside (IPTG) as described in chapter 2.2.4.1 and the recombinant protein purified with Ni^{2+} affinity chromatography followed by gel filtration chromatography.

3.2 Cloning USR1, USR12, DSR1 and DSR12

The template (UT11) cDNA encoding amino acids 2 to 594 of rat Utr [Zuellig *et al.,* 2000] was used to prepare USR1 and USR12 and a pRSV plasmid containing full-length human Dys cDNA was used for DSR1 and DSR12. The spectrin repeat boundaries of Utr and Dys were indentified from the template sequence using the SMART web server (http://smart.embl-heidelberg.de/) [Schultz *et al.,* 1998], and correlated with results from the Pfam database (http://pfam.sanger.ac.uk/) [Finn *et al.,* 2008]. Primers were then designed for all four constructs USR1, USR12, DSR1 and DSR12 for cloning into pPROEx HTb vector (Invitrogen) (Table 3-1.). This enabled recombinant protein expression with an N-terminal $(His)_6$ tag cleavable from the target proteins by a tobacco etch virus (rTEV) protease cleavage site.

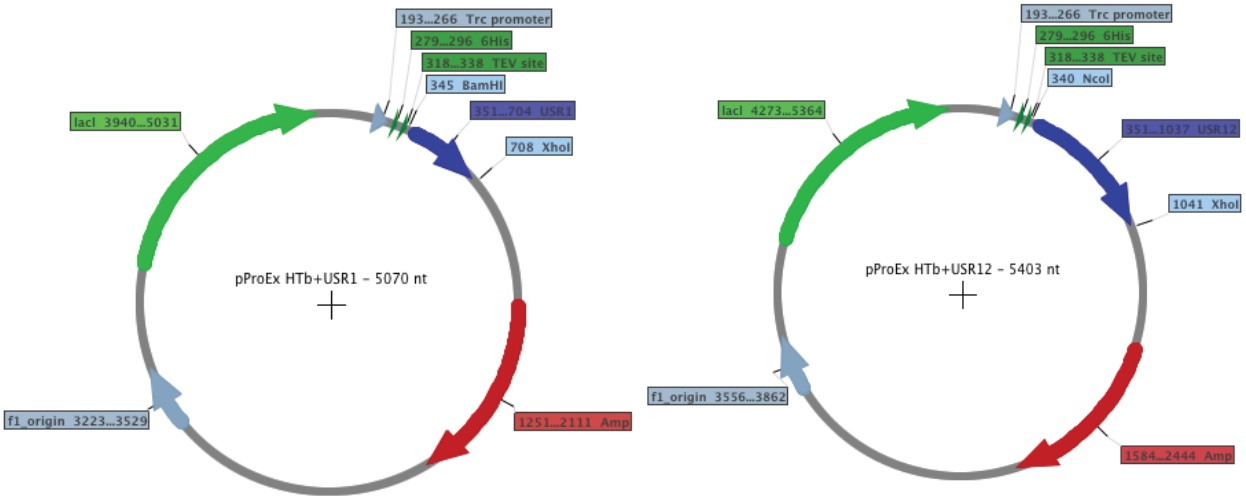
Table 3-1. Primers used for the study.

Name	Sequence	Orientation
USR1	GGTGGTCCATGGCAGACATGGATTTGGACAGCTACCAG	Sense
USR1	ATTGATCCTCGAGTCATTTCTTCTGCAGTTCCATCAGGGCATC	Antisense
USR12	ACCACCAAGCTTCTCGAGTTATTCTTGCAACCTGTTCCAACGT	Sense
USR12	ACCACCAAGCTTCTCGAGTTATTCTTGCAACCTGTTCCAACGT	Antisense
DSR1	GGTGGTCCATGGCAGACATGGATTTGGACAGCTACCAG	Sense
DSR1	ATTGATCCTCGAGTCATTTCTTCTGCAGTTCCATCAGGGCATC	Antisense
DSR12	ACCACCAAGCTTCTCGAGTTATTCTTGCAACCTGTTCCAACGT	Sense
DSR12	ACCACCAAGCTTCTCGAGTTATTCTTGCAACCTGTTCCAACGT	Antisense

3.2.1 Cloning

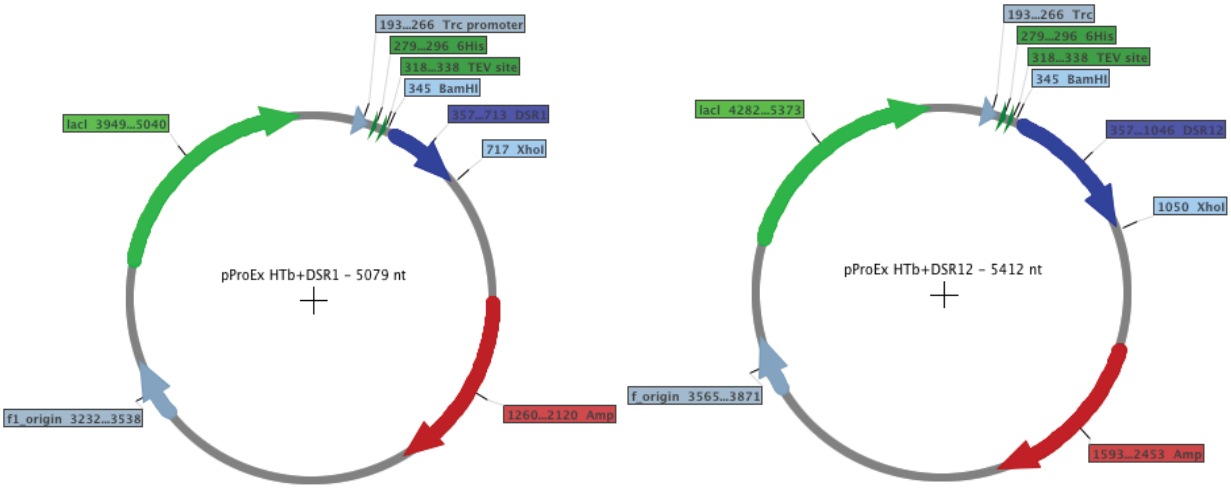
The N-terminal spectrin repeat 1 of rat Utr (USR1, amino acids 308 to 426) was cloned into the *Bam*HI and *Xho*I restriction sites in the multiple cloning sites of the pPROEx HTb vector (Figure 3-1[a]) and the first two N-terminal spectrin repeats (USR12, amino acids 308 to 537) of rat Utr, cloned into the *Nco*I and *Xho*I restriction sites in the multiple cloning site of the pPROEx HTb vector (Figure 3-1[b]).

Similarly, the first N-terminal spectrin repeat of human Dys (DSR1, amino acids 338 to 456) was cloned into *Bam*HI and *Xho*I restriction sites in the multiple cloning sites of the pPROEx HTb vector (Figure 3-2[a]) and the first two N-terminal spectrin repeats (DSR12, amino acids 338 to 567) of human Dys, cloned into the *Bam*HI and *Xho*I restriction sites in pPROEx HTb (Figure 3-2[b]).



[a] **[b]**

Figure 3-1. Plasmid maps containing N-terminal spectrin repeats from rat Utr. **[a]** USR1 was cloned in *Bam* HI and *Xho* I restriction sites in the multiple cloning sites of the pPROEx HTb Vector, **[b]** USR12 was cloned in *Nco* I and *Xho* I restriction sites in the multiple cloning sites of the pPROEx HTb Vector.



[a] **[b]**

Figure 3-2. Plasmid maps containing N-terminal spectrin repeats from human Dys. **[a]** DSR1 was cloned in *Bam* HI and *Xho* I restriction sites in the multiple cloning sites of the pPROEx HTb Vector, **[b]** DSR12 was cloned in *Bam* HI and *Xho* I restriction sites in the multiple cloning sites of the pPROEx HTb Vector.

3.3 Proteins expression in *E. coli*

The recombinant plasmid containing the target sequence was then transformed into chemically competent *E. coli* BL21 (DE3) cells using a heat shock method as described in chapter 2.2.2.6 and the protein was expressed as described in chapter 2.2.4.1.

3.4 Purification of His-tagged proteins (USR1, DSR1, USR12 and DSR12)

The cell pellets was resuspended in 40 ml of lysis buffer (50 mM Na_2HPO_4/NaH_2PO_4 pH 8.0, 150 mM NaCl, 5 mM imidazole, 1 mM β-mercaptoethanol and one EDTA free protease inhibitor tablet dissolved (Roche). The cell suspension was lysed (see chapter 2.2.4.2) by French press, followed by sonication using a Misonix ultrasonic cell disrupter (see chapter 2.2.4.3) to shear the nucleic acid. Samples were then centrifuged at 17,000 x *g* for 20 minutes using a SORVALL SS34 fixed angle rotor at 4 °C to remove insoluble materials. The resultant supernatant was then filtered using a 0.20 μm filter (Sartorius stedim biotech).

The filtered lysate was passed through a 5 ml prepacked HighTrap Ni^{2+}-NTA resin column (GE Healthcare) preequilibrated with lysis buffer at a flow rate of 0.5 ml per minute using an Econo Pump (Bio-Rad) to allow the His-tagged Utr or Dys to bind with the resin. The column was washed twice with 10 ml of 10 mM imidazole in PBS (50 mM Na_2HPO_4/NaH_2PO_4 pH 8.0, 150 mM NaCl) and then three washes with 20 mM imidazole in PBS to remove loosely-bound contaminant proteins. The His-tagged recombinant proteins bound to the resin were eluted using three 10 ml washes each of 50, 100 and 150 mM imidazole in PBS. 20 μl of samples from each fraction was added to 10 μl of 5x SDS loading buffer (see chapter 2.2.3.2) and samples were heated before loading 5 μl on to the SDS-PAGE gel.

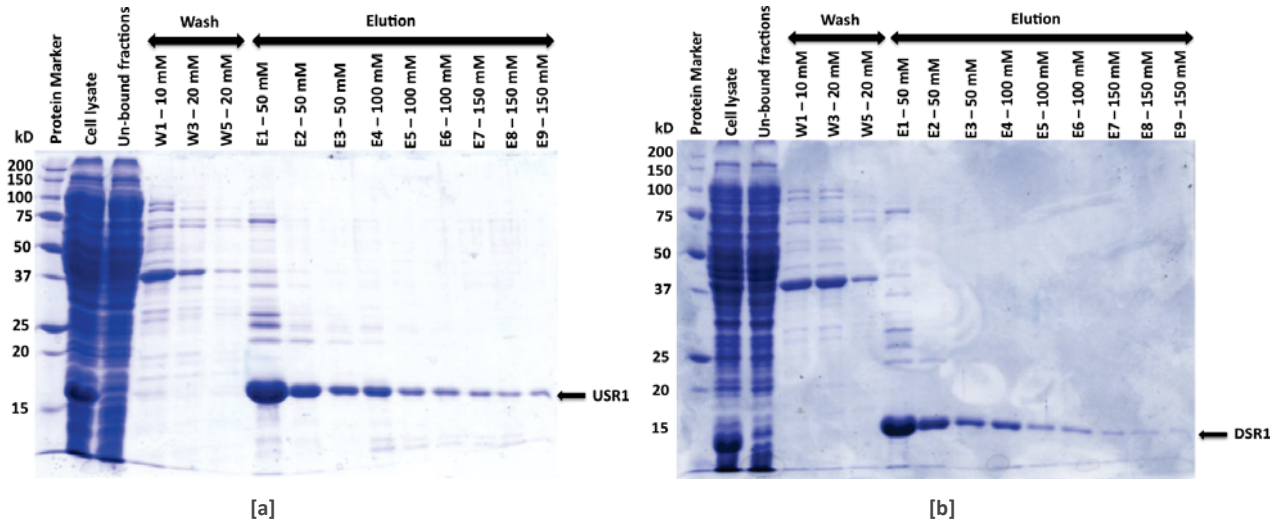


Figure 3-3. SDS-PAGE analysis showing the affinity purification of, **[a]** USR1 and **[b]** DSR1. W1, W3 and W5 are column washes with 10 and 20 mM imidazole in PBS to remove un-bound and non-specific proteins. E1 to E9 are fractions eluted using 50, 100 and 150 mM Imidazole in PBS buffer.

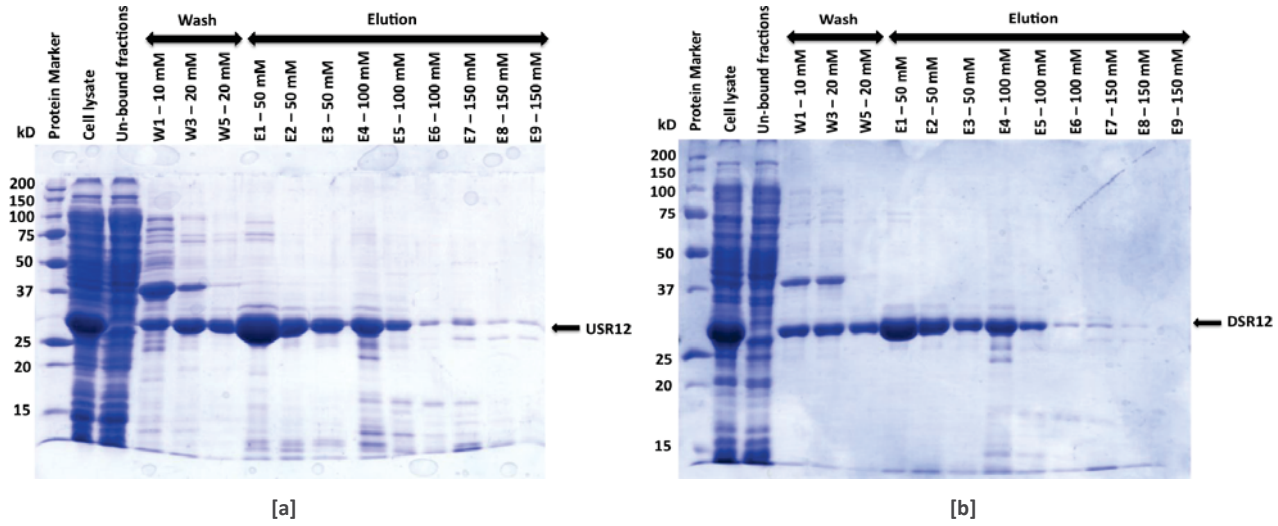


Figure 3-4. SDS-PAGE analysis showing the affinity purification of, **[a]** USR12 and **[b]** DSR12. W1, W3 and W5 are column washes with 10 and 20 mM imidazole in PBS to remove un-bound and non-specific proteins. E1 to E9 are fractions eluted using 50, 100 and 150 mM Imidazole in PBS buffer.

Eluted protein fractions were dialysed against 2.0 L PBS overnight at 4 °C using Spectra/Por dialysis membrane with molecular weight cutoff (MWCO) 6-8000 Da (Spectrum Laboratories, Inc) to reduce the imidazole concentration. The dialysed protein fractions were centrifuged at 14,000 rpm at 4 °C to remove any precipitates formed during dialysis and then concentrated using 20 ml Vivaspin concentrators (GE Healthcare) with MWCO 5000 kDa.

3.5 N-terminal $(His)_6$-tag cleavage

The N-terminal hexa His-tag was cleaved by incubating Utr or Dys protein samples overnight at 4 °C with rTEV protease [Carrington *&* Dougherty*,* 1988]. The digested protein samples were loaded on a Ni^{2+}-NTA column and washed off with PBS. The recombinant proteins without a His-tag were collected during subsequent washing steps (F1 to F8). Un-cleaved Dys or Utr proteins and His-tag rTEV protease were left bound to the HighTrap Ni^{2+}-NTA columns (GE Healthcare) and were eventually eluted using 500 mM imidazole in PBS. The collected fractions was analysed by loading 5 µl on to SDS-PAGE as shown in *Figure 3-5* for USR12. Similar analysis using SDS-PAGE was observed for USR1, DSR1 and DSR12, *data not shown.*

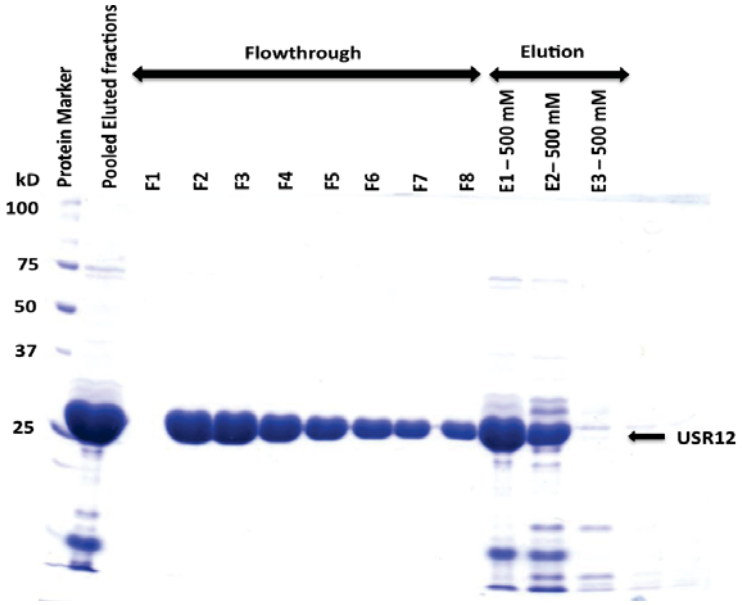
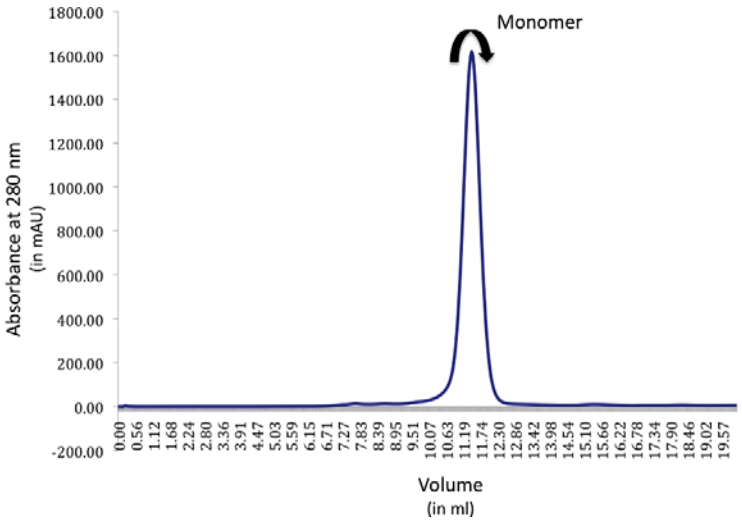


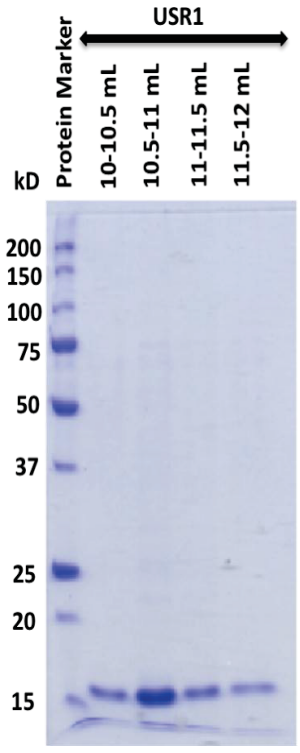
Figure 3-5. SDS-PAGE analysis showing the affinity tag removal of USR12.
Fractions F1 to F8 contain proteins without his-tag. The uncleaved his-tagged proteins including rTEV protease were eluted using 500 mM imidazole in PBS.

The flow through fractions from PBS washing steps (F2 to F8, as shown in Figure 3-5.) were pooled and concentrated using Vivaspin 20 columns (GE Healthcare) with a 5 kDa cutoff, then purified over a preequilibrated (20 mM Tris pH 8.0, 50 mM NaCl) superdex S75 10/300 size exclusion column (GE Healthcare) at 4 °C in batches of 250 µl using a superloop with a flow rate of 0.4 ml per minute and collecting 0.5 ml fractions. A single monomeric peak was observed in solution from the chromatography profile for USR1 and a second peak was observed for the other proteins (DSR1, USR12 and DSR12) that is probably due to the presence of a cysteine in the sequence. However there exists no-equilibrium between monomer and dimeric species in solution. The dimer can be converted to monomer by adding 5 mM DTT as a reducing agent [Cleland, 1964] in Tris-Buffer (20 mM Tris-HCL pH 8.0, 50 mM NaCl) during gel filtration.

[a]

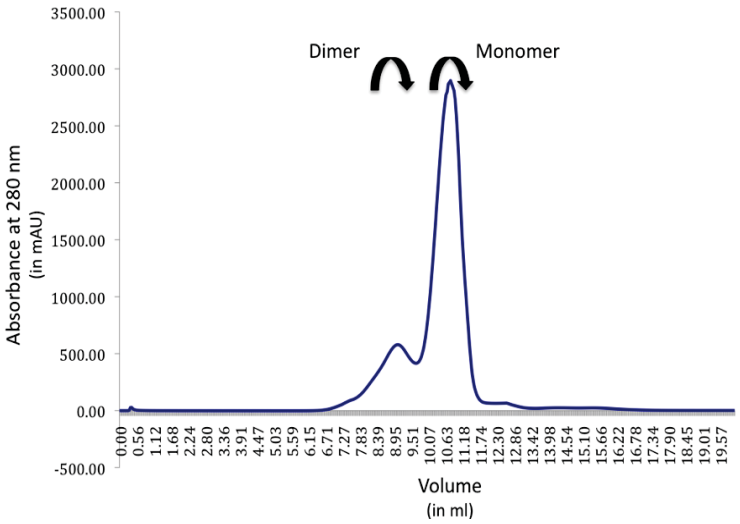


(i) Gel filtration profile of USR1 that elutes as a monomer

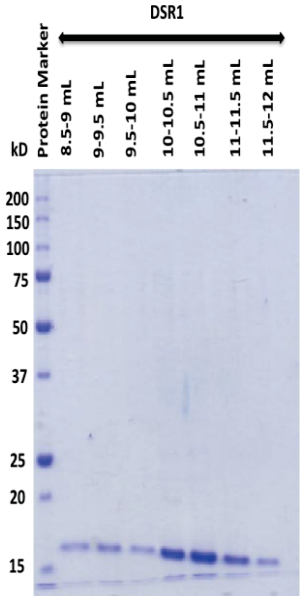


(ii) SDS-PAGE analysis of eluted fractions from gel filtration chromatography.

[b]

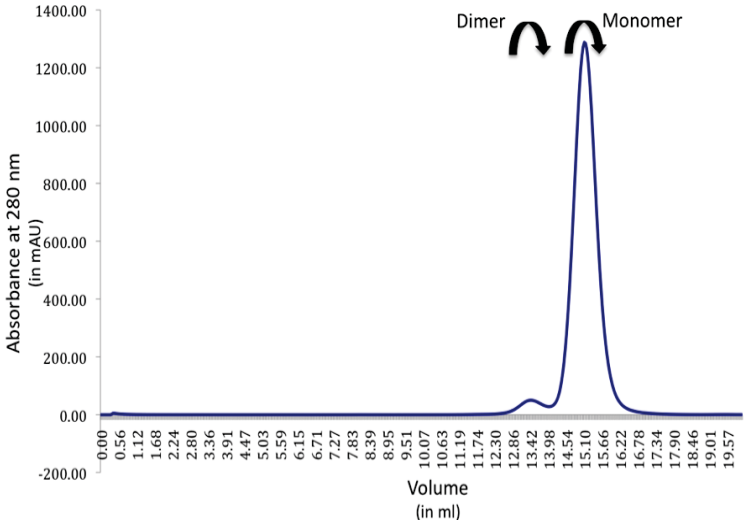


(i) Gel filtration profile of DSR1 that elutes as both monomer and dimer

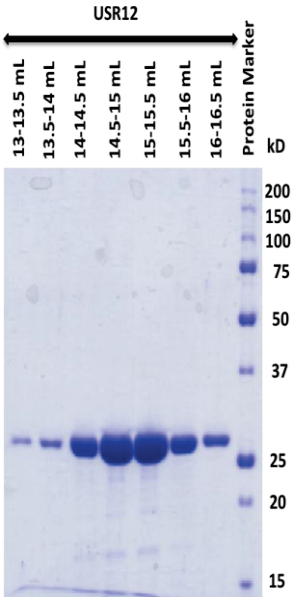


(ii) SDS-PAGE analysis of eluted fractions from gel filtration chromatography.

[c]

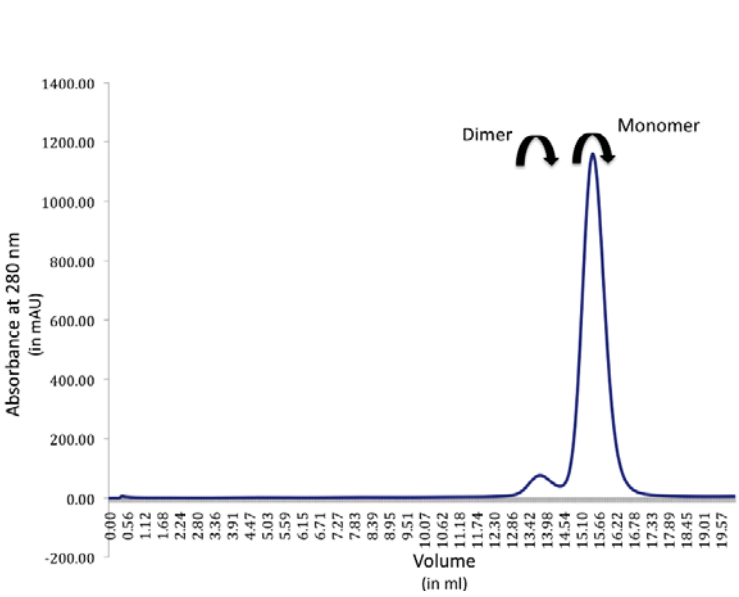


(i) Gel filtration profile of USR12 that elutes as both monomer and dimer.



(ii) SDS-PAGE analysis of eluted fractions from gel filtration chromatography.

[d]



(i) Gel filtration profile of DSR12 that elutes as both monomer and dimer.

(ii) SDS-PAGE analysis of eluted fractions from gel filtration chromatography.

Figure 3-6. Elution profile of the **(i)** recombinant proteins (**[a]** USR1, **[b]** DSR1, **[c]** USR12, **[d]** DSR12) and **(ii)** corresponding SDS-PAGE analysis.

The peak fractions were pooled and concentrated using VivaSpin concentrating columns for further use. Protein concentration was determined as described in chapter 2.2.4.9.

3.6 Summary

cDNA encoding the N-terminal first only and first two spectrin repeats from Dys and Utr was cloned in pPROEx HTb vector and overexpression was achieved in *E. coli* BL21 DE3 cells. Purification involved a two-step process; first using IMAC followed by gel filtration chromatography. Over 95% pure protein was obtained as observed from gel filtration and SDS-PAGE. The concentrated protein samples were stored at -80 °C for further use.

4. Structure determination and analysis

4.1 Introduction

Utrophin (Utr) and Dystrophin (Dys) are large (Utr: 395 kDa and Dys: 427 kDa) multidomain proteins present in humans and other metazoa. The central rod domains, also known as spectrin repeats, comprise ~75% of the whole Utr/Dys protein. Although single and multiple spectrin repeat domain structures of erythroid spectrin [Grum *et al.,* 1999; Kusunoki *et al.,* 2004a; Ipsaro *et al.,* 2009] and α-actinin [Djinovic-Carugo *et al.,* 1999; Ylanne *et al.,* 2001] were solved using NMR [Pascual *et al.,* 1997] and x-ray crystallography, the atomic structures of Utr/Dys spectrin repeat domains remained unclear [Legrand *et al.,* 2011]. The secondary structure of Utr/Dys spectrin repeats has been analysed by circular dichrosium and the dynamics of the conserved tryptophan environment probed by fluorescence spectroscopy [Kahana & Flood*,* 1997; Amann *et al.,* 1999; Rybakova *&* Ervasti, 2005; Saadat *et al.,* 2006; Legradinier *et al.,* 2009b; Mirza *et al.,* 2010; Prochniewicz *et al.,* 2009; Yamashita *et al.,* 2010].

In this chapter, the aim was to determine the structure of N-terminal first spectrin repeat and N-terminal first two-spectrin repeat domains of Utr/Dys using x-ray crystallography. Solving protein structures using x-ray crystallography involved three main approaches, First was to purify a protein which was at least ~95% pure suitable for crystallisation trials, secondly was to obtain diffraction quality crystals and the final step was to determine the crystal structure.

4.2 Protein crystallisation experiments

Recombinant proteins (USR1, DSR1, USR12 and DSR12) were purified (as shown in chapter 3.4) and initial crystal screening was performed using the Sparse matrix screens as listed in Table 2-2. (Experimental procedure) at 4 and 21 °C, employing both hanging drop and sitting drop vapour diffusion systems respectively [McRee *&* Duncan, 1993]. Each drop typically contains 400 nl each of protein and precipitant solution (reservoir solution/mother liquor) pipetted using a Mosquito Nanodrop crystallisation robot (TTP Labtech). The crystallisation conditions were optimised by varying the

concentrations of precipitant and protein solutions, altering the pH of buffers and also by microseeding [Bergfors, 1999]. In some occasions a microbatch method were used for obtaining protein crystals (USR12 and DSR12) [D'Arcy *et al.,* 1996].

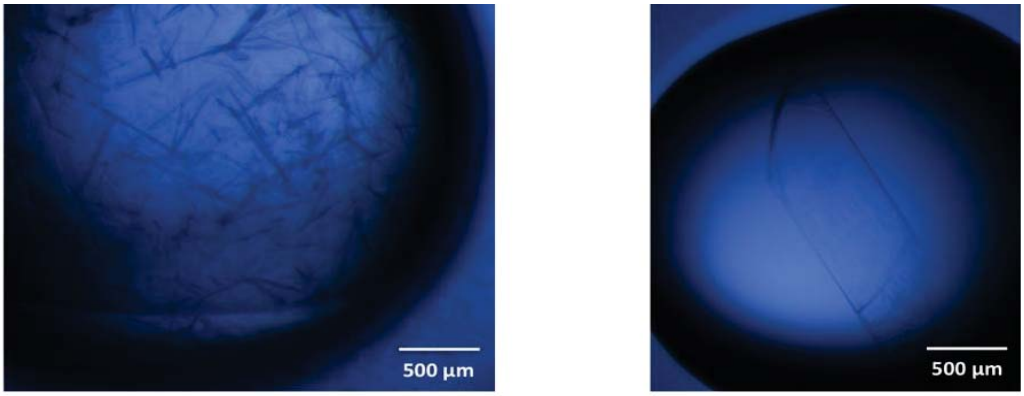
4.2.1 Crystallisation of USR1

4.2.1.1 Initial screening

USR1 protein samples eluted from gel filtration chromatography were concentrated to 7.2 mg/mL using VivaSpin columns with a 5 kDa MWCO. Crystallisation trials were set-up using the sitting drop vapour diffusion technique. Initial crystals were obtained at 4 °C after 2 days from condition A6 (30% Polyethylene glycol 4000, 0.1 M Tris HCl pH 8.5, 0.2 M magnesium chloride) with Hampton crystal screens I & II. The crystals observed from this condition were clustered needles (Figure 4-1[a]). For data collection, a single crystal was required.

4.2.1.2 Optimisation of crystallisation

Crystal optimisation was done based on the initial A6 hit; a 1 µL protein solution drop was mixed with 1 µL of mother liquor/reservoir solution onto the centre of the clear siliconised 22 mm circle glass coverslip (Hampton Research Inc). The coverslips were then sealed with petroleum jelly onto a VDX plate (Hampton Research Inc) for a hanging drop vapour diffusion experiment [McPherson, 1982], and kept at 4 °C. A thin plate-like crystal was observed after 4 days from the optimized crystal screen (30% w/w Polyethylene glycol 4000, 0.1 M Tris HCl pH 8.7, 0.2 M $MgCl_2$) (Figure 4-1[b]). The crystal was briefly soaked into cryoprotectant-containing mother liquor (reservoir solution) and 20% (v/v) glycerol [McFerrin & Snell, 2007] and frozen in a 0.5 mm cryoloop (Hampton Research Inc, USA) by quick immersion in liquid nitrogen and mounted on the R-Axis for data collection at 120 K.



[a] **[b]**

Figure 4-1. USR1 crystals under a light microscope **[a]** Clustered needle-like protein crystals were obtained from the initial screening conditions using Hampton Research crystal screen I&II; **[b]** A single thin plate-like protein crystal was obtained from crystal optimisation for data collection.

4.2.2 Crystallisation of DSR1

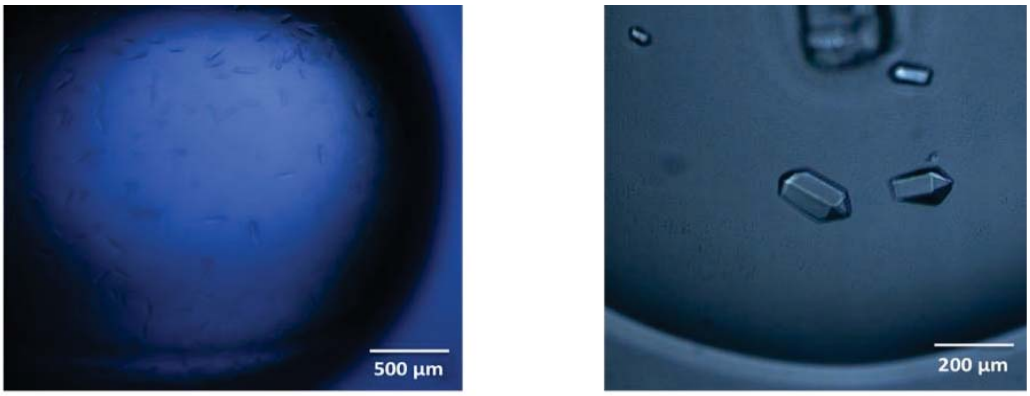
4.2.2.1 Initial screening

DSR1 exists as monomer and dimer in solution as analysed from gel filtration chromatography. Both monomer and dimer protein samples were concentrated to 8.6 mg/mL using VivaSpin columns with 5 kDa MWCO. Crystallisation trials were set-up with the proteins, using the sitting drop vapour diffusion technique in a 96 well INTELLI_PLATE (Art Robbins Instruments) with the Mosquito nano-litre high throughput robot at a 1:1 (v/v) ratio of 0.4 µL protein mixed with 0.4 µL mother liquor. Initial crystals were observed for the dimer species at 21 °C after 4 days, from condition C8 (0.1 M Tris HCl pH 8.5, 2.0 M ammonium sulphate) from the molecular dimensions structural screens I & II. A cube or rod shaped crystals obtained with the above condition diffracted to 3.2 Å.

Microcrystals were obtained from the monomer species (Figure 4-2[a]) but they were not of diffraction quality. Attempts to optimise the crystal conditions were conducted, but unfortunately no protein crystals were observed.

4.2.2.2 Optimisation of crystallisation

Crystal optimisation was done based on the initial hit for dimeric DSR1, a 1 µL protein drop was mixed with a 1 µL of mother liquor/reservoir solution onto the centre of clear siliconised 22 mm circle glass coverslips (Hampton Research Inc). The coverslips were then sealed with petroleum jelly onto a VDX plate (Hampton Research Inc) for hanging drop vapour diffusion, and kept at 21 °C. A cube-shaped crystal was observed after 6 days from the similar crystal condition (0.1 M Tris HCl pH 8.5, 2.0 M $(NH_4)_2SO_4$) to the initial screen that diffracted to 2.2 Å (Figure 4-2[b]). The crystal was briefly soaked into mother liquor containing 15% (v/v) glycerol [McFerrin & Snell, 2007] and frozen in a 0.5 mm cryoloop (Hampton Research Inc, USA) by quick immersion in liquid nitrogen and mounted on the R-Axis beam for data collection at 120 K.



[a] **[b]**

Figure 4-2. DSR1 crystals under a light microscope **[a]** Microcrystals formed from the monomeric DSR1 were not suitable for data collection; **[b]** Diffraction quality protein crystals were obtained from dimeric species.

4.2.3 Crystallisation of USR12

Similar to DSR1, USR12 also existed as monomer and dimer species in solution from gel filtration. Both monomer and dimer species were concentrated separately to 10.2 mg/ml for crystallisation. Protein crystals were obtained after 4-months from the monomeric species using sitting-drop vapour diffusion at 21°C (Figure 4-3[a]) with a 1:1 ratio of protein to mother liquor solution containing 0.2 M magnesium acetate, 20% PEG 8000 and 0.1 M Na cacodylate pH 6.5 from Hampton screens I &

II. The protein crystals obtained after such a long time were likely due to *in-situ* proteolysis in the crystallisation drop [Mandel *et al.,* 2006; Dong *et al.,* 2007]. This was confirmed by dissolving the protein crystal in 5X-SDS loading buffer and analysing using SDS-PAGE as described in chapter 2.2.3.2 (Figure 4-3[b]).

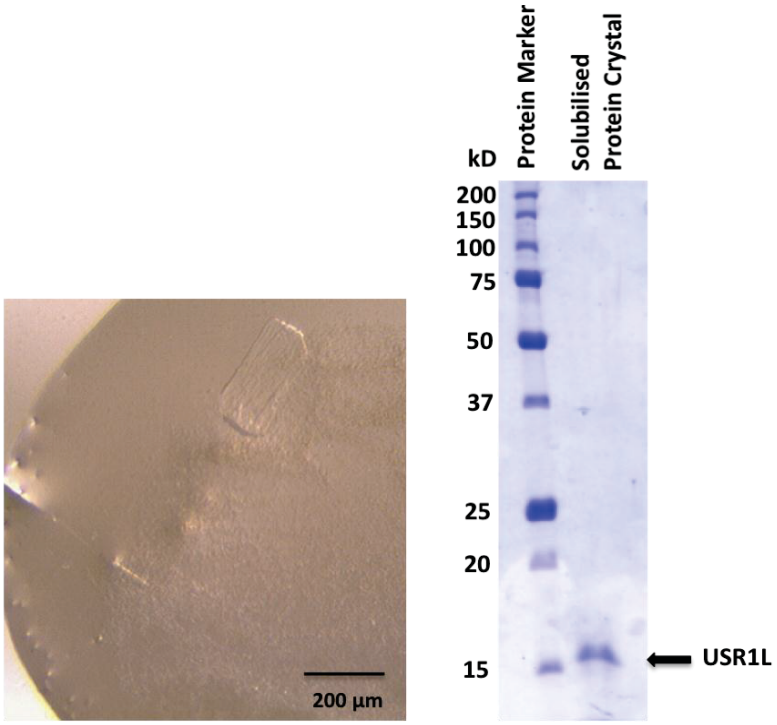


Figure 4-3. [a] Protein crystal appeared after 4-months at 21 °C; **[b]** SDS-PAGE analysis of the protein crystal (USR1L) obtained from *in-situ* proteolysis.

SDS-PAGE analysis of a protein crystal, showed a band at ~15 kDa. This clearly indicated that the crystal could be of a single repeat domain, containing either the first or second repeat domain but not tandem repeats. A crystal was briefly soaked into cryoprotectant containing reservoir solution and 20% glycerol [McFerrin & Snell, 2007], was frozen in a 0.5 mm cryoloop (Hampton Research Inc, USA) by quick immersion in liquid nitrogen and mounted on the beam for data collection at 120 K.

4.2.4 Crystallisation of DSR12

DSR12 also exists in monomer and dimer in solution from gel filtration. No protein crystals were obtained from microbatch and vapour diffusion crystallisation

experiments with the investigated concentrations (5.5, 7.8, 12.5, 28 mg/ml) using crystal screens listed in Table 2-2.

4.3 X-ray diffraction data collection and processing

4.3.1 Structure determination of USR1

4.3.1.1 Data collection and reduction

A single protein crystal was flash frozen in liquid nitrogen and mounted on the goniometer for data collection. X-rays of wavelength (Cu-Kα) of 1.5418 Å were generated using a Rigaku MicroMax-007 rotating anode equipped with Osmic multilayer mono-chromating optics and the diffraction pattern was collected on a RAxis IV^{++} image plate detector placed at a distance of 120 mm from the protein crystal. Initially two diffracted images were collected each at 90° apart using the CrystalClear Software (Rigaku, USA). Initial images were processed using d*TREK [Pflugrath, 1999] and the presumed space group was $P2_1$ corresponding to a monoclinic cell with cell lengths of a= 43.0 Å, b= 58.66 Å, c= 91.45 Å and angles were $\alpha = \beta = \gamma =$ 90°. A total of 260 images were collected based on the strategy recommended by the CrystalClear software and each image was collected after exposing the protein crystal to x-rays for 8 min at a 0.5° oscillation. Raw images collected were indexed and the Laue group and a possible space group were determined to be $P2_12_12_1$ using Pointless [Evans, 2006] and integrated using MOSFLM [Leslie *et al.,* 1990] or alternatively reindexing using SORTMTZ to convert into other possible space groups. Based on the percentage of solvent content in the unit cell calculated from the Matthew's coefficient program [Matthews *et al.,* 1968] it was estimated that there were two molecules in the asymmetric unit with solvent content of 36.6% v/v which falls close to the range typical for protein crystals of ~40 to 60% .

4.3.1.2 Scaling and averaging

The integrated reflections from MOSFLM [Leslie *et al.,* 1990] were scaled and averaged using the SCALA program [Evans, 2006] from the CCP4i software suite to the best possible resolution satisfied from analysis of the values of completeness, multiplicity

and mean intensities (I/sigI) (Table 4-1)). In the case of USR1, 99.9% completeness was the best possible solution observed for a resolution range from 35.92 to 1.95 Å and a value of 3.0 for mean I/sigI at the outer shell, corresponding to a three times signal to noise ratio and hence the best possible resolution cutoff for USR1 to 1.95 Å was taken.

4.3.1.3 Molecular replacement and generating initial density maps

From the x-ray diffraction data, one can determine the Miller indices of a reflection (h,k,l), and the intensities of the reflection I(h,k,l), but we need phases of the reflections in order to calculate the electron density with in the crystal. Phase angles cannot be directly estimated by measuring intensity, this is called phase problem. The phase problem in protein x-ray crystallography can be solved by several techniques: Multiple isomorphous replacement (MIR), multiple wavelength anomalous dispersion (MAD), single wavelength anomalous dispersion (SAD) and molecular replacement (MR) methods. In the present study, we used molecular replacement (MR) [Read, 2001] method to solve the phase problem for all the three proteins.

To solve the structure using molecular replacement method, a homologous protein whose structure has already been determined was used as a search model to determine the structure of the target protein. In the case of USR1, the structure was solved by molecular replacement with PHASER [McCoy *et al.,* 2007] where one molecule of spectrin repeat 2 from α-actinin (PDBid: 1QUU) [Djinovi-Carugo *et al.,* 1999] with sequence identity of 23% (determined by BLAST alignment) was used as a search model. The rotation and translation search using PHASER derived two solution molecules in the asymmetric unit [McCoy *et al.,* 2005] corresponding to a correlation coefficient of >0.6 and an initial R-factor of 40.0%, with sensible crystal packing and good quality electron density maps.

4.3.1.4 Model building and Refinement

The electron density maps calculated were visualized in COOT [Emsley *et al.,* 2004]. The structure model building was a repeating process in which the protein atoms were

fitted manually in the electron density map followed by few cycles of refinement using REFMAC [Murshudov *et al.,* 1997]. After each refinement, new structure factors were calculated with the calculated new phases combined with the original intensities to calculate an improved electron-density map. After each round of manual rebuilding, the structures was subjected to further translation libration screw (TLS) [Winn *et al.,* 2001], non-crystallographic symmetry (NCS) [Kleywegt*,* 1996] restrained refinement against a maximum likelihood target using REFMAC and PHENIX [Adams *et al.,* 2002]. When the map improved, water molecules were added automatically in COOT at 3 sigma positive density as observed in the weighted difference density map ($F_{obs} – F_{calc}$). After refinement, the position of each water molecule was checked in COOT at the corresponding electron-density and the suspected water molecules were removed.

The quality of the final model was evaluated from the R-factor, which is a value of agreement between the amplitudes from observed structure factor (Fobs) and those calculated from the model (Fcalc). To prevent the structure from being over-refined, an Rfree value was calculated during the refinement process using 5 % of the data selected randomly that were omitted from the refinement. The final model for USR1 was refined with REFMAC to a R-factor of 0.20 and an R-free of 0.23.

4.3.1.5 Structure and geometry parameters of USR1

The structure of the Utr N-terminal spectrin repeat showing a triple helical coiled coil structure was solved using x-ray crystallographic methods to 1.95 Å. There are two molecules in the asymmetric unit of the $P2_12_12_1$ space group cell and the final model contains 226 amino acids (310 – 424 in molecule-1 and 308 – 424 in molecule-2), 206 water molecules and a suspected magnesium ion from the crystallisation condition was present. Model geometry was excellent, with 98.7% (223) of amino acids in the preferred region of the Ramachandran plot [Ramachandran *et al.,* 1963], only 1.3% (3) are in the additionally allowed regions and none are in disallowed regions calculated using the MOLPROBITY server [Davis *et al.,* 2004] (Figure 4-4).

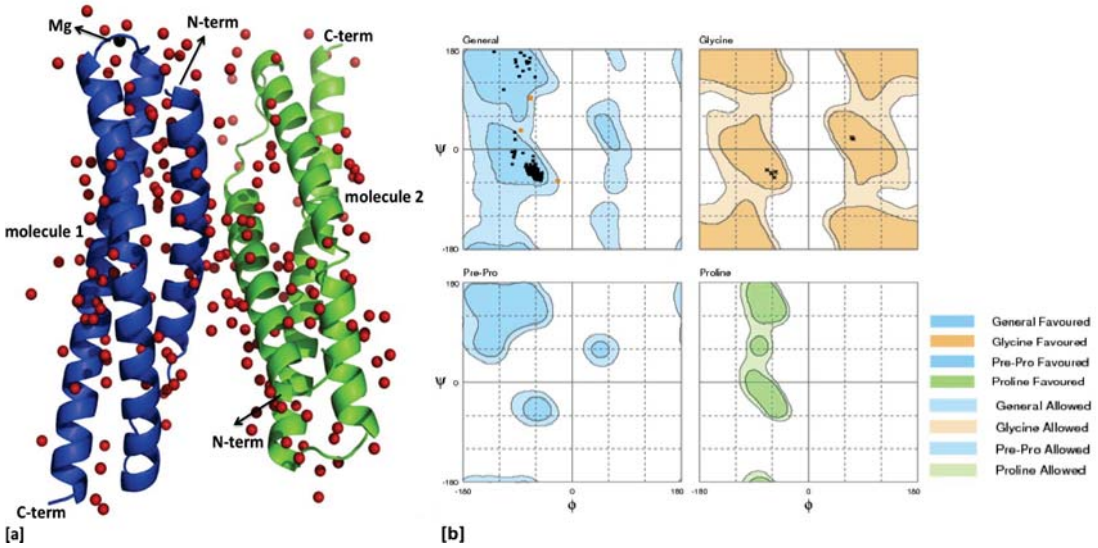


Figure 4-4. [a] Cartoon representation of the N-terminal first spectrin repeat of Utr showing the two molecules in the asymmetric unit (molecule 1 in blue and molecule 2 in green). Water molecules (red) and magnesium ion (Mg^{2+}) (black) are in spheres. **[b]** Structural validation of backbone geometry: Ramachandran plot to visualize backbone dihedral phi (ϕ) and psi (Ψ) angles using MOLPROBITY [Davis *et al.,* 2004] and the figures were made using RAMPAGE server [Lovell *et al.,* 2002].

4.3.2 Structure determination of DSR1

4.3.2.1 Data collection and reduction

A single DSR1 protein crystal was flash frozen in liquid nitrogen and mounted on the goniometer for x-ray diffraction data collection. The x-ray diffraction pattern was collected on a RAxis IV^{++} image plate detector placed at a distance of 120 mm from the protein crystal. Initially two diffracted images were collected each at 90° apart using the CrystalClear control software (Rigaku, USA). Initial images were processed using d*TREK [Pflugrath, 1999] and the space group were identified to be P3 corresponding to a trigonal cell. The unit cell lengths were a= b= 76.04 Å, c= 66.49 Å and angles were $\alpha = \beta = 90°$, $\gamma = 120°$. A total of 240 images were then collected based on the strategy recommended by the CrystalClear software and each image was collected after exposing the protein crystal to x-rays for 10 min at every 0.5° oscillation. Raw images collected were indexed and the Laue group and possible space group were determined to be $P3_221$ based on reflection intensities and systematic absences using the POINTLESS program and then the intensities were integrated using MOSFLM [Leslie *et al.,* 1990]. The percentage of solvent content, was calculated using the Mathew's program [Matthews *et al.,* 1968], in the CCP4 suite, as ~38%.

4.3.2.2 Scaling and averaging

Data scaling and averaging was performed as described for USR1. 98.4% completeness was the best possible coverage observed for a resolution range from 33.24 to 2.3 Å, with an R_{merge} of 0.605 (60%) corresponding to sufficient redundancy with I/sigI = 3.1 in the highest resolution shell, and hence the best possible resolution cutoff for DSR1 to 2.3 Å was taken into consideration.

4.3.2.3 Molecular replacement and generating initial density maps

The structure was solved by molecular replacement with PHASER [McCoy *et al.,* 2007] where one molecule from USR1 was used as a search model. The rotation and translation search using PHASER derived two molecules in the asymmetric unit.

4.3.2.4 Model building and Refinement

The structure was manually rebuilt in COOT [Emsley *et al.,* 2004] using sequence information and calculated electron density maps. After each round of manual rebuilding, the structure was refined using REFMAC5 [Murshudov *et al.,* 1997] reaching the final R and Rfree of 0.19 and 0.26.

Solvent molecules (water) were automatically added in COOT [Emsley *et al.,* 2004] to the spherical electron densities at chemically sensible positions. After refinement each water molecule was checked thoroughly and if no density was observed it was removed in COOT.

4.3.2.4 Structure and geometry parameters of DSR1

The structure of the Dys N-terminal first spectrin repeat was solved using x-ray crystallography to 2.3 Å using the molecular replacement method to a final R-factor of 0.19 and an R-free of 0.26. There are two molecules in the asymmetric unit of the $P3_221$ space group. The cysteine residue in position 433 of molecule-1 forms a disulphide bond with the 433 cysteine residue of molecule-2, the disulphide bonds could be reduced in solution by adding 5 mM DTT as shown in gel filtration analysis, This suggested that DSR1 crystallisation was aided by the formation of the disulphide bond between the two molecules in asymmetric unit that facilitated the crystal contacts as seen previously in other cases [Banatao *et al.,* 2006].

The final model contained 228 amino acids (338 – 453 in molecule-1 and 339 – 452 in molecule-2), and 34 water molecules. Model geometry was excellent, with 98.7% (225) of amino acids in the preferred region of the Ramanachandran plot

[Ramachandran *et al.,* 1963], 1.3% (3) are in the allowed regions and none are in disallowed regions, as calculated using the MOLPROBITY server [Davis *et al.,* 2004] (Figure 4-5).

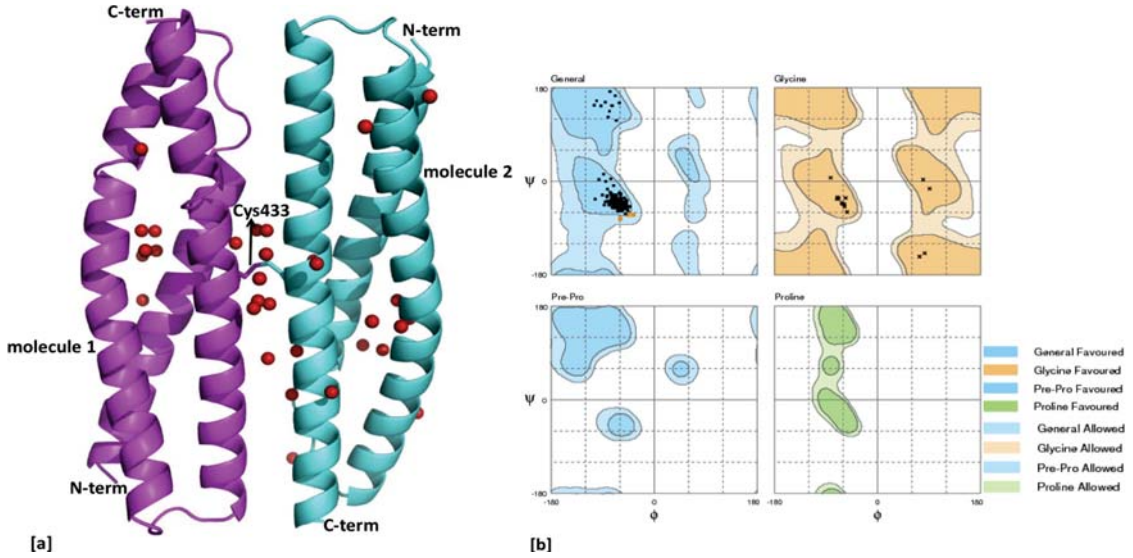


Figure 4-5. [a] Cartoon representation of the Dys N-terminal first spectrin repeat showing the two molecules in the asymmetric unit (molecule 1 in magenta and molecule 2 in cyan). Water molecules (red) are shown as spheres and the disulphide bond between the two molecules is shown in stick representation. **[b]** Structural validation of the protein structure backbone geometry: A Ramachandran plot to visualize backbone dihedral phi (ϕ) and psi (ψ) angles using MOLPROBITY [Davis *et al.,* 2004] with the figures were made using RAMPAGE server [Lovell *et al.,* 2002].

4.3.3 Structure determination of USR1L

4.3.3.1 Data collection and reduction

Data collection was performed similarly to USR1. Initial images were processed using d*TREK [Pflugrath, 1999]. The unit cell lengths were a= 42.74 Å, b= 58.02 Å, c= 91.27 Å and angles were $\alpha = \beta = \gamma =$ 90°. The mosaicity was 1.55°, higher in comparison to USR1 (0.91°) and DSR1 (0.64°). A total of 240 images were collected based on the strategy recommended by the CrystalClear software and each image was collected after exposing the protein crystal to x-rays for 10 min at 0.5° oscillation. Raw images collected were indexed and the Laue group and a possible space group were determined to be $P2_12_12_1$ using POINTLESS and integrated using MOSFLM [Leslie *et al.,* 1990]. The percentage of solvent content was calculated using Mathew's program [Matthews *et al.,* 1968] in the CCP4 suite to be ~34% v/v.

4.3.3.2 Scaling and averaging

Data scaling and averaging were performed as described for USR1. 98.4% completeness was the best possible solution observed for a resolution range from 35.87 to 2.0 Å, with a mean I/sigI of 2.3 for and R_{merge} of 0.72 at the outer shell; hence the best possible resolution cutoff for USR1L to 2.0 Å was taken into consideration.

4.3.3.3 Molecular replacement and generating initial density maps

The structure was solved by molecular replacement with PHASER [McCoy *et al.,* 2007] where one molecule from USR1 was used as a search model. The rotation and translation search using PHASER derived two molecules in the asymmetric unit.

4.3.3.4 Model building and Refinement

The structure was manually rebuilt in COOT [Emsley *et al.,* 2004] using sequence information for the C-terminus and the obtained electron density maps. After each round of manual rebuilding, the structure was refined using REFMAC5 [Murshudov *et al.,* 1997] reaching the final R and Rfree of 0.20 and 0.26.

Solvent molecules (water) were automatically added in COOT [Emsley *et al.,* 2004] to the spherical electron densities in the maps at chemically sensible positions. After refinement each water molecule was checked thoroughly and if modelled where no density was observed it was then removed.

4.3.3.5 Structure and geometry parameters of USR1L

The structure of the Utr N-terminal spectrin repeat six amino acids longer at its C-terminus was determined to 2.0 Å. This protein formed as a result of *in-situ* proteolysis in the crystal drop from USR12 crystallisation experiments. The final model containing 244 amino acids (308 – 430 in molecule-1 and molecule-2), 51 water molecules and a suspected magnesium ion from the crystallisation conditions. The geometry of the final protein structure was excellent, with 99.2% (242) of amino acids in the preferred region of the Ramanachandran plot, 0.8% (2) are in the allowed regions and none are in disallowed regions, as calculated by the MOLPROBITY server [Davis *et al.,* 2004] (Figure 4-6).

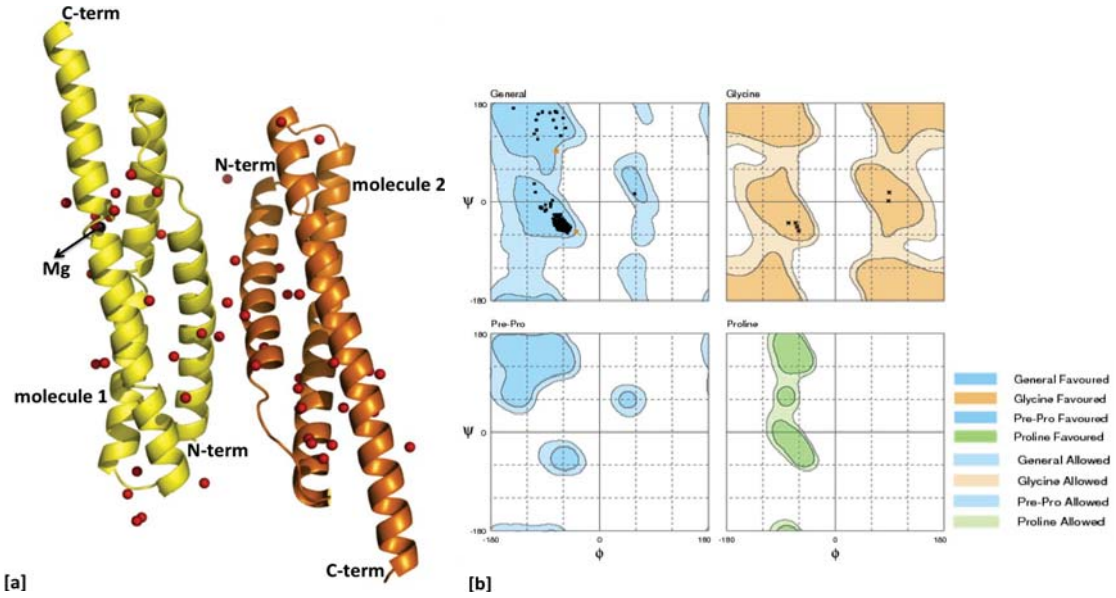


Figure 4-6. [a] Cartoon representation of the six amino acids longer version of N-terminal first spectrin repeat of Utr showing the two molecules in the asymmetric unit (molecule 1 in yellow and molecule 2 in orange), water molecules (red) and magnesium ion (Mg^{2+}) (black) are in spheres. **[b]** Structural validation of backbone geometry: Ramachandran plot to visualize backbone dihedral phi (ϕ) and psi (Ψ) angles using MOLPROBITY [Davis *et al.,* 2004] and the figures were made using RAMPAGE server [Lovell *et al.,* 2002].

Table 4-1. x-ray data collection and refinement statistics

Data Collection			
Protein (Molecular weight, kDa)	USR1 (13.66)	USR1L (14.26)	DSR1 (13.63)
Space group	$P2_12_12_1$	$P2_12_12_1$	$P3_221$
Cell dimensions	a= 43.00 Å b= 58.66 Å c= 91.45 Å α= β= γ= 90°	a= 42.74 Å b= 58.02 Å c= 91.27 Å α= β= γ= 90°	a= b= 76.04 Å c= 66.49 Å α= β= 90° γ= 120°
Resolution range (Outer shell) (Å)	35.92 – 1.95 (2.06 – 1.95)	35.87 – 2.0 (2.11 – 2.0)	33.24 – 2.3 (2,42 – 2.3)
Mosaicity (°)	0.91	1.55	0.64
Unique reflections	17305	15608	9999
Multiplicity (Outer shell)	4.0 (3.9)	5.7 (5.5)	7.3 (7.4)
Completeness (Outer shell) (%)	99.9 (100)	98.4 (97.3)	98.4 (97.4)
R_{merge} (Outer shell)	0.085 (0.406)	0.141 (0.720)	0.113 (0.605)
Mean I/sigI (Outer shell)	10.2 (3.0)	7.6 (2.3)	11.7 (3.1)
Wavelength, Cu Kα (Å)	1.5418	1.5418	1.5418
Mathews Coeff. (% solvent)	36.64	34.83	38.01

$R_{merge}=\Sigma|I_i-\langle I_i\rangle|/\Sigma I_i$ where I_i is the intensity of a single reflection and $\langle I_i\rangle$ is the mean intensity of that reflection

Refinement statistics			
Number of reflections	16320	14759	9035
Number of residues	439	297	264
Number of protein atoms	1857	1988	1866
Number of solvent molecules	206	51	34
R-$factor_{work}$ (%)	20.02	20.06	19.32
R-$factor_{free}$ (%)	23.54	26.81	25.97
R.m.s deviations Bond length (Å) Bond angle (^{o})	 0.015 1.236	 0.019 1.590	 0.017 1.606
Average B-factor for all atoms ($Å^2$)	37.0	37.4	43.3
Wilson B-factor ($Å^2$)	23.05	26.05	35.85
Ramachandran plot (%) Residues in most favoured region Number of outliers	 98.7 0.0	 99.2 0.0	 98.7 0.0
PDB id	3UUL	3UUM	3UUN

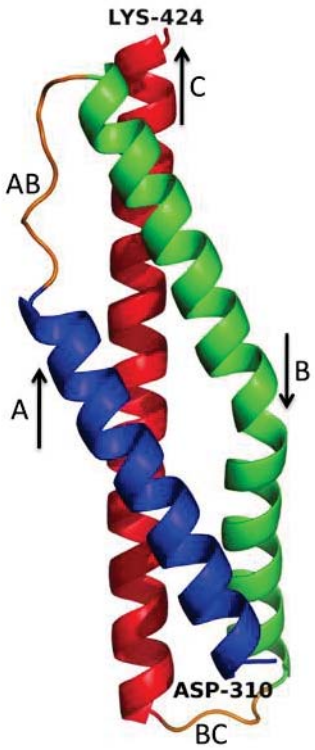
$R_{working}=\Sigma|F_o-F_c|/\Sigma F_o$ R_{free} is the R factor calculated for the cross-validated test set of reflections.

4.3.4 Statistical validation of x-ray crystal structures

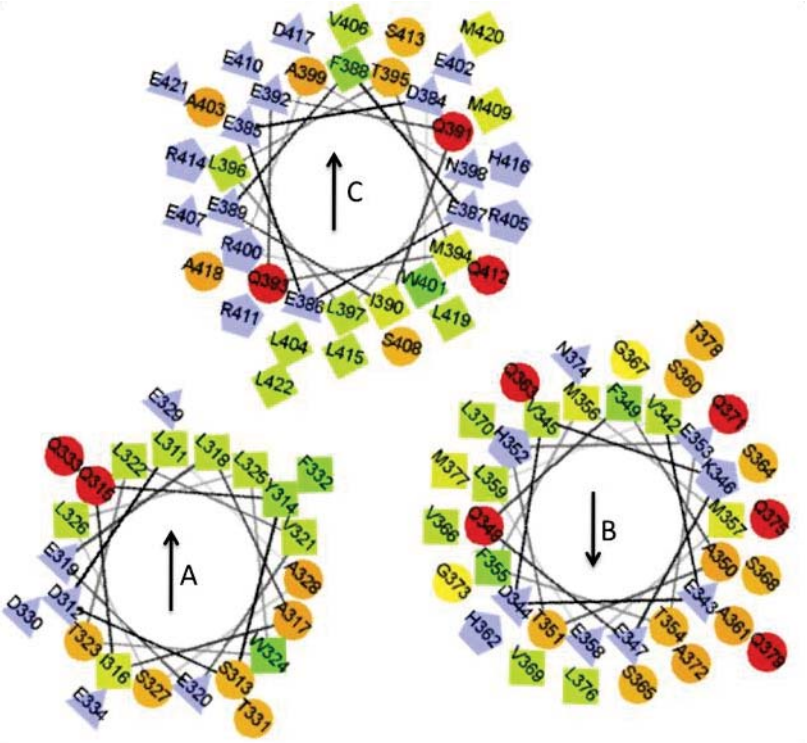
X-ray crystallography is a powerful method for studying protein structure at high resolution. There are number of parameters required to asses the quality of the structure such as good quality x-ray data (data collection strategy and image quality) and proper refinement and validation [Garman, 1999 & 2003]. There are no outliers found in USR1, DSR1 and USR1L (Table 4-1) from MOLPROBITY [Davis *et al.,* 2004]. However, if an outlier exists that doesn't necessarily mean that the structure is wrong, there might be possible unusual conformations associated with biological function [Petock, 2003; Videau, 2004]. The x-ray crystallography data collection and refinement statistics of USR1, USR1L and DSR1 are listed in Table 4-1.

4.4 Structure of Utr/Dys SR1 (USR1/DSR1)

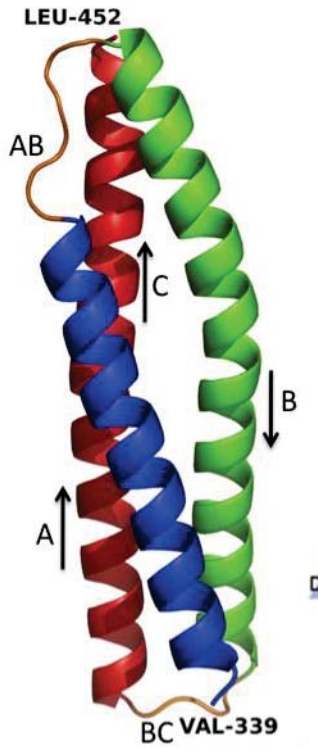
The crystal structures of the N-terminal first spectrin repeat domain from Utr (USR1) and Dys (DSR1) have been determined using x-ray crystallography. The structures show, a triple-helical bundle folded into a left-handed coiled-coil ~55 Å long and ~20 Å in diameter. Each repeat domain consists of three helices- 'A' (residues from 310-334 for USR1, 240-364 for DSR1), -'B' (342-379 for USR1, 372-409 for DSR1) -'C' (384-423 for USR1, 414-453 for DSR1) and are separated by two loops 'AB' and 'BC'. Helices 'A' and 'C' are straight and Helix 'B' has a significant kink at the middle point with the loss of canonical hydrogen bonding between NH and the C=O group of the amino acid 4-residues earlier (n-4) (361-364 for USR1 and 391-394 for DSR1), with the introduction of a distorted helical region. Helix 'A' consists of 24- residues, helix 'B' contains 38-residues and helix 'C' with 39- residues. The loop 'AB' contains 7- residues which is slightly longer when compared to the loop 'BC' with 4- residues. A 6- amino acid longer C-terminus structure of USR1 (USR1L) was also determined as shown in Figure 4-7[e], clearly suggesting that helix 'C' is continuous and nested into helix A' of the second repeat domain (USR2) as observed for spectrin [Grum *et al.,* 1999] and α-actinin [Ylanne *et al.,* 2001].



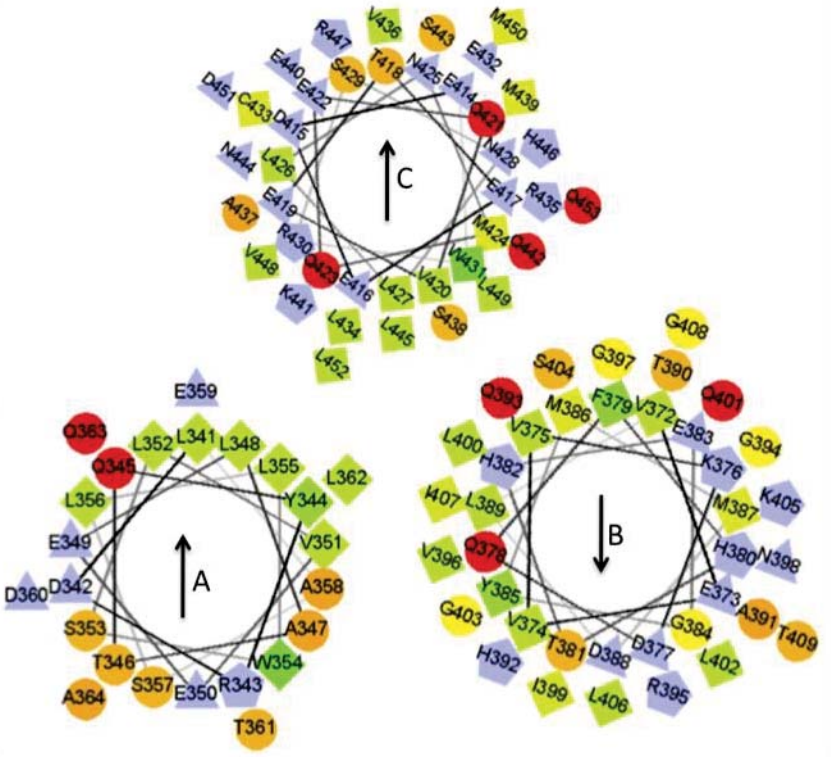
[a]



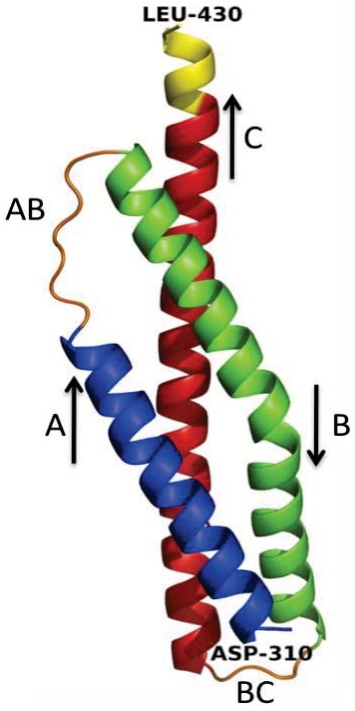
[b]



[c]



[d]



[e]

Figure 4-7. [a] Cartoon representation of USR1, Helix-A In blue, Helix-B in green, Helix-C in red, AB and BC- Loops are in orange. **[b]** Helical projection of USR1 looking down from C-terminus showing buried hydrophobic residues at 'a' and 'd' heptad positions (made using http://rzlab.ucr.edu/scripts/wheel/wheel.cgi). **[c]** Cartoon representation of DSR1, Helix-A In blue, Helix-B in green, Helix-C in red, AB and BC- Loops are in orange. **[d]** Helical projection of DSR1 looking down from C-terminus, showing buried hydrophobic residues at 'a' and 'd' position from the heptad repeat. **[e]** Cartoon representation of USR1L, Helix-A In blue, Helix-B in green, Helix-C in red, AB and BC-Loops are in orange, and the 6-amino acid extension at the C-terminus is coloured in yellow.

The three helices in each repeat domain show a characteristic heptad pattern found in α-helical coiled-coil structures, with positions labelled by the letter 'a' to 'g' [Koenig & Kunkel, 1990; Parry *et al.,* 1992], where positions 'a' and 'd' along the heptad pattern consist of hydrophobic residues, position 'e' and 'g' are charged residues, and the remaining positions are polar residues as shown in Figure 4-7[b], [d]. The interaction between the helices is predominantly through hydrophobic residues at 'a' and 'd' positions facing inwards from the solvent (Figure 4-7[b], 1[d]).

In the spectrin repeat domain helix ‘A’ and ‘B’ are antiparallel with hydrophobic residues in nature at position ‘d’ of one helix at the same level facing position ‘a’ of the other helix packing laterally. This is called ridge-to-ridge packing typically found in coiled-coil structures [Banner *et al.,* 1987]. Positions ‘a’ and ‘d’ of helix ‘C’ were packed roughly in the middle between the hydrophobic interactions between helix ‘A’ and ‘B’; this kind of hydrophobic interactions stabilises each spectrin repeat domain [Yan *et al.,* 1993]. Apart from the hydrophobic residues, charged residues at ‘e’ and ‘g’ position also contribute to the stability of the triple-helical bundle as also shown for repeat 16 from chicken brain α-spectrin [Grum *et al.,* 1999] and the remaining residues in positions ‘b’ and ‘c’ are more likely to be polar in nature and facing towards the solvent.

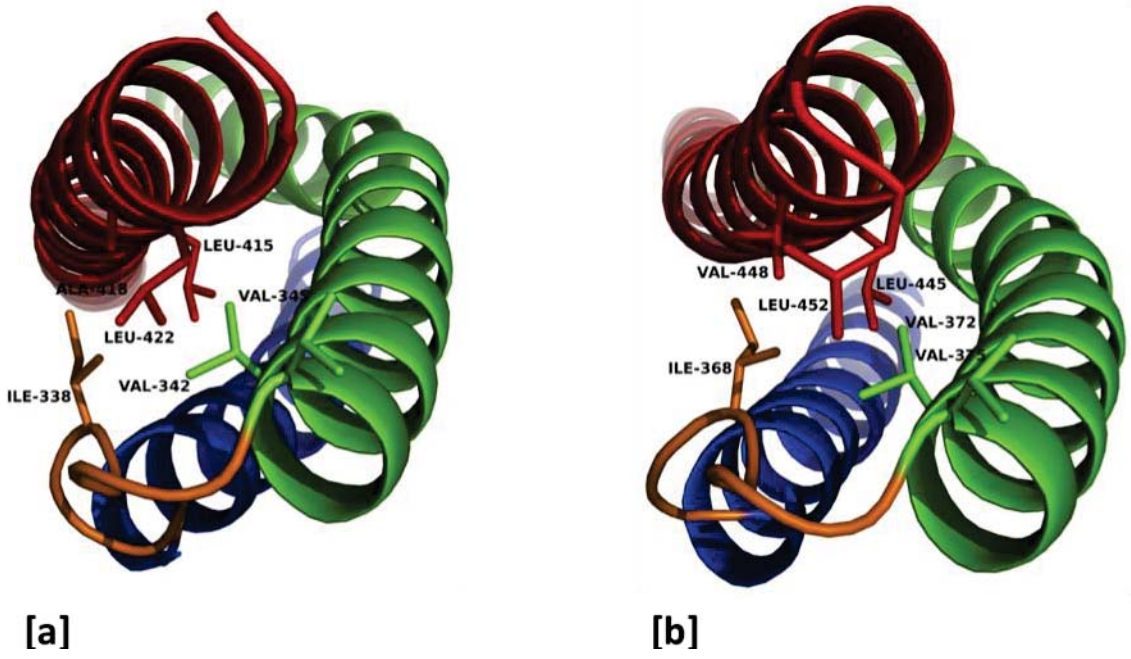


Figure 4-8. Cartoon representation of **[a]** USR1, **[b]** DSR1 looking down from the C-terminus, showing hydrophobic side chains of helix-A in blue, helix-B in green, helix-C in red and loop-AB in orange.

Based on protein sequence analysis, each spectrin repeat domain from Utr and Dys contains 109 amino acids and the heptad pattern of the long helix ‘C’ terminated with histidine residue at 416 for Utr and 446 for Dys [Koenig & Kunkel, 1990; Winder *et al.,* 1995]. The domain structures determined here using x-ray crystallography (USR1 and

DSR1) are 6 amino acids longer at the C-terminus (115 amino acids), a continuity of the heptad repeat pattern until a leucine residue at 422 for USR1 and 452 for DSR1 in the 'a' heptad position. However, attempts to crystallise a 109 amino acid long construct for the Utr N-terminal first spectrin were unsuccessful *(data not shown)*. The 6-amino acid extension at C-terminus is necessary from a structural perspective as these amino acids form interactions with helix 'B' and the loop 'AB'. Interactions due to the extension of 6-amino acids includes residues at 418 (A in USR1)/448 (V in DSR1) at position 'd' and 422 (L in USR1)/452 (L in DSR1) at position 'a' involved in hydrophobic interactions with residues at 342 (V in USR1)/ 372 (V in DSR1) at position 'a', 345 (V in USR1)/375 (V in DSR1) at position 'd' and residue 338 (I in USR1)/368 (I in DSR1) in loop 'AB' as shown in Figure 4-8.

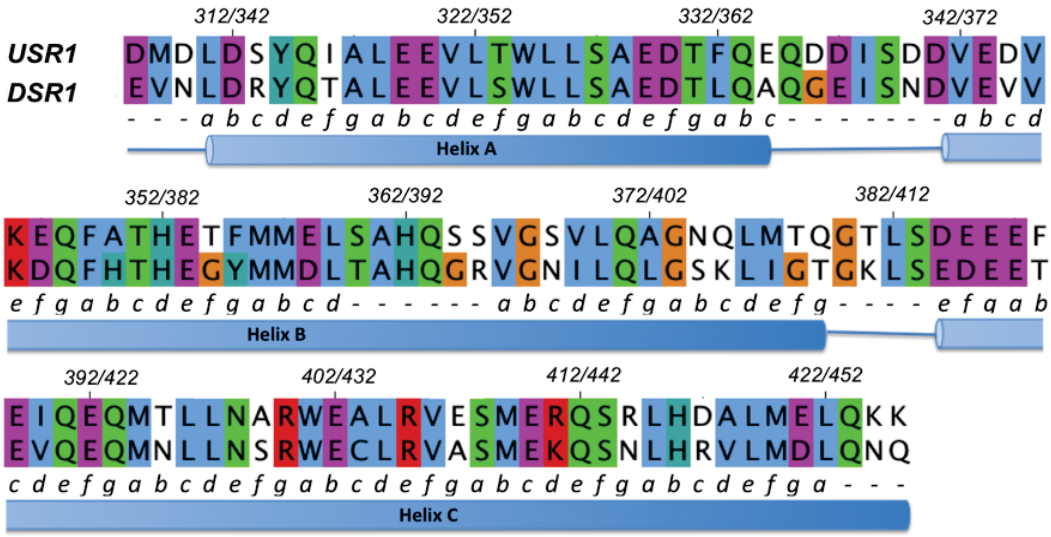


Figure 4-9. Sequence alignment of the N-terminal first spectrin repeat domain from Utr (USR1) and Dys (DSR1) using ClustalW [Thompson *et al.,* 1994]. The colour code is based on conserverd residues and displayed in JalView [Clamp *et al.,* 2004; Waterhouse *et al.,* 2009]. Heptad phasing, helices and loops are shown according to the structure.

Analysis of a structure based sequence alignment (Figure 4-9) revealed that the sequence phasing at the linker region between N-terminal first two spectrin repeats from Utr and Dys was assigned to the first spectrin repeat, as longer repeats are required for stability than those suggested from sequence alignment. This was also

determined from multiple spectrin repeat domain structures from spectrin [Grum *et al.,* 1999; Kusunoki *et al.,* 2004b].

4.4.1 Kink region of USR1 and DSR1

A significant helical kink was observed in both USR1 and DSR1 (Figure 4-7[a] and [c]). This bending in helix 'B' was due to the bulky cyclic residues Trp324 (helix A), His362 (helix B) and Trp401 (helix C) in USR1, these residues are also conserved in DSR1 (Trp354 (helix A), His392 (helix B) and Trp431 (helix C)). The bulky sidechains are wider than the hydrophobic core of the triple helical bundle, which causes helix 'B' to bend. The two Trps at the kink region are conserved and are also found in spectrin repeat domains from spectrin [Kusunoki *et al.,* 2004a], α-actinin [Djinovic-Carugo *et al.,* 1999] and plectin [Sonnenberg *et al.,* 2007]. In the case of the plakin domain from plectin, due to the absence of the bulky groups in repeat-2, there was no kink observed in helix 'B'. As a result helix 'B' is straight and packs more directly against helix 'A' and 'C' [Sonnenberg *et al.,* 2007]. Previous studies have shown that mutating the conserved Trp residues in helix A and C in repeat-16 from spectrin resulted in lower melting temperatures and denaturation under lower concentrations of urea from CD analysis, as a result of destabilising the repeat domain [Pantazatos *&* MacDonald, 1997]. From the sequence alignment [Winder *et al* 1995], His362 at position 'd' in helix 'B' from USR1 (His392 in DSR1) was not conserved in the rest of the repeat domains from Utr and Dys though other repeats do contain Pro and Gly, known helical breakers, that are also likely to bend helix 'B' due to the lack of an amide hydrogen (NH) backbone in proline (Pro) to form a hydrogen bond and also having less degrees of rotation along the ψ and ϕ angles, and glycine (Gly) has no side chain so the structure is less constrained and both these amino acids do not follow a typical Ramachandran plot.

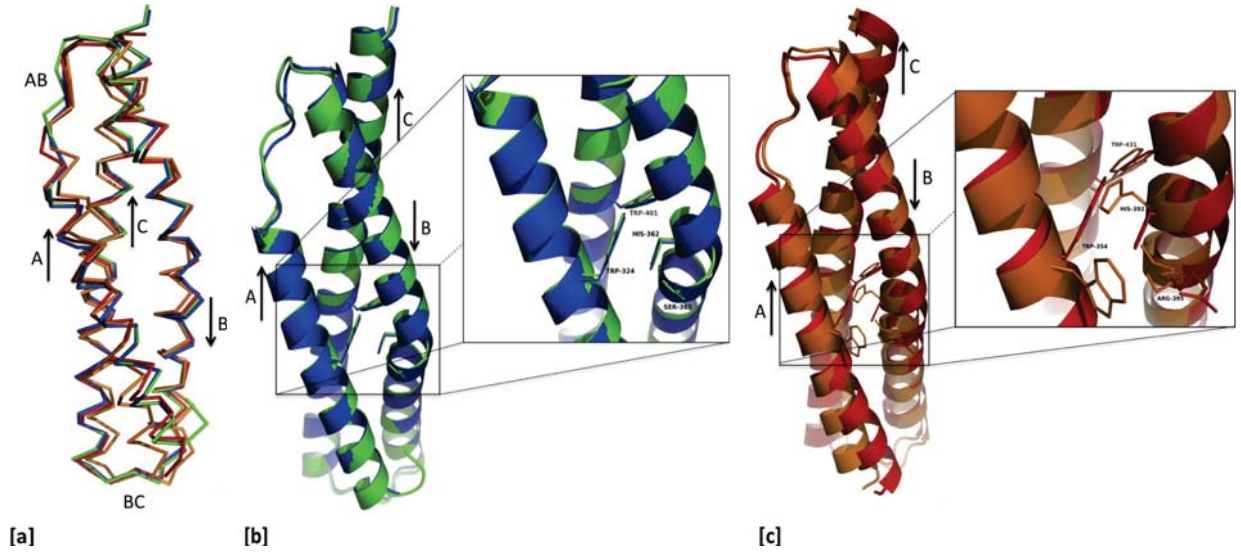
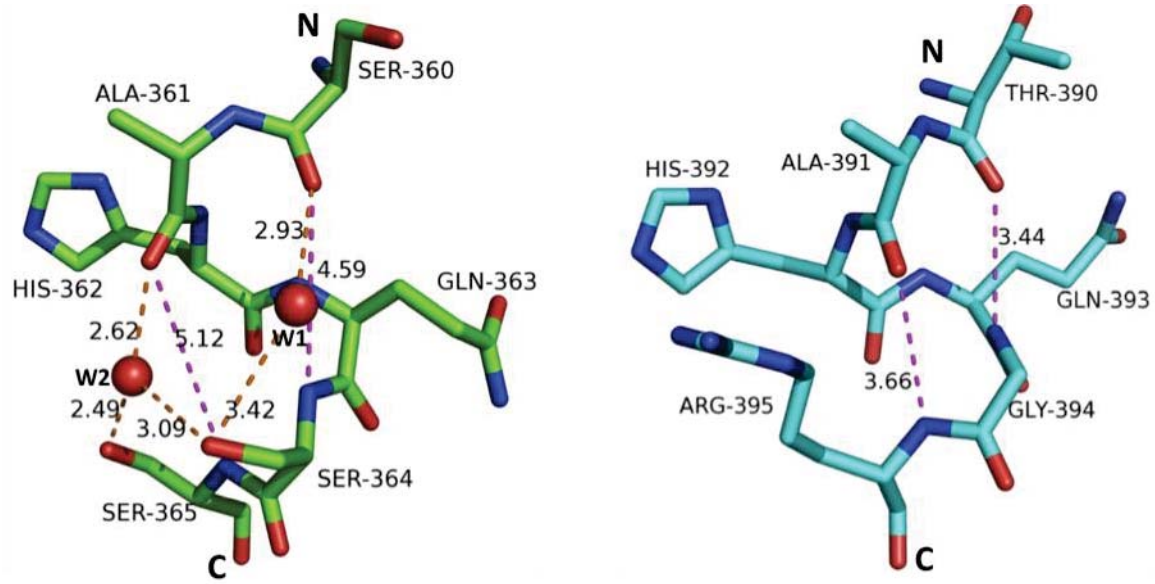


Figure 4-10. Cα- superposition of **[a]** two molecules in asymmetric unit of USR1 (molecule A in blue, molecule B in green) DSR1 (molecule A in red and molecule B in orange) represented in sticks. **[b]** Cartoon representation of two molecules in the asymmetric unit of USR1, highlighting the residues at the kink region and side chains shown as sticks. **[c]** Cartoon representation of the two molecules in the asymmetric unit of DSR1, highlighting the residues at the kink region and side chains shown as sticks.

DaliLite (http://www.ebi.ac.uk/Tools/dalilite/) [Holm & Park, 2000] pair wise structure comparison of the two molecules in the asymmetric unit for USR1 showed a backbone Cα- root mean square deviation (RMSD) of 0.5 Å (115 Cα) with the only difference existing in loop 'AB'. In comparison the two molecules in the asymmetric unit for DSR1 showed a backbone Cα- root rmsd of 1.5 Å (114 Cα), greater compared to the rmsd of USR1. The two molecules in the asymmetric unit for USR1 have similar stacking interactions (π-π/ aromatic-aromatic interactions) for Trp324 in the middle of helix 'A' with His362 in helix 'B' and Trp401 in helix 'C'. Although one molecule (molecule A) for DSR1 at Trp354 in helix 'A' shows similar stacking interactions with His392 in helix 'B' and Trp431 in helix 'C' to USR1, the other molecule (molecule B) at Trp354 has shown a cation-π interaction with Arg395 in helix 'B' as shown in Figure 4-10[c].

4.4.2 Loss of canonical intrahelical hydrogen bonds

Due to the presence of the kink in helix B, the backbone hydrogen bonds (≤ 3.5 Å) were lost at this position. It was observed that two canonical intrahelical hydrogen bonds were lost for USR1 in helix 'B' between Ser360 and Ser364, Ala361 and Ser365. The distance between the backbone oxygen of Ser360 and backbone nitrogen of Ser364 is 4.59 Å and the distance between the backbone oxygen of Ala361 and the backbone nitrogen of Ser365 is 5.12 Å (Figure 4-11[a]). Interestingly, the backbone oxygen of Ser360 forms a hydrogen bond with water molecule 'W1' at a distance of 2.93 Å and the backbone oxygen of Ala361 forms a hydrogen bond with water molecules 'W2' at a distance of 2.62 Å, seen in both the molecules in asymmetric unit. It was also observed that the side chain oxygens of Ser364 and Ser365 were also forming hydrogen bonds with water molecules at a distance of 3.42 Å with 'W1' and 2.49 Å with 'W2'. The presence of these water molecules may contribute the stability of USR1. There were no water molecules observed at the kink region, close to the residues for DSR1 at helix 'B' for both the molecules in asymmetric unit. The intrahelical hydrogen bonding distance between backbone oxygen of Thr390 and backbone nitrogen of Gly394 is 3.44 Å and the backbone oxygen of Ala391 and backbone nitrogen of Arg395 is 3.66 Å for DSR1, more or less present within the hydrogen binding distances (Figure 4-11[b]).



[a] **[b]**

Figure 4-11. Stick representation of **[a]** residues from 360 to 365 for USR1 and **[b]** residues from 390 to 395 for DSR1. Water molecules (W1 and W2) are shown in red spheres. Distance between backbone oxygens and backbone nitrogens are coloured in magenta, and the hydrogen bonding distance between protein oxygens and water molecules are coloured in orange.

4.4.3 Structure prediction

The I-TASSER server was used to predict the N-terminal first two spectrin repeat structures for Utr and Dys. Single molecules from the crystal structures of the N-terminal first spectrin repeat from Utr and Dys were used as model templates. I-TASSER was best suited for predicting protein structures as it uses the combination of various techniques such as threading, *ab initio* modelling followed by structure refinement [Zhang *et al.,* 2008]. The goal of this prediction is to understand the helical linker between the adjacent repeats. A function for the second repeat from Utr is unknown although it has been shown that the second repeat from Utr binds to membrane oleyl-phospholipids [Rumeur *et al.,* 2007].

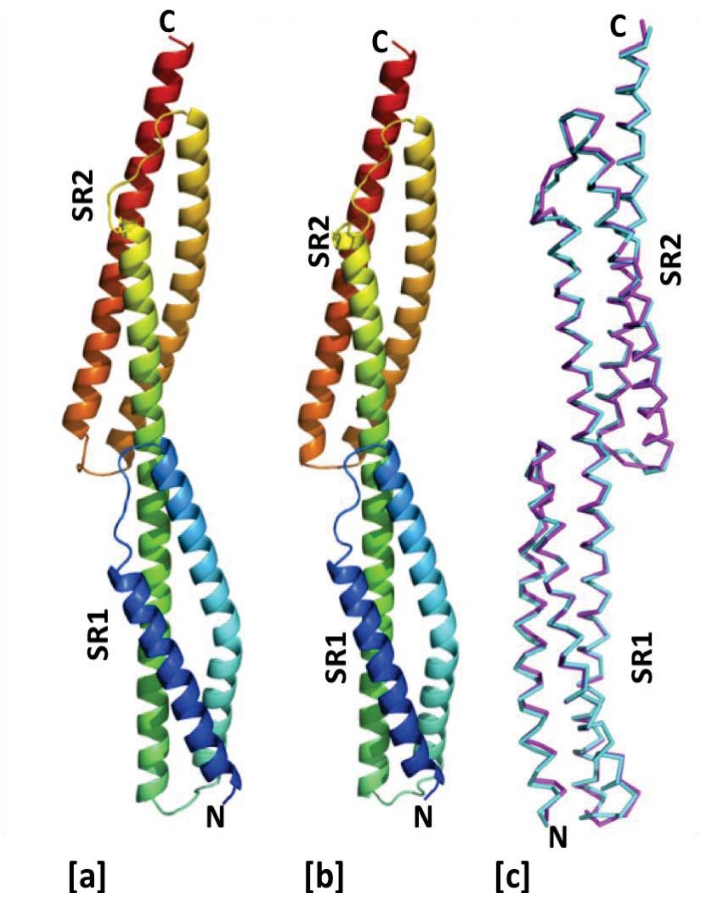


Figure 4-12. Predicted models of **[a]** USR12 from Utr and **[b]** DSR12 from Dys calculated using the I-TASSER server **[c]** Cα trace overlay of USR12 in cyan and DSR12 in pink.

The predicted model for USR12 and DSR12 looks essentially identical to the multiple spectrin repeat from α-actinin [PDBid - 1HCI] and spectrin [PDBid - 1CUN]. For the long USR1L construct Leu427 in position 'a' and Leu430 in position 'd' are oriented towards where we would expect the C-terminus of helix B' and loop B'C' of the second spectrin repeat as shown in Figure 4-13[a]. However the orientation of the residues Leu427 and Leu430 differed from the predicted model; if we overlay the model of Utr second spectrin repeat (SR2) from I-TASSER on to the fragment A' observed in USR1L, the second spectrin repeat has rotated ~180° compared to the predicted model (Figure 4-13[c]) furthermore Leu427 and Leu430 interact with hydrophobic residues in helix 'B' and 'C' of the second spectrin repeat as shown in Figure 4-13[b].

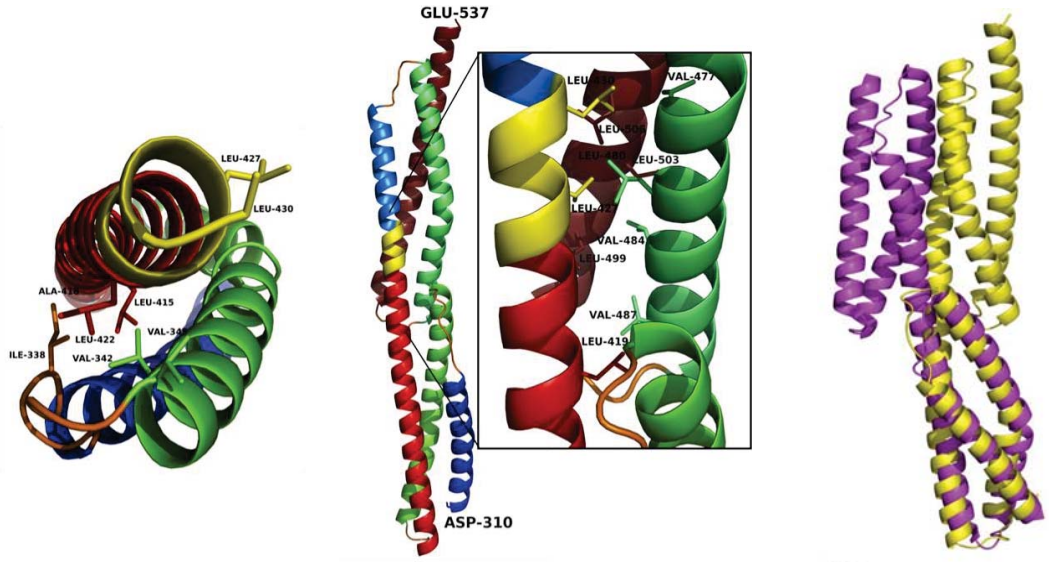


Figure 4-13. [a] Cartoon representation of USR1L, looking down from the C-terminus showing the hydrophobic side chains of helix 'A' in blue, helix 'B' in green, helix 'C' in red, loop 'AB' in orange and the 6 amino acid C-terminus coloured in yellow. **[b]** Cartoon representation of a hybrid USR12 model containing the experimentally-determined USR1L structure and a predicted Utr second repeat structure superposed on the overlapping residues of helix 'A' (yellow), highlighting the nested region where Leu427 and Leu430 interact with the hydrophobic residues in helix 'B' (light green) and 'C' (brown). **[c]** Cartoon representation of a Cα- overlay of the predicted USR12 structure in magenta to a hybrid model of USR1L and predicted SR2.

4.4.4 Comparison with other spectrin repeats

The DALILite server [Holm & Park, 2000] was used to search the protein data bank (PDB) for homologous structures of USR1 and DSR1 with selected results shown in Table 4-2. The structures were found to be most similar to repeat 16 of α-spectrin (PDBid – 1CUN) with rmsd of 1.7 Å to USR1 (111-Cα) and 1.5 Å to DSR1 (110-Cα). Furthermore a consistent rmsd of 1.7 Å and 1.8 Å were seen for USR1 and DSR1 compared to spectrin and α-actinin repeats.

Table 4-2. Comparison of USR1/DSR1 to selected spectrin repeat structures.

Protein molecule (Repeat Number)	**PDB ID**	**USR1**		**DSR1**	
		rmsd Å (lali; Cα)	**% id**	**rmsd Å (lali; Cα)**	**% id**
α-spectrin (Repeat 16)	1CUN	1.7 (111)	22	1.5 (110)	17
Human skeletal muscle α-actinin (Repeat 2)	1QUU	1.8 (111)	18	1.7 (112)	20
Human α-actinin 2 (Repeat 2)	1HCI	1.7 (111)	18	1.6 (112)	21
Erythrocyte α-spectrin (Repeat 1)	3LBX	1.8 (109)	18	1.7 (110)	18
Erythrocyte β-spectrin (Repeat 14)	3F57	1.8 (111)	20	1.8 (112)	18
Erythrocyte β-spectrin (Repeat 8)	1S35	2.1 (111)	18	1.7 (107)	18
Chicken brain α-spectrin (Repeat 15)	1U4Q	1.8 (111)	16	1.8 (112)	16
Chicken brain β-spectrin (Repeat 14)	3EDV	1.8 (111)	18	2.0 (112)	19
Plectin (Repeat 3)	3PDY	2.4 (107)	14	2.1 (106)	14

The root mean squared differences (rmsd) in Å between repeat domains are reported. The number of Cα atoms (length of alignment 'lali') used in calculating the rmsd for each pair wise comparison are included in brackets. '% id' is % sequence identity.

The USR1/DSR1 repeat structures were superposed onto the other spectrin repeats as shown in Figure 4-14. The structure of USR1 and DSR1 are very similar to the spectrin repeats from spectrin, α-actinin and plectin with Cα-rmsds for superposition ranging from 1.5 to 2.4 Å. The main structural difference lies in the loops 'AB' and 'BC', as a result of high conformational variability.

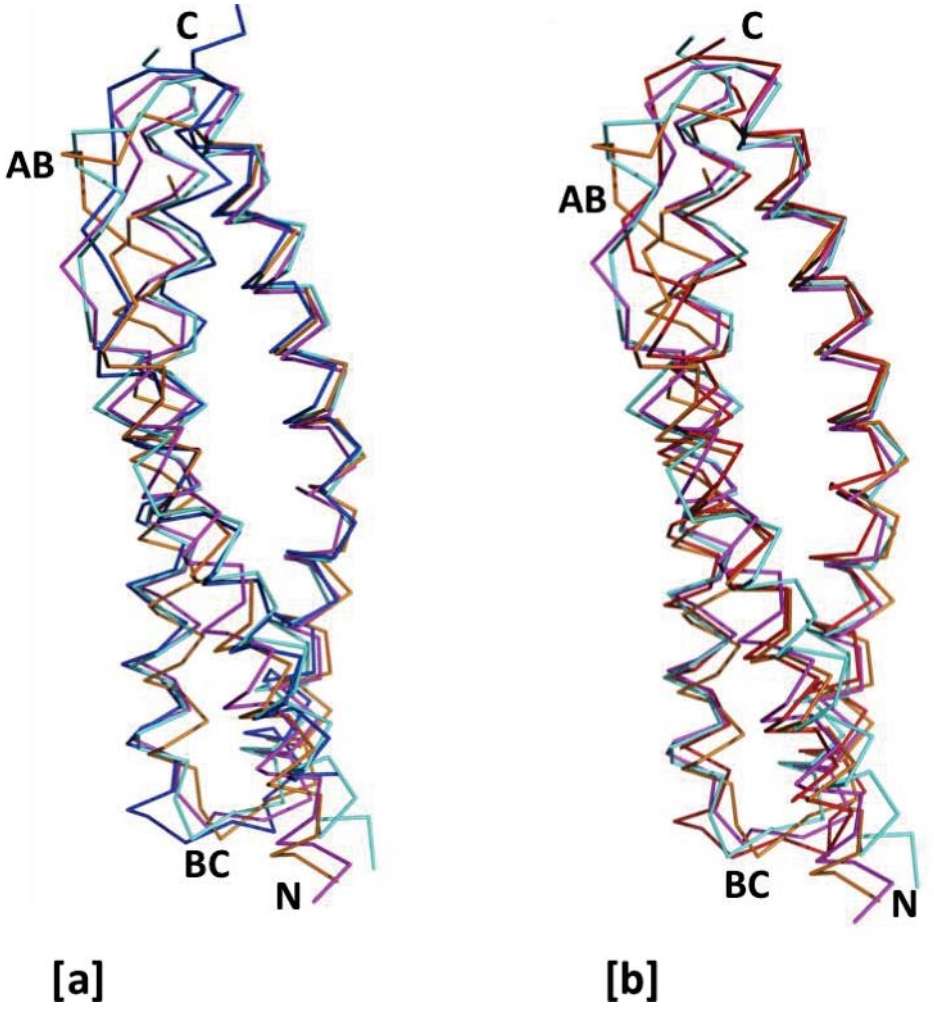
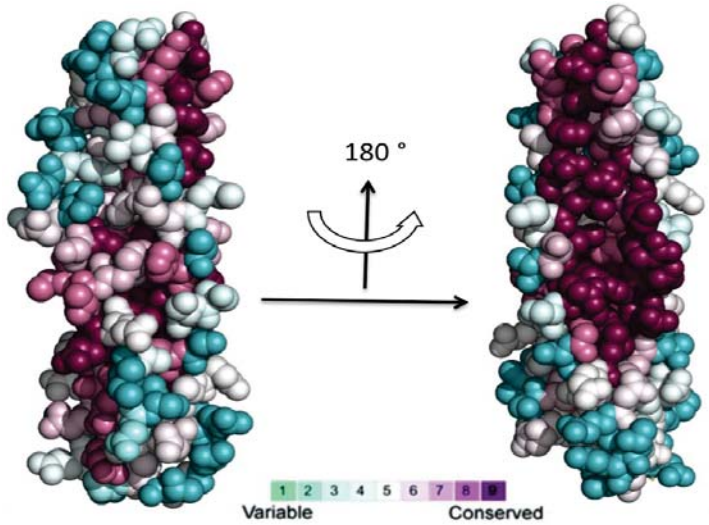


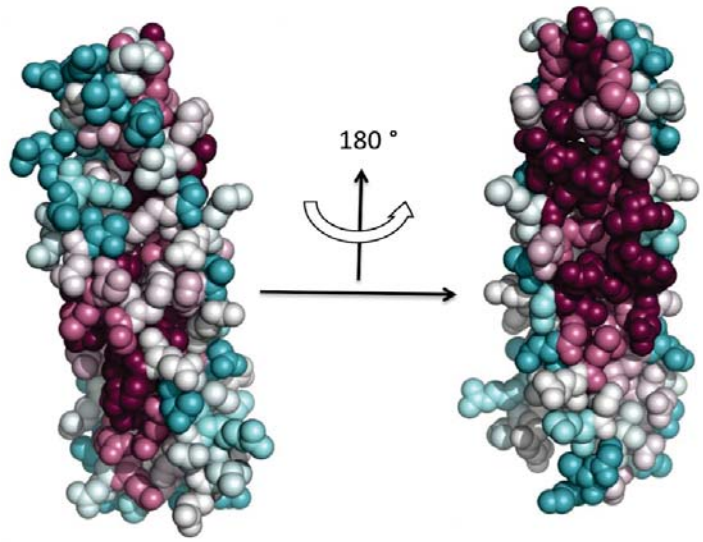
Figure 4-14. Cα trace overlay of **[a]** USR1 in blue, **[b]** DSR1 in red with homologous spectrin repeat structures, repeat 17 from α-spectrin (1CUN) in magenta, repeat 2 from α-actinin (1HCI) in cyan and repeat 4 from plectin (2PDY) in orange.

4.5 Functional analyses of N-terminal first spectrin repeat from Utr and Dys

Any specific function for the N-terminal first spectrin repeat domain from Utr (USR1) and Dys (DSR1) is unknown. In order to predict its function, online servers such as ConSurf [Goldenberg *et al.,* 2009], LIGSITE [Huang *et al.,* 2006] and 3DLigandSite [Wass *et al.,* 2010] were used to predict the conserved residues on the surface and to probe any possible binding sites and binding partners.



[a]



[b]

Figure 4-15. Space-filling models of **[a]** USR1 and **[b]** DSR1 obtained from the CONSURF server, representing conserved residues in purple, variable in cyan and neutral in white. Two views oriented by 180° with respect to each other.

The ConSurf server [Goldenberg *et al.,* 2009] was used to identify the functional regions in proteins based on the evolutionary conserved residues within the protein family. In the case of USR1 and DSR1, a strongly conserved groove on the face of helices 'B' and 'C' was predicted (Figure 4-15). It has been reported that the conserved regions such as these are often associated with protein function [Ashkenazy *et al.,* 2010], and serve as possible binding sites.

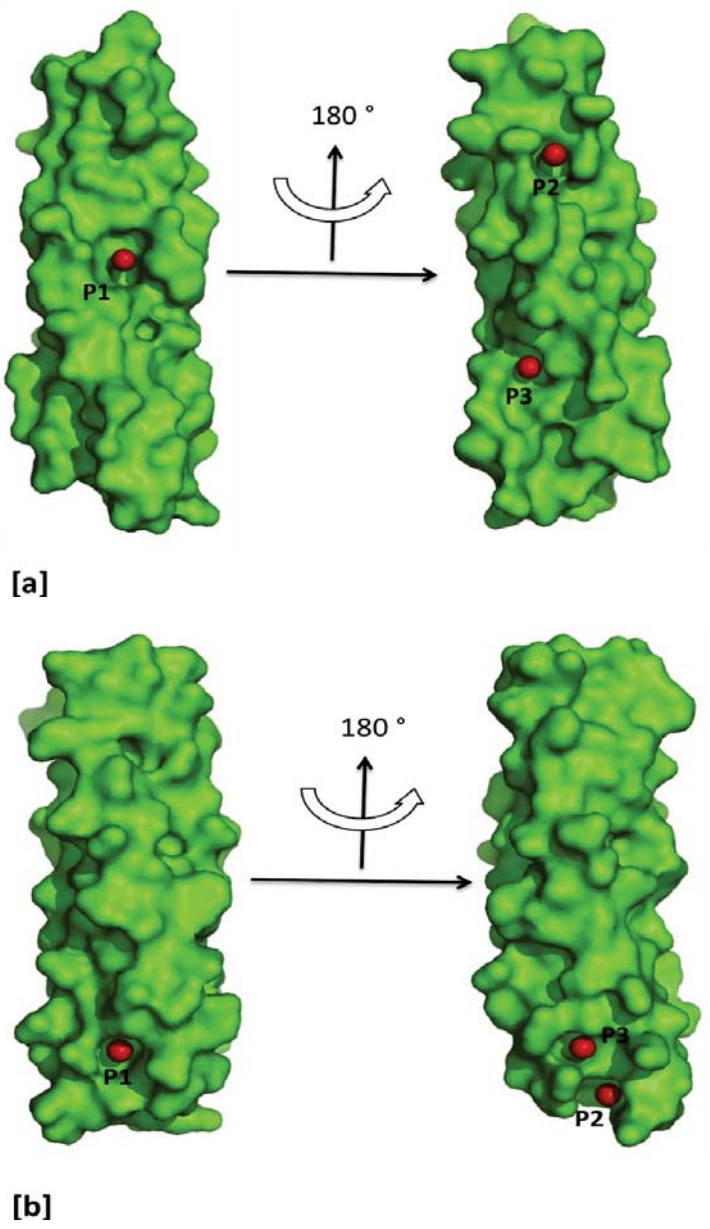


Figure 4-16. Surface-filling models of **[a]** USR1 and **[b]** DSR1 coloured green and potential ligand (in red) binding pockets P1, P2, P3 predicted from the LIGSITE server. Two views oriented by 180° with respect to each other.

The LIGSITE server [Huang *et al.,* 2006] was used to predict the top three ligand binding pockets of at least 3 $Å^3$ on the surface of the UR1 and DSR1 protein structures. These pockets were thought to be the most likely ligand binding sites [Morita *et al.,* 2008]. The top three possible ligand-binding sites for USR1 and DSR1 were identified (Figure 4-16) and interestingly the positions of these pockets are different for USR1 and DSR1. In USR1 a ligand binding pocket was observed in the centre between helix 'A' and helix 'B' (P1), between helix 'B' and helix 'C' (P2) and between helix 'A' and helix 'C' (P3) (Figure 4-16 [a]). Whereas for DSR1, all three ligand-binding pockets were predicted at the N-terminal end of the domain (Figure 4-16[b]).

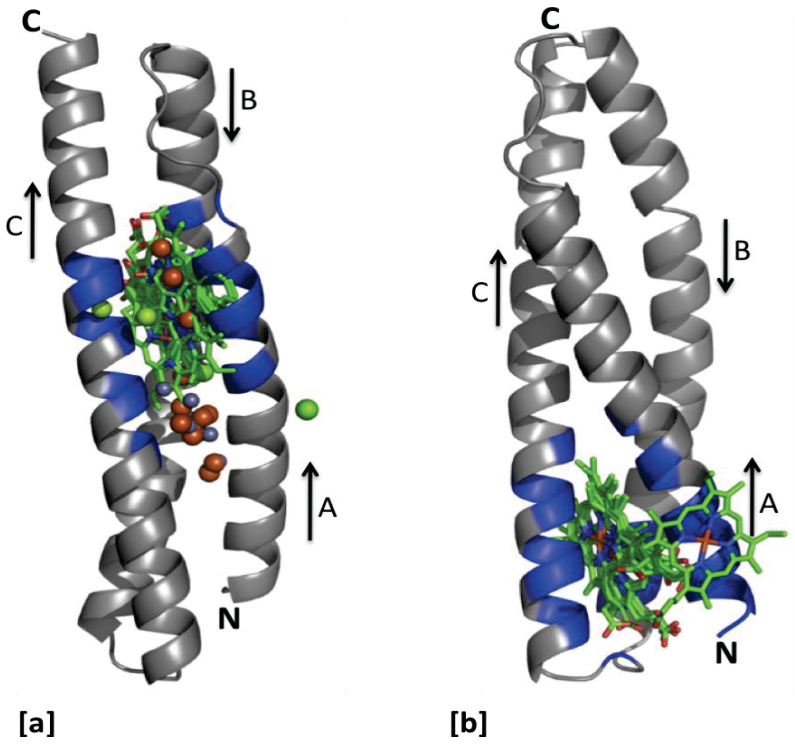


Figure 4-17. Ligand binding sites for **[a]** USR1 and **[b]** DSR1 were predicted using the 3DLigandSite server. Images were rendered using PyMol and predicted binding sites were coloured in blue, organic molecules (heme variants in green) in sticks and ions (Fe^{2+} in brown, Ca^{2+} in green, Mg^{2+} in light green, Zn^{2+} in grey) in spheres.

The 3DLigandSite [Wass *et al.,* 2010] web server was used to predict ligand-binding sites and binding partners for USR1 and DSR1. It utilises ligand-binding data from

similar structures in the protein data bank (PDB) and combines these results with residue conservation to make predictions. From the predicted binding sites, it shows that the USR1/DSR1 proteins were aligned with the structures from erythroid spectrin in the PDB, as it was known that erythroid spectrin binds to the haemoglobin variants [Chakrabarti *et al.,* 2001; Datta *et al.,* 2003] and the ligand-binding for USR1 and DSR1 predicted that they could bind to heme with different binding sites (Figure 4-17). Cations such Fe^{2+}, Ca^{2+}, Mg^{2+} and Zn^{2+} were also predicted to bind USR1, but were absent in DSR1. However there have been no reports showing, whether the Utr and Dys spectrin repeats bind to heme variants.

4.6 Summary

We have determined the first structures of spectrin repeat domains from Utr and Dys proteins. The high resolution structures of the N-terminal first spectrin repeats of Utr and Dys were solved to 1.95 and 2.3 Å respectively, showing a three-helical bundle fold having a left handed coiled-coil fold as previously predicted from sequence [Cross *et al.,* 1990; Parry *et al.,* 1992]. The structures of Utr and Dys spectrin repeat domains were found to be very similar to the repeat domains found in spectrin [Grum *et al.,* 1999] and α-actinin [Djinovic-Corugo *et al.,* 1999]. A six amino acid longer version of Utr N-terminal first spectrin repeat (USR1L) has also been determined to 2.0 Å from a USR1L crystal that was obtained from the consequences of *in-situ* proteolysis in the crystal drop containing equal volumes of N-terminal first two spectrin repeats of Utr and precipitant solution.

4.7 Accession number

Coordinates and structure factors have been deposited in the protein data bank (PDB) [Bernstein *et al.,* 1977] with the following accession codes: 3UUL for USR1, 3UUM for USR1L and 3UUN for DSR1.

5. *In vitro* actin-binding studies of recombinant Utr and Dys proteins.

5.1 Introduction

The actin-binding properties of Utr and Dys spectrin repeats are not completely understood and the mechanism of its binding still remains unclear. Since there was no solved structures of spectrin repeat for both Dys and its paralogue Utr, the mode of interaction with F-actin was poorly understood. Previous studies showed that the actin-binding domain (ABD) of Utr along with the first spectrin repeat bind F-actin more effectively than actin-binding domains alone [Sutherland-Smith *et al.,* 2003], this suggests that the first spectrin repeat of Utr may be involved in binding to F-actin and the model reconstruction from cryo-EM showed that the first spectrin repeat contacts F-actin laterally. In this chapter the *in vitro* actin-binding properties of N-terminal first and N-terminal first two spectrin repeats of both Utr and Dys were compared.

5.2 Actin-binding assay (ABA)

5.2.1 Actin Polymerisation

Actin is a monomeric 42 kDa globular protein, also termed as G-actin, found in all eukaryotic cells. Polymerisation of G-actin was initiated in the presence of ATP, $MgCl_2$ and KCl to form filamentous F-actin in a process called nucleation and assembly.

15.3 µL of monomeric actin (4 mg/mL suspended in water) purchased from Cytoskeleton Inc., was mixed with 1.7 µL of 10x F-buffer (20 mM Tris-HCl pH 8.0, 20 mM $MgCl_2$, 10 mM ATP, 500 mM KCl) with the mixture kept at room temperature for 1 hour, allowing a viscous filamentous actin (F-actin) solution to form.

5.2.2 F-actin binding assay

N-terminal spectrin repeats of both Utr and Dys were transferred into ligand buffer (20 mM Tris-HCl pH 8.0, 6.25 mM DTT, 120 mM NaCl) to a final concentration measured to a final concentration of 20 or 40 µM and then centrifuged at 75,000 x *g* at 25 °C to

remove any precipitate or insoluble protein present. 20 µl of USR1 or DSR1 supernatant was mixed with 5 µl F-actin and 5 µl of actin-binding buffer (ABB) (120 mM Tris-HCl pH 8.0, 1.5 M NaCl, 20 mM $MgCl_2$, 10 mM ATP, 1 mM $CaCl_2$) to a total volume of 50 µl. The assay was incubated at 25 °C in polycarbonate centrifuge tubes *(TLA 100)* (Beckmann Coulter) for 30 min followed by ultracentrifugation at 75,000 x *g.* Any protein that binds to F-actin precipitates (P - pellet fractions) and protein that doesn't bind to F-actin remains in the supernatant (S - soluble fractions). 8 µl of both the soluble and the pellet fractions were loaded onto an SDS-PAGE gel as described in chapter 2.2.3.2.

5.2.3 Interaction of spectrin repeats with F-actin

Recombinant protein concentrations of 20 and 40 µM were used with the 8.2 µM final F-actin concentration for binding studies. A 20 µM concentration sample of Filamin-B actin-binding domain (ABD) kindly donated by Dr. Gregory M. Sawyer [Sawyer *et al.,* 2009] was used as a positive control (Table 5-1). The amount of protein bound to actin was calculated from the known concentration of protein added to the assay compared to the ratio of protein concentrations in supernatant and pellet fractions; this was achieved by quantifying the band intensities for each sample in SDS-PAGE gels (Figure 5-1) using Quantity One® (Biorad) densitometry software. The actin-binding assay was performed more than three times for each protein.

Table 5-1. Contents of the actin-binding assay

TLA 100 **Tube No.**	**Protein Concentrations USR1/DSR1/USR12/DSR12 (µM)**	**Concentration of Filamin-B ABD (µM)**	**F-actin (µM)**	**Actin-binding buffer (ABB) (µl)**
1	20	-	8.2	5
2	40	-	8.2	5
3	-	20	8.2	5

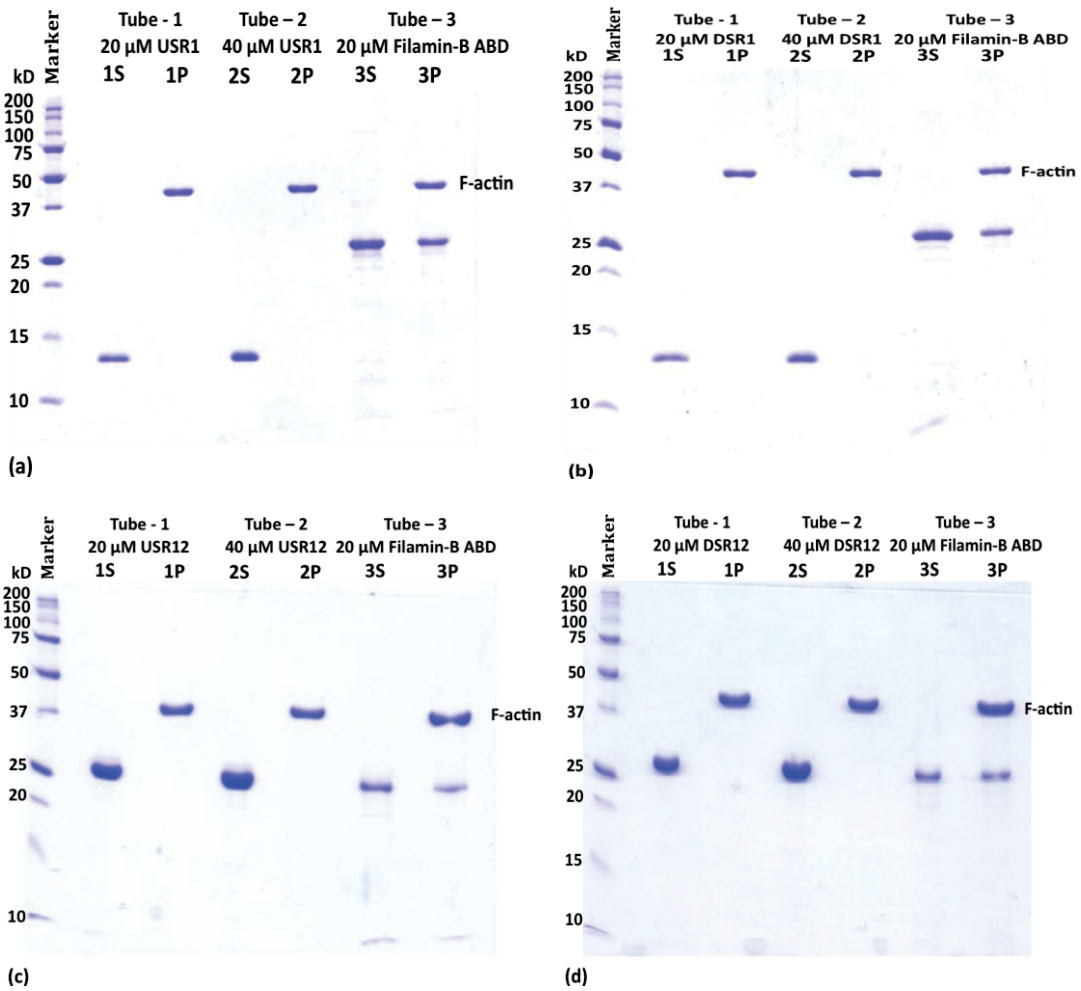


Figure 5-1. Actin co-sedimentation assay where S and P are supernatant and pellet fraction. 3S and 3P are supernatant and pellet fractions of 20 µM Filamin-B ABD (24 kDa) used as a positive control in all the assays. The proteins bands of USR1/DSR1 are of a molecular weight of 14 kDa and USR12/DSR12 of molecular weight of ~25 kDa.

(a) 1S, 1P are the supernatant and pellet fractions of 20 µM USR1, 2S and 2P are from 40 µM USR1.

(b) 1S, 1P are the supernatant and pellet fractions of 20 µM DSR1, 2S and 2P are from 40 µM DSR1.

(c) 1S, 1P are the supernatant and pellet fractions of 20 µM USR12, 2S and 2P are from 40 µM USR12.

(d) 1S, 1P are the supernatant and pellet fractions of 20 µM DSR12, 2S and 2P are from 40 µM DSR12.

The F-actin co-sedimentation assay showed no affinity for the N-terminal spectrin repeats of Utr and Dys for F-actin in the absence of the actin-binding domain (ABD). However the presence of the spectrin repeat combined with the actin-binding domain increases the binding affinity towards F-actin compared to actin-binding domain alone. This may be due to the presence of the spectrin repeat increasing the stability of the ABD (containing CH1 and CH2 domains) by bringing the CH2 domain closer to the inner domain of F-actin [Sutherland-Smith *et al.,* 2003]. Sequence analysis using the ProtParam tool from ExPasy server (http://web.expasy.org/protparam/) showed that the N-terminal first and first two spectrin repeats of Utr and Dys have a predicted isoelectric point (pI) of around 4, this suggests that N-terminal spectrin repeats (USR1, DSR1, USR12 and DSR12) come under the acidic group of repeat domains (Figure 5-2) thus lacking affinity towards acidic F-actin when compared to the basic spectrin repeats (repeat 11 to 17) from Dys [Rybakova *et al.,* 1996; Amman *et al.,* 1999].

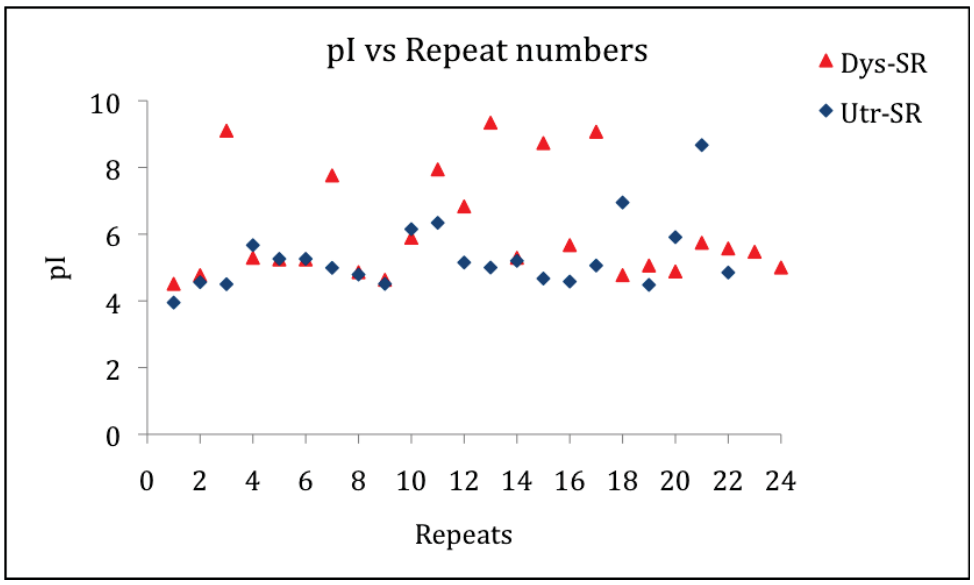


Figure 5-2. Schematic plot of the predicted isoelectric point (pI) vs spectrin repeat number of Utr and Dys. Protein sequence boundaries of each repeat were selected based on Winder *et al.,* 1995 and the pI was predicted using the ProtParam tool from the ExPasy Server.

5.2.4 Protein electrostatic surface analysis

Analysis of the electrostatic surface from the structures of the N-terminal first spectrin repeats of Utr and Dys showed an acidic surface along the face of helices A and B and including the BC-loop (Figure 5-3).

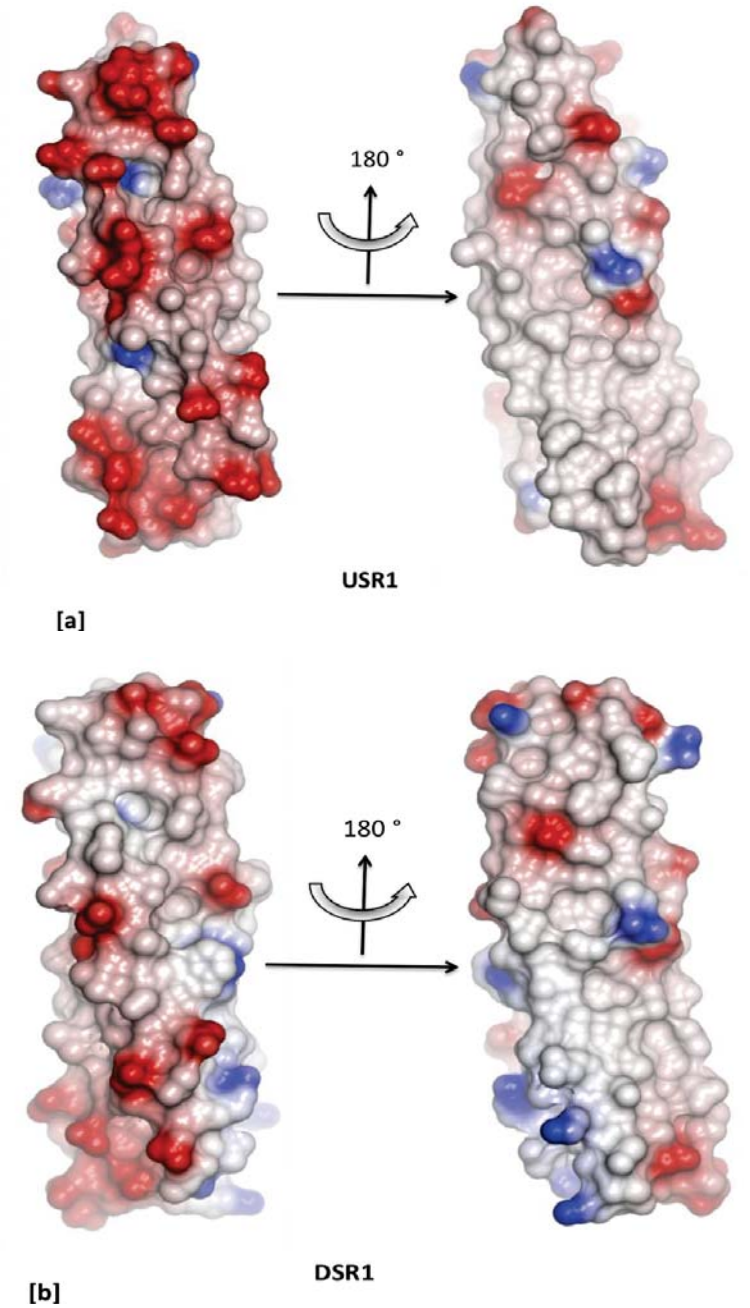


Figure 5-3. Electrostatic surface representation of **[a]** USR1 and **[b]** DSR1 each view pair has a 180° orientation along the vertical axis. The surface has been colour coded from red (-0.5 V) to blue (+0.5 V) by electrostatic potential.

Although the structure of the N-terminal first two spectrin repeats of Utr (USR12) and Dys (DSR12) were not determined, atomic models were predicted from the protein sequences using the I-TASSER Server [Zhang *et al.,* 2008] with the single repeat from the solved structures of USR1 and DSR1 used as template. The surface electrostatic potential from the predicted models do not contain the energy types (charge) for each atoms and hence electrostatic surface for predicted models cannot be generated using the CCP4mg software, however this can be achieved by using the APBS plugin [Baker, 2004; Baker & McCammon, 2005] either in VMD or PYMOL.

5.3 Summary

Electron microscopy image reconstruction of the Utr416 construct, containing the Utr actin-binding domain (ABD) along with the first spectrin repeat showed difference density for the first spectrin repeat associated laterally with F-actin [Sutherland-Smith *et al.,* 2003], where the structure of an α-actinin spectrin repeat was used as a model template in the difference density map. However F-actin co-sedimentation assays of the isolated Utr N-terminal first spectrin repeat domain showed no affinity with F-actin at the investigated concentrations (20 and 40 μM) in the absence of actin-binding domains. A similar conclusion was made for the Dys N-terminal first spectrin repeat domain. Analysis of surface electrostatic potential showed that the presence of acidic surfaces due to the Asp and Glu residues along the helix 'A' and 'B' in USR1 and DSR1 could provide the reason for a lack of binding to F-actin. The negative electrostatic character correlates with the lack of affinity with acidic F-actin as observed for multiple spectrin repeats isolated for Dys and Utr, unlike the Dys basic repeats 11-17 that showed intrinsic actin-binding affinity [Amman *et al.,* 1999]. This suggests that although the N-terminal first spectrin repeats from Utr and Dys have no intrinsic actin-binding activity, they contribute to the ABD to bind F-actin more effectively compared to ABD's alone [Sutherland-Smith *et al.,* 2003]. The N-terminal first two spectrin repeats from Utr and Dys also showed no-affinity for F-actin and analysing the sequence using the ProtParam tool from the ExPASy server has revealed an Isoelectric potential (pI) about 4.0, the most acidic part of Utr and Dys spectrin repeat domains.

6. Protein stability studies using Circular Dichroism (CD)

6.1 Circular Dichroism (CD)

Circular dichroism (CD) is an excellent technique used for secondary structure analysis, stability studies and comparison of macromolecules. Proteins possess chromophores which give raise to a CD signal corresponding to secondary structure features (α-helix and β-sheets) in the far UV region (240 – 180 nm) and the environment around the aromatic side chains in the near UV region (320 – 260 nm) [Kelly & Price, 2000; Kelly *et al.,* 2005; Greenfield, 2007].

6.2 Introduction

In this chapter, I aim to investigate the stability of secondary structures at varying temperatures and urea concentrations for the N-terminal spectrin repeat domain of Utr compared to that of N-terminal spectrin repeat of Dys.

The CD signal was initially recorded in terms of milli degree *or* in degrees ellipticity [θ], and these values were converted to mean residue ellipticity or Δε using the formula below:

Ellipticity, [θ], in deg. · cm^2/dmol = (millidegrees x mean residue weight) / (path length in mm x concentration in mg/ml)

(or)

[θ] = millidegrees / (path length in mm x the concentration of protein x the number of residues)

and mean residue ellipticity, Δε=[θ]/3298 [Kelly & Price, 2000].

6.3 Secondary structure prediction

Protein samples of high purity were used for the CD measurements. Buffers such as Tris are optically active [Kelly *et al.,* 2005] and hence the proteins were buffer exchanged to phosphate buffer (5 mM Na_2HPO_4/NaH_2PO_4 pH 7.4, 10 mM NaCl) using VivaSpin ultrafiltration columns. Ideally water can be used instead of buffers for a minimal noise to signal ratio in CD spectroscopy but not all proteins are stabile in H_2O. 0.5 – 1.0 mg/ml protein concentrations were used for all the proteins (USR1, DSR1, USR12 and DSR12) throughout the CD experiments analysed over a wavelength range of 190-260 nm, averaged over three scans, with a bandwidth of 1 nm, a response time of 1 second, a pitch of 1 nm and collected at 21 °C in a 1 mm quartz cuvette. Each CD-spectra was acquired using the Chirascan spectrometer (Applied Photophysics) and averaged and blanked against buffer alone using Pro-Data Chirascan V4.1.2 software (Applied Photophysics).

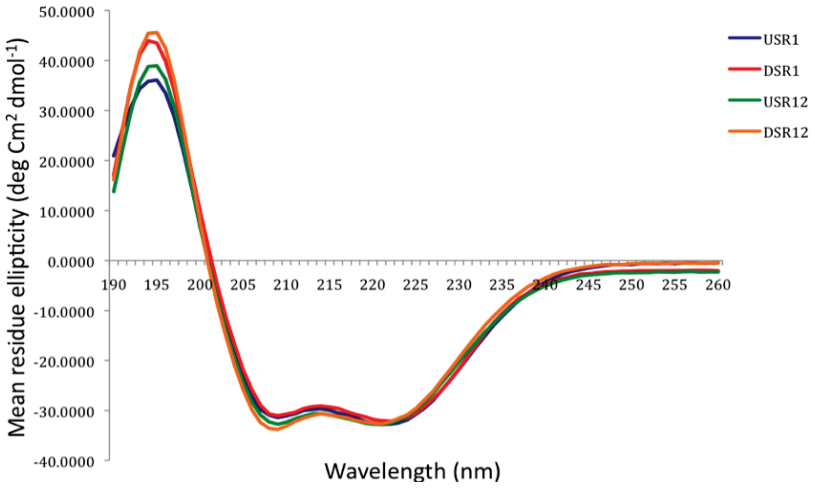


Figure 6-1. CD spectra of USR1 (blue), DSR1 (red), USR12 (green) and DSR12 (orange)

The ϕ and ψ angles give the α-helical CD spectrum a characteristic spectrum with two negative bands at 222 and 208 nm, and a positive band at 192 nm. The negative band values at 222 and 208 nm are normally used for secondary structural analysis for α-helical proteins [Toumadje *et al.,* 1992; Moffit *et al.,* 1957].

6.3.1 Secondary Structure Analysis using CDNN

CDNN is a neural network program widely used to calculate the secondary structure of proteins. A neural network is an artificial intelligence program used to find correlations in data, first the program was trained using the CD signals from a set of known proteins and the trained network was then used to analyze the secondary structure of unknown proteins [Bohm *et al.,* 1992].

CDNN requires a text files as an input format; this was achieved using Pro-Data Viewer [Applied Photophysics], a software program with the CD instrument. Deconvolution using the text file containing wavelengths and associated CD values was used in CDNN software [Applied Photophysics] for calculating secondary structures.

Table 6-1. CDNN deconvolution results from the CD-spectra obtained at 5 °C

Secondary Structure	USR1 (195 – 260 nm)	DSR1 (195 – 260 nm)	USR12 (195 – 260 nm)	DSR12 (195 – 260 nm)
Helix	99.1%	99.5%	99.6%	99.9%
Beta turn	3.1%	3.7%	3.0%	2.1%
Random coil	0.6%	0.4%	0.2%	0%

The results from the CDNN deconvolution (Table 6-1*)* showed that the percentage of helices in the recombinant proteins [USR1, DSR1, USR12 and DSR12] are broadly similar and coincides with the calculated helical content from the crystal structures of Utr and Dys N-terminal first spectrin repeat domains (98%) and predicted models (98.5%). Although there exist small differences between the predicted helices from CD spectra and crystal structure, these differences were within the acceptable error range of ±5% [Kelly *et al.,* 2005].

6.4 Stability studies of spectrin repeat domains using CD

6.4.1 Effect of Temperature on protein stability

CD spectra of the recombinant proteins were obtained at temperatures (Figure 6-2) ranging from 5 to 90 °C with data collected at a bandwidth of 1nm, step size of 0.5 s and wavelength range from 180 to 260 nm. During the course of the experiment the temperature was increased at a rate of 1 °C per minute with the aid of a peltier device supported by a water bath.

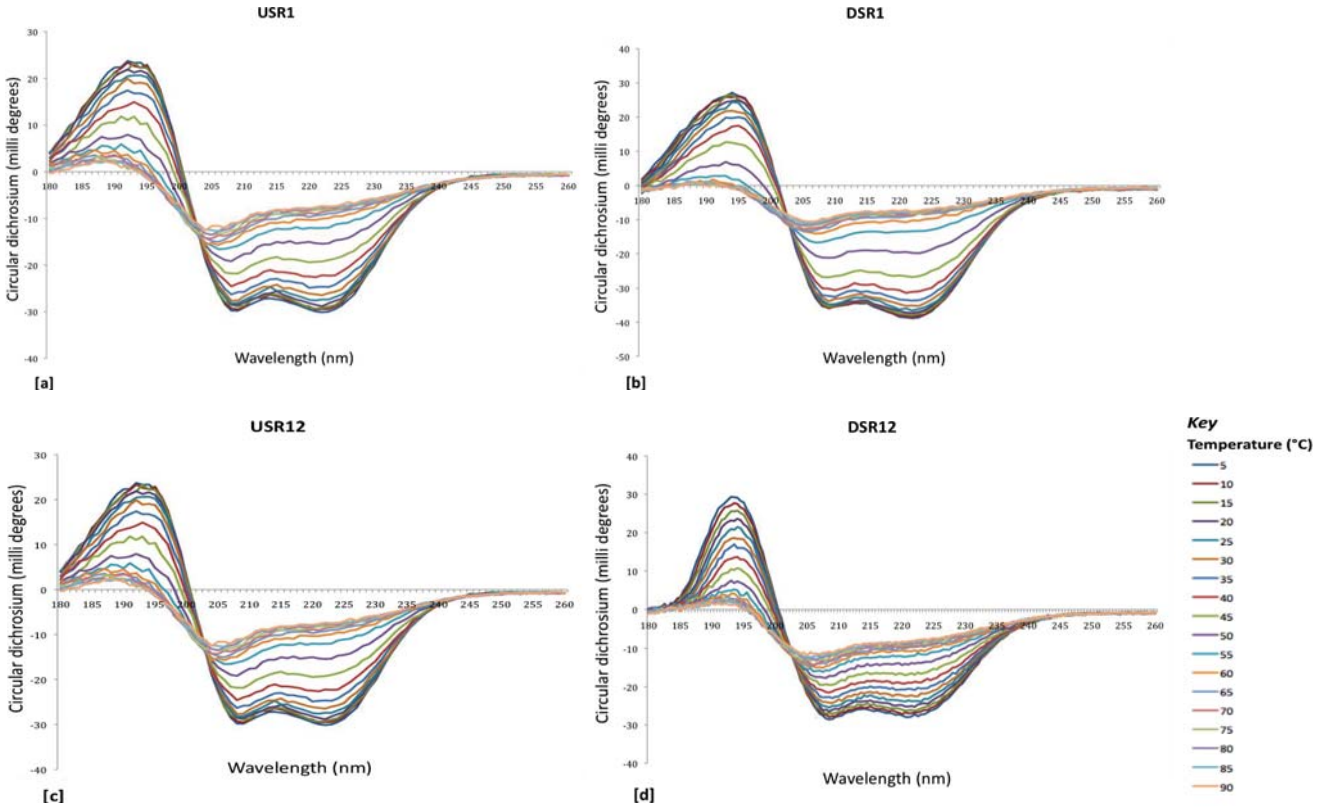


Figure 6-2. Transition observed in CD spectra for recombinant proteins as a function of change in temperature (from 5 to 90 °C). Temperatures are listed in the key. **[a]** USR1, **[b]** DSR1, **[c]** USR12, **[d]** DSR12.

Melting curves were obtained by monitoring the ellipticity at 222 nm as a function of temperature and the percentage of α-helical content was calculated by assuming a mean residue ellipticity $[\theta]_{222}$ at 5 °C corresponding to 100% α-helix [Greenfield & Fasman, 1969]. A strong CD signal was monitored at 5 °C, while a significant loss of secondary structure was observed at increased temperature upto 90 °C where the protein was denatured.

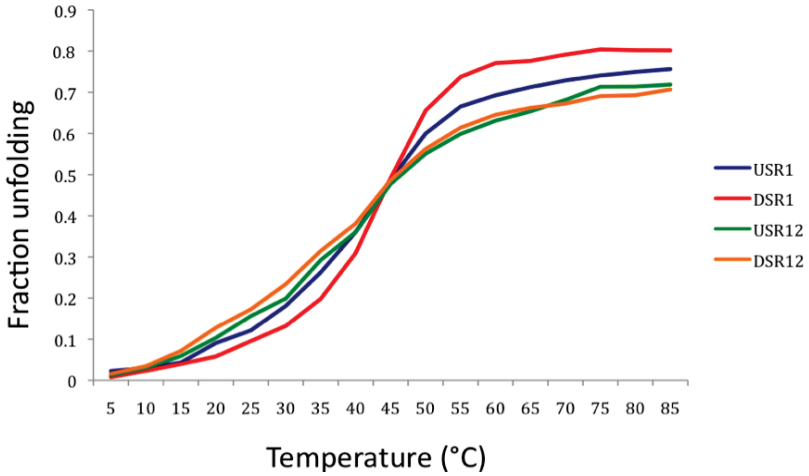


Figure 6-3. CD temperature scans at 222 nm for Utr and Dys proteins in 5 mM phosphate buffer pH 7.4, 10 mM NaCl: USR1 (blue), DSR1 (red), USR12 (green) and DSR12 (orange).

The midpoint between the folded and unfolded transition was taken as the melting temperature *(T_m)*. The T_m value for each protein was determined as 47 °C and there are no significant difference between the N-terminal first and first two spectrin repeat domains of Utr and Dys. Furthermore after thermal denaturation at 90 °C for the N-terminal spectrin repeats from Utr and Dys, proteins were cooled to 5 °C for 5 minutes and refolding spectra were recorded at a bandwidth of 1 nm over a spectral range between 190 nm and 260 nm and the temperature lowered from 90 °C to 5 °C at a rate of 1 °C per minute (Figure 6-4).

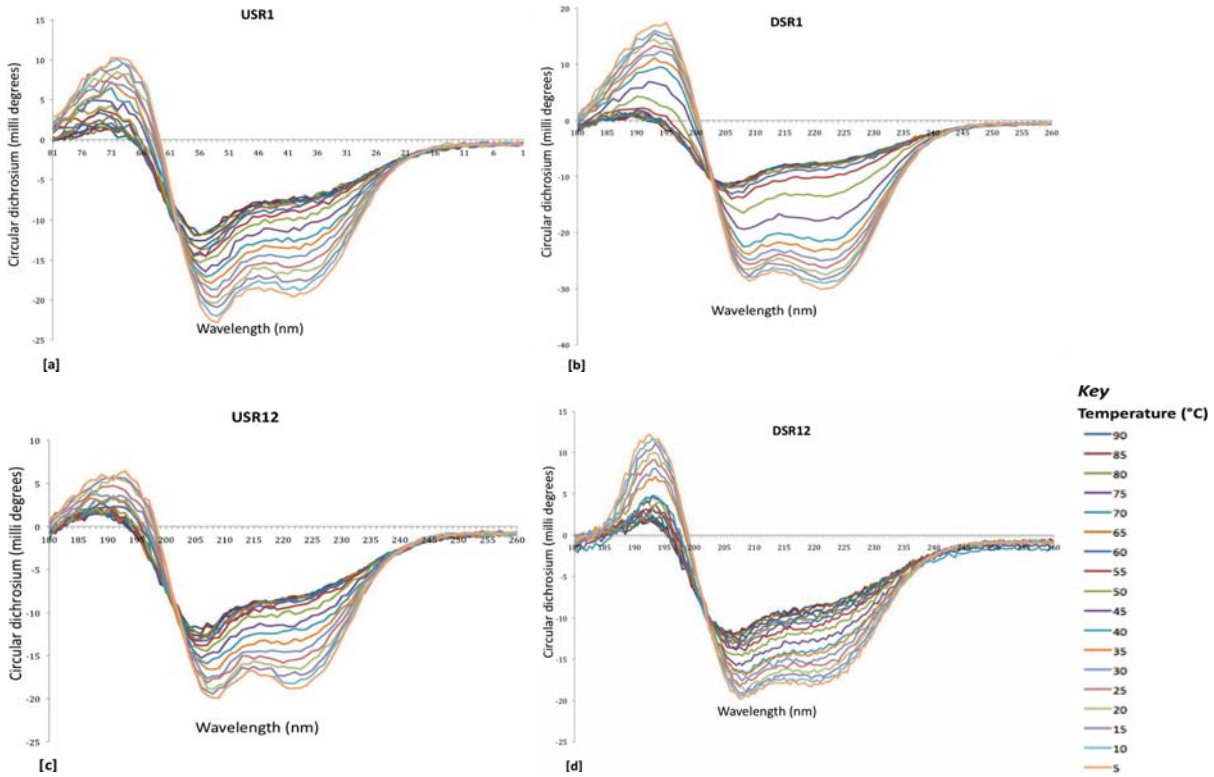


Figure 6-4. Transition observed in CD spectra for recombinant proteins as a function of change in temperature (from 90 to 5 °C). Temperature used is listed in key. **[a]** USR1, **[b]** DSR1, **[c]** USR12, **[d]** DSR12.

6.4.2 Effect of Urea on protein stability

Although urea induced unfolding [Pace, 1986] is an important tool to understand protein folding, the absolute mechanism of denaturation remains still unknown [Bennion & Daggett, 2004; Burgardt *et al.,* 2009]. Similar to the melting temperature, urea induced unfolding of N-terminal spectrin repeat domains was monitored by CD spectroscopy at 222 nm assuming the recombinant proteins without urea corresponding to 100% α-helix.

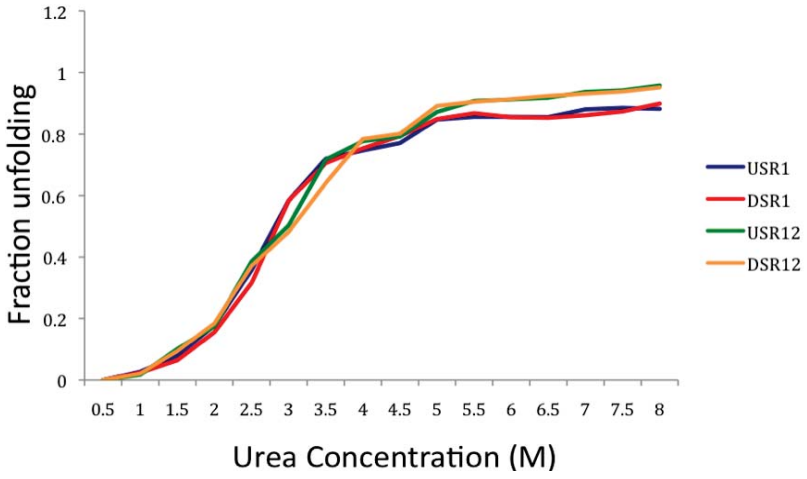


Figure 6-5. The influence of urea concentration on protein stability was monitored at 222 nm using CD-spectra: USR1 (blue), DSR1 (red), USR12 (green), DSR12 (orange).

The concentration of urea required for half unfolding N-terminal spectrin repeat domain of Utr and Dys was determined to be ~3 – 3.5 M. The complete loss of protein secondary structure was observed at 8.0 M urea in protein solution.

6.5 Summary

CD thermal stability scans have suggested that there are no significant differences observed between the N-terminal first alone and first two spectrin repeats from Utr and Dys. A melting temperature (T_m) of ~47 °C was observed. Previous studies from erythroid spectrin repeats have shown that the melting temperatures of multiple repeats (R16R17) are higher (T_m 52 °C) when compared to a single repeat domain (R16, T_m 49 °C) [MacDonald & Pozhanski, 2001]. However a similar difference was not seen

in Utr and Dys N-terminal repeats. Thermal stability measurements of each repeat domain from erythroid spectrin repeats have showed that the melting temperatures ranged from 21 to 72 °C [An *et al.,* 2006], although an average of T_m ~37 °C was observed in most of the single repeat domains. It has also been showed that single mutations at the helical hydrophobic position of repeat-16 from erythroid β-spectrin reduces the T_m from 71 °C to either 61 or 56 °C depending on the mutation, and the double mutation further reduces the T_m from 71 °C to 47 °C. The greater change in the T_m is due to the disruption of helical packing [An *et al.,* 2006]. Recent studies using CD-spectroscopy have shown that the mutation identified in Dys N-terminal first spectrin repeat (L427P) from DMD patients causes the domain to have reduced T_m from 53.4 °C to 50 °C [Acsadi *et al.,* 2012], indicating a lower thermal stability for the mutant protein. Furthermore after thermal denaturation at 90 °C the Utr and Dys N-terminal spectrin repeats were cooled to 5 °C for 5 minutes and refolding spectra were recorded at a bandwidth of 1nm over a spectral range between 190 nm and 260 nm and temperatures lowered from 90 °C to 5 °C at a rate of 1 °C per minute during which almost 85-90% of ellipticity were observed at 50 °C calculated from the percentage difference between unfolding and refolding at 208 and 222 nm CD-spectra, indicating that the proteins were refolding.

Urea unfolding experiments have shown that there is no noticeable difference in the denaturation profile between N-terminal spectrin repeat domains from Utr and Dys to ~3.0 M, although the increase in protein stability was observed in multiple repeats (R16R17) from spectrin to 3.5 M urea concentration compared to 2.2 M for the single repeat domain (R16) [MacDonald *&* Pozharski, 2001]. The difference in stability between single and multiple repeats observed in erythroid spectrin suggested that it is necessary to understand the stability of individual spectrin repeat domains of Utr and Dys.

7. Molecular dynamic studies of the N-terminal first spectrin repeat domains of Utr (USR1) and Dys (DSR1)

7.1 Introduction

The N-terminal spectrin repeat domains from Utr and Dys consist of three antiparallel α-helices separated by a loops and folded into a left-handed coiled coil structure as shown in chapter 4 (Figure 4-4[a], 4-5[a]) structurally similar to that of spectrin repeat domains from spectrin and α-actinin [Pascual *et al.,* 1997; Grum *et al.,* 1999; Djinovic-Carugo *et al.,* 1999]. There are about 500 spectrin repeats found in the human genome [Altmann *et al.,* 2002], most commonly in cytoskeletal actin binding proteins like dystrophin, utrophin, α-actinin and spectrins, where they constantly undergo external mechanical stress. Although, Dys and Utr spectrin repeat domains showed binding towards neuronal nitric oxide synthase, filamentous actin and membrane lipids [Le Rumeur *et al.,* 2010] the most important functional characteristics of the spectrin repeats are to maintain cellular structural rigidity and flexibility. Previous experiments using molecular dynamics simulations [Altmann *et al.,* 2004; Johnson *et al.,* 2007; Legardinier *et al.,* 2009] and atomic force microscopy (AFM) [Reif *et al.,* 1999; Lenne *et al.,* 2000; Lee *&* Discher, 2001; Law *et al.,* 2003] studying tandem spectrin repeat domains from erythroid spectrin and α-actinin suggested that the flexibility of multiple repeats depends upon the linker region between domains. However under higher extension, the triple helical domain undergoes further unfolding and refolding, and thus functions as an elastic element with in the cell [Lee & Discher, 2001].

In the present chapter, molecular dynamics (MD) simulations were performed using NAMD and VMD (Visual Molecular Display) tools [Humphrey *et al.,* 1996] to compare the conformational flexibility, stability and to establish a connection between the structural and the mechanical properties between N-terminal first spectrin repeat domain of Utr (USR1) and Dys (DSR1). In particular through the analysis of the root mean square deviation (RMSD) as a function of simulation time, the root mean square fluctuation for each amino acid (RMSF), and by calculating the solvent accessible

surface area (SASA) and bending-angles of the protein helices followed by analysis of unfolding simulations.

7.2 Molecular dynamic (MD) simulations

The structures of the N-terminal first spectrin repeat domain of Utr (USR1) and Dys (DSR1) were solved using x-ray crystallography to 1.95 and 2.3 Å were used for these simulations. Both the proteins USR1 and DSR1 have the same number of amino acids (112) folded into left-handed coiled coil structure of three alpha helices separated by loops as shown in Figure 4-7 (Chapter 4). For convenience during simulation experiments and analysis, the amino acid residues for both proteins (USR1 and DSR1) were renumbered from 1 to 112.

The USR1 and DSR1 domain structures were aligned such that the coiled coil structures were oriented along the X-axis as shown in Figure 7-1 and the hydrogen atoms not observed in the crystal structure were generated using the *psfgen* plugin in the VMD software [Humphrey *et al.,* 1996]. The aligned proteins were solvated (adding water molecules) with the TIP3P protocol [Jorgensen *et al.,* 1983] using the VMD plugin *solvate* and the system was neutralised with counter ions using the *autoionize* plugin available in VMD to reach an ionic strength of 150 mM NaCl, more closely approximating physiological conditions. Periodic boundary conditions were used in all simulations and a cutoff of 12 Å for Van der Waals interactions was assumed [Lim *et al.,* 2008]. An integrated time step of 2 fs and a constant pressure of 1 atmosphere were used. Langevin dynamics were employed to enforce constant temperature (T = 310K) and the simulations were carried out in a NVT ensemble in which the number of particles (N), the volume (V) and the temperature (T) are kept constant through out the simulations. The entire system containing protein, water and ions was gradually heated to 310 K and equilibrated. During the process of simulation the water molecules get scattered from the cubical system, and hence the water molecules were harmonically restrained to their original position to maintain the shape of a waterbox.

Molecular dynamics (MD) simulations for all the proteins were carried out using the program NAMD v2.6 employing the CHARMM27 force field parameter [Foloppe *et al.,* 2000; MacKerell *et al.,* 2004] and the figures were made using VMD software [Humphrey *et al.,* 1996].

7.2.1 Equilibrium molecular dynamic simulations

The size of the waterbox was set from the origin large enough to accommodate the entire protein, 90 X 54 X 52 Å for USR1 and 88 X 51 X 57 Å for DSR1 during equilibration as shown in Figure 7-1. The complete system contains 23048 atoms for USR1 and 23624 for DSR1. Each equilibrium simulations contain *approx* 1100 frames and each frame was recorded at every 20 ps simulation.

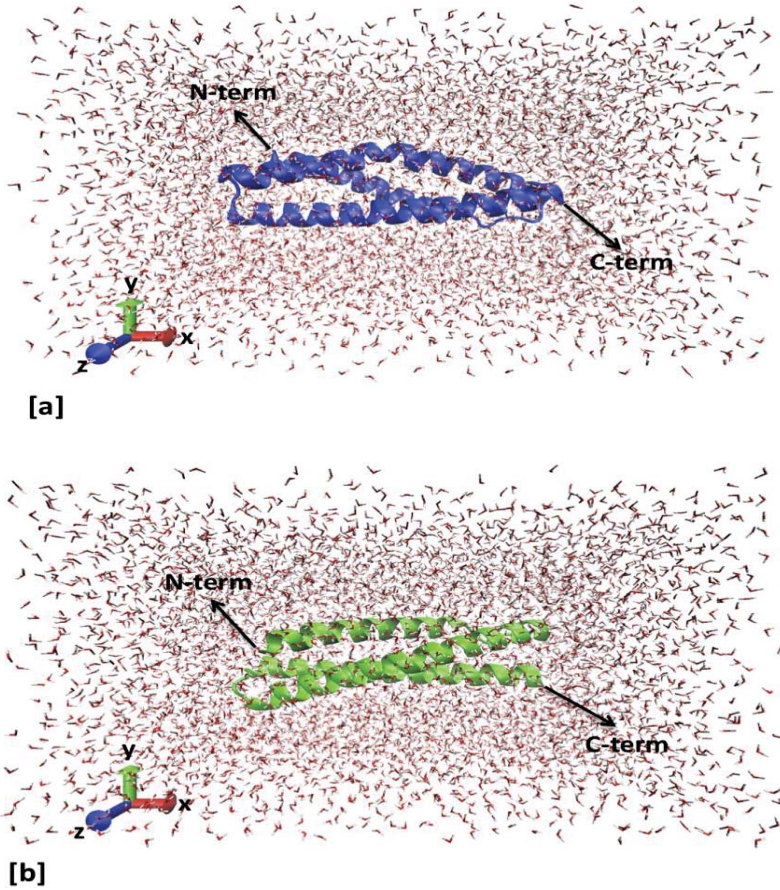


Figure 7-1. [a] Ribbon representation of USR1 (blue) oriented on the *X*-axis in an explicit water (pink/grey) box of size 90 X 54 X 52 Å from the origin; **[b]** Ribbon representation of DSR1 (green) oriented on *X*-axis in an explicit water box of size 88 X 51 X 57 Å from the origin.

7.2.1.1 Trajectory analysis for USR1 and DSR1 minimisation equilibrium

7.2.1.1.1 Calculation of the RMSD

To understand the stability and conformational change of the proteins, the main chain root mean square deviations (Cα-RMSD) were measured from the trajectories created during minimisation-equilibration simulations in explicit solvent with respect to simulation time. The resulting RMSD profiles for USR1 and DSR1 are shown in Figure 7-2.

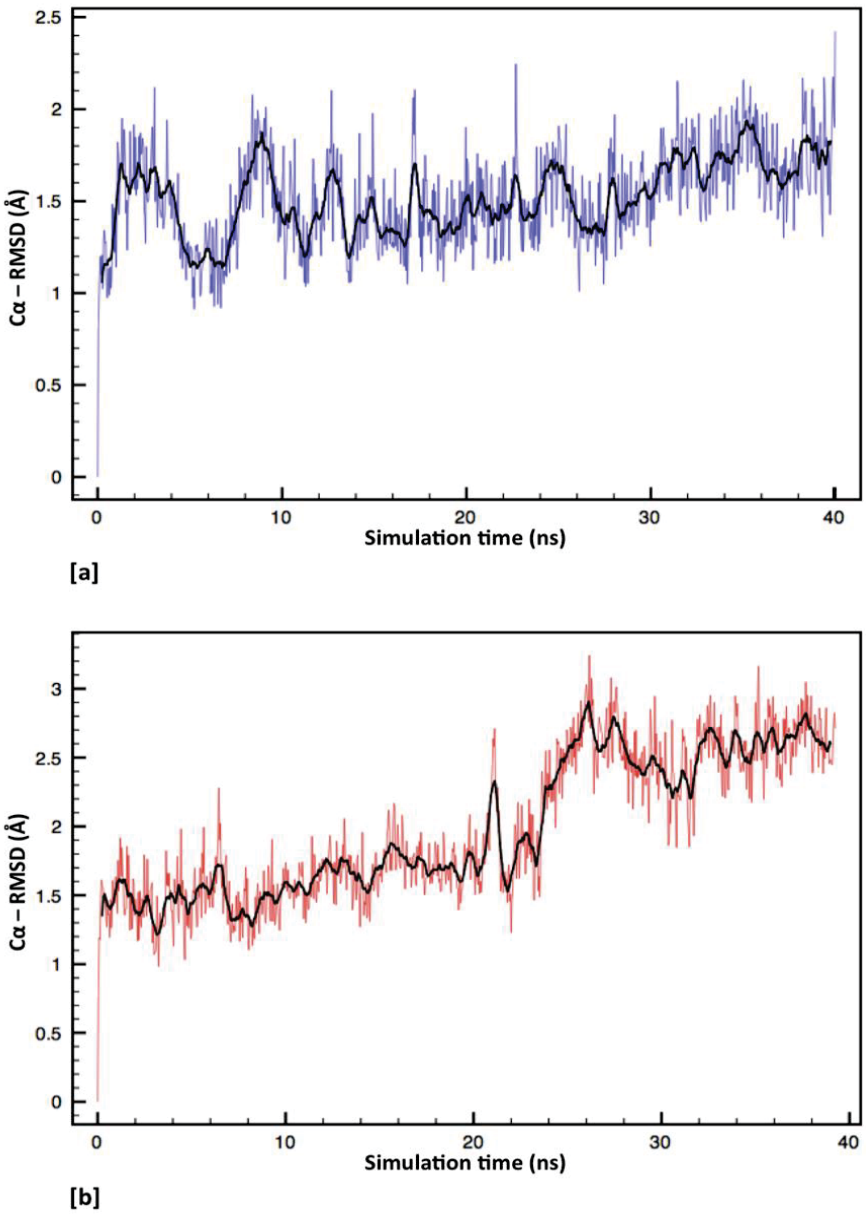
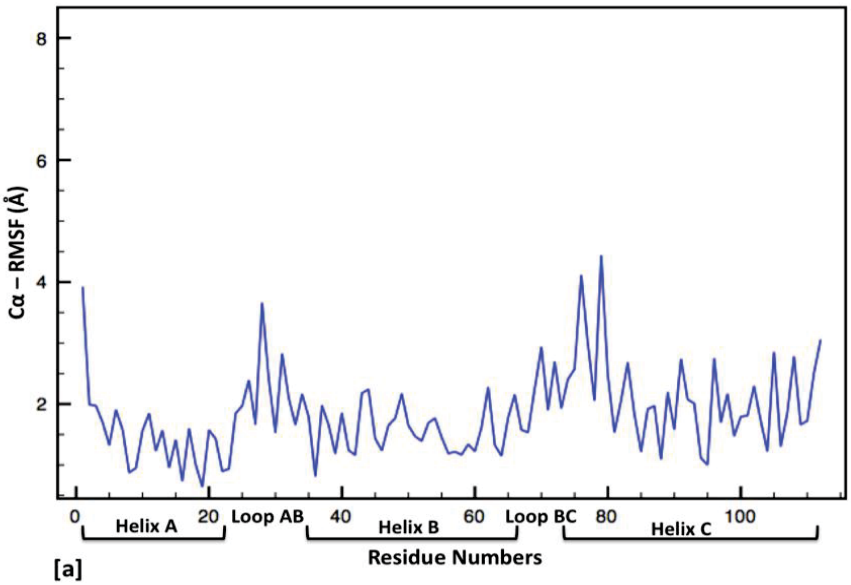


Figure 7-2. Cα-RMSD profile of **[a]** USR1 in blue and **[b]** DSR1 in red, calculated from the trajectory. The running average (black) over 25 ns was plotted using *xmGrace* to improve clarity.

C_α-RMSD trajectories were analysed and compared between USR1 and DSR1. The USR1 system was found to be more rigid throughout the 40 ns simulation time although it has a little conformational change with respect to the initial model. The final C_α-RMSD for USR1 has reached approximately 1.8 Å after 40 ns. On the other hand the DSR1 system has showed a conformational change to approximately 1.8 Å with respect to the initial model after 24 ns simulation time. Furthermore the C_α-RMSD tends to increase to a final value of 2.6 Å after 40 ns, suggesting that the DSR1 system is more flexible when compared to USR1, consistent with structural differences in internal side chains seen in the two crystal structures.

7.2.1.1.2 Calculation of RMSF for each amino acid

The Root Mean Square Fluctuations (RMSF) of Cα atoms are used to measure the deviation of individual residues from their initial structural positions averaged over the trajectory. The resulting RMSF profile of individual amino acids for USR1 and DSR1 are shown in Figure 7-3.



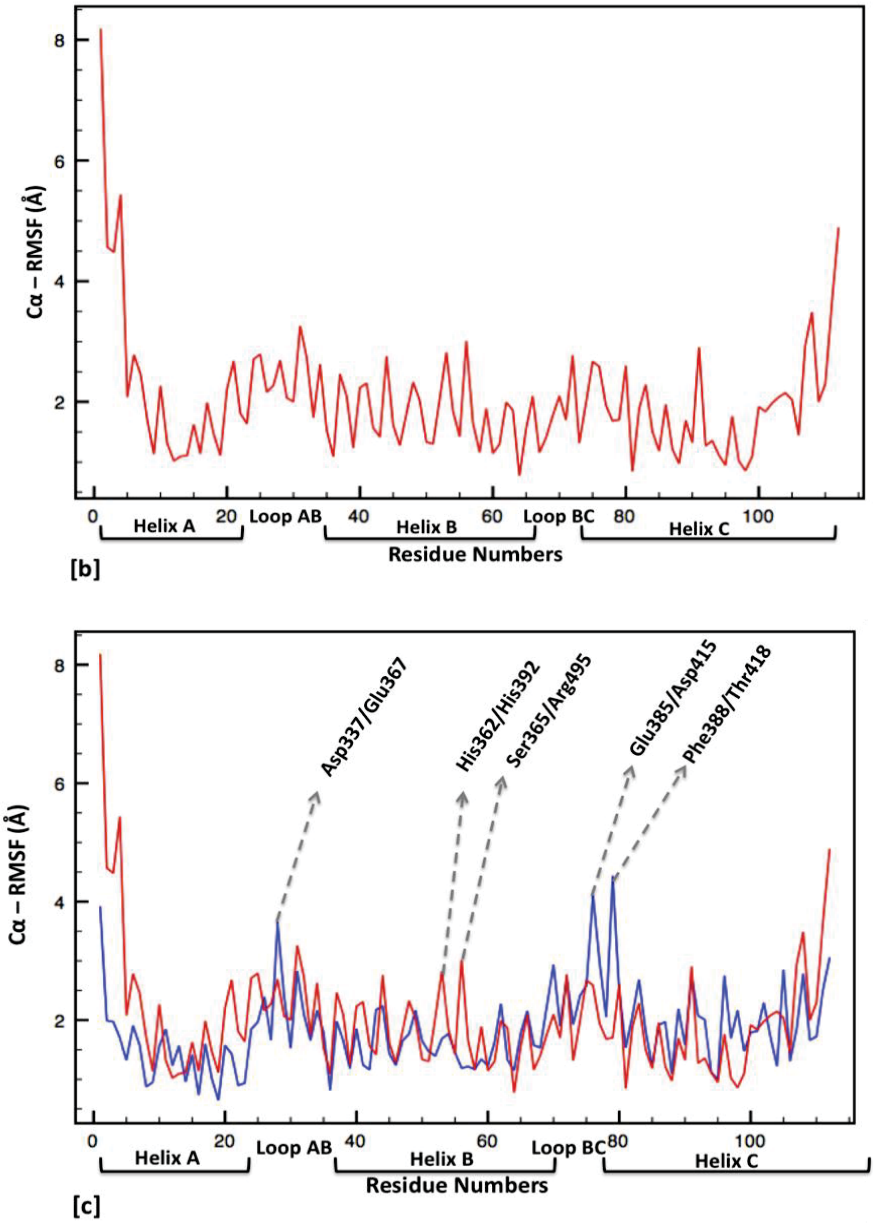


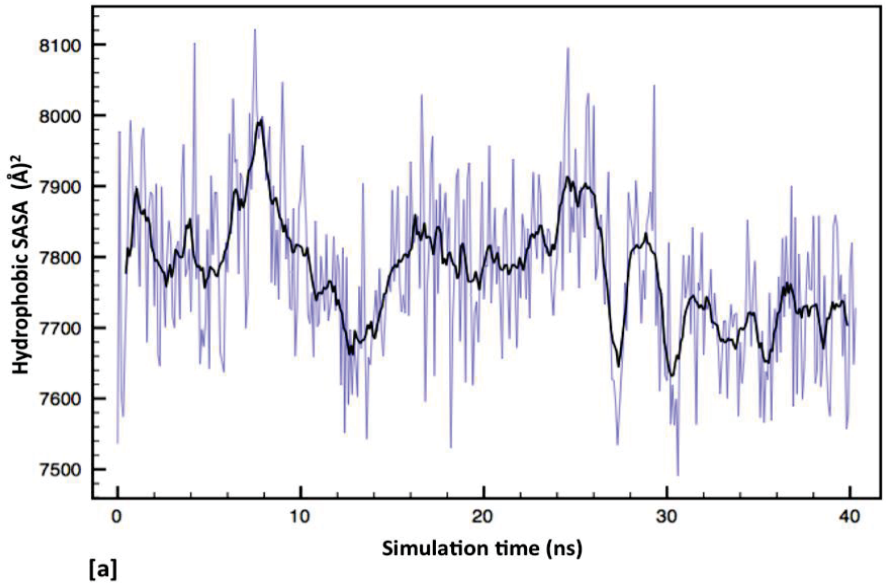
Figure 7-3. RMSF profile of individual Cα-atoms averaged over the trajectory from their original position for **[a]** USR1 in blue, **[b]** DSR1 in red and **[c]** Comparison of USR1 and DSR1. Residues showing greater RMSF are labelled.

The root-mean-square fluctuation (RMSF) of the backbone C_α- atoms is used to measure the fluctuation of individual residues with respect to simulation time. During minimisation equilibrium, the systems (USR1 and DSR1) are subjected to 40 ns simulation at 310 K and the C_α-RMSF profiles were calculated from the trajectories as shown in Figure 7-3. Although C_α-RMSF shows a large conformational change between USR1 and DSR1 for residues in the loop-AB, the kink region in helix B and the N-

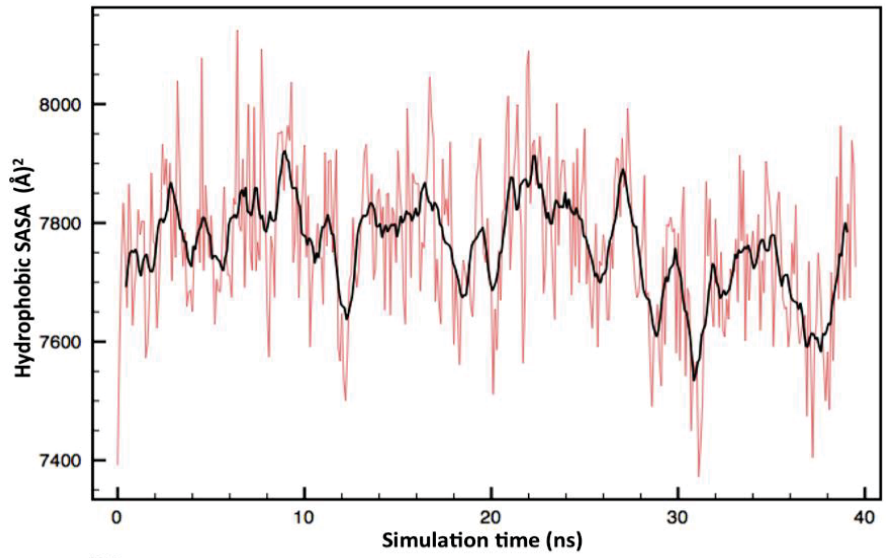
terminal region of helix C. The C_α-RMSF of Asp337 in the loop AB for USR1 is 3.7 Å, greater compared to Glu367 for DSR1 to 2.6 Å. The RMSF of Glu385 for USR1 is 4.1 Å compared to 2.6 Å for Asp415 in DSR1. Due to the presence of the hydrophobic residue (Phe388) for USR1 at the heptad 'b' position in helix 'C' it shows higher RMSF of 4.2 Å when compared to 1.7 Å for a polar Thr418 residue at the N-terminal region of helix 'C' in DSR1, as the position 'b' amino acids in the heptad pattern for coiled-coil proteins are typically polar and exposed to the surrounding solvent molecules. His362 and Ser365 at the kink region in USR1 show a C_α-RMSF of 1.7 and 1.2 Å compared to the residues His392 and Arg495 to 2.8 and 3.0 Å, this may be due to the presence of two water molecules forming hydrogen bonds with the side chain and backbone carbonyl oxygen (seen in the x-ray structure, Figure 4-11) stabilising the main chain along the kink region for USR1 that were missing for DSR1. It was previously reported that the higher RMSF is due to the fluctuation of residues forming hydrogen bonding and not from the hydrophobic core [Mirijianian *et al.,* 2007].

7.2.1.1.3 Calculation of Hydrophobic SASA

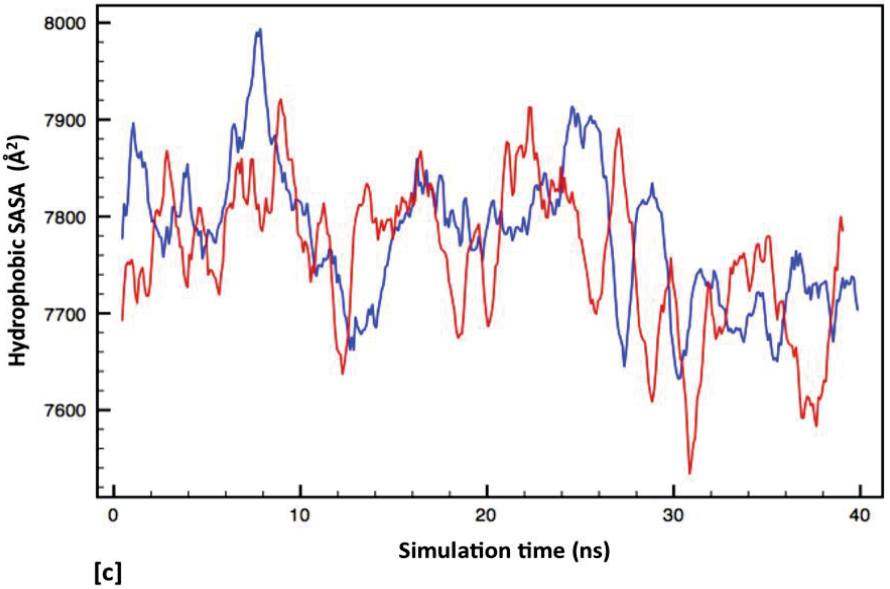
In the current analysis, we have calculated the hydrophobic solvent accessible surface area (SASA) of USR1 and DSR1 with respect to simulation time, shown in Figure 7-4. The behaviour of the protein structure in the presence of solvent can be strongly influenced by the interaction of amino acid side chains with the surrounding solvent molecules since proteins are folded preferentially burying hydrophobic side chains from the external solvent and stabilising the folded state, this is termed the hydrophobic effect [Kauzmann*,* 1959]. From the hydrophobic SASA analysis one can quantitate the exposure of hydrophobic burial with respect to simulation time by calculating the change in free energy of proteins to solvent interactions.


Hydrophobic SASA (Å)²
8100
8000
7900
7800
7700
7600
7500
0
10
20
30
40
Simulation time (ns)

[a]


Hydrophobic SASA (Å)²
8000
7800
7600
7400
0
10
20
30
40
Simulation time (ns)

[b]


Hydrophobic SASA (Å²)
8000
7900
7800
7700
7600
0
10
20
30
40
Simulation time (ns)

[c]

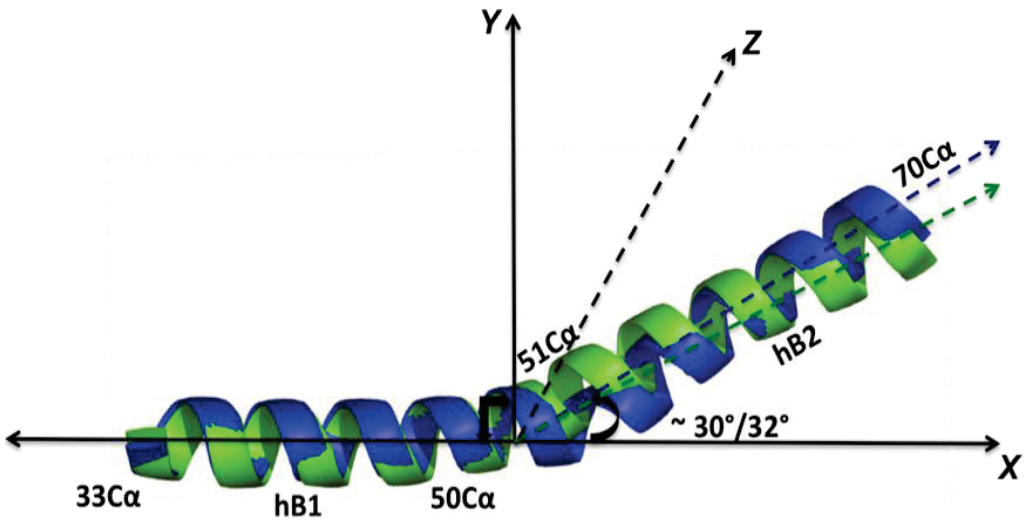
Figure 7-4. The hydrophobic Solvent Access Surface Area profile was calculated for **[a]** USR1 in blue and **[b]** DSR1 in red, at each trajectory and plotted with respect to simulation time. The running average (black) over 25 ns was plotted using *xmGrace* to improve clarity. **[c]** Comparison of hydrophobic- SASA between USR1 (blue) and DSR1 (red).

Water plays an important role in the association of proteins and ligand. The influence of water molecules around the surface of the protein can be examined in more details by monitoring the protein hydrophobic-SASA. This can be achieved by measuring the free energy of solvating polar or non-polar groups [Miller *et al.,* 2009]. In the case of the USR1 and DSR1 systems, the SASAs of hydrophobic side chains were measured with respect to the 40 ns simulation time. The SASA for initial models were measured as approximately 7700 $Å^2$ for the USR1 and DSR1 systems. The SASA for DSR1 increased to 8000 $Å^2$ at 8 ns simulation time (Figure 7-4[c]) and to 7900 $Å^2$ at 10 ns for USR1. A greater decrease in the SASA to approximately 7500 $Å^2$ for USR1 was observed at 30 ns simulation compared to 7700 $Å^2$ for DSR1. A consistent decrease in SASA was observed for USR1 along 40 ns simulation when compared to DSR1, this suggests that USR1 is packed more closely in solution than DSR1.

7.2.1.1.4 Calculation of bending-angle (Helix II)

A significant kink was observed in Helix-II (*referred to* Helix-B in Chapter 4) in the x-ray structure for USR1 and DSR1 as shown in Figure 4-7. In order to understand the stability of the USR1 and DSR1 systems, the bending-angles (hB1, hB2) of helix 'B' were analysed and the degree of bending during the simulation was measured. hB1 corresponds to the residues from V342 to A361 for USR1 and V372 to A391 for DSR1 along the *X*-axis, and hB2 corresponds to the residues from H362 to Q379 for USR1 and H392 to T409 for DSR1 (Figure 7-5[a]). The variation of hB2 along the *YZ*-plane with respect to hB1 along *X*-axis was measured to 40 ns simulation time and the orientation of hB2 in *YZ*-plane corresponds to the helix 'B' bending-angle.

Although there was no difference in the degree of variation of hB2 between the initial models (x-ray structures) for USR1 (~ 30°) and DSR1 (~32°), a significant drop to approximately 26° for USR1 under 2 ns simulation time and a sudden increase in the degree of bending for DSR1 to approximately 38° was observed. It can also be seen from Figure 7-5[b], that the degree of bending between USR1 and DSR1 are not uniformly distributed throughout 40 ns simulation time. The variation in the distribution indicates that hB2 for DSR1 was bending more freely reaching a highest bending-angle of over 38° frequently, compared to twice for USR1 along the *YZ*-plane and the direction of lowest bending-angles reached was approximately 28°, not reached for DSR1. The drop in the bending-angle may enhance the backbone hydrogen bonding at the helix kink region (residues Ser360 to Ser365 for USR1 and Thr390 to Arg395 for DSR1) and packing between the helices as shown from the hydrophobic-SASA analysis. Throughout the simulation experiments the highest and lowest distribution of hB2 for USR1 and DSR1 are not similar and hB2 from DSR1 shows greater flexibility compared to USR1. This suggests that the kink region in DSR1 is more flexible compared to the kink region present in USR1.



[a]

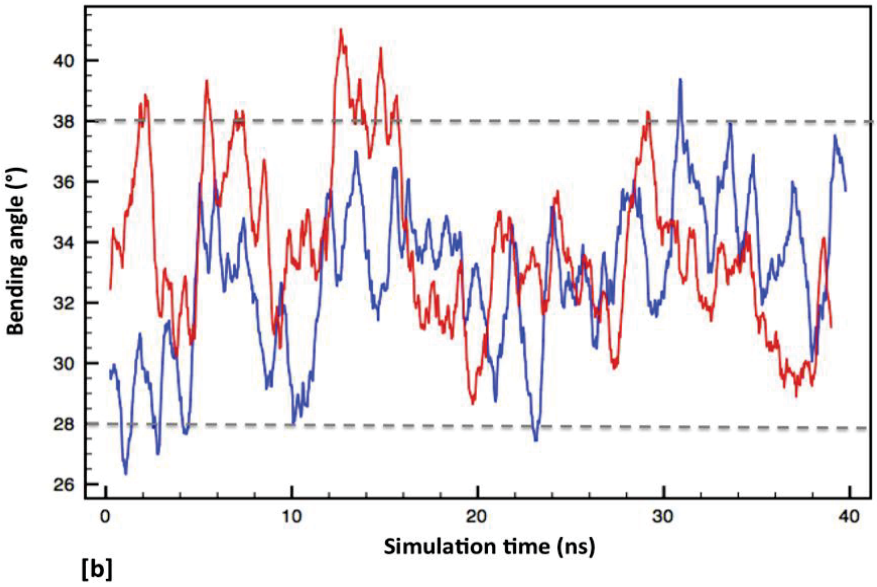


Figure 7-5. [a] Schematic representation of the bending-angle between hB1 and hB2 from USR1 and DSR1. Structural cartoon representation of the backbone Cα overlay structure of helix 'B' alone from USR1 (green) and DSR1 (blue). **[b]** Distribution of the bending-angle between hB2 with respect to hB1 for a 40 ns equilibrium dynamic simulation experiment.

7.2.2 Steered Molecular Dynamics (SMD) *or* Protein un-folding Simulation

Molecular dynamics (MD) protein un-folding simulations have recently became an important tool to understand and analyse the protein folding and un-folding, thus providing a dynamics structural and functional relationship for proteins. Although atomic force microscopy (AFM) provides a dynamic force required for unfolding the protein molecules, the conformational change of the stretched proteins cannot be observed, and x-ray crystallography and NMR spectroscopy are not suitable for this purpose.

Steered molecular dynamics (SMD) is a novel method to observe the intermediate protein trajectories during unfolding and also are useful to study the mechanical properties of proteins [Israelewitz *et al.,* 2001] as a function of simulation time.

SMD simulations at constant velocity stretching (SMD-CV) were implemented for the USR1 and DSR1 systems. The size of the periodic water box was increased when compared to the size of the water box in equilibration analysis, to accommodate the unfolded proteins (shown in Figure 7-6). The complete structures, including solvent, now contain 140213 atoms for USR1 and 146666 for DSR1.

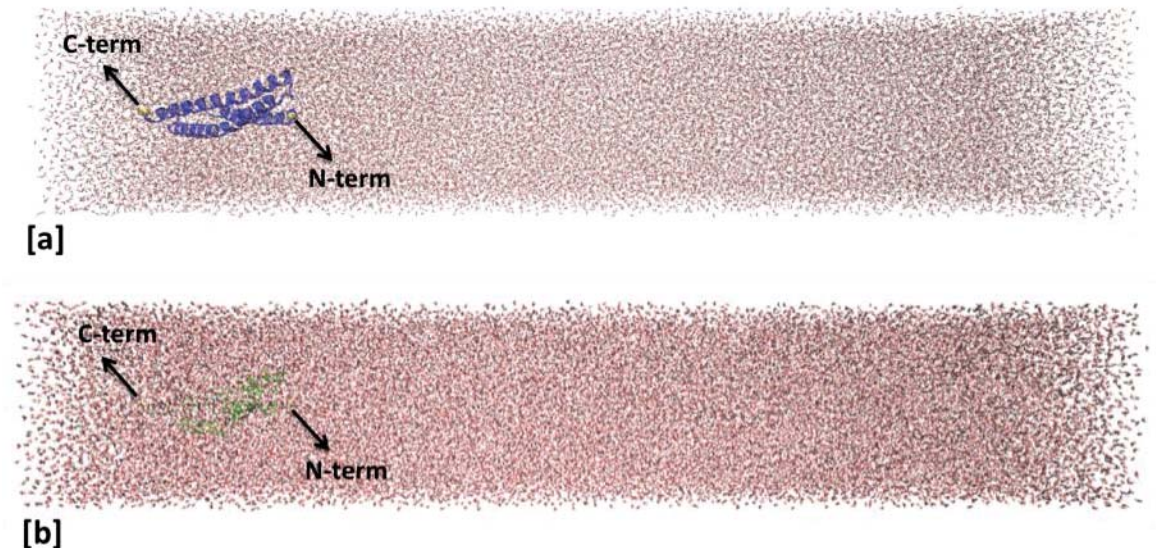


Figure 7-6. VMD ribbon representation of **[a]** USR1 (blue) and **[b]** DSR1 (green) oriented on the *X*-axis in an extended explicit water (pink/grey) box.

Steered molecular dynamics at constant velocity (SMD-CV) were carried out with the solvated (neutralising by adding water and ions) and equilibrated structure, by fixing the Cα atom of the C-terminus (^{112}Cα) of the domain and then applying a force to the Cα atom of the N-terminus (^{1}Cα) in the other direction along the *X-axis*. Each domain was elongated gradually with time (t) until the polypeptide chain was completely straightened as shown in Figure 7-7. Although in molecular dynamics one terminus was fixed during the simulation, the external force applied on the other terminus is translated to the fixed terminus showing equal forces applied in both directions, this is defined as harmonically restrained.

The force experience by the pulled atom was given by the following equation, $F = K(vt - x)$ [Liu *et al.,* 2011] with a constant velocity of 10 Å/ns (1 nm/ns) along the *X-axis* in a time step of 2 fs (femtoseconds). Where *x* is the displacement of the pulled atom from

its original position, *v* is the specified velocity used for pulling the atom and *K* is a harmonic spring constant. From the above equation the force extension profile was obtained from the analysis of trajectories.

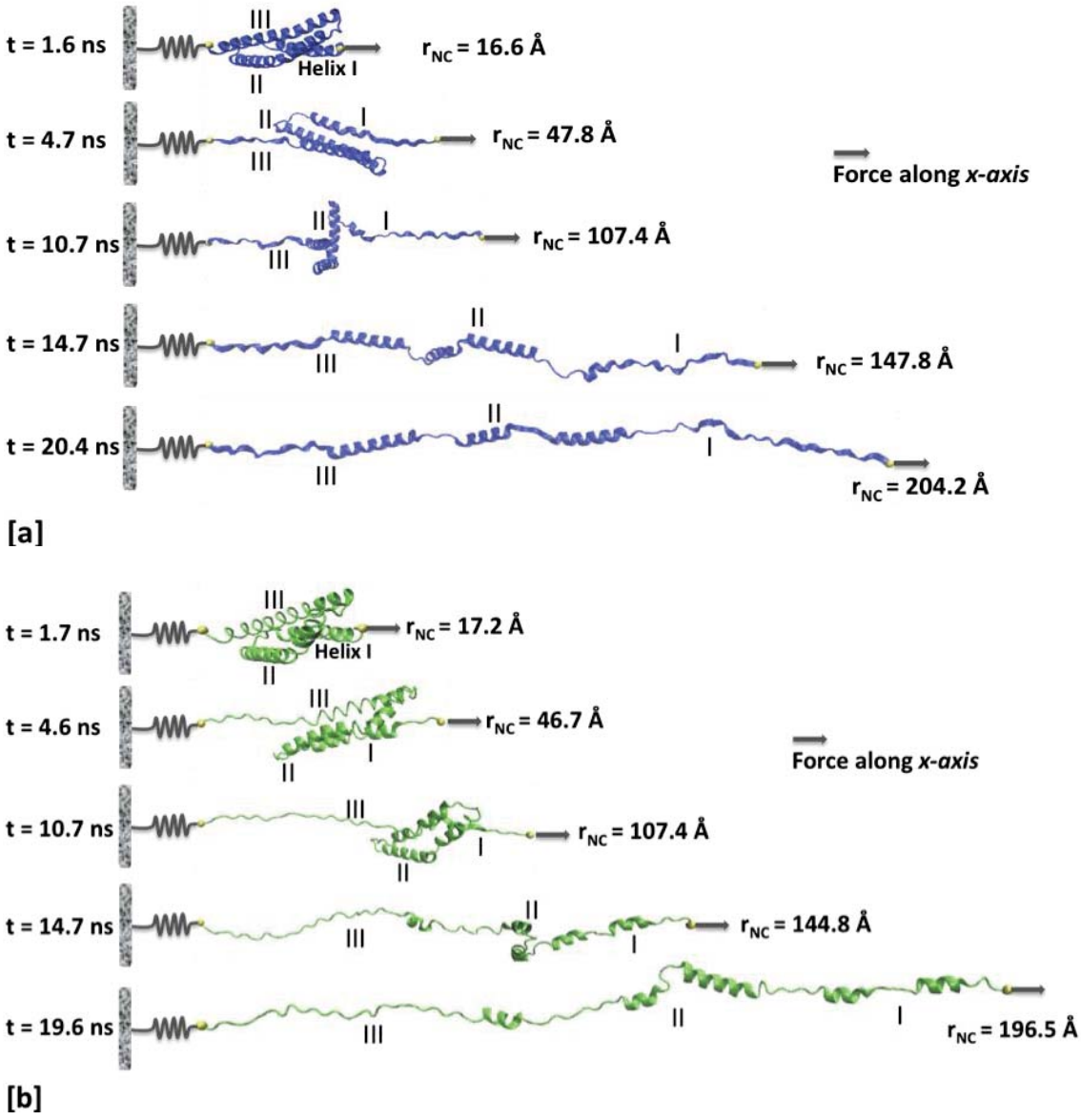
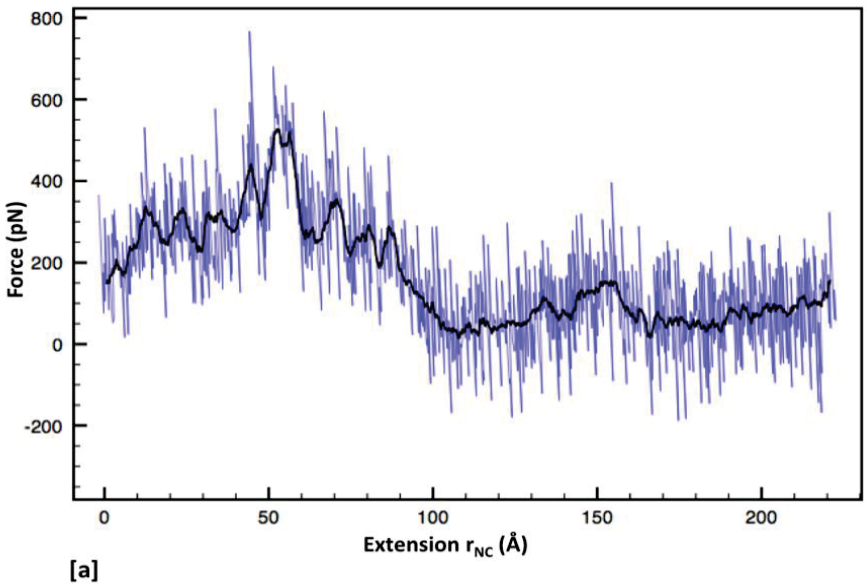


Figure 7-7. Unfolding intermediates of **[a]** USR1 in blue and **[b]** DSR1 in green, obtained at given time 't' ns from SMD simulations. Where r_{NC} is the position of the N-terminal residue as a function of simulation time and the direction of force with which the molecules was pulled was shown. Water molecules are not displayed for the clarity of protein molecules.

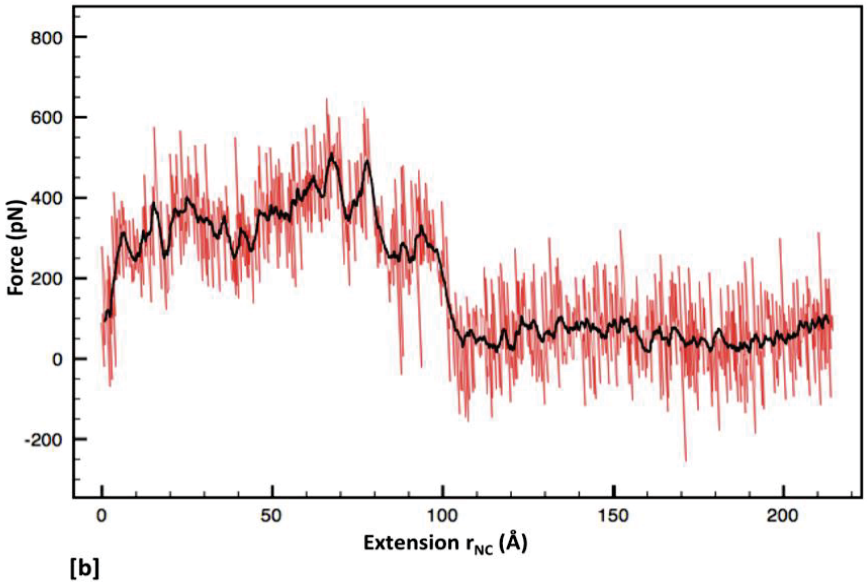
In our SMD simulations, the trajectory file shows the unfolding intermediates of the USR1 and DSR1 systems and as the unfolding occurs, the buried residues are exposed to the surrounding solvent molecules.

7.2.2.1 SMD Trajectory analysis of USR1 and DSR1

Force-extension profiles were calculated from the trajectories as a function of simulation time and the range of unfolding forces are shown in Figure 7-8.



[a]



[b]

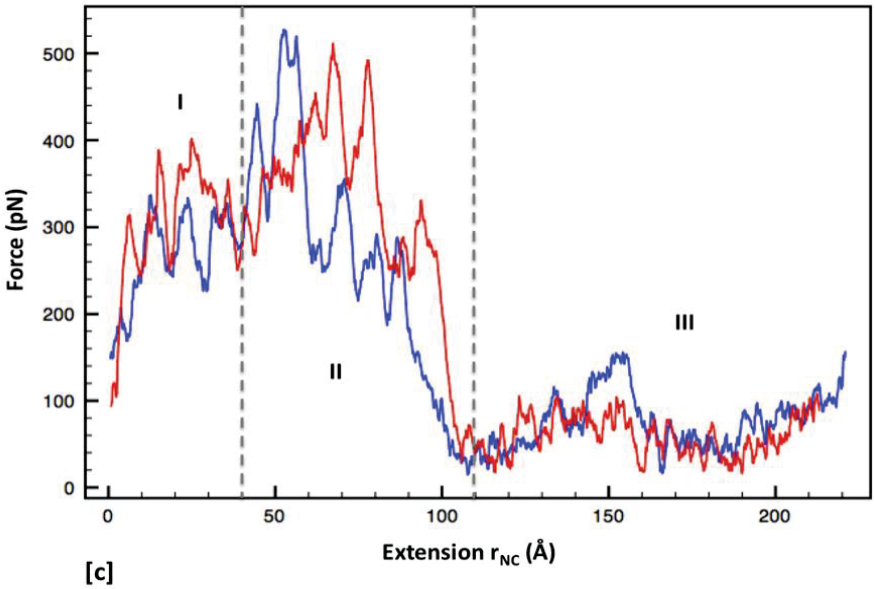


Figure 7-8. Force-extension profile from unfolding simulation of **[a]** USR1 in blue and **[b]** DSR1 in red. Black is the running average calculated from *xmGrace* with a 50 ns window size. **[c]** Comparison of force extension profile of USR1 (blue) and DSR1 (red) observed when the proteins were unfolded at a constant velocity of 10 Å/ns. Where r_{NC} is the position of N-terminal residue as a function of simulation time. The unfolding process was divided into three regions: I) preburst, II) main burst and III) post burst.

During SMD simulation the initial force reaches from ~100 pN to ~320 pN, with a rapid drop to ~240 pN for the DSR1 system with an N-terminal extension of 10 Å, whereas for USR1 the force was increased gradually and reached ~340 pN with an N-terminal extension of 15 Å along the *X*-axis and a rapid drop in force to ~240 pN with in 5 Å was observed.

The unfolding process was divided into three regions: **I)** preburst: from 0 to 40 Å extension, during which the N- and C-terminal backbone hydrogen bonds along with the interaction of the C-terminal helix with loop 'AB' was ruptured with protein reaching a maximum extension of ~100 Å (distance between N- and C- terminal residues). **II)** main burst: from 40 to 110 Å, during which the interactions between the helices begins to breakdown and the protein reaches a maximum extension of ~160 Å. **III)** postburst with extension from 110 to 220 Å during which the proteins are fully extended (>200 Å) with the polypeptide chain approaching its maximum length.

During prebust (I), a consistent force of ~340 pN was observed for USR1 whereas for DSR1 the force varies between ~300 to 400 pN as shown from the Figure 7-8[c]. Regions within the main burst suggest that the hydrophobic residues at heptad 'a' and 'd' positions along the helix 'A', 'B' and 'C' were exposed to the surrounding water molecules and contribute an additional stress to the polypeptide during unfolding, this stress was relieved by the water molecules by increasing hydration around the backbone [Kusunoki *et al.,* 2004a; Ortiz *et al.,* 2005]. During the main burst the higher forces of ~520 and ~500 pN for USR1 and DSR1 system were recorded, however the force consistently drops when the extension reached ~110 Å. A significantly smaller force of ~150 pN was observed in the postburst phase during the extension between ~140 and 160 Å for USR1. This shows that there may still exist partial tertiary structure at the longer extension for USR1 compared to DSR1, suggesting that the helical packing in USR1 is more stable than DSR1.

7.3 Summary

In this chapter, MD simulation was used to study the motion and the characteristics of the USR1 and DSR1 domains. We analysed the data from the trajectories and further discussion are made in chapter 8.6. Comparative analysis of a significant bend in Helix-II (Helix B) of USR1 and DSR1 has been conducted. Steered molecular dynamics (SMD) revealed an intermediate state of protein during forced-unfolding events from which the force-extension profile as a function of simulation time were calculated and the stability and mechanical properties of USR1 and DSR1 domain are compared.

Trajectories were analysed and images were rendered using VMD [Humphrey *et al.,* 1996]. Running averages were calculated and graphs were plotted using *xmGrace* software (http://plasma-gate.weizmann.ac.il/Grace). The distances between the N- and C-termini were measured using VMD.

8. Conclusions & Future Work

8.1 Conclusions

Crystal structures of the N-terminal first spectrin repeat domain from Utr (USR1) and Dys (DSR1) were solved to 1.95 and 2.3 Å. Although there exists similar structures of multiple spectrin repeats from α-actinin [Djinovic-Carugo *et al.,* 1999; Ylanne *et al.,* 2001], spectrin [Grum *et al.,* 1999; Kusunoki *et al.,* 2004a] and plectin [Sonnenberg *et al.,* 2007], these are the first structures determined for any of the spectrin repeats from Utr and Dys. The structure of a 6 amino acid longer version of the N-terminal spectrin repeat domain (USR1L) showed that the linker region between N-terminal multiple spectrin repeat domains are helical and continue into the A' helix of the second repeat domain. These new structures also aid in designing new sequence alignment and boundaries for the remaining spectrin repeat domains from Utr and Dys.

In vitro actin-binding assays showed that the N-terminal spectrin repeat domains from Utr and Dys have no intrinsic affinity towards F-actin, this is likely due to the presence of acidic residues (Asp and Glu) along the surface of the helices 'A' and 'B' from the structure analyses. Sequence analysis showed that the N-terminal domains have a pI of about 4.0, the most acidic part of the repeat domains in Utr and Dys explaining the lack of intrinsic binding to acidic F-actin. However the first N-terminal spectrin repeat from Utr contributes to the Utr actin binding domain's (ABD) ability to bind F-actin more effectively compared to ABD's alone [Sutherland-Smith *et al.,* 2003].

Although it was suggested that multiple spectrin repeat domains are thermodynamically more stable than a single repeat domain [Kahana *&* Marsh*,* 1994; Calvert *et al.,* 1996; Menhart *et al.,* 1996], in our current studies both N-terminal multiple (USR12 and DSR12) and single repeat (USR1 and DSR1) domains from Utr and Dys have similar stabilities with respect to elevated temperature and urea concentration. Interestingly the N-terminal spectrin repeat domains from Utr and Dys showed refolding after having been unfolded at high temperatures.

Furthermore equilibrium molecular dynamics simulation suggests that the kink region in helix 'B' for USR1 is more stable than DSR1, this may be due to the presence of water molecules forming hydrogen bonds with the backbone carbonyl and Ser side chain groups in USR1 and is unavailable at the kink region for DSR1. Steered molecular dynamics simulation showed that DSR1 is more flexible and required less force for unfolding compared to USR1. It was also shown that partial tertiary structure exists for USR1 at longer extension compared to DSR1, revealing that the packing between the helices for USR1 is more compact and structurally more stable than for DSR1.

8.2 Future directions

Although Utr is an autosomal homolog of Dys and can act as a potential therapeutic replacement in *mdx*-mice [Tinsley *et al.,* 1996; Sonnemann *et al.,* 2009], full characterisation of similarities and differences of these two large multidomain proteins is required which includes structure, function and dynamics analysis.

A recent finding of pathogenic mutations in Dys protein spectrin repeats shows that it is important to understand the complete atomic structure of the Utr and Dys proteins. This can be achieved on protein fragments either by using x-ray crystallography or NMR technique. Although x-ray crystallography will show the atomic structure of the proteins, the dynamics of the proteins cannot be directly measured, this can be achieved using NMR spectroscopy. Although the structure of N-terminal multiple spectrin repeat domains was not solved, the 6 amino acids longer version of USR1 suggests that the second spectrin domain for Utr may be oriented ~180° compared to the other homologous multiple spectrin repeats, this suggestion can be examined using small angle x-ray scattering (SAXS) experiments for N-terminal multiple spectrin Utr repeat, as SAXS is a good technique to Investigate proteins for macromolecular folding, linker flexibility between domains, shape and assembly state in solution [Putnam *et al.,* 2007].

A further investigation using molecular dynamic simulations is required to understand the flexibility of multiple repeat domains between Utr and Dys and to compare the

structural geometry, hydrophobic surface analysis, stability of the proteins and packing between the helices. It is also necessary to know the difference between cooperative unfolding of multiple repeat domains to that of single repeats using steered molecular dynamics and compare these calculations with experimental results obtained using atomic force microscopy (AFM).

In our current research I have showed that there is no difference in the stability between N-terminal multiple repeats compared to single repeat domains and furthermore the repeat domains can refold after unfolding at higher temperatures as analysed from the CD spectra. However a complete understanding of the comparative stability of the rod domains between Utr and Dys using CD spectroscopy and tryptophan fluorescence spectroscopy is required.

Although I have showed that the N-terminal acidic spectrin repeat domain alone cannot bind to F-actin a further investigation on to the basic spectrin repeat domains alone from Utr and Dys with F-actin is required.

9. Appendices

9.1 Plasmid Maps

9.1.1 pPROEx HTb (Invitrogen)

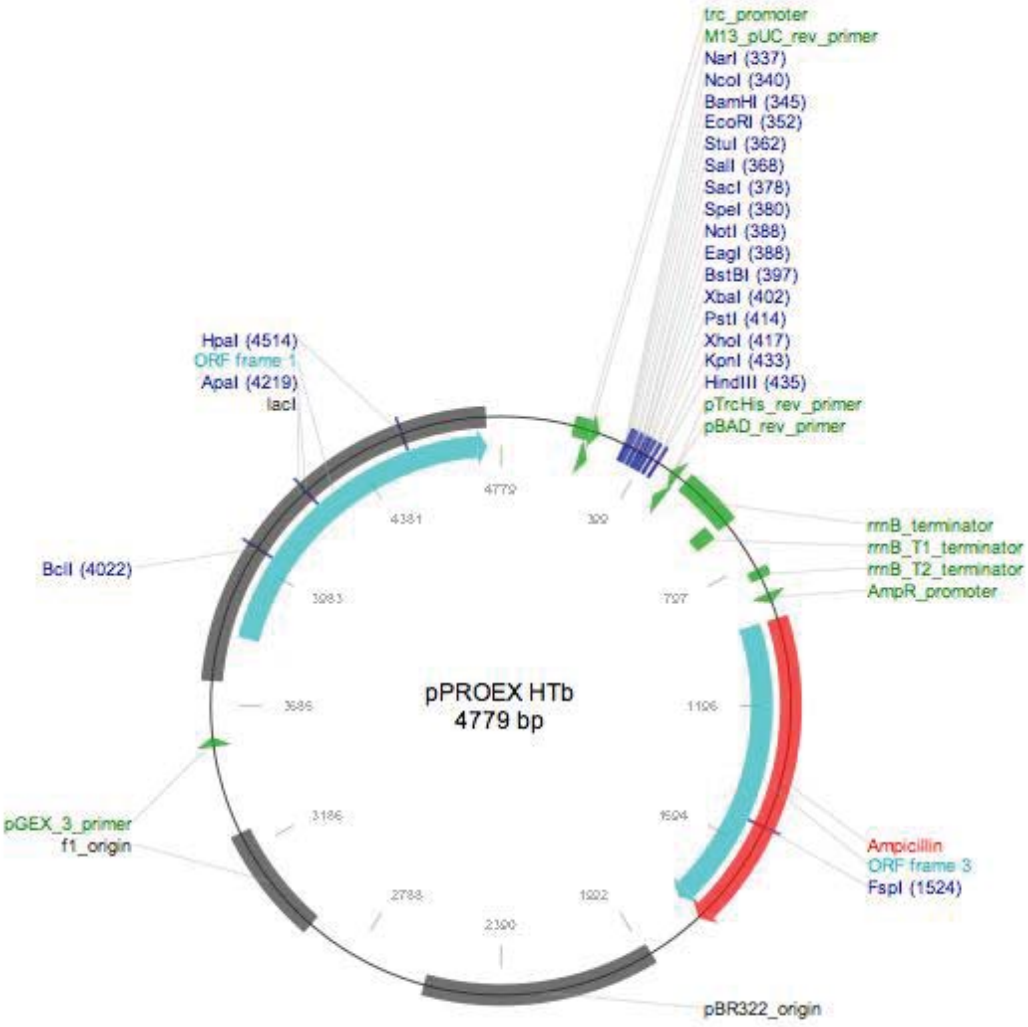


Figure 9-1. Plasmid map of pPROEx HTb (invitrogen)

Constructs

Vector	*Tag*	*Enzyme cleavage*	*gene of interest*
pPROEx HTb	$(HIS)_6$	rTEV protease	Utr N-terminal first spectrin repeat
pPROEx HTb	$(HIS)_6$	rTEV protease	Dys N-terminal first spectrin repeat
pPROEx HTb	$(HIS)_6$	rTEV protease	Utr N-terminal first two spectrin repeats
pPROEx HTb	$(HIS)_6$	rTEV protease	Dys N-terminal first two spectrin repeats

9.1.2 pGEX 4T1 (Invitrogen)

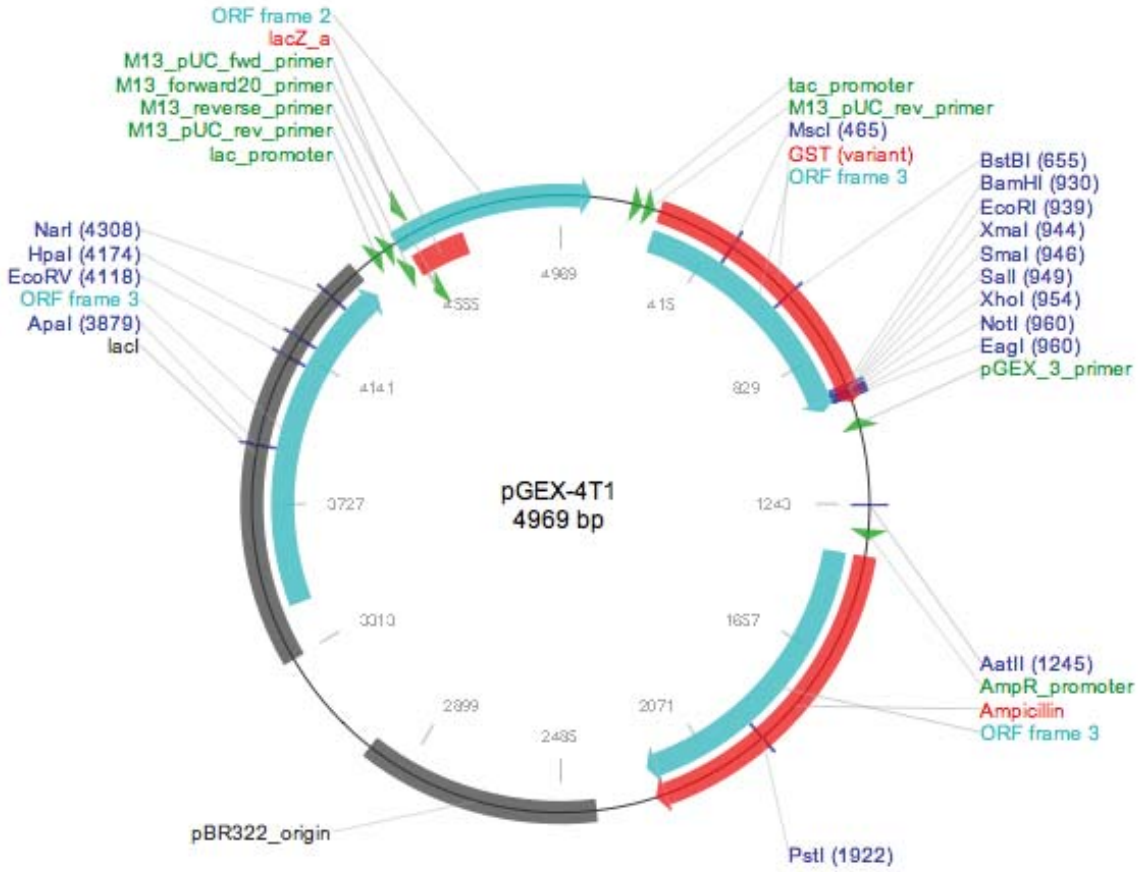


Figure 9-2. Plasmid map of pGEX 4T1 (invitrogen)

9.1.3 Expression constructs made for future studies

Vector	*Tag*	*Enzyme cleavage*	*gene of interest (domain, amino acid)*	*Solubility*	*Purification*	*Crystallisation*
pPROEx HTb	$(HIS)_6$	rTEV protease	Dys spectrin repeat – 11 (basic repeat, 1460 – 1578)	Insoluble	------	------
pPROEx HTb	$(HIS)_6$	rTEV protease	Dys ABD + first spectrin repeat (12 – 455)	Soluble	OK, exists as monomer and dimer in solution	No crystals were obtained
pPROEx HTb	$(HIS)_6$	rTEV protease	Utr ABD + first spectrin repeat (30 – 425)	Soluble	OK, exists as monomer and dimer in solution	No crystals were obtained
pPROEx HTb	$(HIS)_6$	rTEV protease	Utr ABD + first two spectrin repeats (30 – 537)	Soluble	OK, exists as monomer and dimer in solution	No crystals were obtained
pPROEx HTb	$(HIS)_6$	rTEV protease	Dys C-terminal domains (WWEF1EF2ZZ, 3049 – 3356)	Insoluble	------	------
pPROEx HTb	$(HIS)_6$	rTEV protease	Dys ZZ-domains (3304 – 3354)	Insoluble	------	------
pGEX-4T1	gst	Thrombin	Dys ZZ-domains (3304 – 3354)	Soluble	OK, but precipitated after cleaving solubility tag.	------
pGEX-4T1	gst	Thrombin	Dys C-terminal domains (WWEF1EF2ZZ, 3049 – 3356)	Soluble	OK, but precipitated after cleaving solubility tag.	------
pSUMO	$(HIS)_6$	rTEV protease	Dys ZZ- domains (3304 – 3354)	Soluble	OK, but precipitated after cleaving solubility tag.	------

9.2 Gel filtration standard curve

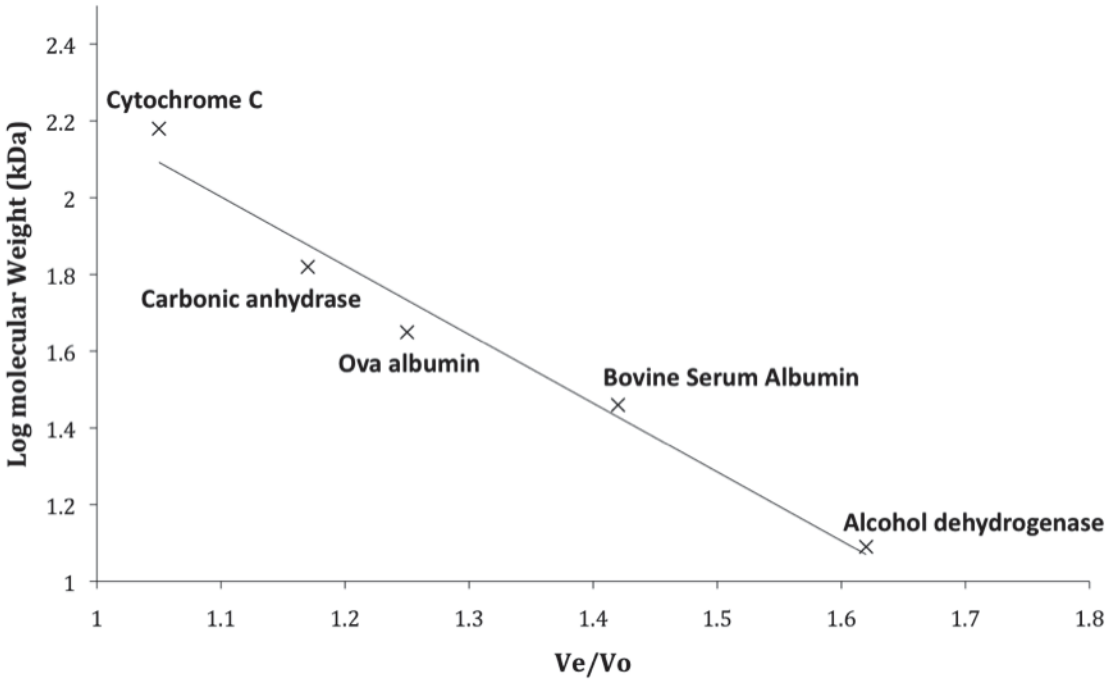


Figure 9-3. Calibration of Superdex S75 high performance columns (GE Healthcare)

The elution volume (V_e) decreases linearly with the logarithm of the molecular hydrodynamic volume. The Superdex S75 column was calibrated using five standard samples (cytochrome C (12.4 kDa), carbonic anhydrase (29 kDa), ova albumin (45 kDa), bovine serum albumin (66 kDa), alcohol dehydrogenase (150 kDa)), and the void volume (V_o) was determined by large standard such as β–amylase (200 kDa). The standards and samples were run in phosphate buffer (50 mM K_2HPO_4/KH_2PO_4 pH 8.0, 150 mM NaCl) with a flow rate 0.4 ml/min at 4 °C. The elution volumes of the standards are divided by the void volume (Ve/Vo) and plotted against the log of the standard's molecular weights.

10. References

Acsadi G, Moore SA, Cheron A, Delalande O, Bennett L, Kupsky W, Rumeur EL, Hubert JF, (2012). Central domain causes protein misfolding and mild Becker muscular dystrophy. *J Biol Chem.,* 287: 18153-18162.

Adams PD, Grosse-Kunstleve RW, Hung LW, Loerger TR, McCoy AJ, (2002). PHENIX: building new software for automated crystallographic structure determination. *Acta Cryst* D58: 1948-1954.

Alderton JM, Steinhardt RA, (2000). Calcium influx through calcium leak channels is responsible for the elevated levels of calcium-dependent proteolysis in dystrophin myotubes. *J Biol Chem.,* 275: 9452-9460.

Altmann SM, Grunberg RG, Lenne PF, Ylanne J, Raae A, Herbert K, (2002). Pathways and intermediates in forced unfolding of spectrin repeats. *Structure.,* 10: 1085–1096.

Amann KJ, Renley BA, Ervasti JM, (1998). A cluster of basic repeats in the dystrophin rod domain binds F-actin through an electrostatic interaction. *J Biol Chem.,* 273: 28419-28423.

Amann KJ, Guo AWX, Ervasti JM, (1999). Utrophin lacks the rod domain actin-binding activity of dystrophin. *J Biol Chem.,* 274: 35375-35380.

An X, Guo X, Zhang X, Baines AJ, Debnath G, Moyo D, Salomao M, Bhasin N, Johnson C, Discher D, Gratzer WB, Mohandas N, (2006). Conformational stabilities of the structural repeats of erythroid spectrin and their functional implications. *J Biol Chem.,* 281: 10527-10532.

Armstrong S, Shivell C, (2001). Ischemic loss of sarcolemmal dystrophin and spectrin: correlation with myocardial injury. *J Mol & Cell Card.,* 33(6): A4-1179.

Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N, (2010). ConSurf: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucl Acids Res.,* 38: 529-533.

Baker NA, (2004). Poisson-Boltzmann methods for biomolecular electrostatics. *Meth Enzymol.,* 383: 94-118.

Baker NA, McCammon JA, (2005). Electrostatic interactions. In, structural bioinformatics. *Volume 44 (eds Weissig H and Bourne PE), John Willey & Sons, Inc, Hoboken, NJ, USA.*

Banner DW, Kokkinidis M, Tsernoglou D, (1987). Structure of the ColE1 rop protein at 1.7 A resolution. *J Mol Biol.,* 196(30): 657-675.

Banuelos S, Saraste M, (1998). Structural comparisons of calponin homology domains: implications for actin binding. *Structure.,* 6(11): 1419-1431.

Baron MD, Davison MD, Jones P, Critchley DR, (1987). The sequence of chick actinin reveals homologies to spectrin and calmodulin. *J Biol Chem.,* 262: 17623–17629.

Batchelor CL and Winder SJ (2006). Sparks, signals and shock absorbers: how dystrophin loss causes muscular dystrophy. *Trends Cell Biol.,* 16(4): 198-205.

Bartlett JMS, Stirling D, (2003). “A short history of the polymerase chain reaction". *PCR Protocols.,* 226: 3–6.

Batty TGG, Kontogiannis L, Johnson O, Powell HR, Leslie AWG, (1990). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Cryst* D67: 271-281.

Beggs AH, Hoffman EP, Snyder JR, Arahata K, Kunkel LM (1991). Exploring the molecular basis for variability among patients with Becker muscular dystrophy:

dystrophin gene and protein studies: *Am J of Hum Gen.,* 49: 54-67.

Bennion BJ, Daggett V, (2004). Counteraction of urea-induced protein denaturation by trimethylamine N-oxide: a chemical chaperone at atomic resolution. *Proc. Natl. Acad. Sci USA.,* 101(17): 6433-6438.

Bergfors TM, (1999). Protein crystallography: technique, strategies and tips- a l*aboratory, manual. Ed.1. International University Line*.

Bernstein FC, Koetzle TF, Williams GJB, Meyer EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, M. Tasumi M, (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. *J Mol Biol.,* 112: 535–542.

Bewick GS, Nicholson LV, Young C, O'Donnell E, Slater CR, (1992). Different distribution of dystrophin and related proteins at nerve-muscle junctions. *Neurorepo.,* 3: 857-860.

Bhasin N, Law R, Liao G, Safer D, Ellmer J, (2005). Molecular extensibility of mini-dystrophins and a dystrophin rod construct. *J Mol Biol.,* 352: 795-806.

Bhosle RC, Michele DE, Campbell KP, Li Z, Robson RM, (2006). Interactions of intermediate filament protein synemin with dystrophin and utrophin. *Biochem Biophys Res Commun.,* 346: 768-777.

Biggar WD, Klamut HJ, Demacio PC, Stevens DJ, Ray PN, (2002). Duchenne muscular dystrophy: current knowledge, treatment, and future prospects. *Clin Orthop.,* 401: 88-106.

Blake DJ, Tinsley JM, (1994). "The utrophin and dystrophin genes share similarities in genomic structure." *Hum Mol Genet.,* 2(11): 1765–72.

Blake DJ, Tinsley JM, Davies KE, Knight AE, Winder SJ, (1995), Coiled-coil region in the carboxy-terminal domains of dystrophin and related proteins: potentials for protein-protein interactions. *Trends Biochem.,* 20: 133-135.

Blake DJ, Tinsley JM, Davies KE, (1996). Utrophin: a structural and functional comparison to dystrophin. *Brain Pathol.,* 6: 37-47.

Blake DJ, Weir A, Newey SE, Davies KE, (2002). Function and genetics of dystrophin and dystrophin-related proteins in muscle. *Physiol Rev.,* 82: 291-329.

Blankinship AJ, Gregorevic, Chamberlain, (2006). Gene therapy strategies for Duchenne muscular dystrophy utilizing recombinant adeno-associated virus vectors. *Mol Ther.,* 13: 241-249.

Bloch RJ, Pumplin DW, (1992). A model of spectrin as a concertina in the erythrocyte membrane skeleton. *Trends Cell Biol.,* 2: 186-189.

Bohm G, Muhr R, Jaenicke R, (1992). Quantitative-analysis of protein far UV circular-dichroism spectra by neutral networks. *Prot Eng.,* 5: 191-195.

Bonnet D, Begard E, (1984). Interaction of anilinonaphtyl labeled spectrin with fatty acids and phospholipids: a fluorescence study, *Biochem Biophys Res Commun.,* 120: 344–350.

Bork P, Sudol M, (1994). The WW domain: a signalling site in dystrophin? *Trends Biochem Sci* 19: 531–533.

Broderick MJF, Winder SJ, (2002). Towards a complete atomic structure of spectrin family proteins. *J Stru Biol.,* 137 (2): 184-193.

Brunger AT, (1992). X-PLOR, Version 3.1: a system for x-ray crystallography and NMR. New Haven: The Howard Hughes medical institute and department of molecular

biophysics and biochemistry.

Bryson K, McGuffin LJ, Marsden RL, Ward JJ, Sodhi JS, (2005). Protein structure prediction servers at University College London. *Nucl Acids Res.,* 33: 36–38.

Burgardt NI, Ferreyra RG, Falomir-Lockhart L, Corsico B, Ermacora MR, Ceolin M, (2009). Biophysical characterisation and urea-induced unfolding of recombinant Yarrowia lipolytica sterol carrier protein-2. *Biochim et Biophys Acta.,* 1794(8): 1115-1122.

Calvert R, Kahana E, Gratzer WB, (1996). Stability of the dystrophin rod domain fold: evidence for nested repeating units. *Biophys J.,* 71: 1605-1610.

Campbell KP, (1995). Three muscular dystrophies: loss of cytoskeleton-extra-cellular matrix linkage. *Cell.,* 80: 675-679.

Carrington JC, Dougherty WG, (1988). A viral cleavage site cassette: identification of amino acid sequences required for tobacco etch virus polyprotein processing. *Proc Natl Acad Sci USA.,* 85: 3391-3395.

Castresana J, Saraste M, (1995). Does Vav bind to F-actin through a CH domain? *FEBS Lett.,* 374, 149–151.

Chakrabarti A, Bhattacharya S, Ray S, Bhattacharyya M, (2001). Binding of a denatured heme protein and ATP to erythroid spectrin. *Biochem Biophys Res Commun.,* 282: 1189–1193

Clamp M, Cuff J, Searle SM, Barton GJ, (2004). “The Jalview: java Alignment Editor”. *Bioinfo.,* 20: 426-427

Clark AR, Sawyer GM, Robertson SP, Sutherland-Smith AJ, (2009). Skeletal dysplasias

due to Filamin A mutations result from a gain-of-function mechanism distinct from allelic neurological disorders. *Hum Mol Genet.,* 18: 4791-4800.

Cleland WW, (1964). "Dithiothreitol, a new protective reagent for SH groups". *Biochem.,* 3: 480–482.

Clerk A, Morris GE, Dubowitz V, Davies KE, Sewry CA, (1993). Dystrophin-related protein, utrophin, in normal and dystrophic human fetal skeletal muscle. *Histochem J.,* 25: 554-561.

Cohn RD, Campbell KP, (2000). Molecular basis od muscular dystrophies. *Musc Ner.,* 23: 1456 – 1471

Cross RA, Stewart M, Kendrick-Jones J, (1990). Structural predictions for the central domain of dystrophin. *FEBS Lett.,* 262(1): 87-92.

Culligan KG, Ohlendieck K (2002): Abnormal calcium handling in muscular dystrophy. *J Basic Appl Myol.,* 12:147-157

Davis I, Leaver-Fay A, Chen V, Block J, Kapral G, Wang X, Murray L, Arendall W, Snoeyink J, Richardson J, Richardson D, (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucl Acids Res.,* 35: 375-383.

Davison MD, Baron MD, Critchley DR, Wootton JC, (1988). Structural analysis of homologous repeated domains in [alpha]-actinin and spectrin. *Int J Biol Macromol.,* 11 (2): 81-90.

D'Arcy A, Elmore C, Stihle M, Johnston JE, (1996). "A novel approach to crystallising proteins under oil." *J Crys Grow*., 168: 175-180.

Datta P, Chakrabarty SB, Chakrabarty A, Chakrabarti A, (2003). Interaction of erythroid spectrin with hemoglobin variants: implications in b-thalassemia. *Blood Cells Mol Dis.,* 30: 248–253.

Deconinck AE, Potter AC, Tinsley JM, Wood SJ, Vater R, Young C, Metzinger L, Vincent A, Slater CR, Davies KE, (1997a). Postsynaptic abnormalities at the neuro-muscular junction of utrophin deficient mice. *J Cell Biol.,* 136: 883-894.

Deconinck AE, Rafael JA, Skinner JA, Brown SC, Potter AC, Metzinger L, Watt DJ, Dickson JG, Tinsley JM, Davies KE, (1997b). Utrophin-Dystrophin deficient mice as a model for Duchenne muscular dystrophy. *Cell.,* 90: 717-727.

DeWolf C, McCauley P, Sikorski AF, Winlove CP, Bailey AI, Kahana E, Pinder JC, Gratzer WB, (1997). Interaction of dystrophin fragments with model membranes, *Biophys J.,* 72: 2599–2604.

Diakowski W, Prychidny A, Swistak M, Nietubyc M, Bialkowska K, Szopa J, Sikorski AF, (1999). Brain spectrin (Fodrin) interacts with phospholipids as revealed by intrinsic fluorescence quenching and monolayer experiments. *Biochem J.,* 338: 83–90.

Djinovic-Carugo K, Young P, Gautel M, Saraste M, (1999). Structure of the alpha2 actinin rod: molecular basis for cross-linking of actin filaments. *Cell.,* 98: 537-546.

Djinovic-Carugo K, Gautel M, Ylanne J, Young P, (2002). The spectrin repeat: a structural platform for cytoskeletal protein assemblies. *FEBS Lett.,* 513: 119-123.

Dong A, Xu X, Edwards AM, (2007). In situ proteolysis for protein crystallization and structure determination. *Nature Meth.,* 4, 1019-1021.

Dubreuil RR, Byers TJ, Sillman AL, Bar-Zvi D, Goldstein LS, Branton D, (1989). The complete sequence of drosophila alpha-spectrin: conservation of structural domains between alpha-spectrins and alpha-actinin. *J Cell Biol.,* 109: 2197-2205.

Dubreuil RR, (1991). Structure and evolution of the actin crosslinking proteins. *Bioassays.,* 13: 219-226.

Ehmsen J, Poon E, Davies K, (2002). The dystrophin associated protein complex. *J Cell Sci.,* 115: 2801-2803.

Emery AE, (1991). Population frequencies of inherited neuromuscular diseases – A world survey. *Neuromuscul Disord.,* 1(1): 19-29.

Emery AEH, (1993). Duchenne muscular dystrophy, 2nd edition *oxford press, oxford,* pp. 80-107.

Emsley P, Cowtan, (2004). Coot: model-building tools for molecular graphics. *Acta Cryst* D60: 2126-2132.

England SB, Nicholson LV, Johnson MA, Forrest SM, Love DR, (1990). Very mild muscular dystrophy associated with the deletion of 46% of dystrophin. *Nature.,* 343: 180-182.

Ervasti JM, (1990). Deficiency of a glycoprotein component of the dystrophin complex in dystrophic muscle. *Nature.,* 345: 315-319.

Ervasti JM. (2007). Dystrophin, its interactions with other proteins, and implications for muscular dystrophy. *Biochim et Biophys Acta.,* 1772(2): 108-117.

Evans P, (2006). Scaling and assessment of data quality. *Acta Cryst.,* D62: 72-82.

Ferlini A, Sewry C, Melis MA, Mateddu A, Muntoni F, (1999). X-linked dilated cardiomyopathy and the dystrophin gene. *Neuromuscul Disord.,* 9(5): 339-46.

Finn R, Tate J, Mistry J, Cogill P, Sammut J, Hotz H, Ceric G, Forslund K, Eddy S, Sonnhammer E, Bateman A, (2008). The Pfam protein families database. *Nucl Acids Res* 36: 281-288.

Gale R (1993). Crystallography Made Crystal Clear. San Diego. *Academic Press:* 8(10): 29-38.

Galkin VE, Orlova A, VanLoock MS, Rybakova IN, Ervasti JM, Egelman EH (2002) The utrophin actin-binding domain binds F-actin in two different modes: implications for the spectrin superfamily of proteins. *J Cell Biol.,* 157: 243-251.

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A, (2005). Protein Identification and Analysis Tools on the ExPASy Server. *The Proteomics Protocols Handbook, Humana Press*. 571-607.

Gimona M, Mital R, 1998. The single CH-domain of calponin is neither sufficient nor necessary for F-actin binding. *J Cell Sci.,* 111: 1813-1821.

Gimona M, Djinovic-Carugo K, Kranewitter WJ, Winder SJ, (2002). Functional plasticity of CH domains. *FEBS Lett.,* 513(1): 98-106.

Goldenberg O, Erez E, Nimrod G, Ben-Tal N, (2009). The ConSurf-DB: Pre-calculated evolutionary conservation profiles of protein structures. *Nucl Acids Res.,* 37: 323-327.

Gramolini AO, Jasmin BJ (1999), Expression of the utrophin gene during myogenic differentiation. *Nucl Acids Res.,* 27: 3603-3609.

Greenfield N, Fasman GD, (1969). Computed circular dichroism spectra for the evalution of protein conformation. *Biochem.,* 8: 4108-4116.

Greenfield NJ, (2007). Using circular dichroism spectra to estimate protein secondary structure. *Nat Proto.,* 1: 2876-2890.

Grum VL, Li D, MacDonald RI, Mondragon A, (1999). Structures of two repeats of spectrin suggest models of flexibility. *Cell.,* 98: 523-535.

Holm L, Park J (2000) DaliLite workbench for protein structure comparison. *Bioinfor.,* 16: 566-567.

Hnia K, Zouiten D, Hugon G, Diment A, Bramham J, Winder SJ, (2007). ZZ domain of dystrophin and utrophin: topology and mapping of a β-dystroglycan interaction site. *Biochem J.,* 410:667-677.

Huang X, Poy F, Zhang A, Sudol M, Eck MJ, (2000). Structure of a WW domain containing fragment of dystrophin in complex with β-dystroglycan. *Nat Struct Biol.,* 7: 634–638.

Huang B, Schroeder M, (2006). LIGSITE: predicting protein binding sites using the Connolly surface and degree of conservation. *BMC Struc Biol.,* 6-19.

Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. *J Mol Graph.,* 14: 33–38.

Imbert N, Cognard C, Duport G, Guillou C, Raymond G, (1995). Abnormal calcium homeostasis in Duchenne muscular dystrophy myotubes contracting in vitro. *Cell.,* 18: 177-186.

Kabsch W, (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J Appl Cryst.,* 26: 795-800.

Kahana E, Marsh PJ (1994). Conformation and phasing of dystrophin structural repeats. *J Mol Biol.,* 235(4): 1271-1277.

Kahana E, Gratzer WB (1995). "Minimum folding unit of dystrophin rod domain." *Biochem.,* 34(25): 8110-8110.

Kahana E, Flood G, (1997). Physical properties of dystrophin rod domain. *Cell Mot & Cytoskel.,* 36(3): 246-252.

Kay BK, Williamson MP, Sudol M, (2000). The importance of being proline: The interaction of proline-rich motifs in signalling proteins with their cognate domains. *FASEB J.,* 14: 231–241.

Kauzmann W, (1959). Some factors in the interpretation of protein denaturation, *Adv Prot Chem.,* 14: 1–63.

Keep NH, Norwood FL, Moores CA, (1999). "The 2.0 A structure of the second calponin homology domain from the actin-binding region of the dystrophin homologue utrophin". *J Mol Biol.,* 285 (3): 1257-64.

Kelly SM, Price NX, (2000). The use of circular dichroism in the investigation of protein structure and function. *Curr Prot Pept Sci.,* 1: 349–384

Kelly SM, Jess TJ, Price NC, (2005). How to study proteins by circular dichroism. *Biochim et Biophys Acta.,* 1751:119-139.

Kleywegt GJ, (1996). Use of non-crystallographic symmetry in protein structure refinement. *Acta Cryst* D52: 842-857.

Koenig M, Hoffman EP, Bertelson CJ, Monaco AP, Kunkel LM, (1987). Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomics organization of the DMD gene in normal and affected individuals. *Cell.,* 50: 509-517.

Koenig M, Monaco AP, Kunkel, (1988). The complete sequence of dystrophin predicts a rod-shaped cytoskeletal protein. *Cell.,* 53: 219-228.

Koenig M, Kunkel LM, (1990). Detailed analysis of the repeat domain of dystrophin reveals four potential hinge segments that may confer flexibility. *J Biol Chem.,* 265: 4560-4566.

Krissinel E, (2009). Crystal contacts as nature's docking solutions. *J Comp Chem.,* 31(1): 133-43.

Kuhlman PA, Hemmings L, Critchley DR, (1992). The identification and characterization of an actin-binding site in [alpha]-actinin by mutagenesis. *FEBS Lett.,* 304(2-3), 201-206.

Kusunoki H, Minasov G, MacDonald RI, Mondragon A, (2004a). Independent movement, dimerisation and stability of tandem repeats of chicken brian alpha-spectrin. *J Mol Biol.,* 344: 495-511.

Kusunoki H, MacDonald RI, Mondragon A, (2004b). Structural insights into the stability and flexibility of unusual erythroid spectrin repeats. *Structure.,* 12: 645-656.

Ipsaro JJ, Harper SL, Messick TE, Marmorstein R, Mondragon A, Speicher DW, (2010). Crsytal structure and functional interpretation of the erythrocyte spectrin tetramerisation domain complex. *Blood.,* 115(23): 4843-4852.

Isralewitz B, Gao M, Schulten K, (2001). Steered molecular dynamics and mechanical functions of proteins. *Curr Opin Struct Biol.,* 11: 224-230.

Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML, (1983). Comparison of simple potential functions for simulating liquid water. *J Chem Phys.,* 79: 926-935.

Lai Y, Thomas GD, Yue Y, yang HT, Li D, Long C, Judge L, Bostick B, Chamerlain JS, Terjung RL, Duan D, (2009). Dystrophin carrying spectrin-like repeats 16 and 17 anchor nNOS to the sarcolemma and enhance exercise performance in a mouse model of muscular dystrophy. *J Clini Invest.,* 119: 624-635.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, (2007). ClustalW and ClustalX version 2.0. *Bioinfo.,* 23: 2947-1948.

Laskowski R, MacArthur M, Moss D, Thornton J, (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst.,* 26: 283-291.

Law DJ, 1994. Talin, Vinculin and DRP (utrophin) concentrations are increased at mdx myotendinous junctions following onset of necrosis. *J Cell Sci.,* 107: 1477 – 1483.

Law R, Carl P, Harper S, Dalhaimer P, Speicher DW, Discher DE, (2003). Cooperativity in forced unfolding of tandem spectrin repeats. *Biophys J.,* 84: 533–544.

Lee JC, Discher DE, (2001). Deformation-enhanced fluctuations in the red cell skeleton with theoretical relations to elasticity, connectivity, and spectrin unfolding. *Biophys J.,* 81(6): 3178-3192.

Legardinier S, Tascon C, Rocher C, Hardy S, Hubert JF, Rumeur EL, (2009a). Mapping of the lipid-binding and stability properties of the central rod domain of human dystrophin. *J Mol Bio.,* 389: 546-558.

Legardinier S, Legrand B, Bondan A, Rocher C, Rumeur EL, Hubert JF, (2009b). A two-amino acid mutation encountered in Duchenne muscular dystrophy decreases stability of the rod domain 23 (R23) spectrin-like repeat of dystrophin. *J Bio Chem.,* 284(13): 8833-8832.

Legrand B, Giudice E, Nicolas A, Delalande O, Rumeur EL, (2011). Computational study of the human dystrophin repeats: interaction properties and molecular dynamics. *PLoS ONE.,* 6(8): e23819.

Lehman W, Craig R, Kendrick-Jones J, Sutherland-Smith AJ, (2004). An open or closed case for the conformation of calponin homology domains of F-actin? *J Mus Res & Cell Moti.,* 25: 351-358.

Lenne PF, Raae AJ, Altmann SM, Saraste M, Hörber JKH, (2000). States and transitions during forced unfolding of a single spectrin repeat. *FEBS Lett.,* 476(3): 124-128.

Leslie AG, (1990). Refined crystal structure of type III chloramphenicol acetyltransferase at 1.75 Å resolution. *J Mol Biol.,* 213(1): 167-186.

Levine BA, Moir AJG, Patchell VB, Perry SV, (1990). The interaction of actin with dystrophin. *FEBS Lett.,* 263: 159-162.

Levine BA, Moir AJG, Patchell VB, Perry SV, (1992). Binding sites involved in the interaction of actin with the N-terminal region of dystrophin. *FEBS Lett.,* 298(1): 44-48.

Lim BB, Lee EH, Sotomayor M, Schulten K, (2008). Molecular basis of fibrin clot elasticity. *Structure.,* 16(3): 449-459.

Liu Y, Hsin J, Kim H, Selvin PR, Schulten K, (2011). Extension of a three-helix bundle domain of myosin VI and key role of calmodulins. *Biophys J.,* 100(12): 2964-2973.

Love DR, Hill DF, Dickson G, Spurr NK, Byth BC, Marsden RF, Walsh FS, Edwards YH, Davies KE, (1989). An autosomal transcript in skeletal muscle with homology to dystrophin. *Nature.,* 339: 55–58.

Lovell SC, Davis IW, Arendall WB, De Bakker PIW, Word JM, Prisant MG, Richardson JS, Richardson DC, (2002). Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins: Struct, Func & Gene.,* 50: 437-450.

Lovejoy B, Choe S, Cascio D, McRorie DK, DeGrado WF, Eisenberg D, (1993). Crystal structure of a synthetic triple-standard alpha-helical bundle. *Science.,* 259: 1288-1293.

Lin AY, Prochniewicz E, James ZW, Thomas DD, (2011). Large-scale opening of utrophin's tandem calponin homology (CH) domains upon actin binding by an induced-fit mechanism. *Proc Natl Acad Sci USA.,* 108(31): 12729-12733.

Lux SE, Palek, J, (1995). Disorders of the red cell membrane. *Blood*., 1701-818.

MacDonald RI, Pozharski EV, (2001). Free energies of urea and of thermal unfolding show that two tandem repeats of spectrin are thermodynamically more stable than a single repeat. *Biochem.,* 40: 3974–3984.

MacKerell Jr AD, Bashford D, Bellott M, (1998). All-hydrogen empirical potential for molecular modeling and dynamics studies of proteins using the CHARMM22 force field. *J Phys Chem Bio.,* 102: 3586–3616.

MacKerell Jr AD, Feig M, Brooks CL, (2004). Extensing the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulation. *J Comput Chem* 25(11)., 1400-1415.

Macias MJ, Wiesner S, Sudol M (2002). "WW and SH3 domains, two different scaffolds to recognize proline-rich ligands". *FEBS Lett.,* 513(1): 30–7.

Man NT, Morris GE, Davies KE, (1992a). The dystrophin-related protein, utrophin, is expressed on the sarcolemma of regenerating human skeletal muscle fibres in dystrophies and inflammatory myopathies*. Neuromusc. Disord.,* 2: 177-184.

Man NT, Thankh LT, Blake DJ, Davies KE, Morris GE (1992b): Utrophin, the autosomal homolog of dystrophin, is widely expressed and membrane-associated in cultured cell lines. *FEBS Lett.,* 313, 19-22.

Mandel CR, Gebauer D, Zhang H, Tong L, (2006). A serendipitous discovery that in situ proteolysis is essential for the crystallization of yeast CPSF-100 (Ydh1p). *Acta Cryst* F62: 1041-1045.

Masuda-Hirata M, Suzuki A, Amano Y, Yamashita K, Ide M, Yamanaka T, Sakai M, Imamura M, Ohno S, (2009). Intracellular polarity protein PAR-1 regulates extracellular

laminin assembly by regulating the dystroglycan complex. *Genes Cells.,* 14: 835–850.

Matthews B, (1968). Solvent content of protein crystals. *J Mol Biol.,* 33: 491-497.

Mattson MP, (2001). Pathogenisis of neurodegradative disorders. In: *Contemprary neuroscience*. *Humana press, Totowa, NJ*, pp 290 and 294.

Matsumura K, Ervasti JM, Ohlendieck K, Kahl SD, Campbell KP (1992): Association of dystrophin-related protein with dystrophin-association proteins in mdx mouse muscle. *Nature.,* 360: 588-591.

Matsumura K, Ervasti JM, Ohlendieck K, Kahl SD, Campbell KP, (1993). Association of dystrophin-related protein with dystrophin-association proteins in mdx mouse muscle. *Nature* 360: 588-591.

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ, (2007). Phaser crystallographic software. *J Appl Cryst.,* 40: 658-674.

McFerrin MB, Snell EH, (2007). The development and application of a method to quantify the quality of cryoprotectant solutions using standard area-detector X-ray images. *J Appl Cryst.,* 35: 538-545.

McPherson A, (1982). The preparation and analysis of protein crystals, *John Wiley & Sons, New York.*

McPhie P, (1971). Dialysis. Methods in enzymology. *Academic Press Inc., New York* 22: 23-32.

McRee, Duncan E, (1993). Practical protein crystallography. *San Diego: Academic Press Inc:* pp. 1-23.

Menhart N, Mitchell T, Lusitani D, Topouzian N, Fung LW, (1996). Peptides with more than one 106-amino acid sequence motif are needed to mimic the structural stability of spectrin. *J Biol Chem.,* 271: 30410–30416.

Menhart N, (2006). Hybrid spectrin type repeats produced by exon-skipping in dystrophin. *Biochim et Biophys Acta.,* 1764: 993-999.

Menke A, Jockusch H, (1991). Decreased osmotic stability of dystrophin less muscle cells from the mdx mouse. *Nature.,* 349: 69-71.

Miller CA, Gellman SH, Abbott NL, de Pablo JJ, (2009). Association of Helical β-Peptides and their Aggregation Behavior from the Potential of Mean Force in Explicit Solvent. *Biophys J.,* 96(11): 4349-4362.

Mirijianian DT, Chu JW, Ayton GS, Voth GA, (2007). Atomistic and coarse-gained analysis of double spectrin repeat units: the molecular origin of flexibility. . *J Mol Biol.,* 365(2): 523-534.

Mirza A, Sagathevan M, Sahni N, Choi L, Menhart N, (2010). A biophysical map of the dystrophin rod. *Biochim et Biophys Acta.,* 1840: 1796-1809.

Miura P, Jasmin B, (2006). Utrophin upregulation for treating Duchenne or Becker muscular dystrophy: how close are we? *Tren mole med.,* 12(3): 122-129.

Mizumo Y, Nonaka I, Hirai S and Ozawa E, (1993). Reciprocal expression of dystrophin and utrophin in muscles of Duchenne muscular dystrophy patients, female DMD carriers and control subjects. *J Neu Sci.,* 119: 43-52.

Moffit W, Fitts DD, Kirkwood JG, (1957). Critique of the theory of optical activity of helical polymers. *Proc. Natl. Acad. Sci. USA.,* 43(2): 213-222.

Monaco AP, Neve RL, Colletti-Fenner C, Bertelson CJ, Kurnit DM, Kunkel LM, (1986). Isolation of candidate cDNAs for portions of the Duchenne muscular dystrophy gene. *Nature.,* 323:646-650.

Morita M, Nakamura S, Shimizu K, (2008). Highly accurate method for ligand-binding site prediction in unbound state (apo) protein structures. *Proteins.,* 73: 468–479.

Moores CA, Keep NH, Kendrick-Jones J, (2000). Structure of the utrophin actin-binding domain bound to F-actin reveals binding by an induced fit mechanism. *J Mol Biol.,* 297(2): 465-480.

Murshudov GN, Vagin AA, Dodson EJ, (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst* D53: 240-255.

Nakayama S, Kretsinger R, (1994). Evolution of the EF-hand family of proteins. *Annu Rev Biophys Biomol Struct.,* 23: 473–507.

Nelson M, Humphrey W, Gursoy A, (1996). NAMD—A parallel, object-oriented molecular dynamics program. *J Supercomp App.,* 10: 251–268.

Nicolas G, Pedroni S, Fournier C, Gautero H, Craescu C, Dhermy D, Lecomte MC, (1998). Spectrin self-association site characterization and study of beta-spectrin mutations associated with hereditary elliptocytosis. *Biochem J.,* 332: 81–89.

Norwood FLM, Sutherland-Smith AJ, Keep NH, Kendrick-Jones J, (2000). The structure of the N-terminal actin-binding domain of human dystrophin and how mutations in this domain may cause Duchenne or Becker muscular dystrophy. *Structure.,* 8(5): 481-491.

Nuttall JM, LLsley K, Tinsley JN, Sudol M, Winder SJ, (2000). Adhesion-dependent tyrosine phosphorylation of β-dystroglycan regulates its interaction with utrophin. *J Cell Sci.,* 113: 1717–1726.

Ortiz , Nielsen SO, Klein ML, Discher DE, (2005). Unfolding a linker between helical repeats. *J Mol Biol.,* 349: 638-647.

Pace CN, (1986). Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Meth in Enzy.,* 131: 266-280.

Pantazatos DP, MacDonald RI, (1997). Site-directed mutagenesis of either the highly conserved Trp-22 or the moderately conserved Trp-95 to a large, hydrophobic residue thermodynamic stability of a spectrin repeating unit. *J Biol Chem.,* 272: 21052-21059.

Parry DA, Dixon TW, Cohen C, (1992). Analysis of the three-alpha-helix motif in the spectrin superfamily of proteins. *Biophys J.,* 61: 858-867.

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. *J Mol Biol.,* 273: 740-751.

Pearce M, Blake DJ, Tinsley JM, Byth BC, Campbell L, Monaco AP, Davies KE, (1993). The utrophin and dystrophin genes share similarities in genomic structure. *Hum Mol Genet.,* 2: 1765–1772

Perkins SJ, Sim RB, (1986). Molecular modelling of human complement component C3 and its fragments by solution scattering. *Eur J Biochem.,* 157: 155–168.

Petock JM, (2003). Analysis of protein structures reveals regions of rare backbone conformation at functional sites. *Proteins.,* 53: 872–879.

Pflugrath JW, (1999). The finer things in X-ray diffraction data collection. *Acta Cryst* D55: 1718-1725.

Pichavant C, Aartsmus-Rus A, Clemens PR, Davies KE, Dickson G, Takeda S, Wilton SD, Wolff JA, Wooddell CI, Xiao X, Tremblay JP, (2011). Current status of pharmaceutical and genetic therapeutic approaches to treat DMD. *Mol Ther.,* 19: 830-840.

Pons F, Augier N, Leger JO, Robert A, Tome FM, Fardeau M, Voit T, Nicholson LV, Mornet D, Leger JJ, (1991). A homologue of dystrophin is expressed at the neuromuscular junctions of normal individuals and DMD patients, and of normal and mdx mice: immunological evidence. *FEBS Lett.,* 282: 161-165

Pons F, Robert A, Marini JF, Leger JJ, (1994). Does utrophin expression in muscles of mdx mice during postnatal development functionally compensate for dystrophin deficiency? *J Neu Sci.,* 122: 162-170.

Ponting CP, Blake DJ, Davies KE, Kendrick-Jones J, Winder SJ, (1996). ZZ and TAZ: New putative zinc fingers in dystrophin and other proteins. *Tren Biochem Sci.,* 21: 11–13.

Pozzoli U, Sironi M, Caglini R, Comi GP, Bardoni A, Bresolin N, (2002). Comparative analysis of the human dystrophin and utrophin gene structures. *Genetics.,* 160: 793-798.

Prochniewicz E, Henderson D, Ervasti JM, Thomass DD, (2009). Dystrophin and utrophin have distinct effects on the structural dynamics of actin. *Proc Natl Acad Sci USA.,* 106: 7822-7827.

Putnam CD, Hammel M, Hura GL, Tainer JA, (2007). X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q Rev Biophys.,* 40: 191–285.

Ramachandran GN, Ramakrishnan C, Sasisekharan V, (1963). Stereochemistry of polypeptide chain configurations. *J Mol Biol.,* 7: 95-99.

Rando TA, (2001). The dystrophin-glycoprotein complex, cellular signalling and the regulation of cell survival in the muscular dystrophies. *Muscl Ner.,* 24(12): 1575 – 1594

Read RJ, (2001). "Pushing the boundaries of molecular replacement with maximum likelihood." *Acta Cryst* D57: 1373-1382.

Rentschler S, Linn H, Deininger K, Bedford MT, Espanel X, Sudol M, (1999). The WW domain of dystrophin requires EF-hands region to interact with β-dystroglycan. *Biol Chem.,* 380: 431–442.

Rief M, Pascual J, Saraste M, Gaub HE, (1999). Single molecule force spectroscopy of spectrin repeats: low-unfolding forces in helix bundles. *J Mol Biol.,* 286(2): 553-561.

Rumeur EL, Pottier S, Costa GD, Metzinger L, Mouret L, Rocher C, Fourage M, Rondeau-Mouro C, Bondon A, (2007). Binding of the dystrophin second repeat to membrane di-oleyl phospholipids is dependent upon lipid packing. *Biochim et Biophys Acta – Biomembranes.,* 1768(3): 648-654.

Rumeur EL, Winder SJ, Hubert JF, (2010). Dystrophin: more than just the sum of its parts. *Biochim et Biophys Acta.,* 1840: 1713-1722.

Rybakova IN, Amann KJ, Ervasti JM, (1996). A new model for the interaction of dystrophin with F-actin. *J Cell Biol.,* 135: 661-672.

Rybakova IN, Pate JR, Davies KE, Yurchenco PD, Ervasti JM, (2002). Utrophin binds laterally along actin filaments and can couple costameric actin and sarcolemma when overexpressed in dystrophin-deficient muscle. *Mol of Cell.,* 13: 1512-1521.

Rybakova IN, Ervasti JM, (2005). Identification of spectrin-1 like repeats required for high affinity utrophin-actin interaction. *J Biol Chem.,* 280: 23018-23023.

Saadat L, Pittman L, Menhart N, (2006). Structural cooperativity in spectrin type repeats motifs of dystrophin. *Biochim et Biophys Acta.,* 1764: 943-954.

Sahr K, Laurila P, Kotula L, Scarpa A, Coupal E, Leto T, Linnebach A, Winklemann J, Speicher D, Marchesi V, Curtis P, Forget B, (1990). The complete cDNA and polypeptide sequences of human erythroid spectrin. *J Biol Chem.,* 265: 4434–4443.

Sambrook J, Fritsch EF, Maniatis T, (1989). Molecular cloning: a laboratory manual, 2nd ed. *Cold Spring Harbor Laboratory Press*, Cold Spring Harbor, N.Y.

Sawyer GM, Clark AR, Robertson SP, Sutherland-Smith AJ, (2009). Disease-associated substitutions in the Filamin B actin binding domain confer enchanced actin binding affinity in the absence of major structural disturbance: Insights from the crystal structures of Filamin B actin binding domains. *J Mol Biol.,* 390: 1030-1047.

Schultz J, Milpetz F, Bork P, Ponting CP, (1998). SMART, a simple modular architecture research tool: identification of signalling domains. *Proc Natl Acad Sci USA.,* 95: 5857–5864.

Sicinski P, Geng Y, Ryder-Cook AS, Barnard EA, Darlison MG, Barnard PJ, (1989). The molecular basis of muscular dystrophy in the mdx mouse: a point mutation. *Science.,* 244: 1578-1580.

Sonnenberg A, Rojas AM, de Pereda JM, (2007). The structure of a tandem pair of spectrin repeats of plectin reveals a modular organization of the plakin domain. *J Mol Biol.,* 368: 1379-1391.

Sonnemann KJ, Heun-Johnson H, Turner AJ, Baltgalvis KA, Love DA, Ervasti JM, (2009). Functional substitution by TATA-utrophin in dystrophin-deficient mice. *PLOS Med.,* 6(5): e1000083.

Speicher DW, Marchesi VT, (1984). Erythrocyte spectrin is comprised of many homologous triple helical segments. *Nature.,* 31(1): 177-80.

Stamler JS, Meissner G, (2001). Physiology of nitric oxide in skeletal muscle. *Physiol Rev.,* 81: 209-237.

Stedman HH, Sweeney HL, Shrager JB, Mguire HC, Panettieri RA, Petrof B, Narusawa M, Leferovich JM, Sladky JT and Kelly AM, (1991). The mdx mouse diaphragm reproduces the degenerative changes of Duchenne muscular dystrophy. *Nature (London).,* 352: 536-539.

Stone MR, Neill AO, Catino D, Bloch RJ, (2005). Specific interaction of the actin-binding domain of dystrophin with intermediate filaments containing keratin 19. *Mol Bio Cell.,* 9: 4280-4293.

Sutherland-Smith, AJ, Moores CA, Kendrick-Jones J, (2003). An Atomic model for actin binding by the CH domains and spectrin-repeat modules of utrophin and dystrophin. *J Mol Biol.,* 329(1): 15-33.

Thompson JD, Higgins DG, Gibson TJ, (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucl Acids Res.,* 22: 4673-80.

Tinsley JM, Blake DJ, Zuellig RA, Davies KE (1994). Increasing complexity of the dystrophin-associated protein complex. *Proc Natl Acad Sci USA.,* 91: 8307-8313.

Tinsley JM, Potter AC, Phelps SR, Fisher R, Trickett JI, Davies KE, (1996). Amelioration of the dystrophic phenotype of mdx mice using a truncated utrophin transgene. *Nature.,* 384: 349-353.

Tinsley J, Deconinck N, Fisher R, Kahn D, Phelps S, Gillis JM, Davies K, (1998). Expression of full length dystrophin prevents muscular dystrophin in mdx mice. *Nature*

Medi., 4: 1441 – 1444.

Toumadje A, Alcorn SW, Johnson WC, (1992). Extending CD spectra of proteins to 168 nm improves the analysis for secondary structures. *Anal Biochem.,* 200: 321-331.

Tufty RM, Kretsinger RH, (1975). Troponin and parvalbumin calcium binding regions predicted in myosin light chain and T4 Lysozyme. *Science.,* 187: 167-169.

Videau LL, (2004). The cis-Pro touch-turn: A rare motif preferred at functional sites. *Proteins.,* 56: 298–309.

Wass MN, Kelley LA, Sternberg MJ, (2010). 3DLigandSite: predicting ligand-binding sites using similar structures. *Nucl Acids Res.,* 38: 469-473.

Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ, (2009). JalView Ver 2.0: a multiple sequence alignment and analysis workbench, bioinformatics. *Bioinfor.,* 25(9): 1189-1191.

Way M, Pope B, Cross RA, Kendrick Jones J, Weeds AG (1992), Expression of N-terminal domain of dystrophin in *Escherichia coli* and demonstrates of binding F-actin. *FEBS Lett.,* 301: 243-245.

Weir AP, Morgan JE, Davies KE, (2004), A utrophin up-regulation in mdx skeletal muscle is independent of regeneration. *Neuro Disord.,* 14: 19-23.

Winder SJ, Walsh MP. (1990), Structure and functional characterization of calponin fragments. *Biochem Int.,* 22: 335-341.

Winder SJ, Gibson TJ, Hemmings L, Maciver SK, Boalton SJ, Tinsel JM, Davies KE, Critchley DR, (1995). Dystrophin and utrophin: the missing links! *FEBS Lett.,* 369(1): 27-33.

Winder SJ, (1997). The membrane-cytoskeleton interface: the role of dystrophin and utrophin. *J Mus Res Cell Mot., 18*: 617-629.

Winder SJ, (2001). The complexities of dystroglycan. *Tren in Biochem Sci.,* 26: 118–124.

Winn M, Isupov M, Murshudov G, (2000). Use of TLS parameters to model anisotropic displacement in macromolecular refinement. *Acta Cryst* D57: 122-133.

Worton R, (1995). Muscular dystrophies: diseases of the dystrophin-glycoprotein complex. *Science.,* 270: 755-756.

Yamashita K, Suzuki A, Satoh Y, Ide M, Amano Y, (2010). The 8th and 9th tandem spectrin-like repeats of utrophin cooperatively form a functional unit to interact with polarity-regulating kinase PAR-1b. *Biochem Biophys Res Comm.,* 391: 812-817.

Yan Y, Winograd E, Viel A, Cronin T, Harrison SC, Branton D, (1993). Crystal structure of the repetitive segments of spectrin. *Science.,* 262: 2027-2030.

Ylanne J, Scheffzek K, Young P, Saraste M, (2001). Crystal structure of the alpha4 actinin rod reveals an extensive torsional twist. *Structure.,* 9: 597-604.

Zhang Y, (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinf.,* 9: 1-8.

Zubrzycka-Gaarn EE, Bulma DE, Karpati G, Burghes AHM, Belfall B, Klamut HJ, Talbot J, Hodges RS, Ray PN, Worton RG, (1988). The Duchenne muscular dystrophy gene product is localized in sarcolemma of human skeletal muscle. *Nature (London).,* 333: 466-469.

Zuellig RA, Bornhauser BC, Knuesel I, Heller F, Fritschy JM, Schaub MC, (2000). Identification and characterisation of transcript and protein of a new short N-terminal utrophin isoform. *J Cell Biochem.,* 77: 418-431.