

**STRUCTURE AND FUNCTION OF THE CARDIAC  
STRESS RESPONSE PROTEIN MS1**

Thesis submitted for the degree of

Doctor of Philosophy

University of Leicester

by

Claudia Liliane Fiona Fogl, BSc (Leicester)

Department of Biochemistry

University of Leicester

September 2010

## **Abstract**

Myocyte Stress 1 (ms1)/Striated muscle Activator of Rho Signalling (STARS), also known as Actin Binding Rho Activator (ABRA), is a 375 amino acid protein. Its expression increases one hour after the induction of stress in rat hearts through aortic banding. This expression precedes that of early response genes such as c-fos and c-jun. This process has been implicated in the development of left ventricular hypertrophy. ms1/STARS binds to actin, and deletion mutations had shown that residues 234-375 were necessary for actin binding. A mixture of combinatorial domain hunting and rational domain design gave a series of possible domains. Circular dichroism and nuclear magnetic resonance spectroscopy were used to characterise these domains. The first three domains, MSD1 (residues 2-118), MSD2 (40-196) and actin binding domain 1 (ABD1, 193-296) were unfolded, while ABD2 (294-375) was folded. Actin co-sedimentation assays showed that only ABD1 and ABD2 bound to actin. They bound to actin independently. The structure of ABD2 was determined using NMR. Mutation studies, based on the NMR structure and on data about the conservation of positively-charged regions in ms1/STARS homologues, were used to identify the actin binding surface of ABD2. The structure of ABD2 was shown to be a winged helix-turn-helix domain. These domains are often DNA binding domains. When DNA binding was attempted, it was shown that ABD1, ABD2 and the tandem of ABD1 and ABD2 bound to DNA. The identification of the actin binding domain and the discovery of a novel DNA binding ability open up many more possible functions of ms1/STARS.

## **Acknowledgements**

I would like to thank my supervisor, Dr. Mark Pfuhl, for enabling me to perform this research and for all his help and guidance. I would also like to thank the members of my thesis committee, Dr. Mohammed El-Mezgueldi and Dr. Peter Nielsen, for their help, support and advice. I am also grateful to the BBSRC for funding this research.

I would like to thank those people who have helped with experimental techniques described in this thesis, Rebecca Croasdale and Dr. Xiowen Yang for their assistance with the cloning, Dr. El-Mezgueldi, Zimna Wazeer and Sammeh Al-Sarayreh for their invaluable help with the actin binding experiments, Dr. Fred Muskett for his support with the NMR experiments and structure calculation and Dr. Louise Fairall for her help with the DNA binding experiments.

Further thanks must go to the members of Dr. Pfuhl's group, Dr. Didier Philippe, Dr. Sam Schroeder and Dr. Joyce Ratti, and the members of Professor Mark Carr's group for their support and guidance with many aspects of this thesis.

I would like to thank my mother and grandmother for their steadfast encouragement and support throughout my education.

## **Contents**

<b>Abstract.....</b>	<b>2</b>
<b>Acknowledgements .....</b>	<b>3</b>
<b>Chapter 1 - Introduction .....</b>	<b>10</b>
<b>1.1 Left Ventricular Hypertrophy .....</b>	<b>10</b>
<b>1.2 Physiological Changes in Left Ventricular Hypertrophy .....</b>	<b>14</b>
<b>1.3 The Pathway To Left Ventricular Hypertrophy.....</b>	<b>14</b>
<b>1.4 Myocyte Stress 1 (ms1) .....</b>	<b>15</b>
<b>1.5 Proposed Mode of Action of ms1/STARS.....</b>	<b>17</b>
<b>1.5.1 Serum Response Factor (SRF).....</b>	<b>19</b>
<b>1.5.2 Myocardin Related Transcription Factors.....</b>	<b>20</b>
<b>1.5.3 ms1/STARS and MEF2 and MyoD .....</b>	<b>21</b>
<b>1.6 The Actin Binding Activity of ms1/STARS .....</b>	<b>22</b>
<b>1.7 Actin-binding LIM Protein 2 and 3 (ABLIM-2 and -3) .....</b>	<b>23</b>
<b>1.7.1 LIM Domains .....</b>	<b>24</b>
<b>1.8 How We Aim To Study ms1/STARS.....</b>	<b>26</b>
<b>Chapter 2 – Materials And Methods.....</b>	<b>30</b>
<b>2.1 Bioinformatics .....</b>	<b>30</b>
<b>2.2 Construction of Primers .....</b>	<b>30</b>
<b>2.3 Polymerase Chain Reaction (PCR) .....</b>	<b>33</b>
<b>2.3.1 Agarose DNA Gels .....</b>	<b>34</b>
<b>2.4 Cloning .....</b>	<b>34</b>
<b>2.4.1 Restriction Digest of Plasmid Vector .....</b>	<b>34</b>
<b>2.4.2 Gel Extraction of Restriction Products.....</b>	<b>34</b>
<b>2.4.3 Preparation of PCR Products For Cloning .....</b>	<b>34</b>
<b>2.4.4 Ligation of the Restricted Vector and PCR Product.....</b>	<b>35</b>
<b>2.4.5 PCR Screening of Ligated Products.....</b>	<b>35</b>
<b>2.5 Determination of DNA Concentration .....</b>	<b>36</b>
<b>2.6 Cloning by PROTEX .....</b>	<b>36</b>
<b>2.7 DNA Sequencing .....</b>	<b>38</b>
<b>2.8 Transformation of Vector into Cells .....</b>	<b>38</b>
<b>2.9 Protein Expression.....</b>	<b>38</b>
<b>2.9.1 Pilot Expression Trials .....</b>	<b>38</b>
<b>2.9.2 Full-Scale Protein Expression.....</b>	<b>39</b>
<b>2.9.3 Protein Expression in M9 Minimal Media .....</b>	<b>40</b>
<b>2.9.3.1 Media.....</b>	<b>40</b>
<b>2.9.3.2 Expression in M9 Minimal Media.....</b>	<b>40</b>

2.9.4 Glycerol Stocks .....	41
2.9.5 Sodium Dodecylsulphate Polyacrylamide Gel Electrophoresis Gels .....	41
2.9.6 Protein Purification .....	42
2.9.6.1 Purification Using Gel Filtration.....	43
2.9.7 Protein Stabilisation .....	43
2.9.8 Concentrating Protein .....	43
2.9.9 Measuring Protein Concentrations .....	43
2.10 Buffer Exchange.....	44
2.11 Circular Dichroism (CD).....	44
2.11.1 Investigating Secondary Structure with CD .....	45
2.11.2 Measuring Protein Stability with CD.....	46
2.12 Fluorescence Spectroscopy .....	47
2.13 Preparation of Actin from Skeletal Muscle Acetone Powder .....	48
2.13.1 Preparation of Skeletal Muscle Acetone Powder .....	48
2.13.2 Extracting actin from Skeletal Muscle Acetone Powder .....	48
2.13.3 Labelling Actin with Pyrene Iodoacetamide .....	49
2.14 Actin Binding Co-Sedimentation Assay .....	50
2.15 Pyrene-labelled Actin Binding Fluorescence Assay .....	51
2.16 Actin binding by Isothermal Titration Calorimetry (ITC).....	51
2.17 Investigating the Multimeric State of Domains .....	52
2.17.1 Analytical Gel Filtration .....	53
2.17.2 Analytical Ultra-Centrifugation .....	53
2.17.3 Analysing the Multimeric State of ABD2 by NMR .....	53
2.18 Nuclear Magnetic Resonance Spectroscopy .....	54
2.18.1 1D NMR .....	58
2.18.2 Residue Identification .....	58
2.18.3 Secondary Structure Determination .....	60
2.18.4 Tertiary Structure Information.....	61
2.18.5 Structure Calculation Using CYANA .....	62
2.18.5.1 CYANA Inputs .....	63
2.18.5.2 CYANA Structure Calculation.....	65
2.18.5.3 CYANA Outputs .....	70
2.18.5.4 CYANA Refinement and Structure Validation .....	71
2.18.6 Structure Refinement .....	74
2.18.6.1 Residual Dipolar Coupling.....	74
2.18.6.2 Preparation of Bicelles for RDCs .....	76
2.18.6.3 Preparation of Non-Ionic Liquid Crystalline Media for RDCs.....	77

2.18.6.4 Preparation of Filamentous Phage for RDCs.....	77
2.18.6.5 Use of Media to Produce RDCs .....	77
2.18.7 Structure Comparison .....	78
2.19 Protein Interaction Studies .....	78
2.19.1 Following Binding Through Chemical Shift Changes in NMR Spectra.....	78
2.19.2 Gel Shift Assay to Study DNA Binding.....	78
Chapter 3 - The Domains of ms1/STARS .....	81
3.1 Protein Sequence Alignment of ms1/STARS.....	81
3.2 Characterisation of the Other Putative Domains of ms1/STARS Using CD and NMR .....	85
3.3 Discussion of the Domain Organisation of ms1/STARS.....	90
Chapter 4 – Characterisation of ABD2.....	95
4.1 ABD2 - Cloning, Expression, Purification, Biophysical Data and Structure Determination.....	95
4.1.1 Cloning of ABD2 .....	95
4.1.2 Pilot Expression of ABD2 .....	96
4.1.3 Purification of ABD2 .....	98
4.1.4 Mass Spectrometry .....	101
4.2 Characterisation of ABD2.....	102
4.2.1 Characterisation of ABD2 using circular dichroism .....	102
4.2.1 Characterisation of ABD2 Using Circular Dichroism.....	102
4.2.2 Characterisation of ABD2 Using Fluorescence.....	105
4.2.3 Determination of the Multimeric State of ABD2 .....	106
4.2.3.1 Results from Analytical Ultra-Centrifugation (AUC).....	106
4.2.3.2 Results From Analytical Gel Filtration .....	106
4.2.3.3 Results From NMR.....	107
4.3 Structure Determination Using NMR.....	109
4.3.1 Residue Identification Using HNCACB and HNCOCACB Spectra.....	109
4.3.2 Side Chain Assignment.....	114
4.3.3 TALOS output.....	115
4.3.4 Structural Calculation.....	117
4.3.5 The Structure of ABD2.....	120
4.3.6 Structure Validation by Residual Dipolar Coupling (RDC).....	121
4.3.7 PROCHECK Validation .....	123
4.3.8 PDB Structure Submission .....	126
4.3.9 Structure Comparison.....	126
4.3.9.1 Winged Helix-Turn-Helix Domains .....	129
4.4 Actin Binding of ABD2.....	134

4.5 Characterisation of the Actin Binding Domain of ABD2.....	136
4.6 Discussion - Location of Actin Binding Surface of ABD2.....	147
<b>Chapter 5 - Actin Binding Properties of the Other Putative Domains of ms1/STARS.....</b>	<b>150</b>
5.1 Following the Binding of ABD1 to Actin Using NMR.....	155
5.2 Residue Assignment of ABD1 .....	156
5.3 Discussion of the Actin Binding of ms1/STARS.....	156
<b>Chapter 6 - ms1/STARS Binding to ABLIM-2 .....</b>	<b>161</b>
6.1 Characterisation of the Structure of the Domains of ABLIM-2 .....	163
6.2 Binding of the Domains of ABLIM-2 to ms1/STARS.....	176
6.3 Discussion of ABLIM Binding to ms1/STARS.....	182
<b>Chapter 7 - DNA Binding by ms1/STARS.....</b>	<b>185</b>
7.1 Conservation of Residues Between the Recognition Helix of Structurally Similar Winged Helix-Turn-Helix Domains and Helix-3 of ABD2.....	185
7.2 DNA Binding to ABD2 .....	187
7.3 DNA Binding to Tandem (ABD1 and ABD2), ABD1 and ABD2 .....	188
7.4 Interesting Features of ABD1 Revealed By DNA Binding.....	190
7.5 The Binding of the Tandem (ABD1 and ABD2) to DNA Using NMR .....	192
7.6 Discussion of the DNA Binding Activity of ABD1, ABD2 and the Tandem of ABD1 and ABD2 .....	195
<b>Chapter 8 - Discussion .....</b>	<b>201</b>
8.1 - Introduction .....	201
8.2 The Domain Organisation of ms1/STARS.....	201
8.3 The Structure of ABD2.....	203
8.4 Actin binding of ms1/STARS.....	204
8.5 The Binding of ms1/STARS to ABLIM2 .....	205
8.6 The Binding of ms1/STARS to DNA .....	205
8.7 How These Results Aid Our Understanding of ms1/STARS, its Role and What This Could Mean for the Rho Signalling Pathway.....	206
<b>Appendix 1 – Sequences and Primers Used to Produce Constructs .....</b>	<b>214</b>
<b>Appendix 2 – Vectors and Cell Strains .....</b>	<b>219</b>
<b>Appendix 3 – Buffer Recipes.....</b>	<b>220</b>
<b>Appendix 4 – Domains Produced by DomainX.....</b>	<b>224</b>
<b>Appendix 5 – Sequence Specific Assignments .....</b>	<b>225</b>
<b>Appendix 6 – Top 100 DALI Results for ABD2.....</b>	<b>237</b>
<b>Bibliography .....</b>	<b>241</b>

## **Abbreviations**

ABRA	Actin binding rho activator
CD	Circular dichroism
ECG	Electrocardiogram
HSQC	Heteronuclear single quantum coherence
IPTG	Isopropyl $\beta$ -D-1 thiogalactopyranoside
LIC	Ligation independent cloning
LVH	Left ventricular hypertrophy
MEFs	MADS box myocyte enhancer factors
MRTFs	Myocardin related transcription factors
ms1	Myocyte stress 1
NOESY	Nuclear Overhauser Enhancement Spectroscopy
OD	Optical density
SDS PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SRF	Serum response factor
STARS	Striated muscle activator of rho signalling
TCF	Ternary complex factor
TROSY	Transverse Relaxation Optimized Spectroscopy
WT	Wildtype

# **Chapter 1 - Introduction**

## **Chapter 1 - Introduction**

### **1.1 Left Ventricular Hypertrophy**

Left ventricular hypertrophy (LVH) is a syndrome of the heart that occurs in response to increases in blood pressure, particularly those increases of over 180 mm Hg systolic pressure (Kannel, 1983). Systolic pressure is a much more important factor than diastolic pressure, and LVH increases with severity of and exposure to increased blood pressure (Kannel, 1983). On electrocardiograms, it presents as S-T segment and T wave repolarisation abnormalities and/or as large QRS complexes. A figure showing normal S-T segment and T wave repolarisation and a normally sized QRS complex can be seen in figure 1.1.

The Sokolow-Lyon index (Sokolow & Lyon, 1949) is used to determine the extent of LVH and to diagnose it. For LVH to be diagnosed through estimates using the Sokolow-Lyon index, the sum of the S wave in the V1 lead and the R wave in the V6 lead must be more than 35 mV (V1 and V6 refer to the electrodes used to measure the electrocardiogram, and their location is shown in figure 1.2). When a Sokolow-Lyon index value of more than 35 mV is present, the patient can be diagnosed with voltage criteria LVH. The Sokolow-Lyon index and ECG measurements are the two most commonly used ways of diagnosing LVH, although there are other methods, based on either electrocardiograms, radiological techniques or a mixture of the two (Ashley *et al.*, 2000). Figure 1.2 is a description of the basic pattern of electrical activity of the heart which is seen in electrocardiograms.

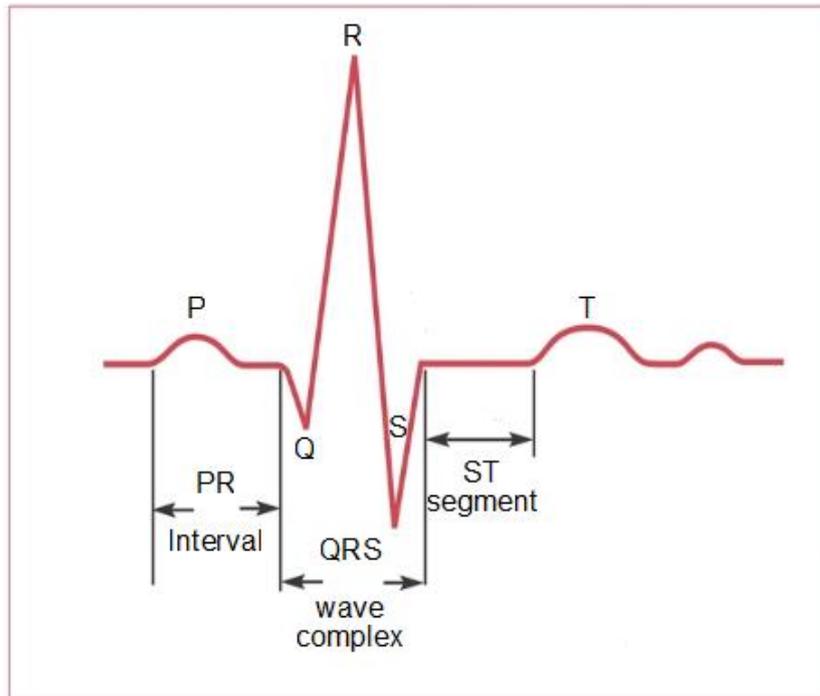


Figure 1.1 – The basic pattern of the electrical activity of the heart. P = atrial depolarisation. PR interval = time between the first deflection of the P wave and the first deflection of the QRS complex. QRS wave complex = ventricular polarisation. Q wave = depolarisation of the interventricular septum. R wave = depolarisation of most of the ventricles. S wave = depolarisation of the base of the heart. ST segment = time between the end of the QRS complex and the start of the T wave. T wave = ventricular repolarisation. Figure taken from ‘Cardiology Explained’ by Ashley and Niebauer (2004).

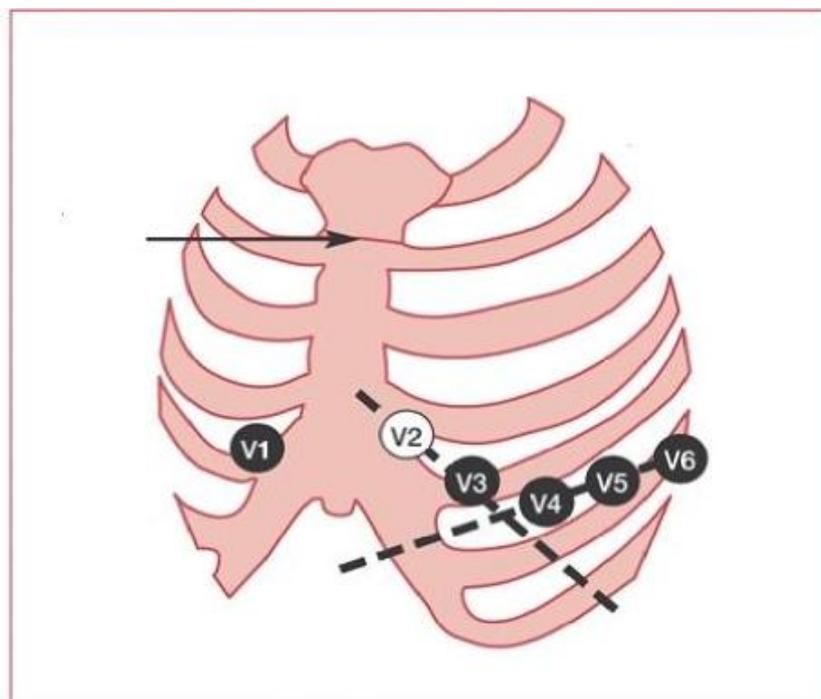


Figure 1.2 – Position of the six V leads used to produce an ECG. Figure taken from ‘Cardiology Explained’ by Ashley and Niebauer (2004).

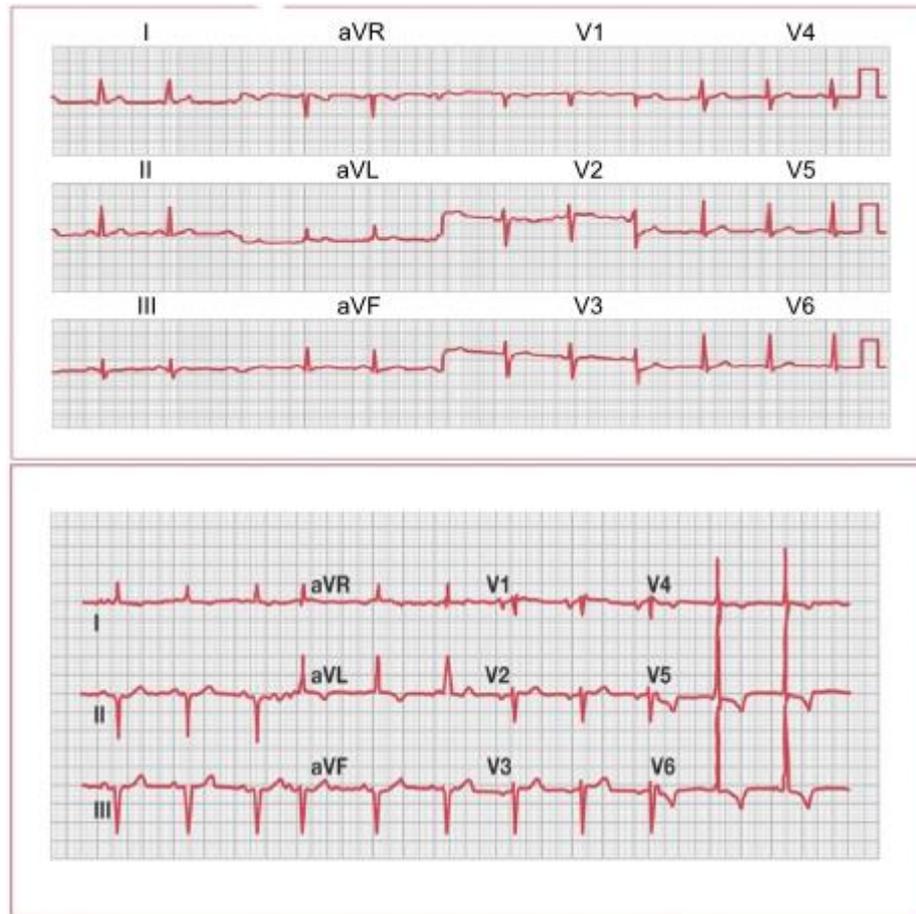


Figure 1.3 – The top figure is that of an ECG of a healthy heart. The bottom figure is an ECG of someone with LVH. The Sokolow-Lyon index is the most commonly calculated index of estimation. If the sum of the S wave in lead V1 (SV1) and the R wave in V6 (RV6) add up to more than 3.5 mV (35 small or seven big squares) then the patient has LVH by voltage criterion. Figures taken from ‘Cardiology Explained’ by Ashley and Niebauer (2004).

In figure 1.3, aVR, aVL and aVF refer to the electrodes used to produce the estimate of potential difference that is measured by leads I, II and III. The aVR electrode is placed on the right arm, the aVL electrode on the left arm and the aVF electrode on the left leg. Lead I measures the difference between the right arm and left arm leads, Lead II measures the difference between the right arm and the left leg leads, while Lead III measures the difference between the left arm and left leg leads. This enables the heart to be measured in three different orientations so that any anomalies are discovered.

In the heart itself, LVH presents as a thickening of the left ventricle of the heart, causing a reduction in the size of the chamber and an increase in its requirement for oxygen.

A patient with electrocardiographic presentation has a greater risk of developing heart disease and having strokes than someone with physiological presentation alone. Compared to someone with a normal heart, patients with electrocardiographic LVH have an eight fold increase in deaths due to cardiovascular disease and a six fold increase in deaths due to coronary diseases (Kannel, 1983).

LVH has a slight predominance in males and a steep increase in incidence with age (Kannel, 1983).

LVH as seen by roentgenograms seems to cause less severe problems than when it is seen on electrocardiograms, and only half the people with LVH by roentgenogram have electrocardiographic LVH. Only 16% of the half of the population that do not have electrocardiographic LVH go on to develop it (Ashley *et al.*, 2000). Having both anatomic and ECG LVH leads to a greater risk of heart failure than either do separately (Ashley *et al.*, 2000). Having LVH with ST depression on an ECG is more dangerous than having high R-wave LVH as seen on an ECG, doubling the danger of death (Ashley *et al.*, 2000). It is thought that this is because electrocardiographic LVH picks up a later stage of LVH, with voltage criteria LVH being found when the concentric hypertrophic stage develops and repolarization changes have occurred after ischemia and/or enlargement by eccentric dilation occurs (Kannel, 1983). In a similar way, high R-wave LVH could be a marker of the heart's response to higher blood pressure, with ST depression being a marker for clinically important poor prognosis, as, if it is included in LVH criteria, the risk of cardiac death increases 15-fold (Ashley *et al.*, 2000).

In the Framingham Study, LVH was found to be a lethal risk attribute, in that it preceded 30 percent of all deaths and 45 percent of cardiovascular deaths. Within 5 years of its diagnosis 35% of men and 20% of women with it had died; a similar number to those with coronary heart disease. It also increases the mortality rate of other heart diseases if it is found with them compared to the mortality rate of those diseases on their own.

### **1.2 Physiological Changes in Left Ventricular Hypertrophy**

In LVH the walls of the ventricle thicken (Kannel, 1983). Other differences that have been noted between healthy hearts and hearts with LVH are that LVH hearts have a lower capillary density (Hudlicka *et al.*, 1992 and Anversa *et al.*, 1986). In physiological left ventricular hypertrophy, such as that which occurs in athletes and does not cause pathological symptoms, the capillary density is not reduced (Richey & Brown, 1998).

### **1.3 The Pathway To Left Ventricular Hypertrophy**

When left ventricular hypertrophy (LVH) was induced in rat hearts using aortic banding the expression of a number of genes increased (Mahadeva *et al.*, 2002). Initially, the expression of *ms1/STARS* increases after 1 hour (Mahadeva *et al.*, 2002). Following this there is expression of early response genes, *c-fos*, *c-myc* and *c-jun*. These are followed by the re-expression of foetal isoforms of cytoskeletal proteins such as foetal  $\alpha$ -actinin and  $\beta$ -myosin heavy-chains (figure 1.4). This characteristic pattern of early response gene and protein foetal isoform expression had been seen previously (Glennon *et al.*, 1995 and Sadoshima *et al.*, 1992) but the expression of *ms1/STARS* which occurs before all of them had not been observed previously.

Expression of these early response genes and the foetal isoforms of heart muscle proteins causes remodelling of the heart, which leads to hypertrophic growth of the wall of the left ventricle of the heart. This growth is due to hypertrophy of the cells not hyperplasia.

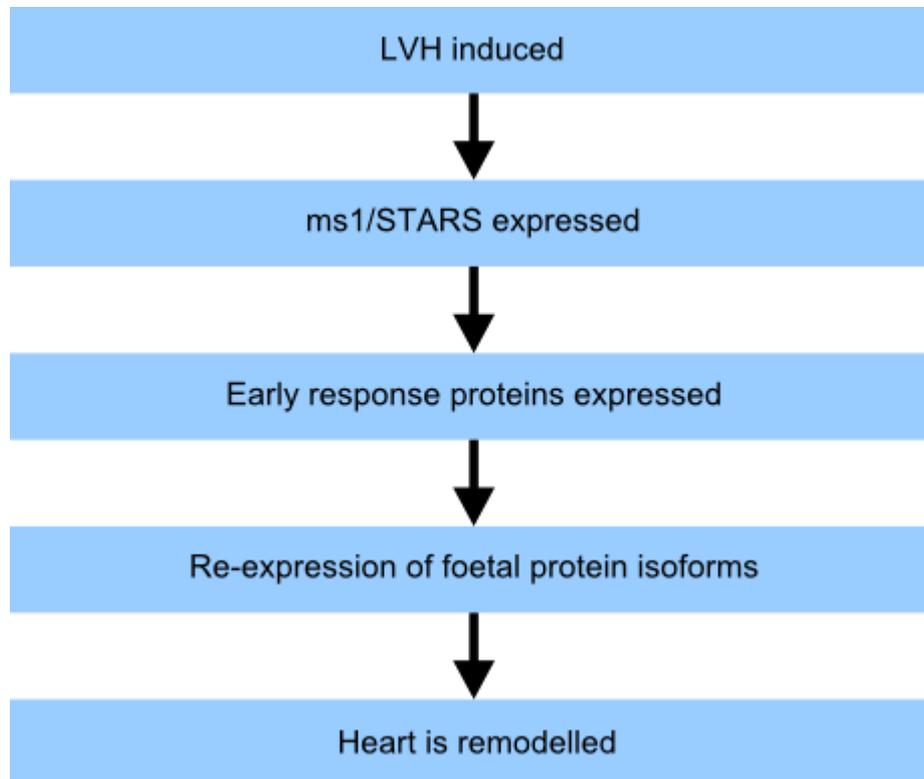


Figure 1.4 – The stages of the pathway of the development of left ventricular hypertrophy.

### **1.4 Myocyte Stress 1 (ms1)**

Myocyte stress 1 (ms1) (Mahadeva *et al.*, 2002) is also known as STARS (striated muscle activator of rho signalling) (Arai *et al.*, 2002) or ABRA (actin binding rho activator) (Uniprot). Its accession number is NP\_631905. The gene that encodes human ABRA is located on chromosome 8, position 8q23.1 (Arai *et al.*, 2002). Human ABRA is highly homologous to rat ABRA (figure 1.5). The rat homologue was used for our experiments.

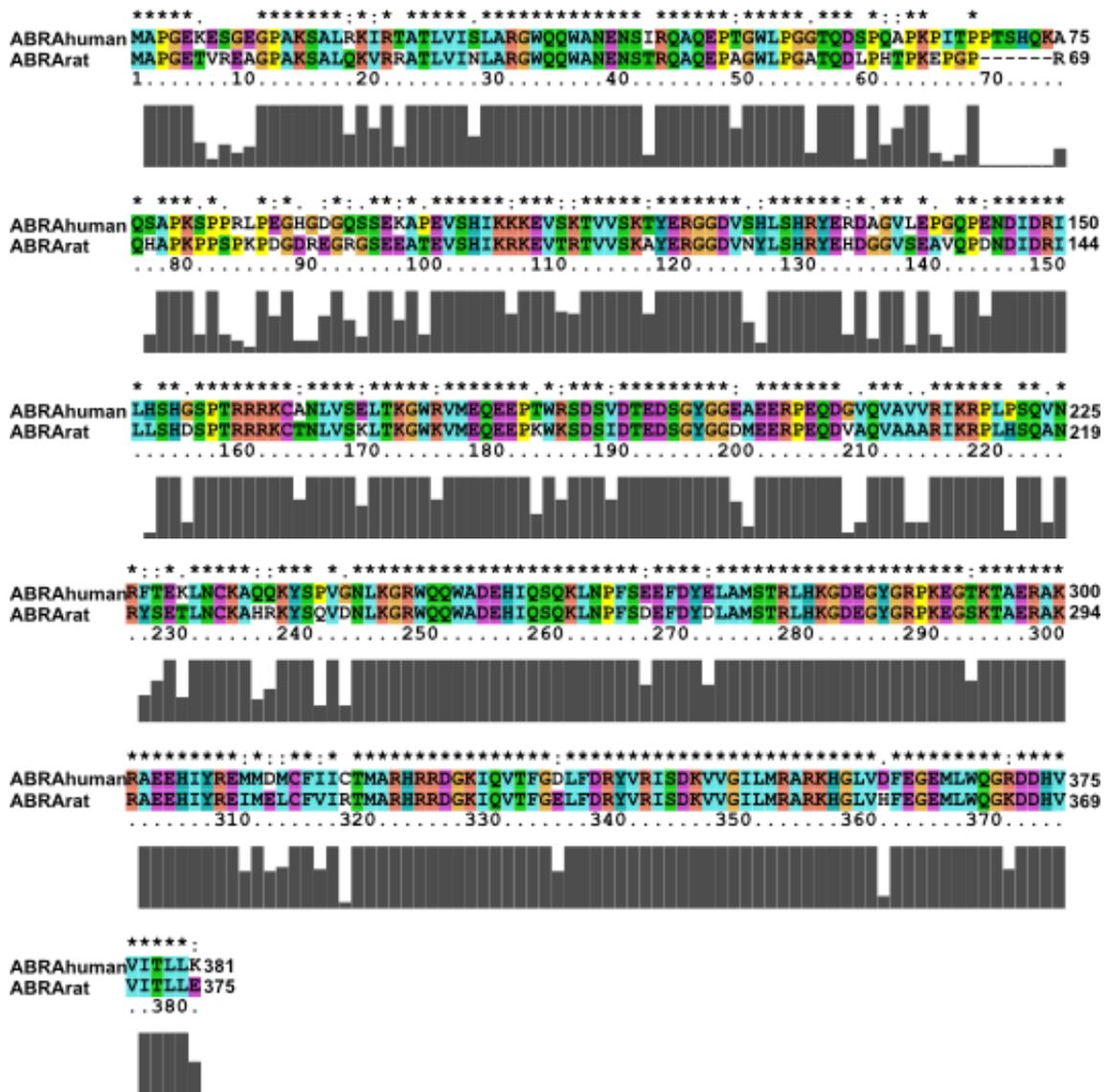


Figure 1.5 – Sequence alignment of human and rat ABRA. Their accession numbers are NP\_631905 for human ABRA and NP\_787038 for rat ABRA.

The gene for rat ms1/STARS (accession number NP\_787038) is located on chromosome 7, position 7q31. The protein expressed is 375 amino acids, and contains no well established domain organisation. It is predominantly expressed in the heart and skeletal muscle (Mahadeva *et al.*, 2002), and it is expressed in the left ventricle of the embryonic heart. In particular, it localises to the I band and M line in rat cardiomyocytes (figure 1.6) (Arai *et al.*, 2002). Expression is then upregulated in adult

hearts in response to stress (Mahadeva *et al.*, 2002). The timing and location of its expression in the heart suggest that ms1/STARS could be a muscle specific regulator of transcription for proteins that could then possibly control differentiation, growth and sarcomeric organisation (Mahadeva *et al.*, 2002).

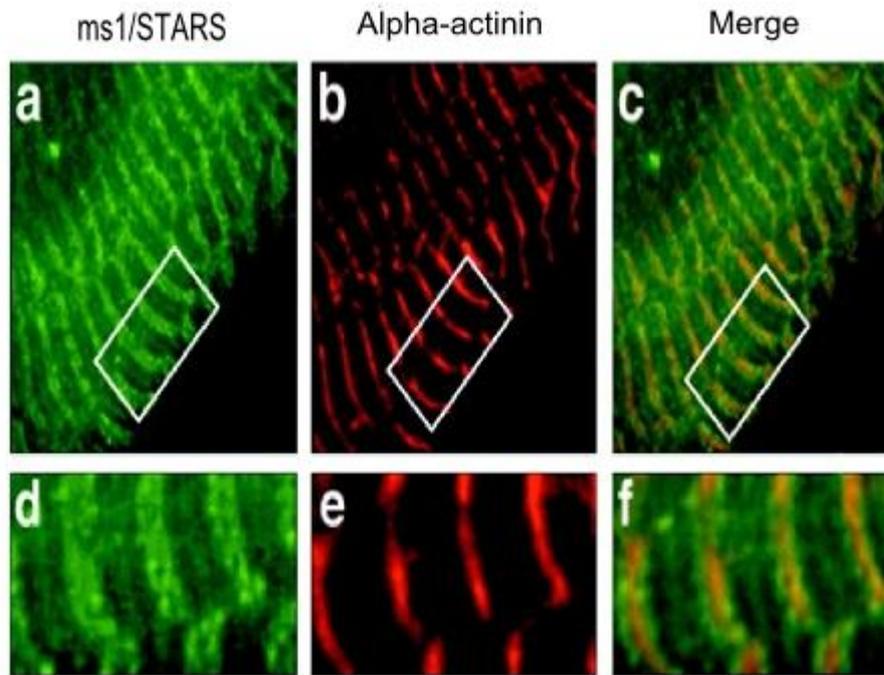


Figure 1.6 – Staining of ms1/STARS and  $\alpha$ -actinin shows that they co-localise on the I band of rat neonatal myocytes, with some staining in the M-line. Figure taken from Arai *et al.*, 2002.

### **1.5 Proposed Mode of Action of ms1/STARS**

It was proposed that ms1/STARS is activated by Rho signalling, and that, once activated, it stabilises actin filaments (Kawahara *et al.*, 2005). This then reduces the available pool of G-actin. This reduction removes the inhibition of SRF through G-actin, enabling SRF to affect the transcription of the genes whose expression it organises. A diagrammatic representation of this pathway is shown in figure 1.7.

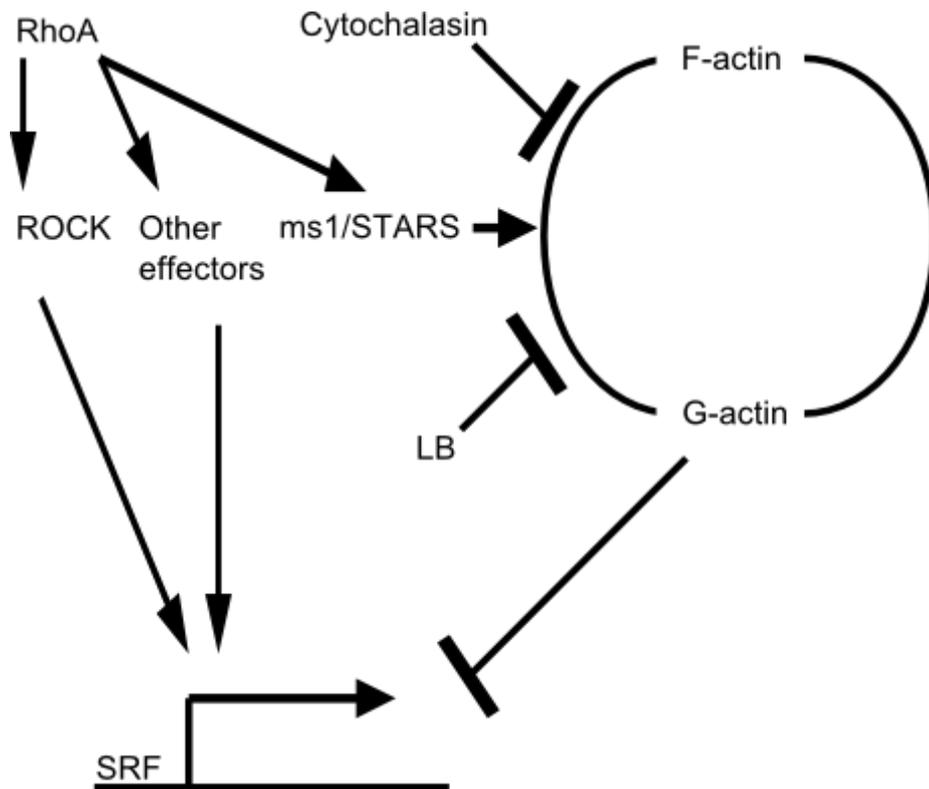


Figure 1.7 – A diagrammatic representation of the putative RhoA-activated pathway mediated by ms1/STARS. LB = Latrunculin B is an inhibitor of actin polymerisation.

Rho is a small GTP-ase that regulates the assembly of focal adhesions and actin fibres when cells are exposed to growth factors (Ridley & Hall, 1992). Actin itself has an influence over many other cellular processes such as contraction, cytokinesis, endocytosis, mitosis, motility and secretion (Burrige & Chrzanowksa-Wodnicka, 1996; Schmidt & Hall, 1998 and Pantaloni *et al.*, 2001). Rho also stimulates SRF through other actin dynamics-related mechanisms, including the activation of Rho kinase and mDia (Sotiropoulos *et al.*, 1999 and Mack *et al.*, 2001), not just through ms1/STARS. Rho is also required for muscle cell differentiation (Takano *et al.*, 1998). Because Rho acts through many other non-muscle-specific pathways, it must require a muscle-specific regulator of its actions, which could be ms1/STARS (Arai *et al.*, 2002).

The link between ms1/STARS binding to actin and the activation of SRF downstream was the myocardin-related transcription factors (MRTFs) (Kuwahara *et al.*, 2005). MRTFs are co-activators of SRF (Wang *et al.*, 2002), and can be induced to relocate to the nucleus by the polymerisation of actin, which can be induced by Rho itself or ms1/STARS. This was shown (Kuwahara *et al.*, 2005) using a modified luciferase protein whose promoter has been engineered to have a CArG box (a DNA sequence of CC(A/T)<sub>6</sub>GG), the site where SRFs bind (Pellegrini *et al.*, 1995 and Treisman, 1986). This is transfected into cells which do not naturally express ms1/STARS. Without it, there is no expression of luciferase. When ms1/STARS is transfected into these cells, luciferase is expressed. This shows that ms1/STARS can be an activator of SRF activity. Because ms1/STARS is expressed specifically in striated muscle cells, RhoA could use it to respond to specific stimuli in a cell-type specific manner. In the case of ms1/STARS this stimuli is increased blood pressure, the cell type is heart muscle cells and the response is hypertrophic growth in the left ventricular cell wall. Interestingly, RNAi-mediated knock downs of MRTF-B have a similar phenotype to those of SRF knockout mice, which suggest they affect similar processes (Li *et al.*, 2005).

The link with SRF is significant as SRF regulates the transcription of early response genes like c-fos and c-jun whose expression is increased following an increase in blood pressure. Transgenic mice with permanently upregulated SRF also develop cardiomyopathy (Zhang *et al.*, 2001), showing that increased SRF is part of a pathway leading to hypertrophy.

### **1.5.1 Serum Response Factor (SRF)**

Serum response factor (SRF) is part of the MADS (MCM1, AG, DEFA and SRF) box transcription factor family, related to the MEFs (MADS box myocyte enhancer factors).

Through combination with different co-factors it can direct the expression of different groups of genes. For instance, when it combines with ternary complex factor (TCF) it responds to mitogen-activated protein kinase signalling (Buchwalter *et al.*, 2004, Dalton & Treisman, 1992 and Hippskind *et al.*, 1991), while interacting with myocardin and MRTFs enables it to respond to actin signalling (Parmacek, 2007, Pipes *et al.*, 2006, Wang *et al.*, 2001 and Wang *et al.*, 2002). These interactions are mutually exclusive (Wang *et al.*, 2004 and Zaromytidou *et al.*, 2006).

SRF is involved in a wide range of signalling, and is important in gastrulation, development of the cardiovascular system, the function of smooth, striated and cardiac muscle cells, endothelial cells, vascularisation, the development and regeneration of the liver, T cell and B cell function and the function of neuronal cells. SRF target genes include actin, cofilin 1 and talin 1.

SRF regulates skeletal muscle gene expression through binding of the CArG box, also known as the serum response element (Cen *et al.*, 2004, Owens *et al.*, 2004 and Miano, 2003). It acts in a muscle-cell specific manner and disrupting its action causes impairment of myoblast fusion and differentiation (Vandromme *et al.*, 1992, Wei *et al.*, 1998 and Croissant *et al.*, 1996). Disruption of ms1/STARS via RNAi reduces muscle-specific SRF activity which further suggests a link between the two (Kuwahara *et al.*, 2005).

### **1.5.2 Myocardin Related Transcription Factors**

Myocardin, MRTF-A (also known as MAL, MKL1 and BASC) and MRTF-B (also known as MKL2 and MAL16), the myocardin related transcription factors (MRTFs) have been shown to be related. They have a series of conserved domains (Wang *et al.*,

2002) but different expression patterns, with myocardin being expressed specifically in the cardiovascular system (Parmacek, 2007, Pipes *et al.*, 2006 and Wang *et al.*, 2001) while the other two are expressed more widely. They associate and co-activate with SRF through a basic region and the adjacent Glu-rich domain.

They bind to G-actin using three RPEL domains, forming a stable complex (Arsenian *et al.*, 1998, Guettler *et al.*, 2008, Miralles *et al.*, 2003 and Zaromytidou *et al.*, 2006), which keeps the MRTFs in the cytoplasm. Myocardin contains RPEL domains which do not bind to G-actin. If myocardin is in the form of homodimers it cannot be affected by G-actin signalling (Kuwahara *et al.*, 2005 and Wang *et al.*, 2001). However, if it forms heterodimers with the other MRTFs, myocardin may be influenced by G-actin signalling acting on the other component of the heterodimer (Wang *et al.*, 2003).

ms1/STARS has been shown to promote the movement of MRTF-A into the nucleus (Kuwahara *et al.*, 2007), which positively regulates the MRTF-A-SRF pathway. Other positive regulators include the Nkx2-5 family (Chen & Schwartz, 1996), some GATA zinc-finger proteins (Belaguli *et al.*, 2000 and Sepulveda *et al.*, 2002) and the CRP family of cysteine-rich LIM-only proteins (Chang *et al.*, 2003).

Negative regulators include LIM-only protein FHL2 (Philippar *et al.*, 2004 and Neuman *et al.*, 2009), histone deacetylase 4 (HDAC4) (Davis *et al.*, 2003) and krueppel-like factor 4 (KLF4) (Liu *et al.*, 2005).

### **1.5.3 ms1/STARS and MEF2 and MyoD**

ms1/STARS is thought to be activated by myocyte enhancer factor-2 (MEF-2) binding to two promoter sites downstream of the ms1/STARS transcription start site (Kuwahara *et al.*, 2007). MEF-2 itself is a stress-responsive transcriptional activator, and signalling

through it could be how stress causes the expression of ms1/STARS. However, as MEF-2 null mice still express ms1/STARS, just at a much lower level, there must also be other transcription factors that play a role. The promoter of ms1/STARS contained binding sites for MyoD (Ouszain *et al.*, 2002).

MyoD is one of the myogenic regulatory factors (MRFs), a group which includes four basic helix-loop-helix E box binding proteins, MyoD, Myf5, Myogenin and MRF4 (Molkentin *et al.*, 1995). The MEFs (MADS-box myocyte enhancer factor) and the MRFs work together to express muscle-specific genes in the embryo (Black & Olson, 1998 and Molkentin & Olson, 1996). SRF and MyoD had been shown to interact and act together to activate certain promoters (Blais & Dynlacht, 2005). However, the ms1/STARS promoter only contains binding sites for MyoD (Ouszain *et al.*, 2008). Mef2C was shown to bind to the ms1/STARS promoter, but appears to be needed for late embryogenesis. This suggests a pattern of expression of ms1/STARS under the control of different transcription factors at different times, and that this expression is controlled overall by MyoD, supported by myogenin and MEF2, which are induced by MyoD (Ohkawa *et al.*, 2006 and Ohkawa *et al.*, 2007).

## **1.6 The Actin Binding Activity of ms1/STARS**

Deletion mutations (Arai *et al.*, 2002) showed that deletion of either aa 234-279 or aa 346-375 abolished binding of actin by ms1/STARS. Previous studies by the group at Leicester University using negatively stained cryo-electron microscopy (figure 1.8) (Fock *et al.*, unpublished) showed that this region (aa 234-375) of ms1/STARS caused actin bundling. This can only occur if there is more than one actin binding site in the region, as bundling requires more than one actin binding site on the same molecule of the binding protein.

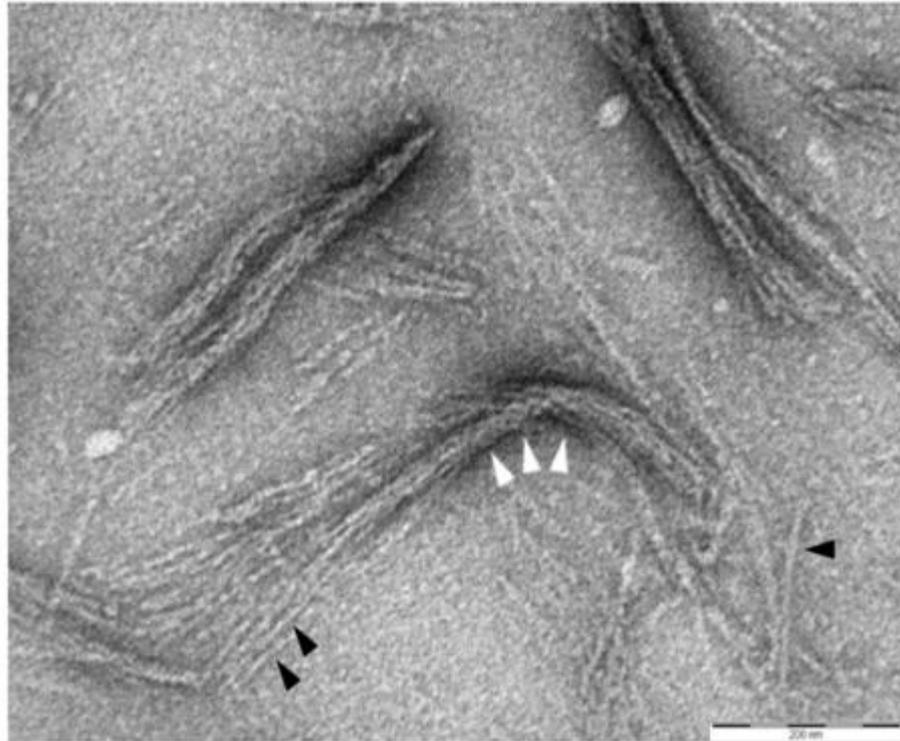


Figure 1.8 – Negative stained cryo-electron microscopy image showing bundling of actin fibres by the C-terminal (aa 234-375) of ms1/STARS (Fock *et al.*, unpublished). White arrows denote actin bundles; black arrows denote individual actin fibres. The bar = 200 nm.

Deletion mutants have shown that residues 234-375 are required for actin binding but not whether it is independently able to bind to actin.

The actin binding activity of ms1/STARS can be studied through the identification of the boundaries of the actin binding domains. The domains can then be independently expressed in *E. coli* and, following purification, their ability to bind to actin can be assayed using both qualitative and quantitative assays.

### **1.7 Actin-binding LIM Protein 2 and 3 (ABLIM-2 and -3)**

Only two other ms1/STARS binding proteins have been found, ABLIM-2 and ABLIM-3 (Barrientos *et al.*, 2007). They are novel members of the ABLIM (actin-binding LIM protein) protein family and contain four LIM domains, followed by a linker region and

then a villin headpiece domain (figure 1.9). This means they contain a LIM consensus sequence. ABLIM-1 was found in the retina in humans and in heart tissue in mice and has been suggested to have a function in actin-dependent cell signalling (Roof *et al.*, 1997). ABLIM-2 and 3 meanwhile were also shown to bind to F-actin and to stimulate the activity of ms1/STARS (Barrientos *et al.*, 2007). ABLIM-2 is expressed in brain, spleen and skeletal muscle but not in the heart, while ABLIM-3 is expressed predominantly in the heart, along with some expression in the liver and lungs (Barrientos *et al.*, 2007). The interaction between ms1/STARS and ABLIM-2 and -3 was found using a yeast two hybrid screen of a skeletal muscle cDNA library. In the screen, the C-terminus of ms1/STARS (residues 234-375) was used as the bait (Barrientos *et al.*, 2007).

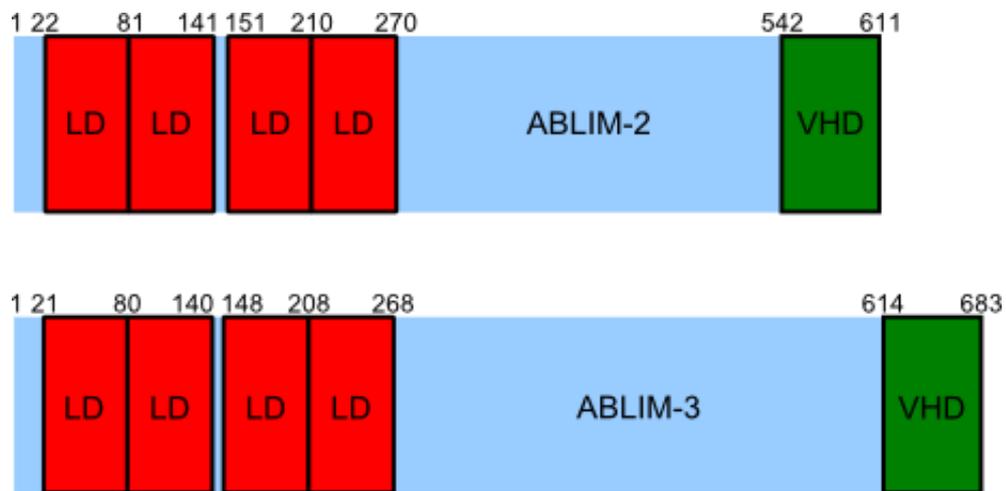


Figure 1.9 – The domain organisation of ABLIM-2 and ABLIM-3. LD = LIM domain, VHD = vilin headpiece domain.

### 1.7.1 LIM Domains

The name of the LIM domains comes from the first three proteins in which the domain was discovered. These proteins were the *C. elegans* cell-lineage protein LIN-11 (Freyd

*et al.*, 1990), rat insulin gene-enhancer-binding protein Isl1 (Karlsson *et al.*, 1990) and *C. elegans* mechanosensory neuron-specifying protein MEC-3 (Way & Chalfie, 1989).

Each LIM domain contains a LIM consensus sequence. The LIM consensus sequence is  $CX_2CX_{16-23}CX_2CX_2CX_2CX_{16-21}CX_2(C/H/D)$  with X denoting any amino acid (Schmeichel & Beckerle, 1994). This forms 2 zinc fingers, with cysteines 1-4 co-ordinating the first zinc atom and cysteines 5-8 co-ordinating the second (figure 1.10) (Michelsen *et al.*, 1994). Each zinc finger forms two anti-parallel  $\beta$ -hairpins (Perez-Alvarado *et al.*, 1994 and Kosa *et al.*, 1994) as can be seen in figure 1.11.

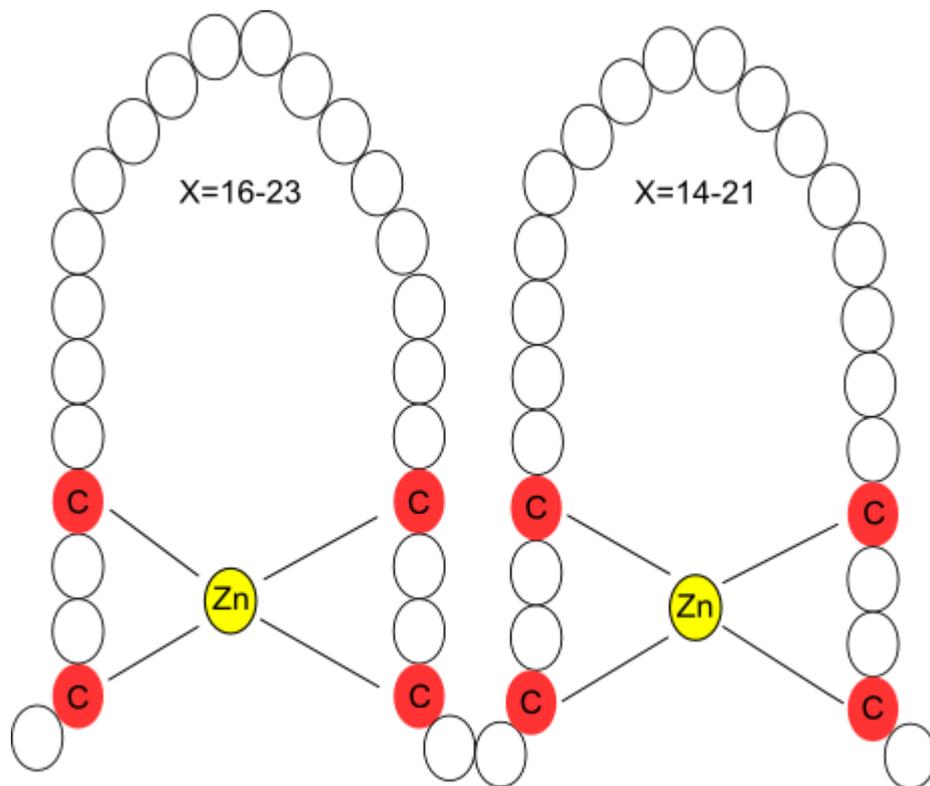


Figure 1.10 – Graphical representation of the conserved sequence of a typical LIM domain. The 4 cysteines co-ordinate one zinc atom each, forming 2 zinc finger domains inside each LIM domain.

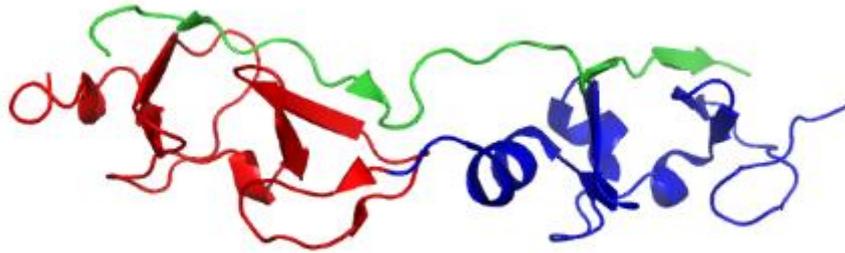


Figure 1.11 – Structure of the LMO4 LIM domains (red and blue) bound to the LIM-interaction domain of LIM-domain-binding protein 1 (green) (Jeffries *et al.*, 2006).

LIM domains are found in a variety of eukaryotic proteins, with a variety of functions. ABLIM-2 and -3 are members of the actin associated family of LIM proteins, which also includes paxillin and zyxin.

LIM domains are protein-protein interaction domains (Schmeichel & Beckerle, 1994, Arber & Caroni, 1996 and Feuerstein *et al.*, 1994); however their preferred binding partners are unknown. Some LIM domains act as simple domains, whereas others act as tandems. Mutations in the *C. elegans* homologue of ABLIM-1, unc-115, lead to a developmental defect in some neurons (Lundquist *et al.*, 1998). Interestingly, unc-115 also functions downstream of a Rho-family GTPase, Rac, and is also possibly a regulator of actin dynamics (Struckhoff & Lundquist, 2003).

### **1.8 How We Aim To Study ms1/STARS**

To study ms1/STARS, its domain organisation must first be understood. Domains are being used rather than the whole protein, as we have discovered that the whole protein cannot be expressed in *E. coli* without degradation. That ms1/STARS degrades into domains can be seen by running gels of the degraded full-length protein (figure 1.12). Instead of there being one large smear, as would occur if there were total degradation, there are distinct bands, suggesting it degrades into domains. After the identification of

the domain boundaries of ms1/STARS, using both sequence homology and combinatorial domain hunting (Reich *et al.*, 2006), the domains will be characterised using biophysical methods including circular dichroism, fluorescence spectroscopy and NMR.

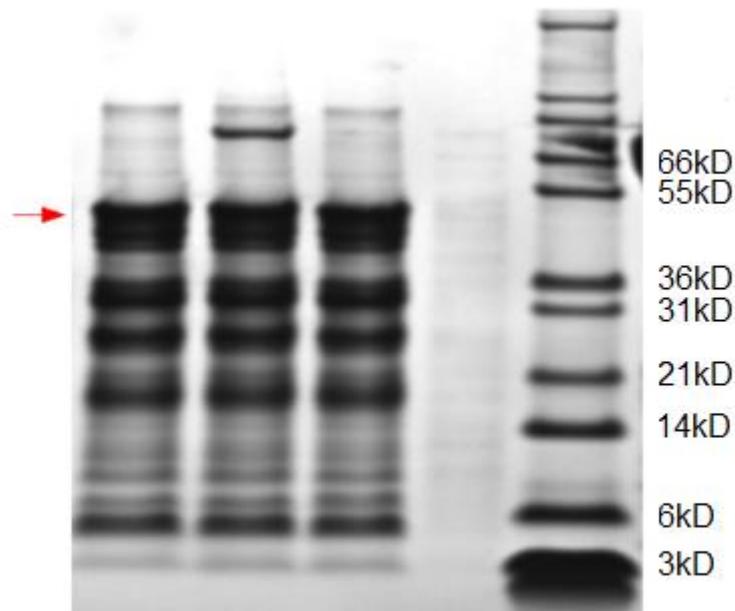


Figure 1.12 – SDS PAGE gel of the full length ms1/STARS. It degrades into distinct segments suggesting that it is made of multiple domains. Full length ms1/STARS is indicated by the red arrow.

If the actin-binding domain is folded and the right size to be studied by NMR (nuclear magnetic resonance spectrometry), the structure of the identified actin-binding domain will then be studied using NMR.

NMR can also be used to follow binding of proteins. This can be done by comparing a  $^{15}\text{N}$ -HSQC (heteronuclear single quantum coherence) spectrum of the protein plus the putative binding protein with a  $^{15}\text{N}$ -HSQC of the protein alone. An HSQC spectrum shows a peak for the amide group of every residue in the protein, except for prolines which have imino groups instead of amide groups. Binding to a ligand causes

movement of these peaks because the interaction changes the chemical environment of the protons that produce the peaks changing their chemical shifts. If the residue that the amide groups belong to can be identified by combining the data from the  $^{15}\text{N}$ -HSQC spectrum with those of HNCA, HNCACB, HN(CO)CA and HN(CO)CACB spectra, the residues involved in binding can be discovered. While this technique is suitable for testing binding of ABLIM-2 and -3 to the C-terminal of ms1/STARS, it is not suitable for testing the binding of a protein to actin. When binding occurs, the rotation time of the whole bound protein and ligand must be small enough for it to tumble at a rate that is fast enough for the NMR machine to be able to interpret the data. F-actin on its own is larger than this, and therefore when anything binds to it, the combined actin and ligand is above this limit, and therefore the peaks become too broad to be seen. This occurs when the protein that binds to actin has a defined secondary structure. If a protein does not have a defined secondary structure and the binding area is small, only those residues that actually bind disappear as the rest, because they are unstructured, remain dynamic enough to tumble quickly enough that the peaks do not broaden and therefore can still be seen.

Instead, the residues important for actin binding in the actin binding domains will be identified by mutation of residues that were presumed to be important once a 3D structure has been calculated.

There is also an interest in where ms1/STARS binds to ABLIM-2 and -3 as the same region (amino acids 234-279) has been implicated in binding to both actin and ABLIM2 and 3 (Arai *et al.*, 2002 and Barrientos *et al.*, 2007).

# **Chapter 2 – Materials and Methods**

## **Chapter 2 – Materials And Methods**

### **2.1 Bioinformatics**

A Blast-p search was performed using the amino acid sequence of ms1/STARS as the search parameter; the Blast-p algorithm was used for this (<http://www.ncbi.nlm.nih.gov>) (Altschul *et al.*, 1990). The database used was the non-redundant protein sequence database.

Following this ClustalW (<http://www.ebi.ac.uk/Tools/clustalw/>) (Thompson *et al.*, 1994) was used to align the sequences using the default settings on the website.

ClustalX (Larkin *et al.*, 2007) was used to provide the alignment figures in this thesis.

RONN (regional order neural network) (Yang *et al.*, 2005) was used to try to predict which regions of ms1/STARS would be ordered and disordered.

### **2.2 Construction of Primers**

Using the cDNA sequence (Accession ID: BC158575) of the most conserved region of ms1/STARS as a template, as determined by comparison to and alignment with its homologues using ClustalW, primers were designed for the last 81 residues of ms1/STARS, residues 294-375 (primers and sequence can be found in appendix 1). As ClustalW aligned the sequences of the similar short proteins and the C-terminal of ms1/STARS in two different ways, two sets of primers were designed, one to produce a construct from residue 294-375 and the other to produce a construct from residue 297-375. Two further smaller constructs were also made (residues 314-375 and 317-375, the primers and sequence of these can be found in appendix 1), because this region was even more strongly conserved. The melting temperatures of the primers were calculated

using OligoCalc (<http://www.basic.northwestern.edu/biotools/oligocalc.html>) (Kibbe, 2007) and the primers were modified so that they had similar melting temperatures. The sequences of the designed primers were then sent to Invitrogen to be produced (see section 4.1.2 for details). This DNA was then cloned into pETM-11 vector (figure 2.1).

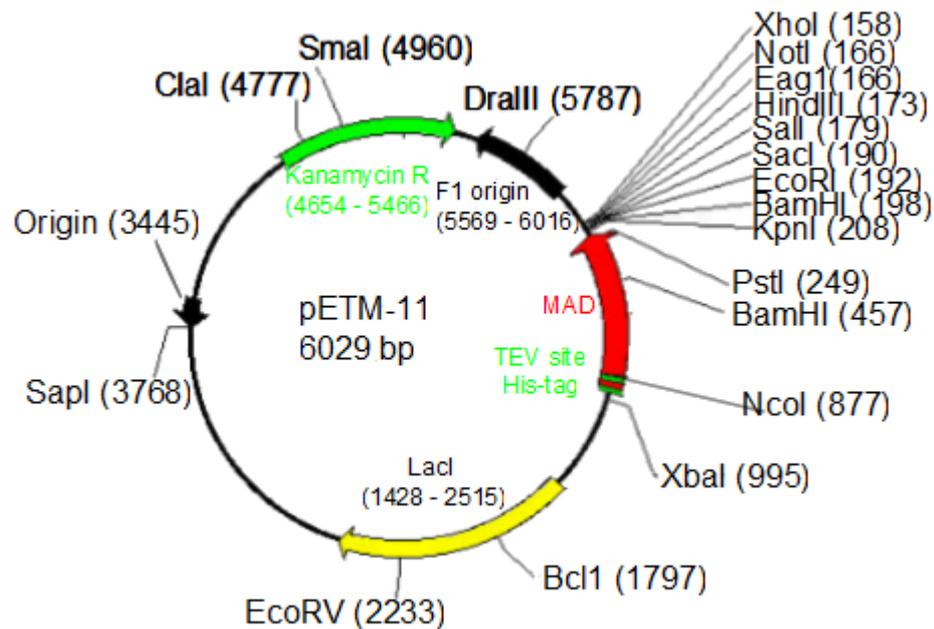


Figure 2.1 – Vector map of pETM-11 with major features annotated. The numbers in brackets are the location in base pairs of the start of the features. Kanamycin R is the kanamycin resistance gene.

The same procedure was used to produce primers for the mutation analysis of the second actin binding site of ms1/STARS (ABD2 (residues 294-375)) which involved expressing the ABD2 mutants in pLeics03 (figure 2.2), for sub-cloning the DomainX constructs (see table 2.1) that were in a pDXV plasmid into pLeics07 (figure 2.3) and for the sub-cloning of the individual LIM domains of ABLIM-2, producing tandem domains of ABLIM-2, and for production of constructs containing LIM domain 4, the linker and the VHD of ABLIM2, the VHD on its own and of LIM1 to the VHD. These

ABLIM-2 domain constructs were also sub-cloned into pLeics07. pLeics03 and pLeics07 were both developed by Dr. Xiowen Yang at the University of Leicester.

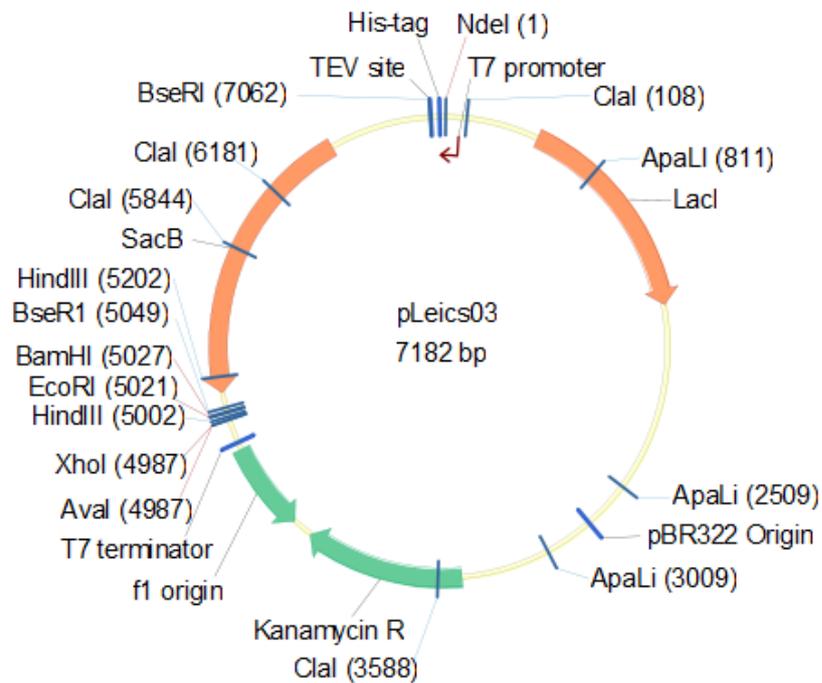


Figure 2.2 – Vector map of pLeics03 with major features annotated. The numbers in brackets are the location in base pairs of the start of the features. Kanamycin R is the kanamycin resistance gene.

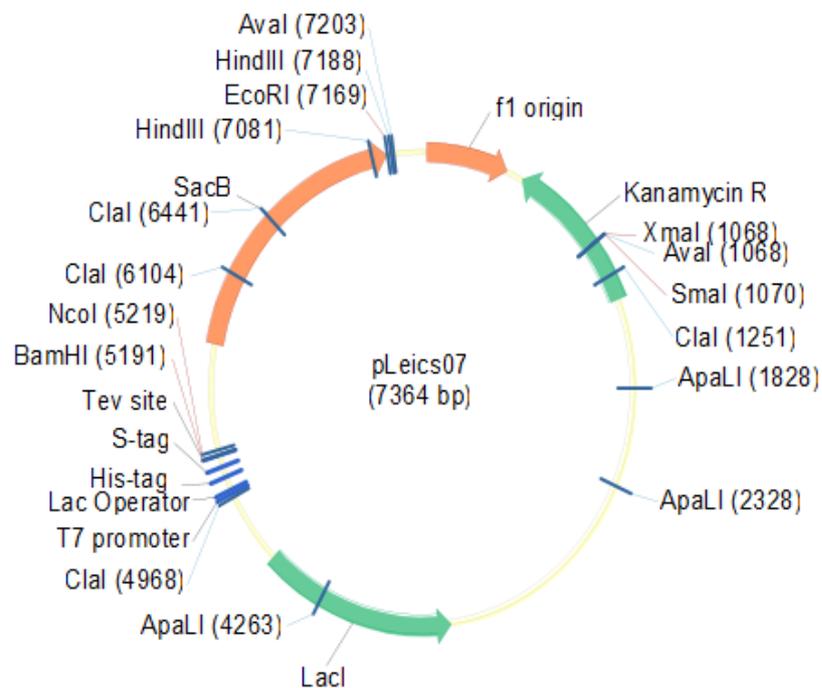


Figure 2.3 – Vector map of pLeics07 with major features annotated. The numbers in brackets are the location in base pairs of the start of the features. Kanamycin R is the kanamycin resistance gene.

DomainX Construct	Residues of ms1/STARS that it contains
MSD1	2-118
MSD2	40-196
ABD1	193-295
ABD2	294-375

Table 2.1 – DomainX construct names and which residues of ms1/STARS they contain. The constructs overlap so that the whole protein is covered by them.

### **2.3 Polymerase Chain Reaction (PCR)**

All media and gel recipes can be found in appendix 3.

Primers are diluted to 25 ng/ $\mu$ l. The template DNA is diluted to 200 ng. These are added to the following PCR reaction mixture - 5  $\mu$ l enhancer (Invitrogen), 5  $\mu$ l amplification buffer (Roche), 1.5  $\mu$ l dNTPs (Promega), 1  $\mu$ l MgSO<sub>4</sub>, 200 ng template DNA, 1  $\mu$ l primers (forward and reverse), 1  $\mu$ l of each pfx polymerase (Invitrogen) and 33.5  $\mu$ l water.

This mixture is then heated at 94°C for 2 minutes followed by 30 cycles of the following times and temperatures: 94 °C for 15 seconds, 60 °C for 30 seconds and 68 °C for 30 seconds. After this, the next step is an annealing step at 72 °C for 12 minutes. The mixture is then held at 4 °C.

Changes can be made to the contents of the PCR mixture, the temperature of the cycle steps and the duration of the cycle steps in order to increase the amount of PCR product produced. Any changes to this basic process used are recorded in the results.

### **2.3.1 Agarose DNA Gels**

1% agarose gels were made. 3 µl of ethidium bromide (Sigma) or Sybr Safe DNA gel stain (Invitrogen) was added to allow visualisation of the DNA. Samples were run for 100 minutes at 80 volts in 1 x TAE buffer.

## **2.4 Cloning**

### **2.4.1 Restriction Digest of Plasmid Vector**

Digestion for the cloning of ABD2 into pETM-11 was performed using BamHI and NcoI restriction enzymes (Roche). For this, 2 µg of pETM11 DNA (29 µl), 4 µl of 10x restriction enzyme buffer (Roche), 1 µl of NcoI, 1 µl of BamHI and 5 µl of distilled water were used. This was then incubated at 37°C for 2 hours.

### **2.4.2 Gel Extraction of Restriction Products**

The products of the restriction digest were cut out of the agarose gel and then extracted using the Qiagen QIAquick Gel Extraction Kit according to the manufacturer's instructions.

### **2.4.3 Preparation of PCR Products For Cloning**

PCR products were run on a gel and then extracted using the Qiagen QIAquick Gel Extraction Kit according to the manufacturer's instructions.

The products of the gel extraction were digested with NcoI and BamHI restriction endonucleases (48 µl PCR product, 6 µl buffer B (Invitrogen), 3 µl NcoI, 3µl BamHI) and incubated at 37°C for 2 hours.

The products of the restriction digest of the PCR products were purified using a Qiagen PCR purification kit according to the manufacturer's instructions.

#### **2.4.4 Ligation of the Restricted Vector and PCR Product**

The concentration of digested vector and PCR product were estimated from an agarose gel by comparison with the brightness of the markers (Ready-Load 1Kb Plus DNA ladder, Invitrogen). They were then put into ligation reactions in 1:1 and 1:3 vector to construct ratios along with 8.5 µl of 1x DNA dilution buffer (Roche), 40 µl of 1x T4 ligation buffer (Roche) and 1 µl of T4 ligase (Roche). The mixtures were incubated on ice for 30 minutes. 2 µl of the ligated DNA was then transformed into 100 µl of Top10 One-Shot (Invitrogen) competent cells.

#### **2.4.5 PCR Screening of Ligated Products**

Colonies of transformed cells were picked from LB agar plates and resuspended in 50 µl of H<sub>2</sub>O. The material on the pipette tip used to do this was then inserted into another Luria broth (LB) agar plate containing the appropriate antibiotic. This was repeated for all colonies. 10 µl of the resuspended colonies were used in a PCR reaction as the template, along with 1 µl of forward primers, 1 µl of reverse primers, 1 µl of H<sub>2</sub>O and 12.5 µl Sigma Red-Taq (Sigma Aldrich).

The PCR cycle used was: heat to 94°C for 30 seconds, then 30 cycles of the following: 94°C for 30 seconds, 60°C for 30 seconds and 72°C for 30 seconds. The mixture was then kept at 72°C for 8 minutes and held at 4°C until needed. The PCR products were then run on an agarose gel.

100 µl of LB medium was added to the remaining 40 µl of the resuspended colony mixture and incubated at 37°C until the PCR reaction finished. These cultures then had 60µl 50% glycerol added to them and were stored in the –80°C freezer.

## **2.5 Determination of DNA Concentration**

UV photospectrometry was used to determine the concentration of the DNA produced by the PCR screen.

Concentration (µg/ml) = absorbance at 260 nm x dilution factor x 50 µg/ml

## **2.6 Cloning by PROTEX**

The cloning of the constructs containing domains of ABLIM-2 and the cloning of the domains of ms1/STARS determined by DomainX was performed by Dr. Xiowen Yang of the University of Leicester's Protein Expression Laboratory (PROTEX). Primers were produced as previously described, and were submitted to Dr. Yang. As the constructs were being cloned into pLeics03 or pLeics07, the primers had to include the following homology regions:

5': TACTTCCAATCCATG...

3': TATCCACCTTTACTGTCA...

followed by the construct-specific residues of the primer. The homology region is shared by all of the group C vectors produced by PROTEX. Clontech's In-Fusion™ PCR cloning kit was used for the cloning. This method, ligation-independent cloning or LIC (Haun *et al.*, 1992), does not use restriction enzymes, ligases or phosphatases. Instead, cloning is performed using recombination which occurs at the sites in the vector homologous to the homology regions in the primers. This means insertion is

always correctly orientated, and the DNA of interest can be transformed into any vector with those homology regions. The vectors contain an antibiotic resistance gene, so LB agar plates containing the antibiotic that the vector is resistant to can be used to screen for uptake of the vector. Successful recombination can also be screened for, as it disrupts the SacB gene present in the vector. SacB is poisonous to *E. coli*, if a vector that has not undergone recombination is taken up by a cell, the SacB product will kill that cell when it is grown on LB agar plates. This means that the only colonies produced by the cloning procedure that will grow on an LB agar plate with the correct antibiotic will be from those cells that have taken up a colony containing the DNA of interest.

To produce mutated proteins, primers were designed with an 18-25 bp sequence homologous to that found in the sequence of interest, then the base being mutated, followed by another 18-25 bp. These were then combined with normal cloning primers to clone the mutated sequence into an expression vector. The procedure is described in figure 2.4.

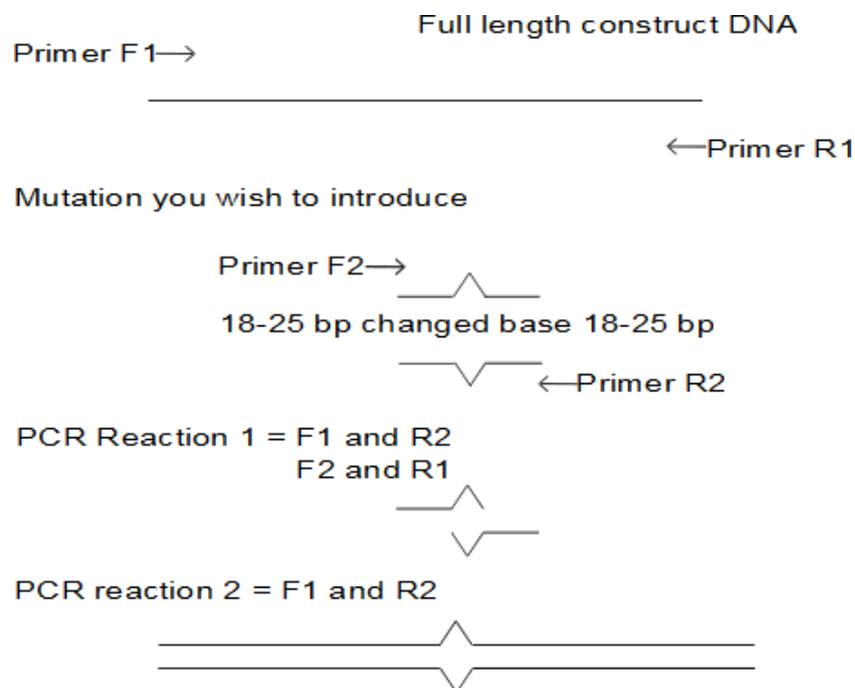


Figure 2.4 – How mutations were introduced into ABD2 using LIC.

## **2.7 DNA Sequencing**

DNA sequencing was performed by the University of Leicester's Protein and Nuclear Acid Chemistry Laboratory (PNAACL). For this at least 8 µl of 62.5 ng / µl DNA was required.

## **2.8 Transformation of Vector into Cells**

Cells from the protein expression cell strain of choice were thawed on ice. 5-10 ng of DNA was added to 25 µl of cells. This mixture was incubated on ice for 30 minutes, then heat-shocked for 30 seconds, then returned to ice for 2 minutes. 250 µl of LB was added, and this mixture was incubated in a shaking incubator at 37°C for 1 hour. 20 µl was spread on LB agar plates containing the appropriate antibiotic. The plates were incubated overnight at 37°C. Details of the antibiotic resistance of the various vectors and cell lines can be found in tables 1 and 2 in appendix 2.

## **2.9 Protein Expression**

### **2.9.1 Pilot Expression Trials**

The recipe for LB media can be found in appendix 3.

10 µl of the appropriate antibiotic was added to 10 ml of LB. 10 µl of the glycerol stock of the cells containing the construct or 1 colony from an LB agar plate of freshly transformed bacteria was added to the LB.

The cells were induced with 750 µM IPTG (Isopropyl β-D-1 thiogalactopyranoside, Melford Laboratories) once they reached an optical density (OD) of 0.7. Optical density is measured by absorption spectrophotometry at 550 nm.

After incubation, the cells were centrifuged at 5000 rpm (4444 rcf in an Eppendorf centrifuge 5804 R) for 15 minutes. The pellet was resuspended in 5 ml of FF6 Wash Buffer and then the cells were opened using sonication. A 10 µl sample was taken; this is the total protein sample. The sample was spun down again, at 5000 rpm (4444 rcf Eppendorf centrifuge 5804 R) for 15 minutes. A 10 µl sample of the supernatant was taken; this is the soluble protein sample.

These samples were run on a sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS PAGE) gel (Invitrogen) (see section 2.12.6) to compare the amounts of soluble and total protein, in order to find the best expression conditions.

### **2.9.2 Full-Scale Protein Expression**

10 µl of the appropriate antibiotic was added to 10 ml of LB. 10 µl of the glycerol stock of the cells containing the construct or 1 colony from an LB agar plate of freshly transformed bacteria was also added to the LB. The cells were incubated at 37°C in a shaking incubator, shaking at 180 rpm.

1 ml from this 10 ml day culture was added to 200 ml LB with 200 µl of the appropriate antibiotic. This was incubated overnight at 37°C in a shaking incubator, shaking at 180 rpm.

20 ml of the overnight culture was added into 4 x 1 litre of LB. The litre cultures were induced with 750 µM IPTG once they reached an OD of 0.7. After this, they were incubated according to the conditions found to produce the highest expression of the protein of interest in the expression trials.

After incubation, the cultures were spun down at 5000 rpm (5471 rcf using an SLC 6000 Sorvall rotor in a Sorvall Evolution RC centrifuge) for 15 minutes. The pellets were resuspended in 8 ml FF6 wash buffer per litre of LB.

There were no problems when this method was used to scale up protein expression.

### **2.9.3 Protein Expression in M9 Minimal Media**

#### **2.9.3.1 Media**

The M9 minimal medium used to express labelled protein contained 42 mM Na<sub>2</sub>HPO<sub>4</sub> (Fisher Scientific), 22 mM KH<sub>2</sub>PO<sub>4</sub> (Fisher Scientific) and 8.5 mM NaCl (Fisher Scientific), at pH 7.4. The micronutrient solution used during expression in minimal media contains 3 x 10<sup>-6</sup> M NaMoO<sub>4</sub> (Sigma Aldrich), 4 x 10<sup>-4</sup> M H<sub>3</sub>BO<sub>3</sub> (Fisher Scientific), 3 x 10<sup>-5</sup> M CoCl<sub>2</sub> (Sigma Aldrich), 1 x 10<sup>-5</sup> M CuSO<sub>4</sub> (Acros), 8 x 10<sup>-5</sup> M MnCl<sub>2</sub> (Sigma Aldrich) and 1 x 10<sup>-5</sup> M ZnSO<sub>4</sub> (Sigma Aldrich). The vitamin solution used contains 0.4 g Choline chloride (Sigma Aldrich), 0.5 g Folic acid (Acros), 0.5 g Pantothenic acid (Acros), 0.5 g Nicotineamide (Acros), 1.0 g Myo-inositol (Sigma Aldrich), 0.5 g Pyridoxal HCl, 0.5 g Thiamine HCl, 0.05 g Riboflavin (Sigma Aldrich) and 1.0 g Biotin per litre of medium.

For complete M9 media, add 1 ml of micronutrient solution (1/1000 dilution), 1 ml of vitamin solution (1/1000 dilution), 2 ml of MgSO<sub>4</sub>, 100 µl of 0.1M CaCl<sub>2</sub>, 6 g of glucose, 1 g of NH<sub>4</sub>Cl and 1 ml of appropriate antibiotic to 1 litre of M9 salt solution.

#### **2.9.3.2 Expression in M9 Minimal Media**

Cell cultures were grown as described in section 2.12.3, until the point at which they are incubated in litre flasks. For M9 minimal media expression, the cultures were grown

until the OD is 0.8 in the litre flask growths. The cells were then spun down at 5000 rpm for 5 minutes (5471 rcf using an SLC 6000 Sorvall rotor in a Sorvall Evolution RC centrifuge) and then resuspended in M9 salts solution by vortex to avoid cell lysis. The resuspended solution was then spun down again at 5000 rpm for 5 minutes (5471 rcf using an SLC 6000 Sorvall rotor in a Sorvall Evolution RC centrifuge). The pellet was then resuspended in complete M9 media by vortex. The solution was incubated for 1 hour at 15°C, and then induced with 750 µM of IPTG and incubated at the conditions shown to provide best growth in the pilot expression trials.

#### **2.9.4 Glycerol Stocks**

300 µl of 50% glycerol was added to 700 µl of LB containing cells that contain the plasmid of interest. These glycerol stocks were then stored in a freezer at -80°C.

#### **2.9.5 Sodium Dodecylsulphate Polyacrylamide Gel Electrophoresis Gels**

Sodium dodecylsulphate polyacrylamide gel electrophoresis (SDS PAGE) gels (NuPage Novex 4-12% pre-cast bis-tris gels (Invitrogen)) were run at 200 V for 35 minutes in NuPage (TM) MES SDS running buffer (Invitrogen). The samples consisted of 10 µl of sample and 5 µl of loading dye (Mark12 Standard (Invitrogen)). The gels were then washed 3 times in water. After the third wash, the gels were covered in Simply Blue Safe Stain (Invitrogen) and heated in a microwave for 10 seconds. They were then incubated on a shaker for at least 1 hour. After incubation, the stain was removed and the gels washed 3 times with water.

### **2.9.6 Protein Purification**

Columns containing 2 ml of Nickel Sepharose 6 Fast Flow resin (GE Healthcare), two for each protein being purified, were washed with 4 column volumes (4 x 12 ml) of FF6 Wash buffer (see appendix 3 for buffer recipe). The cell pellets produced from the large scale protein expression were opened in a French press and were spun down for 90 minutes at 18000 rpm (39191 x g using a Beckman JA25.50 rotor in a Beckman Coulter Avanti J-30 I centrifuge). The supernatant from this centrifugation was diluted by 50% with FF6 wash buffer and poured over the prepared nickel sepharose columns. The flow-through was collected and a 10 µl sample taken to run on an SDS PAGE gel. The column was washed 4 times with FF6 wash buffer and the wash fractions collected. A 10 µl sample was taken for the SDS PAGE gel. 1 column volume of FF6 elution buffer (see appendix 3 for buffer recipe) was poured down the nickel sepharose column and collected. A 10 µl sample was taken to run on an SDS PAGE gel.

The samples on the SDS page gel show the level of purification of the protein. If the protein is pure after this first stage, it can be used directly in experiments; if not, a second column purification stage must be performed. The protein of interest should be in the elution fraction. TEV protease was added to this fraction to cleave the His tag from the protein. 40 µl of AcTEV protease (Invitrogen) was added to the elution fraction. The mixture was put into 3000 MW cut-off dialysis tubing and dialysed overnight against 1 litre of FF6 wash buffer, at 4°C.

After this, the nickel column purification steps are repeated to recover the purified and cleaved protein.

### **2.9.6.1 Purification Using Gel Filtration**

Further purification, if required, can be obtained by use of size exclusion chromatography with a Superdex 75 HiLoad 16/60 pre-packed column (GE Healthcare). The column volume is 120 ml. The machine used to perform this is an Äkta Purifier 9 (GE Healthcare), using Unicorn software. 5 ml aliquots of the protein are run on the column, at a flow rate of 1 ml/min for 120 ml. Flow-through was collected in 5 ml aliquots after 0.23 column volumes (the void volume) for another 0.8 column volumes. The recipe for the buffer used is in appendix 3.

### **2.9.7 Protein Stabilisation**

For proteins which showed excessive precipitation, NVOY (Expedeon) was used according to manufacturer's instructions to stabilise the proteins in solution. It was used at 5 µg / ml per 1 µg / ml of protein.

### **2.9.8 Concentrating Protein**

Vivaspin concentrators of the appropriate molecular weight cut off for the size of the protein were used to concentrate proteins. The concentrators were spun at 2000 rcf at 4°C in an Eppendorf 5804 R centrifuge.

### **2.9.9 Measuring Protein Concentrations**

Two methods were used to measure protein concentrations. The first is a fluorescence based method. For this the Qubit Quantitation Platform (Invitrogen) was used according to the manufacturer's instructions. The second method used ultra-violet

spectroscopy. A scan from 240 nm to 310 nm was run on a sample of the protein. The absorption at 280 nm is used to obtain the concentration.

$$\text{Concentration} = (A_{280\text{nm}} \times \text{dilution factor}) / (\text{extinction co-efficient} \times \text{path length})$$

This gives the concentration in M.

## **2.10 Buffer Exchange**

Two methods were used to exchange proteins between different buffers. In the first method, buffer exchange was performed using Nap5 and PD10 columns (GE Healthcare). Proteins were exchanged into different buffers using these columns according to the manufacturer's instructions. The second method used dialysis. Proteins were put into dialysis membrane (Spectrapore) of an appropriate molecular weight cut off to their molecular weight. The protein was dialysed against the desired buffer, with at least three changes of buffer before the process was finished.

## **2.11 Circular Dichroism (CD)**

Circular dichroism (CD) is a technique based on the polarisation of light in a plane. Polarised light consists of two parts, left and right. If both parts of polarised light pass through a sample and are equally absorbed, there is no change in the polarisation of the light. However, if the components are absorbed in different amounts, the light becomes elliptically polarised. This elliptical polarisation occurs when the sample being studied is chiral, is linked to a chiral molecule or has become asymmetric due to the 3D structure of the molecule.

CD values are measured as the change in absorbance, or left component absorbance minus right component absorbance, but are expressed as the ellipticity in degrees, as the change is plotted in terms of how it affects the polarised light.

Peptide bonds absorb polarised light below 240 nm, aromatic amino acid side chains absorb between 260 nm and 320 nm and disulphide bonds absorb weakly around 260 nm.

All CD experiments were performed using a Jasco J-715 spectropolarimeter.

### **2.11.1 Investigating Secondary Structure with CD**

Different secondary structures absorb polarised light differently.  $\alpha$ -helices have 2 minima when a CD spectrum is taken between 190 and 250 nm.  $\beta$ -sheets produce different spectra depending on whether they are parallel or anti-parallel. This means the signals from  $\beta$ -sheet proteins are smaller. Random coil and unstructured proteins do not have a real minimum in the spectra produced between 190 and 250 nm.

When performing CD spectroscopy the protein must first be exchanged into a low salt buffer to reduce noise. The recipe for the measuring buffer can be found in appendix 3.

To measure the secondary structure of the protein, a spectrum is taken between 180 and 250 nm, using a two-piece cuvette with a 1 mm path-length. It holds 60  $\mu$ l of sample. This spectrum is measured at 5°C. This data is then analysed by fitting it against the curves produced by  $\alpha$ -helical,  $\beta$ -sheet and random coil proteins using deconvolution

methods. This produces a percentage value for each kind of secondary structure found in the protein.

### **2.11.2 Measuring Protein Stability with CD**

As structured proteins produce CD spectra with visible minimums, it is possible to follow the unfolding of proteins using CD spectroscopy. To follow denaturation by temperature, the wavelength being scanned is fixed at the value of the minimum for that protein and then the ellipticity reading at this minimum point is followed as the temperature increases. In a folded protein, the reading remains the same until the protein denatures, at which point there is a sharp increase in the value of the reading at this wavelength.

The same experiment can be performed with increasing amounts of a chemical denaturant instead of heat. In this project, urea was used as the chemical denaturant.

These experiments give valuable information about the stability of a protein or a domain.

To measure the melting curve of the protein, a temperature course was measured from 5°C to 85°C. At each degree a reading was taken at a set wavelength, which depends on the structure of the protein. For ABD2, the wavelength used was 218 nm.

These readings were plotted, as a percentage value in terms of total folding, to find the melting temperature of the protein. For this, either the two-piece cuvette or a one-piece cuvette with a 5 mm path-length can be used.

Another measure of protein stability is how resistant the protein is to chemical denaturation. Denaturation using urea was used to measure this. Protein samples of consistent concentration (30  $\mu$ M for ABD2) were measured in a series of increasing urea concentrations (0 – 8 M urea) to track when they denatured.

## **2.12 Fluorescence Spectroscopy**

Fluorescence spectroscopy of unlabelled proteins relies on the intrinsic fluorescence of the aromatic side chain amino acids in a protein, in particular that of the side chain of tryptophan.

Tryptophan side chains are excited by light at a wavelength of 280 nm, and they emit light between 300 and 350 nm, depending on the environment of the tryptophans in the protein.

A tryptophan that is buried in the hydrophobic core of a protein will have an emission wavelength 10-20 nm shorter than that of a tryptophan exposed on the surface of a protein. This makes it possible to follow denaturation of the protein by following the change in emission wavelength as the protein denatures.

To perform fluorescence spectrometry a 3-5  $\mu$ M sample in 1 ml of buffer is used in a 1 ml quartz cuvette, with a pathlength of 1 cm. The machine used is a LS50B Luminescence Spectrometer (Perkin Elmer). The excitation wavelength is set at 280 nm in order to excite tryptophans, the emissions are scanned from 300 - 450 nm at 150 nm / min and the size of the excitation/emission slit is set at 4.5 nm. Following this, 10

accumulation scans are collected. A spectrum of the buffer alone and the protein is collected. The fluorescence due to the protein is obtained by subtraction of the buffer value from the fluorescence of the protein sample. For information on the chemical denaturation of the protein by urea, place the protein sample in a buffer containing 8M urea, and repeat the procedure.

## **2.13 Preparation of Actin from Skeletal Muscle Acetone Powder**

### **2.13.1 Preparation of Skeletal Muscle Acetone Powder**

Actin is extracted from rabbit skeletal muscle. After myosin was removed from the muscle, the actin remains in a pellet. The pellet was washed in 10 volumes of H<sub>2</sub>O twice, followed by one wash in 4 volumes of 0.4% NaHCO<sub>3</sub>. For this wash it was incubated at room temperature for 45 minutes with stirring. It was then washed twice in 2 volumes of 1 mM Tris-HCl at pH 8.5 at 4°C. It was then washed 4 times in 4 volumes of acetone, with 10 minutes of stirring each time. It was then dried for a minimum of 60 minutes or overnight in the fume cupboard and stored at -20°C.

### **2.13.2 Extracting actin from Skeletal Muscle Acetone Powder**

50 ml of G-Actin buffer (for recipe see appendix 3) was added to 2.5g of skeletal muscle acetone powder and stirred slowly for 20-30 minutes. The mixture was centrifuged at 48000 x g (20000 rpm) for 20 minutes. The supernatant was filtered through glass wool. The pellet was resuspended in G-actin buffer and the action repeated. The two supernatants were pooled together.

The G-actin was polymerised with 0.1 volumes of 10 x KME buffer (for recipe see appendix 3) and incubated in a water bath at 30°C for 45 minutes.

The solution was brought to a concentration of 0.8M KCl and centrifuged at 42000 rpm for 2 hours at 4°C. The pellets were washed with H<sub>2</sub>O and then dissolved in G-actin buffer and dialysed against the G-actin buffer for 48 hours.

The solution was repolymerised and centrifuged at 42000 x g (20000 rpm) for 2 hrs at 4°C. The supernatant was filtered with glass wool, and then loaded onto an S300 filtration column pre-equilibrated with G-actin buffer. Fractions were collected every 4 minutes, while monitoring their absorbance at 290 nm. The fractions that produced a peak in absorbance at 290 nm were pooled together and polymerised using 0.1V of 10x KME buffer, and then incubated at 30°C for 45 minutes. The mixture was then centrifuge at 42000 rpm using a T1250 rotor for 2 hrs at 4°C. The pellets were resuspended in 5 ml of ATPase buffer and stored at 4°C.

### **2.13.3 Labelling Actin with Pyrene Iodoacetamide**

Actin was purified as detailed above until the end of the first 42000 rpm spin step. At this point, the pellet was washed and then soaked in 6.25 ml of pyrene labelling buffer (for recipe see appendix 3). Pre-equilibrate the S300 column as above.

A solution of 30 mM pyrene iodoactamide at 7 mol pyrene/actin was slowly added to the resuspended actin pellet, and left stirring at room temperature for 3 hours. The solution was centrifuged at 12000 rpm for 10 minutes at 4°C in a Sorval SS34

centrifuge rotor. The pellet was soaked in 10 ml of G-actin buffer and dialysed against G-actin buffer for 48 hours and purified as described for unlabelled actin (section 2.13.2).

### **2.14 Actin Binding Co-Sedimentation Assay**

The actin pellet was resuspended in actin binding buffer (for recipe, see appendix 3).

50  $\mu$ M aliquots of actin were taken. 50  $\mu$ M of the positive control protein was added to one, the negative control was an aliquot without any protein added, and then the samples were 50  $\mu$ M of the protein(s) of interest added to the separate aliquots. Separately, a 50  $\mu$ M positive control sample and 50  $\mu$ M sample(s) of the protein(s) of interest were made and not mixed with actin. Calponin was used as the positive control in these experiments.

The mixtures were incubated in Beckman 7 x 20 mm polycarbonate centrifuge tubes at room temperature for 10 minutes and then centrifuged for 30 minutes at 85000 rpm using a TL-100 rotor (Beckman) in a Beckman Optima Max E ultracentrifuge.

The supernatants were put into micro-centrifuge tubes and lyophilised. They were then resuspended in 50  $\mu$ l of SDS loading buffer.

The pellets were washed with actin binding assay buffer and then resuspended in 1 ml of SDS loading buffer. Samples of both the pellets and the supernatants were loaded on an SDS PAGE gel.

## **2.15 Pyrene-labelled Actin Binding Fluorescence Assay**

An emission spectrum of a 6  $\mu\text{M}$  sample of pyrene-labelled actin was taken at 25°C. The sample was excited at 343 nm and emission measured at 385 nm, with a slit width of 1 mm/1 mm.

The protein of interest was added to pyrene-labelled actin in increasing amounts. The amount of protein required and added depends on the binding co-efficient of the protein with actin, which can be roughly estimated from the co-sedimentation assay detailed above. Enough protein must be added to reach a plateau where the actin is saturated by the protein.

A titration of pyrene-actin against buffer was performed to provide a baseline to measure changes against.

## **2.16 Actin binding by Isothermal Titration Calorimetry (ITC)**

Isothermal titration calorimetry (ITC) is a technique that can be used to determine the binding affinity of one macromolecule for another. In this project it was used to try to determine the  $K_D$  for actin binding by the domains of ms1/STARS.

ITC determines the binding affinity by measuring the amount of energy that has to be added to or removed from the system in order to maintain the baseline temperature when a ligand is added to a protein which binds to it. If the concentration of the ligand

and of the binding protein is known, the  $K_D$  can be calculated from the energy required to maintain the baseline temperature.

Before beginning ITC, all protein samples must be in exactly the same buffer so that any temperature changes are due to protein-ligand interaction, not due to difference in the buffers.

The programme used for collecting this ITC data was VPviewer 2000 ITC. Once the computer and the ITC machine (VP-ITC, MicroCalorimeter) were switched on, the experiment could be set up. Jacket temperature was set to 15°C, and the concentration of the protein in the syringe (the ligand) and in the cell (the binding protein) were entered as inputs into the programme. The cell should contain 1.8 ml of the protein, and the glass tube feeding the syringe should contain 600 µl of the ligand. For the ABD2/actin system, actin was placed in the syringe as the ligand and ABD2 was placed in the cell. Once set up, the system automatically adds the ligand in stages and stirs the mixture. The amount of energy required by the system to return to equilibrium after each injection of ligand is measured to determine the binding energy of the ligand to the binding protein. The dissociation constant ( $K_D$ ) of the binding reaction can be calculated from this data.

### **2.17 Investigating the Multimeric State of Domains**

This was carried out using analytical ultra-centrifugation (AUC), analytical gel filtration and NMR spectroscopy.

### **2.17.1 Analytical Gel Filtration**

200  $\mu\text{l}$  of 100  $\mu\text{M}$  protein samples were run on a Superdex<sup>(TM)</sup> 75 column (GE Healthcare) prepared with 2 column volumes of gel filtration buffer. Gel filtration buffer was run through the column at 1 ml/min, fractions were collected every 1 ml, and UV spectrum was recorded at 280 nm. Peaks in this spectrum were compared to the peaks produced when proteins of a known molecular weight were run down the column. These spectra can then be compared, by plotting a graph of log MW against log (elution volume/void volume). The log (elution volume/void volume) of the protein of interest can be plotted on this graph of the standards and the MW that produces that elution volume can be calculated. This is compared to the expected size of the protein in monomeric, dimeric or multimeric states and the state of the protein is determined.

### **2.17.2 Analytical Ultra-Centrifugation**

Samples of 50  $\mu\text{M}$  and 100  $\mu\text{M}$  were provided for analytical ultra-centrifugation (AUC) which was performed by Professor David Scott at the University of Nottingham.

### **2.17.3 Analysing the Multimeric State of ABD2 by NMR**

After assignment of HSQC peaks, further experiments can be carried out to determine the correlation time ( $\tau_c$ ) of the protein. This is related to the size and shape of the protein. It can be determined by performing an experiment to measure T1 and T2 relaxation, and a heteronuclear NOE experiment. The experiment was performed on a 200  $\mu\text{M}$   $^{15}\text{N}$ -labelled sample of ABD2, pH 7.2, on an 800 MHz Bruker spectrometer with a cryoprobe.

## **2.18 Nuclear Magnetic Resonance Spectroscopy**

Nuclear magnetic resonance spectroscopy (NMR) uses the properties of certain nuclei,  $^1\text{H}$ ,  $^{15}\text{N}$ ,  $^{13}\text{C}$  and  $^{32}\text{F}$ , to explore the structure of proteins.

These nuclei have spins that can be one of 2 orientations, meaning that they can be aligned to a magnetic field and their decay from this alignment can be measured and be used to provide information about the structure of the protein they are in. Different NMR experiments provide different information about the structure of the protein.

All the following experiments were performed on Bruker NMR Spectrometers using Bruker's Topspin software. The protein samples were exchanged into measuring buffer (see appendix 3) before all experiments. Table 2.2 contains a summary of the conditions (concentration and pH) and the NMR spectrometers used to perform the 1D NMR experiments performed to determine the folded state of the putative domains produced by DomainX, of ABD2, of the ABD2 mutants and of the expressed domains of ABLIM2. Table 2.3 contains a summary of the conditions of the samples used for residue identification and structure determination for ABD2, and on which NMR spectrometer the experiments were performed. Table 2.4 contains a summary of the conditions of the samples used for the binding experiments performed by NMR and what spectrometers were used to do this.

Table 2.2:

Sample	Concentration ( $\mu\text{M}$ )	pH	Field Strength of Spectrometer (MHz)	Cryoprobe
MSD1	100	7.2	600	Yes
MSD2	100	7.2	600	No
ABD1	100	7.2	600	Yes
ABD2	100	7.2	600	No
LIM2	100	7.2	600	Yes
LIM3	100	7.2	600	Yes
LIM4	100	7.2	600	Yes
VHD	100	7.2	800	Yes
ABD2 R302A	50	7.2	500	No
ABD2 R318A	50	7.2	500	No
ABD2 R347A	50	7.2	800	Yes
ABD2 R349A	50	7.2	800	Yes
ABD2 K350A	50	7.2	500	No
ABD2 H351A	50	7.2	500	No

Table 2.3:

Experiment	Concentration ( $\mu\text{M}$ )	pH	Field Strength of Spectrometer (MHz)	Cryoprobe
HSQC	250	7.2	600	No
$^{15}\text{N}$ NOESY	500	7.2	600	Yes
HNCA	300	7.2	600	No
HNCACB	300	7.2	800	Yes
HN(CO)CA	188	7.2	600	Yes
HN(CO)CACB	300	7.2	600	Yes
CT-HSQC	300	7.2	600	No
HCCH-TOCSY	300	7.2	600	Yes
$^{13}\text{C}$ NOESY	200	7.2	800	Yes
Aromatic TROSY	250	7.2	600	No
Homonuclear NOESY	200	7.2	800	Yes
T1 relaxation experiment	200	7.2	800	Yes
T2 relaxation experiment	200	7.2	800	Yes
Heteronuclear NOE experiment (HSQCNOE)	200	7.2	800	Yes
IPAP	120	7.2	800	Yes

Table 2.4:

Experiment	Protein	Ligand	pH	Spectrometer	Cryoprobe
Actin binding	ABD1 (50 $\mu$ M)	-	7.29	600	Yes
	ABD1 (50 $\mu$ M)	Actin (10 $\mu$ M)	7.29	600	Yes
	ABD1 (50 $\mu$ M)	Actin (20 $\mu$ M)	7.2	600	Yes
	ABD1 (50 $\mu$ M)	Actin (30 $\mu$ M)	7.25	600	Yes
ABLIM binding	ABD2 (25 $\mu$ M)	-	7.22	600	Yes
	ABD2 (25 $\mu$ M)	LIM2 (75 $\mu$ M)	7.22	600	Yes
	ABD2 (25 $\mu$ M)	LIM3 (75 $\mu$ M)	7.22	600	Yes
	ABD2 (25 $\mu$ M)	LIM4 (75 $\mu$ M)	7.22	600	Yes
	Tandem of ABD1 and ABD2 (25 $\mu$ M)	-	7.2	600	Yes
	Tandem of ABD1 and ABD2 (25 $\mu$ M)	LIM 1 and 2 (50 $\mu$ M)	7.2	600	Yes
	Tandem of ABD1 and ABD2 (25 $\mu$ M)	LIM 2 and 3 (50 $\mu$ M)	7.1	600	Yes
	Tandem of ABD1 and ABD2 (25 $\mu$ M)	LIM 3 and 4 (50 $\mu$ M)	7.16	600	Yes
	VHD (20 $\mu$ M)	-	7	800	Yes
	VHD (20 $\mu$ M)	Tandem of ABD1 and ABD2 (40 $\mu$ M)	6.95	800	Yes
DNA binding	Tandem of ABD1 and ABD2 (50 $\mu$ M)	-	7.2	800	Yes
	Tandem of ABD1 and ABD2 (50 $\mu$ M)	DNA (50 $\mu$ M)	7.1	800	Yes
	Tandem of ABD1 and ABD2 (50 $\mu$ M)	DNA (100 $\mu$ M)	7.22	800	Yes
	Tandem of ABD1 and ABD2 (50 $\mu$ M)	DNA (200 $\mu$ M)	7.22	800	Yes

The temperature was set at 298 K for all the experiments.

In each experiment the sample was locked against the D<sub>2</sub>O signal. As the experiment is running, a second experiment is run, which measures the signal from the D<sub>2</sub>O in the sample. This is the lock signal. If the signal changes it means the magnet has drifted.

Magnetic drift could lead to errors in the magnetic pulses used in the NMR experiments. Therefore the magnet is corrected so the value of the D<sub>2</sub>O signal returns to its original value. This means errors due to magnetic drift should not occur.

Following this, the samples were tuned and matched, so that the best performance could be obtained from the electronics of the NMR machine. The magnets were then shimmed to remove any localised areas of difference in magnetisation as the applied magnetic field must be homogeneous during NMR experiments.

The 90° pulse must be optimised so that water suppression can be performed. Water suppression must be used unless the sample is 100% deuterated as the signal from water will be a lot stronger than the signal from the protein because the concentration of water in the sample will be higher than the concentration of the protein in the sample. The length of the 90° pulse is measured in  $\mu\text{s}$ . To calculate the 90° pulse value, the 360° pulse is calculated. This value is 4 x the 90° pulse. It is easier to optimise the 90° pulse this way because smaller adjustments can be made so the final value of the pulse is more exact. The value of the pulse is increased or decreased in order to reduce the water peak. Once the water peak is at its smallest, the 360° pulse value is divided by 4 to give the optimised value of the 90° pulse that will be used.

Once the optimal 90° pulse has been determined, the NMR experiment as a whole can be optimised for the protein being studied. The written pulse programme must be manually checked for any problems. The parameters to be checked include the length and power of the water suppression pulse, the recovery time, plus the value of the phase correction pulse being used for this experiment. The experiment-specific factors to

check can be found in the pulse programme for each experiment. Once these parameters are optimised and correct, the experiment can be run.

Once the spectrum has been acquired, it needs to be processed. First, it must be Fourier transformed to convert the free induction decay signals into a spectrum. Then the peaks must be phased and baseline adjusted.

### **2.18.1 1D NMR**

In order to find the structure of a protein, first a 1D spectrum must be obtained to show if the protein of interest is folded.

### **2.18.2 Residue Identification**

When attempting to determine the structure of a protein, there are two stages. The first is sequence-specific residue identification and assignment, followed by use of the restraints produced by these identified residues to calculate the structure that best fits these restraints.

A 2D Heteronuclear Single Quantum Coherence (HSQC) (Bodenhausen & Ruben, 1980) spectrum is taken to act as a reference for all the other spectra. In a HSQC spectrum, the amide group of each residue (except for prolines which have imino groups instead of amide groups due to their ring structure) produces a peak. In an HSQC, the magnetisation is transferred from the NH proton of the residue to the proton attached to the backbone C of the residue and back to the NH proton to be measured.

HSQCs were also used to follow protein-protein and protein-DNA interaction as described in section 2.22.1.

Once the HSQC has been produced, the residue types must be identified. This is done by combining the data from HNCA (Ikura *et al.*, 1990) and HNCACB (Wittekind & Mueller, 1993) experiments. The characteristic chemical shifts for the alpha and beta carbons of each residue type allow the identification of each residue in the sample by residue type. This is not always possible for all residues in a protein, and the residue must be identified by finding its place in the sequence.

This can be done by using the HNCA and the HNCOCA (Ikura *et al.*, 1990) experiments. These experiments enable the identification of both the inter-residue NH to C $\alpha$  bond and also the intra-residue NH to C $\alpha$  bond to the previous residue. Magnetism in HNCA experiments travels from the H of the NH group of the amino acid through to the internal  $\alpha$ -carbon and then back for detection (figure 2.5). This provides the intra-residue C $\alpha$  chemical shift. The HNCACB experiment follows the same path but continues on to the beta carbon (figure 2.5). The HNCOCA (figure 2.5) experiment goes from the HN of the residue, through the carbon of the CO group of the preceding residue to its  $\alpha$ -carbon. A feature of the HNCA experiment causes half the signal strength to follow this path also; therefore, that experiment shows both the inter- and intra-residue HN to C $\alpha$  chemical shift. Peaks that appear in both spectra must be the  $\alpha$ -carbon of the I-1 residue. This method also enables identification of those residues whose  $\alpha$  and  $\beta$  carbon shifts were not specific enough to identify them. To help with this, HNCOCACB experiments (Grzesiek, 1992) can also be performed. These follow

the same path as the HNCOCA experiments but continue past the alpha carbon to the beta carbon (figure 2.5).

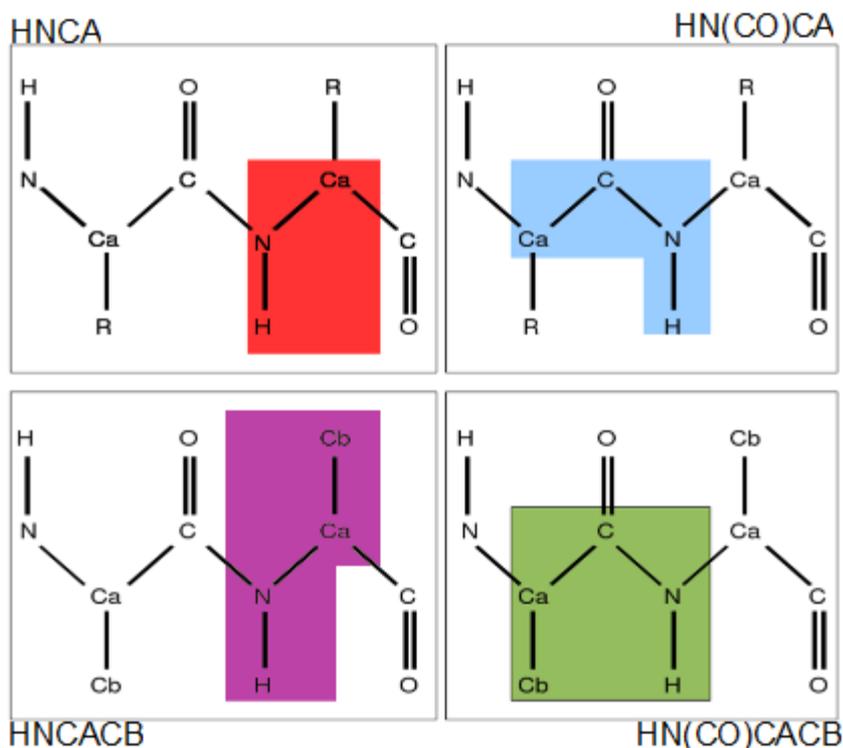


Figure 2.5 – Diagrams of the path of magnetisation in the NMR experiments used to identify residues.

### **2.18.3 Secondary Structure Determination**

<sup>15</sup>Nitrogen-edited NOESY spectra identify protons that are within 5 Ångstroms of the NH group of a residue. HBHACONH spectra allow identification of the hydrogens attached to the alpha and beta carbons and this can be superimposed onto the <sup>15</sup>N-NOESY to allow identification of the side chain hydrogens on the NOESY. Cross-peaks between residues allow identification of the secondary structure of the protein because  $\alpha$ -helices and  $\beta$ -sheets have different patterns of cross-peaks.  $\alpha$ -helices have cross-peaks between the residue I and residues up to -4 or +4 away from it, whereas residues in  $\beta$ -sheets have cross-peaks to residues much further away.

### 2.18.4 Tertiary Structure Information

In order to obtain a 3D structure for a protein, a NOESY spectrum must also be obtained for the side chains of the residues. This 13-carbon-edited NOESY will only be of use if we are first able to identify the side chain carbons. To do this, two other experiments must first be performed. These are a 13-C-edited HSQC (Medvedeva *et al.*, 1993), which shows all the protons linked to carbon atoms in the side chain of the amino acids, as long as the carbon is linked to another carbon. This enables the identification of the type of side chain carbon, with  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$  and  $\epsilon$  carbons all being in distinct regions of the spectrum. These results are combined with the results of an HCCH TOCSY. This second experiment has a transfer of magnetisation from a carbon-linked proton to others within 3 JHN. The magnetisation pathway can be seen in figure 2.6.

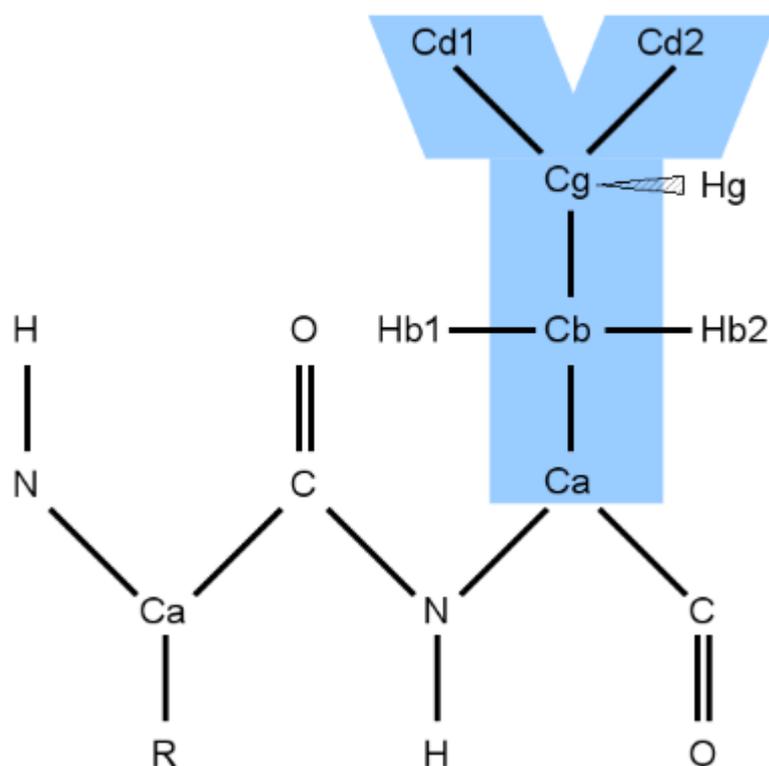


Figure 2.6 – An example of the transfer of magnetisation in an HCCH TOCSY experiment. The residue used in the example is a leucine. The magnetisation can travel from Ca to either Cd1 or Cd2.

By combining the results of these experiments, the side chains can be identified, which means that when a  $^{13}\text{C}$  NOESY-HSQC is performed, data can be gathered on which protons are within 5 Ångstroms of carbon-linked protons.

This provides data for all of the carbons in side chains except for those in the aromatic rings of the aromatic amino acids. For these side chains, different experiments must be used. The experiments used were a TROSY experiment (Meissner & Sorensen, 1999), combined with a homonuclear NOESY. This enabled identification of the carbon-linked protons in the aromatic side chains and the finding of residues within 5 Ångstroms of them.

CCPN Analysis (Fogh *et al.*, 2002) was used to analyse the data, pick the peaks and to co-ordinate and collect the information about the peaks.

### **2.18.5 Structure Calculation Using CYANA**

The data produced by the NMR experiments can be combined and interpreted to produce a tertiary structure for a protein or, in this case a domain, using software such as CYANA (Herrmann *et al.*, 2002). CYANA is an automated method for performing this calculation. Using a computer programme to calculate the tertiary structure of proteins has many advantages, including faster structure calculation and lack of user bias. CYANA calculates the structure based on dihedral angle constraints and distance geometry. These are based, in the case of dihedral angle constraints, on the sequence of the amino acids in protein and the possible position of their backbone given the amount of twisting possible, the distance they can be apart and the possibility of steric

hindrance. In the case of distance geometry, this is a function of the NOE peaks collected for the protein. NOE peaks give a maximum distance apart that two residues, or parts of that residue (in the case of side chains), can be, because if they were further away than 5 Å, the peaks would not be seen. A structure for a protein must not break any of these restraints. Due to the overlap of parts of the 13-C NOESY-HSQC, it is easy to misidentify the peaks that have been picked. CYANA avoids this problem by removing assignments before it calculates the structure, calculating the structure that best fits the data and then assigning the peaks.

### **2.18.5.1 CYANA Inputs**

CYANA requires the following inputs to calculate a structure:

- Backbone torsion angle restraints which are derived from TALOS (Cornilescu *et al.*, 1999). TALOS compares the sequence of the protein of interest with twenty proteins of known structure in its database. It examines residue  $x$  in the context of the string of residues it is in,  $x-1$ ,  $x$  and  $x+1$ . It then compares the backbone torsion angles,  $\phi$  and  $\psi$ , of these triads with triads of the same amino acid composition from the crystal structures of the twenty proteins of known structure. The results of the comparison are presented with each residue being colour-coded as either green for good, red for bad or yellow for ambiguous. The result for each residue is compared to the average of 10 strings of the same amino acid combinations in the database. Positive results are those where 9 out of 10 values are in the same region of the Ramachandran plot as those of the residues in the proteins of known structure to which it is being compared and

where the central residue does not have a positive value of  $\phi$ . Ambiguous results can be manually checked.

- The protein sequence containing the sequence number and residue type in a .seq format.
- The assignments of the residues in the sequence in .prot format, both of which are exported from CCPN Analysis (Fogh *et al.*, 2002).
- For structure determination without previously assigned NOE distances, NOE peak lists from the carbon NOESY, nitrogen NOESY and homonuclear NOESY were also imported from CCPN Analysis.

CYANA also requires two scripts in order to run. They are a CALC.cya and init.cya. The CALC.cya script contains details of the peak lists to be used in the calculation, the assignment file and dihedral angle constraints file to be used, as well as the tolerances to be used in the calculation, the number of structures to be calculated in each cycle and the number of those calculated structures which are to be used as the basis for the next cycle, the number of steps in the simulated cooling, the residue range that the residual mean standard deviation (RMSD) will be calculated for and the random seed used to start the process.

The init.cya file contains the details of which file the log will be exported as, the name of the protein, the residue range that the r.m.s.d. will be calculated for, the CYANA libraries to use, cut off points for the restraints and which sequence file to use.

### **2.18.5.2 CYANA Structure Calculation**

These inputs are used with no assignment of the peaks in the NOESY spectra. This is done because many of the peaks have lots of possible assignments. This makes it hard to gain a definite distance constraint for that peak. CYANA uses CANDID (Herrmann *et al.*, 2002) to overcome this problem.

Each CANDID cycle begins by generating an assignment list of those hydrogen atoms that could contribute to a peak. For each cross peak, these initial assignments are weighted using filtering criteria, including agreement between values of chemical shifts and peak position, self-consistency within the entire network of NOEs (called network anchoring) and, when available, compatibility with the 3D structure calculated in the preceding cycle. There is further discussion of network anchoring below.

In the first cycle, this preceding 3D structure is not available, and network anchoring is the most heavily weighted factor. Initial assignments with a low overall score are discarded.

The retained assignments are interpreted as upper distance limits derived from the cross peak volume for each cross peak. Upper distance constraints can be obtained from a single assigned cross peak or can be ambiguous distance constraints if more than one assignment is retained by a cross peak. The addition of this ambiguity is what made CANDID revolutionary, as before; only unambiguous assignments could be used. As the majority of NOEs cannot be unambiguously assigned from chemical shifts alone, this was a severe limitation of previous automated NOE assignment programmes. This process is also discussed in greater detail below.

If ambiguous distance constraints are included, each NOESY cross peak with  $n > 1$  possible assignments can be seen as the superimposition of  $n$  degenerate signals.

This assignment can be aided by network anchoring and constraint combination.

### **Network Anchoring:**

A network of distance constraints big enough to be used to determine a protein 3D structure will also make a self-consistent subset of correctly assigned constraints. Network anchoring evaluates the consistency of NOE assignments independently from anything known about the 3D structure of the protein of interest and can therefore be used in the first cycle of NOE assignment when a 3D structure is not available to check the strength of the assignments. As each NOE assignment must be contained within a network of all other assignments, network anchoring is a powerful and useful approach for detecting constraints that have no reinforcement from the consistent NOEs around it.

Network anchoring for two atoms  $\alpha$  and  $\beta$  consists of looking for all atoms  $\gamma$  present in either the same or neighbouring residues which are connected simultaneously to  $\alpha$  and  $\beta$ . This connection can be an assignment or that the structure suggests the corresponding distance between them must be short enough to give an observable NOE. All peaks from the peak list contribute simultaneously to the network anchored assignment.

### **Constraint Combination:**

Artefacts and misinterpretation of noise signals can lead to mistakes in structure calculations. This is even more likely for the first cycles of NOE assignment because filtering of constraints based on the 3D structure is not possible at that stage. Constraint combination is applied during the first two cycles of CANDID to reduce the impact of the experimental errors in the spectra on the structure produced, at the cost of temporarily reducing the amount of information used.

Constraint combination consists of generating distance constraints with the combined assignments from different, unrelated cross peaks. An ambiguous constraint is one which will fit into a correct structure as long as at least one of its possible assignments is correct, even if there are additional assignments associated with that constraint that are incorrect. This means that the combined constraints have a lower probability of being wrong than the corresponding original constraints, provided that fewer than half of the original constraints that take part in the combination are wrong.

CANDID has two methods of constraint combination:

- 2→1 in which assignments A and B are replaced by a single ambiguous constraint with assignment AUB.
- 4→4 is a pairwise combination that replaces four assignments A, B, C and D with four distance constraints AUB, AUC, AUD and BUC.

In cases of peaks that have been wrongly assigned, the effect on the global fold is more pronounced with errors in long range NOE assignment, because they involve larger

areas of the protein. Therefore, constraint combination is applied only on long-range NOEs, not on short- or medium-range NOEs.

For 2→1 constraint combination, the number of long-range constraints is halved, whereas in the 4→4 method, more of the original information is conserved. The latter method also takes into account that some peaks have assignments that are more reliable than others, as the four peaks A, B, C and D are used 3, 2, 2 and 1 times respectively to form the final combined constraints.

An example of the advantages of this method can be seen below in figure 2.7. In figure 2.7, B shows what occurs if an NOE peak is wrongly assigned and constraint combination is not applied. A peak has been wrongly assigned as being between C and D, meaning that the calculation must take it into account. This assignment error means that the structure calculated is not the real tertiary structure of the protein. In C, the same peak has been mis-assigned, but constraint combination is used in the calculation. This means that the value the programme uses combines the distance constraints produced by all four residues. Use of combined constraints means that the calculated structure is not forced into the wrong shape by one wrongly assigned peak.

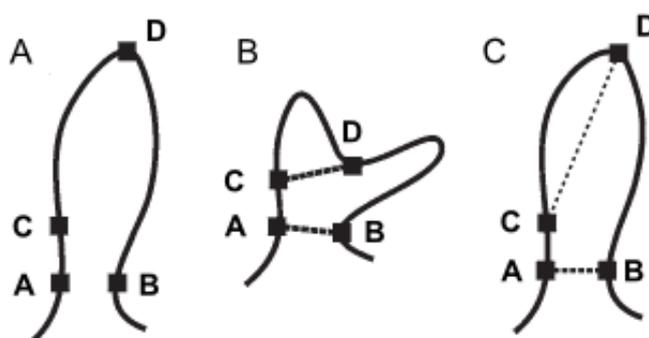


Figure 2.7 – Figure showing the advantages of constraint combination. A is the correct structure. B shows what happens if a peak is wrongly assigned as being from C to D when no constraint combination is applied. C has the same wrongly assigned peak, but because constraint combination has been used this does not affect the calculated structure. Figure taken from Güntert, 2004.

The .prot file is used to provide possible assignments for these peaks. The distance constraints produced by the process are then used to provide 100 structures, which are created using the minimisation of a target function that measures the agreement between a structure and its constraints. The target function for CYANA (Güntert *et al.*, 1997) is such that it is zero if all experimental distances and torsion angle constraints are fulfilled and all non-bounded atom pairs satisfy a check for the steric overlap; and a conformation that satisfies more of these criteria will have a lower value. In torsion angle dynamics the only degree of freedom is the rotation about the covalent bonds, so that the conformation of the protein is defined by all the values of all the torsion angles. The twenty lowest energy structures from this hundred are then taken to be used in the next round of energy minimisation.

The minimisation algorithm used is based on simulated annealing (Kirkpatrick *et al.*, 1983) by molecular dynamic simulation in torsion angle space. This is a combinatorial optimisation method. The concept is taken from the crystallisation procedures applied to metals, where a metal is melted at high temperature and then slowly cooled down in order to minimise the structural defects.

The standard simulated annealing protocol in CYANA starts from a conformation with all torsion angles treated as independent and it consists of five steps. The first is an initial minimisation step, including only those constraints between atoms up to three residues apart in the sequence, and it is performed to reduce high energy interactions that could easily disrupt the torsion angle dynamics algorithm. It is then followed by a high temperature phase, and then a torsion angle dynamics calculation carried out at constant high temperature to avoid local energy minima. The third step is a slow

cooling to close to zero temperature. Up until this step, the hydrogen atoms have not been included in the calculation. In this step they are included as individual hydrogen atoms. The final minimisation step is then carried out, which has 1000 conjugate gradient steps.

Torsion angle dynamics consists of molecular dynamic simulation using torsion angles instead of Cartesian coordinates as degrees of freedom. Compared to the minimisation of the target function, the molecular dynamics simulation also includes the kinetic energy that allows the calculation to overcome barriers of potential that could otherwise lead to the function becoming stuck in a local energy minimum.

This process is repeated for seven cycles.

### **2.18.5.3 CYANA Outputs**

After each round of structure determination and energy minimisation, the following are produced:

- An upper limits file, with the furthest distance apart for the protons of the amino acids in the protein, as well as the residues I-1 and I+1.
- An overview file, which provides the target function, the RMSD, and maximum for the upper limits, the total and maximum for Van der Waals energy and the torsion angle RMSD. It also provides a minimum, a maximum and an average for all of these features for all 20 structures. It lists constraints that are violated

in 6 or more structures. It provides the value of the RMSD over a preset range of residues. For ABD2 this range was residues 295-375. This information gives the average difference from the mean for both backbone and heavy atoms.

- A PDB file containing the twenty lowest energy structures.
- An NOA file showing the assignments given to the NOESY peaks in the various NOESY spectra by CYANA. It also states the number of selected peaks, which spectrum they are in, the number of assigned and unassigned peaks, those with diagonal assignments, and the number of cross peaks and their varieties.

These numbers enable a determination of the quality of the structure produced by CYANA. In 'Automated NMR Structure Calculation With CYANA' (Güntert, 2004), Güntert lists five quality control checks. In section 4.1.9.4, the values for these checks are shown for the structure determined by CYANA for ABD2.

#### **2.18.5.4 CYANA Refinement and Structure Validation**

The overview file produced at the end of each cycle can be read to see which constraints are violated in 6 or more structures. If a constraint is violated this often, it suggests that there is a problem with it. If the violated constraint is a dihedral angle restraint, it can be loosened, and if the violated constraint is a peak, then it is possible to return to the spectrum to see if the peak has been correctly located or if there is something wrong with the peak, for instance that it is an artefact rather than a real peak.

After a structure has been determined, it must be evaluated to understand how reliable the result is. CYANA implements five criteria (Güntert, 2003) that allow evaluation of the structures. These are:

1. The average target function for cycle 1 is below 250Å
2. The average target function for cycle 7 is below 10Å
3. There are less than 20% unassigned NOEs.
4. RMSD for cycle 1 is below 3Å
5. RMSD difference between cycles 1 and 7 is below 3Å

Other structure validation software includes PROCHECK (Laskowski *et al.*, 1993) which validates the geometric quality of the determined structure.

PROCHECK uses reference values of the geometric parameters being considered to judge if a structure is possible. Validation criteria include local geometry such as bond lengths and angles, chirality, tetrahedral geometry and side chain planarity, or the overall quality of the geometry, evaluated by the Ramachandran plot, rotameric states, backbone conformation and packing of the side chains in the hydrophobic core. Additional criteria include inter-atomic bonds, hydrogen bonds and electrostatic interactions.

In this evaluation, the most relevant of the local geometry criteria is held to be the  $\chi_1$  and  $\chi_2$  angles that define the rotamers. Side chain rotamers have a preferred, staggered conformation with dihedral angles of  $-60^\circ$ ,  $60^\circ$  and  $180^\circ$ ; this preference is due to the difference in energy between the eclipsed conformation, which has more steric interactions and therefore a higher energy, and the staggered conformation, which is

energetically favoured because it has fewer steric interactions. Both states can occur in a protein and it is possible for a residue to switch between the two. The relative population in each state varies for each amino acid type, secondary structure elements and environment. In PROCHECK, the  $\chi_1$  and  $\chi_2$  side chain rotamer distribution is considered for each amino acid in the protein sequence and is displayed graphically, with allowed and disallowed regions depicted.

Similar evaluations are performed for the backbone dihedral angles  $\phi$  and  $\psi$ . These are based on steric considerations (Ramachandran *et al.*, 1963) which state that the combinations of the two angles in a polypeptide chain are restricted to certain ranges, and that these can be visualised in a Ramachandran plot. Along with steric restrictions,  $\psi$  and  $\phi$  have different values depending on residue type and secondary structural elements. The Ramachandran plot is the most widely used method to evaluate the quality of a protein structure (Hooft *et al.*, 1997). The plot is divided into four regions: favoured, additionally allowed, generously allowed and disallowed. PROCHECK also used this method.

$\omega$ , the torsion angle about the peptide bond is also considered. As a result of its partially double-bonded nature, it is unable to rotate freely and maintains a position close to planarity. It is mostly found in the trans conformation ( $180^\circ$ ) but can also be found in the less favourite cis conformation ( $0^\circ$ ). Most of the cis peptide bonds, when seen, involve a proline, but rare cases without proline involvement are also known.

PROCHECK also considers non-bonded interactions such as inter-atomic bumps. These occur when the distance between the centres of two atoms is less than is

physically realistic. PROCHECK identifies atoms that are closer in space than 2.6 Å as bumping. This only includes non-hydrogen atoms and those pairs that do not form hydrogen bonds.

Packing of the side chains in the hydrophobic core is another element that can suggest the value of any structure calculated. Different amino acids have different preferences for neighbouring residues and these preferences affect the global fold of the protein. This is particularly true for aromatic and hydrophobic residues that should pack against each other in the hydrophobic core. If this is not present, it can be a sign of something wrong with the calculated structure.

## **2.18.6 Structure Refinement**

It was also desired that a secondary set of data could be used to provide structure refinement and validation in an analogous manner to the  $R_{\text{free}}$  value used in X-ray crystallography. The method used was residual dipolar coupling (RDC).

### **2.18.6.1 Residual Dipolar Coupling**

The residual dipolar coupling method was developed to work around the problems of extending solution state NMR to larger proteins. It is difficult to unambiguously identify NOEs in larger systems, and it is more difficult to measure them as the effect of transverse relaxation grows.

Residual dipolar coupling uses the couplings produced by either the weak alignment of the protein itself or its alignment if it is suspended in dilute liquid crystalline media

(Gronenborn, 2002). This data can be used to obtain information about the orientation of an inter-nuclear vector of the protein. These can be used to refine the overall structure of the protein, both in terms of the local secondary structure and the overall tertiary structure.

Different media can be used to produce this effect, including liquid crystalline media based on phospholipids (Tjandra & Bax, 1997), nematic phases of rod-shaped viruses and filamentous phages (Clare *et al.*, 1998 and Hansen *et al.*, 1998), the liquid crystal phases of surfactants (Rückert & Otting, 2000), purple membrane fragments (Koenig *et al.*, 1999 and Sass *et al.*, 1999) and strained gels (Tycko *et al.*, 2000 and Sass *et al.*, 2000).

It is important to have multiple possible media because all of the media may not work for all systems, for instance, a protein that binds to membranes cannot be studied using purple membrane fragments because the alignment is too strong. There are also advantages to studying the same system with multiple media as each medium may orient the protein differently with respect to the magnetic field, which means that it becomes possible to narrow down the alignment of the tensor even more.

Residual dipolar couplings can be used to refine structures because they add an extra constraint on the orientation of bonds. However, if they are held in reserve until after the initial structural calculation, they can also be used to validate the calculated structure, by comparing experimental RDCs to ones derived from the calculated structure, in a manner similar to that of  $R_{\text{free}}$  when used in X-ray crystallography

structure calculation. If the calculated structure is accurate then the experimental and derived RDCs will agree.

The three methods used for this project were bicelles made from dimyristoyl phosphatidylcholine (DMPC) and dihexanoyl phosphatidyl phosphatidylcholine (DHPC), hexanol and PEG-5000 mixed to produce a medium and Pf1 filamentous phage.

#### **2.18.6.2 Preparation of Bicelles for RDCs**

DHPC and DMPC are combined at a molar ratio of 1:2.5 DHPC: DMPC.

This is done by first solubilising the DHPCs in bicelle buffer (see appendix 3) and mixing using a vortex machine. The mixture is then left rocking at 4°C for 1 hr.

The DMPCs are then added. This mixture is mixed by vortex and left rocking overnight at 4°C. More bicelle buffer is added to reach the final volume. This is then heated to 40°C for 2 minutes and the mixture is transferred directly on to ice for 2 minutes. This step can be repeated up to 6 times. The alignment of the media is then checked visually and in an NMR machine where an aligned medium will cause quadrupolar splitting of the D<sub>2</sub>O signal. If the medium is aligned, it can then be added to the protein to make up a final volume of 350 µl.

### **2.18.6.3 Preparation of Non-Ionic Liquid Crystalline Media for RDCs**

Of the various non-ionic liquid crystalline media outlined in “Alignment of Biological Macromolecules in Novel Non-ionic Liquid Crystalline Media for NMR experiments” (Rückert & Otting, 2000), the mixture which best suited the system being used was 5% w/v C8E5 and hexanol. Two times the amount needed was produced, with half being used to test the alignment of the medium and the other half to align the protein. This medium was added in a 1:1 ratio to the protein.

### **2.18.6.4 Preparation of Filamentous Phage for RDCs**

Filamentous phage was exchanged into working buffer (see appendix 3) and concentrated to 50 mg/ml in a 5000 MWCO Vivaspin column. Alignment of the medium was tested using 1D NMR, as the aligned medium will cause quadrupolar splitting of the D<sub>2</sub>O signal. If the medium is aligned, it can then be added to the protein to make up a final volume of 350 µl.

### **2.18.6.5 Use of Media to Produce RDCs**

Alignment of the media can be seen by monitoring the deuterium quadrupolar splitting of the deuterium lock signal, when the sample is prepared in an aqueous solution of 95% H<sub>2</sub>O, 5% D<sub>2</sub>O.

Once the media is aligned an IPAP experiment (Ottiger *et al.*, 1998) is performed. This experiment enables the calculation of the residual dipolar couplings of the aligned

protein by subtracting the values of an IPAP-HSQC with no <sup>15</sup>N refocusing period from the values of an IPAP-HSQC with a <sup>15</sup>N refocusing period.

These values can then be interpreted using MODULE (Dosset *et al.*, 2001). MODULE allows the back-calculation of the expected RDCs from a structure and if the calculated structure is correct these will agree with the experimental RDCs, giving a further validation tool for the structure.

### **2.18.7 Structure Comparison**

The structure, once it had been refined and validated, was compared to all other structures in the PDB using DALI (Holm *et al.*, 2008).

## **2.19 Protein Interaction Studies**

### **2.19.1 Following Binding Through Chemical Shift Changes in NMR Spectra**

To follow possible binding by NMR, first a reference HSQC spectrum must be taken. Following this, the putative interacting partner is added in increments until saturation is achieved. An HSQC spectrum is taken at each concentration. The spectra were then overlaid on the reference HSQC. Any changes were measured to show which residues were involved in binding.

### **2.19.2 Gel Shift Assay to Study DNA Binding**

Protein in measurement buffer (see appendix 3) plus 0.1% Triton-X100 (Fisher) is added to DNA in the following ratios – 1:1, 1:3, 1:10, 1:30, 1:100, along with a sample

with 0:1 ratio DNA: protein to act as a negative control. This is done with DNA at 1  $\mu\text{M}$  and 3  $\mu\text{M}$ . These solutions are made up to a final volume of 20  $\mu\text{l}$  using measurement buffer. The mixture is incubated for 10 minutes at room temperature. It is then loaded on a 0.7% gel made with  $\frac{1}{2}$  TB buffer (see appendix 3). The gel is run at a constant rate of 30 mA, with voltage varying. No stain is added to the gel until after it is run.

# **Chapter 3 – The Domains of ms1/STARS**

## **Chapter 3 - The Domains of ms1/STARS**

### **3.1 Protein Sequence Alignment of ms1/STARS**

The alignment of the amino acid sequence of ms1/STARS against the most similar proteins found by a Blast search (figure 3.1) shows that it is the C-terminal of ms1/STARS which contains the only area of sequence homology.

The C-terminal alignment can be seen in more detail in figure 3.2.

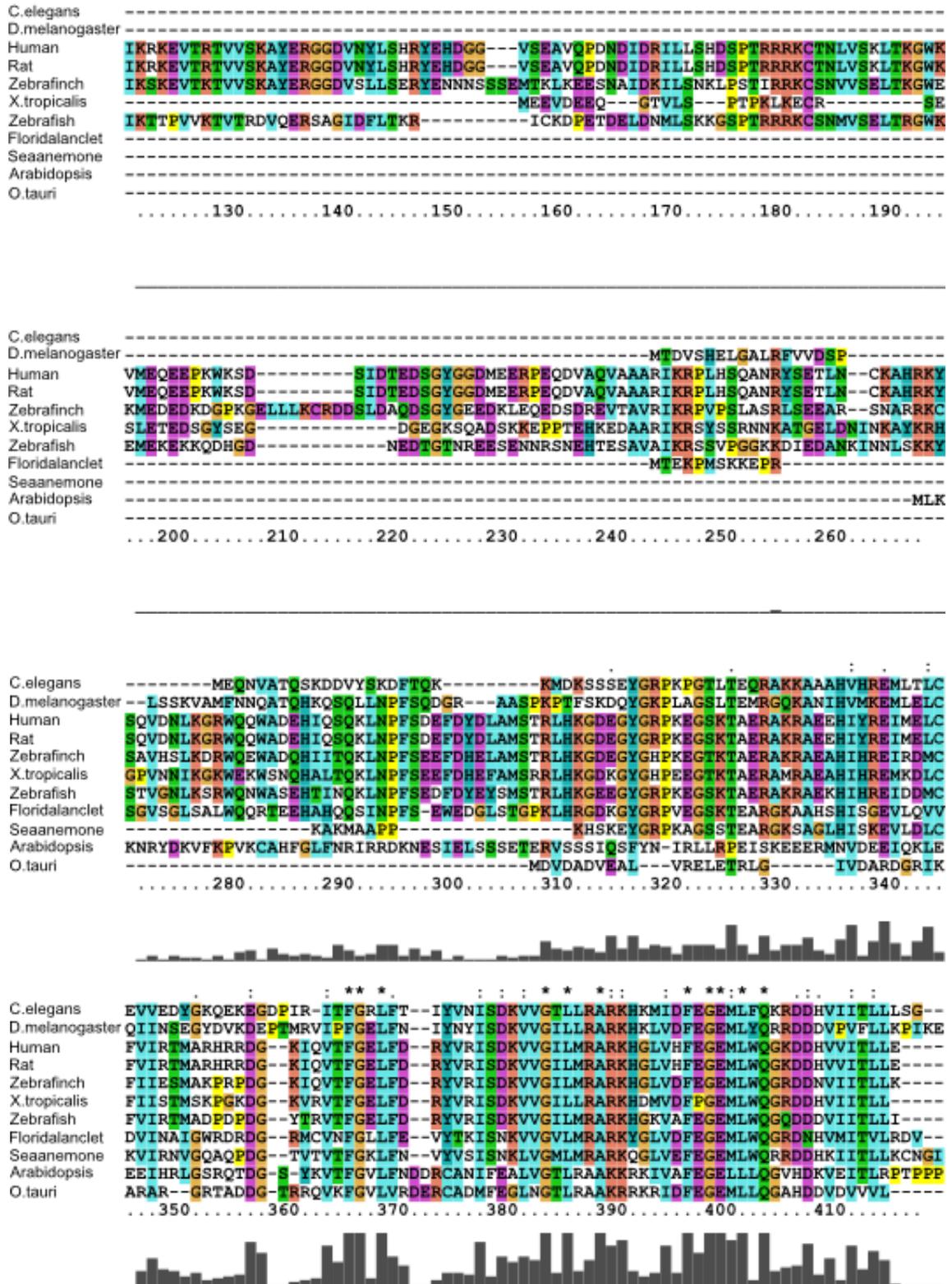


Figure 3.1 – An alignment of examples of proteins homologous to ms1/STARS. The C-terminal is the most conserved. The aligned proteins are C. elegans - NP001129772, D. melanogaster - NP569983, Human – NP631905, Rat NP787038, Zebrafinch XP002199130, Xenopus tropicalis NP001106590, Zebrafish CAK03875, Florida lancllet XP002605450, Sea anemone XP001619805, Arabidopsis CAA20577 and O. tauri CAL57253.

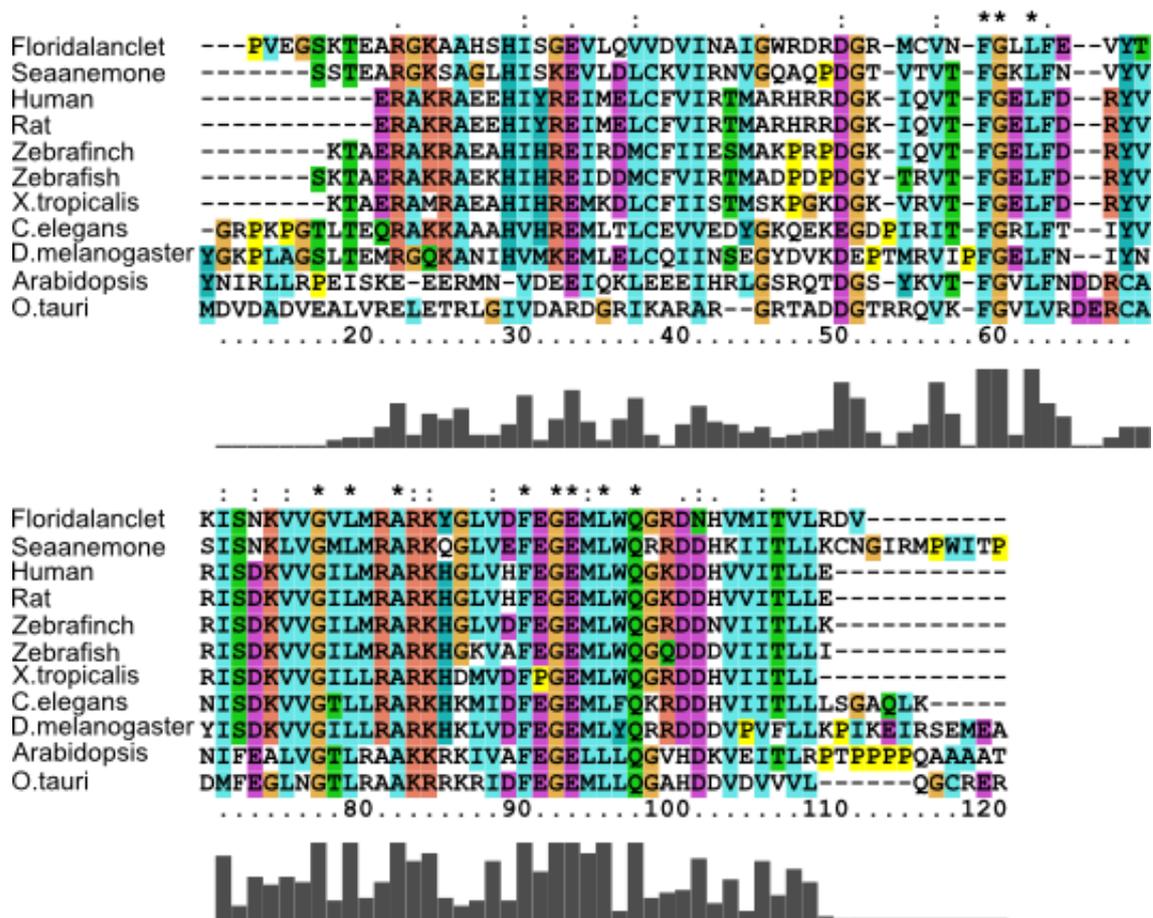


Figure 3.2 – An alignment of the C-terminal of proteins with sequence homology to ms1/STARS. The aligned proteins are *C. elegans* NP001129772, *D. melanogaster* NP569983, Human NP631905, Rat NP787038, Zebrafinch XP002199130, *Xenopus tropicalis* NP001106590, Zebrafish CAK03875, Florida lancllet XP002605450, Sea anemone XP001619805, Arabidopsis CAA20577 and *O. tauri* CAL57253.

As this was a conserved region, and as it is in the region that had been identified as containing an actin binding site (Arai *et al.*, 2002), it was an ideal candidate for attempts to produce a construct containing the second actin binding site of ms1/STARS.

RONN (Zhang *et al.*, 2005), a protein disorder prediction software, was used to predict the disorder of ms1/STARS (figure 3.3). Any region with a probability of disorder greater than 0.5 is considered to be disordered by RONN. The only part of ms1/STARS that is below 0.5, and thus ordered and folded, is the highly conserved C-terminal region of ms1/STARS.

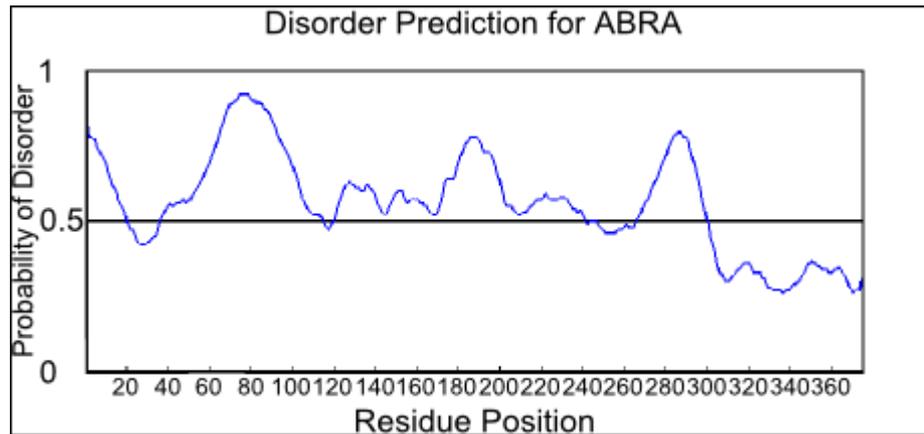


Figure 3.3 – Results of RONN prediction of the disorder of ms1/STARS. Any region with a probability of disorder greater than 0.5 is considered to be disordered.

Domain hunting (by DomainX (Reich *et al.*, 2006)) produced several putative domains (see appendix 4); three of these were chosen for further study as they covered the rest of ms1/STARS not covered by ABD2 (figure 3.4).

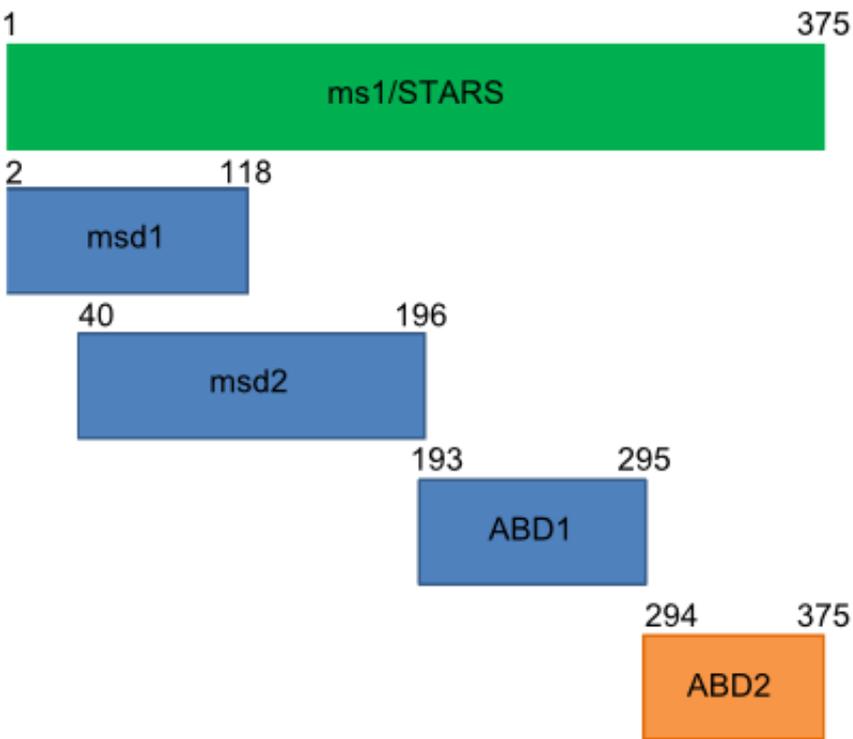


Figure 3.4 – Position, with residue numbering, of the domains found by DomainX.

The putative domains were expressed and characterised using CD and NMR. For the data for ABD2, see chapter 4.

### **3.2 Characterisation of the Other Putative Domains of ms1/STARS Using CD and NMR**

Protein domains are parts of a protein that are independently active. Protein domains are also sometimes independently folded. This means that putative domains need to be characterised in terms of their structure and their function. The methods used to do this were circular dichroism (CD) and NMR as they can be used to report on the secondary structure of proteins.

In the CD spectra, the values shown are in mdeg, which gives ellipticity per molecule and therefore takes the concentration of the protein into account. This means that the values can be compared to each other. The CD data is interpreted iteratively by comparing the values found experimentally to standard CD spectra of  $\alpha$ -helical proteins,  $\beta$ -sheet proteins and random coil proteins. The fitting of the CD data to the standards is directly proportional to the percentage of each secondary structure found in that protein.

Figure 3.5 is the CD spectrum of 30  $\mu$ M msd1. The raw data is shown with a dotted line, while the fitted data is shown with a solid line. There is a peak at 202 nm. This is indicative of a random coil protein. When the data is interpreted, msd1 is shown to contain 23%  $\alpha$ -helical, 7%  $\beta$ -sheet and 70% random coil.

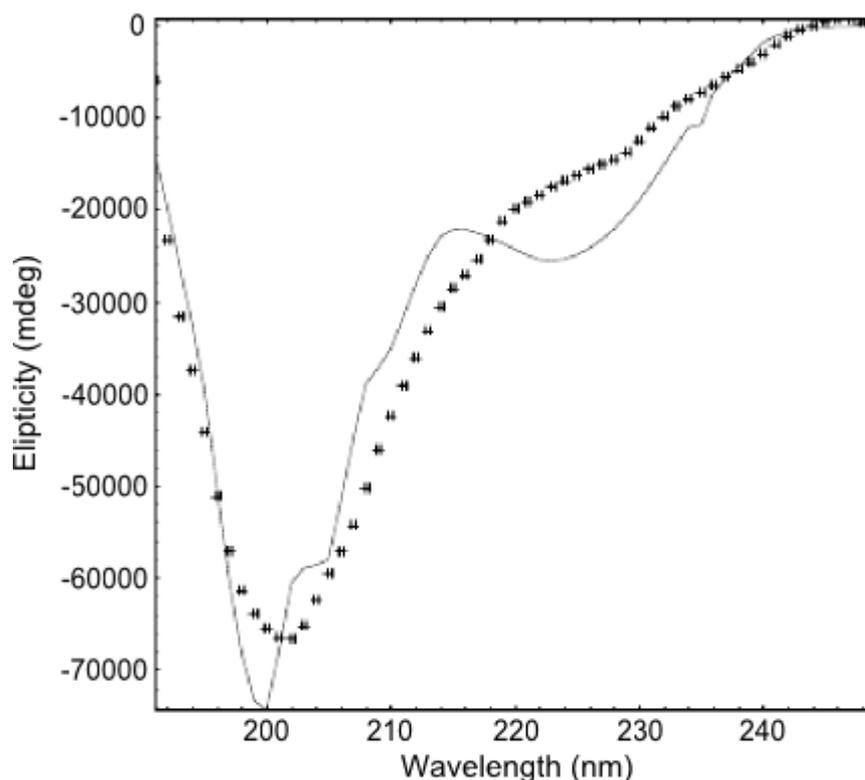


Figure 3.5 – CD spectrum of msd1 (30  $\mu$ M). The raw data is the dotted line and the fitted data is the solid line. It is 23% alpha-helical, 7 % beta-sheet and 70 % random coil.

An additional method capable of detecting protein secondary structures is nuclear magnetic resonance spectroscopy (NMR). Only the aliphatic region of the NMR spectra is shown for the putative domains, as this region shows if the domain is folded or unfolded. This region, between 0 and -2 ppm, will only contain a peak if methyl groups are shifted by proximity to the side chains of aromatic amino acids. This will only occur if the domain is folded. If the domain is not folded, no side chain methyl groups will be close to aromatic side chain groups. This means that there will be no peaks between 0 and -2 ppm. Figure 3.6 shows the aliphatic region of the 1D NMR spectrum of msd1 (100  $\mu$ M). There is no peak between 0 and -2 ppm. This shows that msd1 is unfolded.

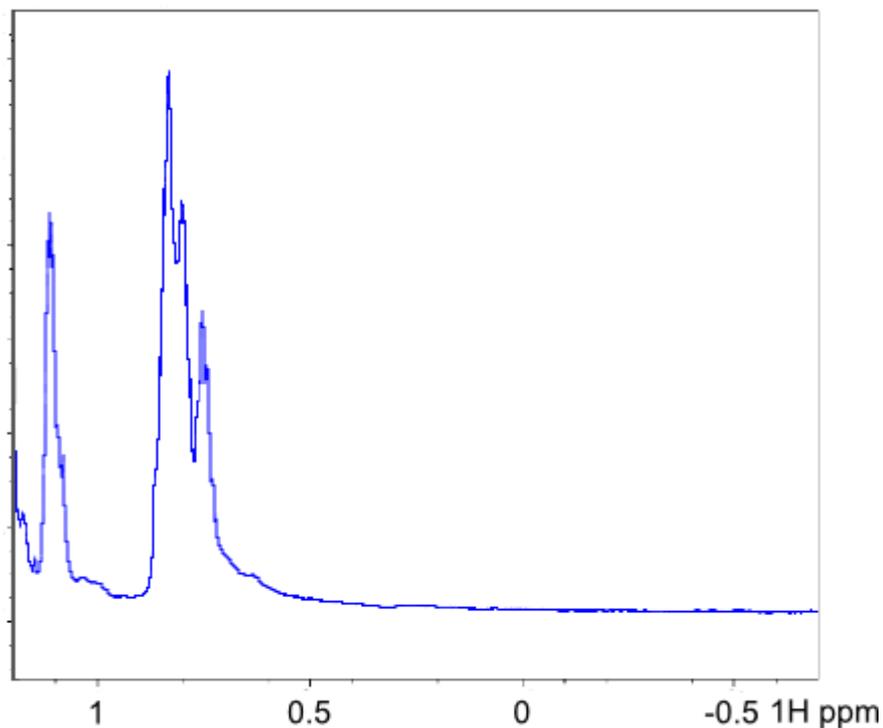


Figure 3.6 – Aliphatic region of a 1D NMR spectrum of msd1 (100 μM). The lack of peaks below 0 indicates that msd1 is unfolded.

Figure 3.7 shows the CD spectrum of 30 μM msd2. It has a peak at 205 nm. Data analysis suggests that this domain contains about 25%  $\alpha$ -helices, 14%  $\beta$ -sheet and 61% random coil. Figure 3.8 shows the aliphatic region of the 1D NMR spectrum of msd2 (100 μM). There is no peak between 0 and -2 ppm. This suggests that msd2 is unfolded.

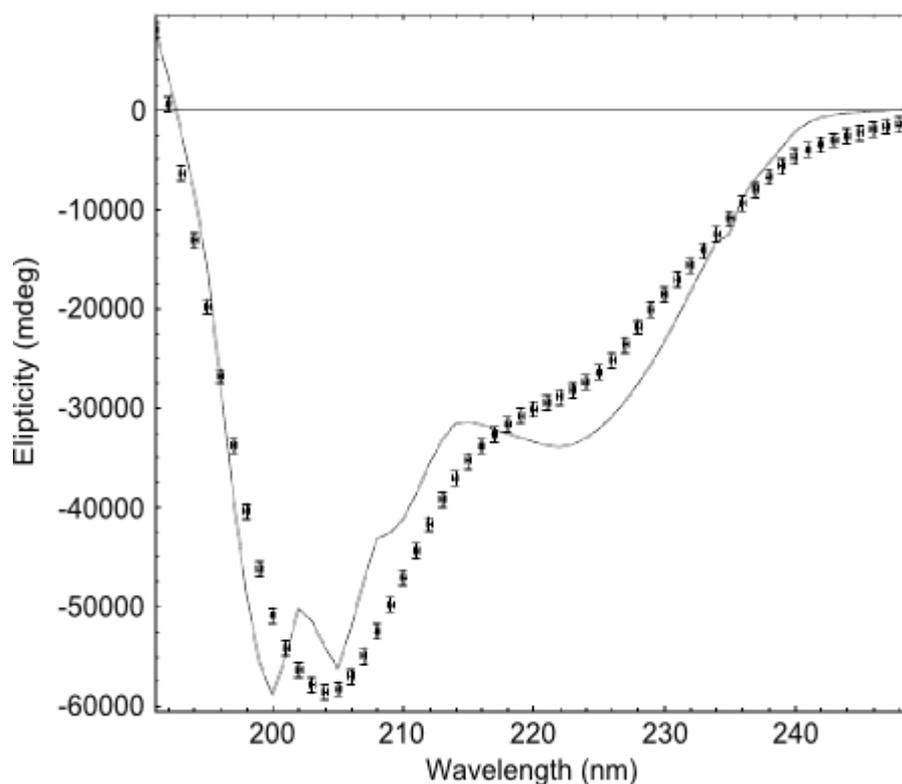


Figure 3.7 – CD spectrum of msd2 (30  $\mu$ M). The raw data is the dotted line; the fitted data is the solid line. It is 25% alpha-helical, 14% beta sheet and 61 % random coil.

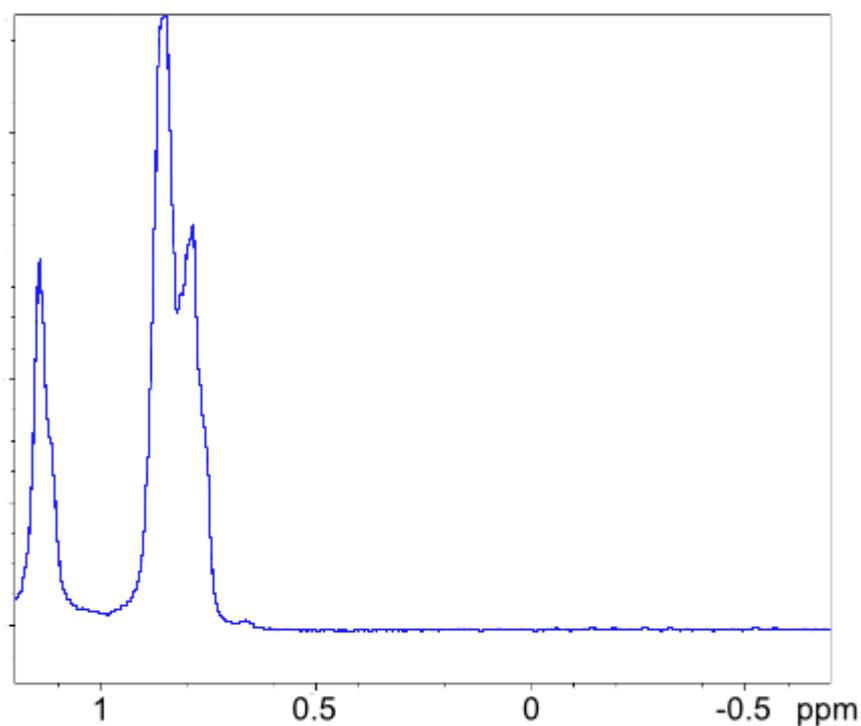


Figure 3.8 – Aliphatic region of a 1D NMR spectrum of msd2 (100  $\mu$ M). The lack of peaks below 0 indicates that ms1d1 is unfolded.

Figure 3.9 shows the CD spectrum of 30  $\mu\text{M}$  ABD1. It has two peaks, one at 203 nm and the other at 208 nm, and when the data is interpreted, the domain is said to be 19%  $\alpha$ -helical, 23%  $\beta$ -sheet and 58% random coil. Figure 3.10 shows the aliphatic region of the 1D NMR spectrum of ABD1 (100  $\mu\text{M}$ ). There is no peak between 0 and -2 ppm. This shows that ABD1 is unfolded.

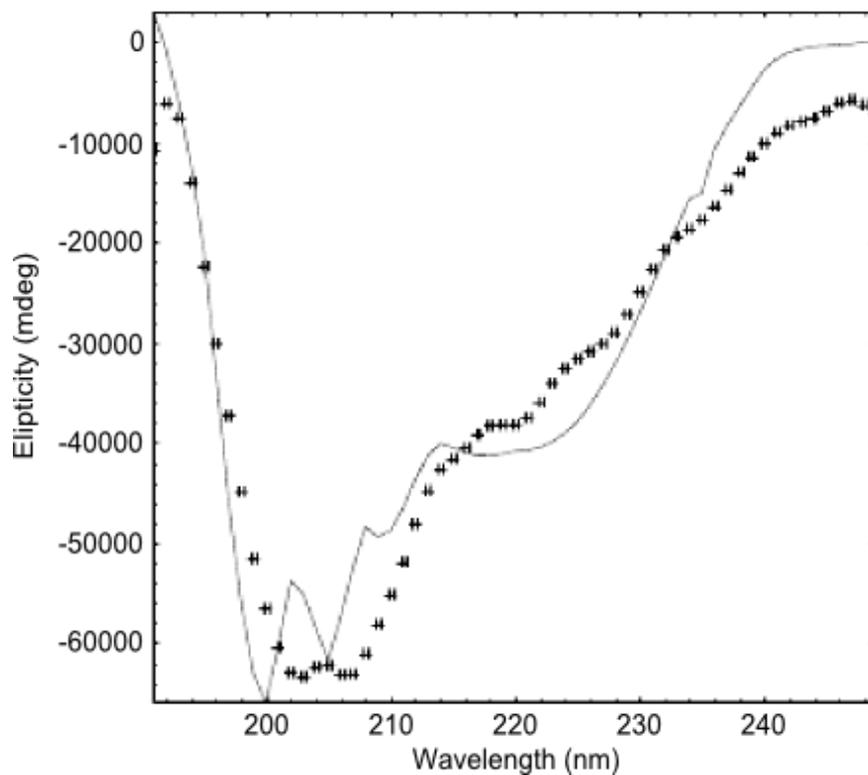


Figure 3.9 – CD spectrum of ABD1 (30  $\mu\text{M}$ ). The raw data is the solid line, and the fitted data is the dotted line. It is 19 % alpha-helical, 23% beta-sheet and 58 % random coil.

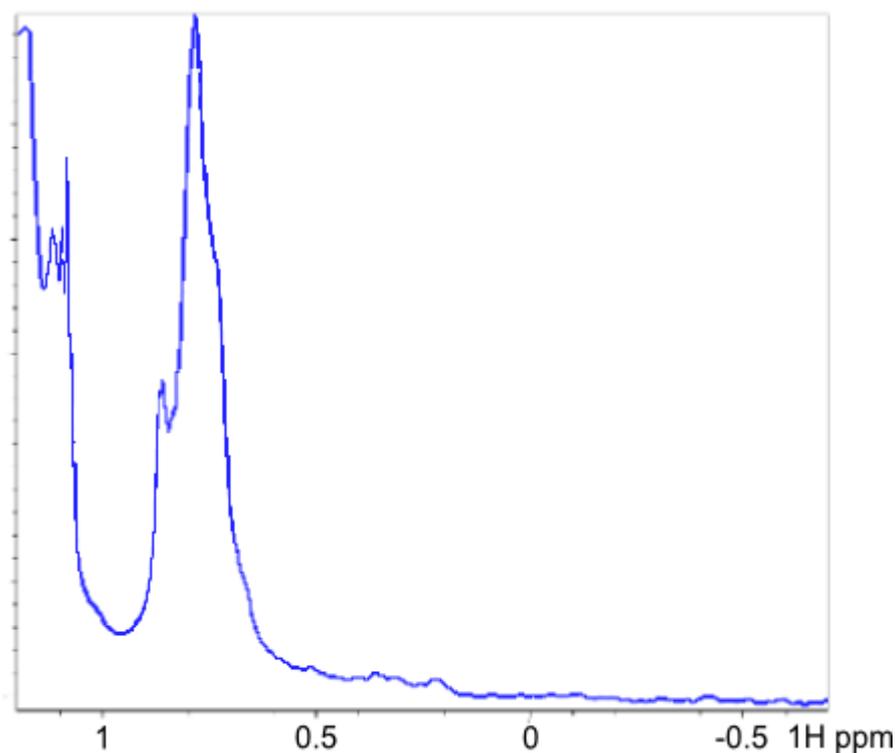


Figure 3.10 – Aliphatic region of a 1D NMR spectrum of ABD1 (100  $\mu$ M). The lack of peaks below 0 indicates that ms1d1 is unfolded.

All of the non-ABD2 domains of ms1/STARS are unfolded as predicted by RONN (see figure 3.3). ABD2 was also characterised using CD and NMR, and the results can be seen in chapter 4.

### **3.3 Discussion of the Domain Organisation of ms1/STARS**

Little was known about the organisation of ms1/STARS at the start of this project. It was believed that ms1/STARS was comprised of domains since, when expression of full-length ms1/STARS was attempted in *E. coli*, it degraded into discrete fragments (see figure 3.1), which suggests that there are domains joined by cleavable linker regions. Rational construct design was used to suggest the boundaries of a C-terminal

domain of ms1/STARS, residues 294-375. Since this is the only conserved region of the protein, it was believed to be a possible location for an actin binding domain.

Another point in favour of this hypothesis was that this conserved region was only found in eukaryotes, and in all kinds of eukaryotes, and not found at all in prokaryotes. Only eukaryotes use actin as part of their cytoskeletal system, whereas prokaryotes use MreB (Jones *et al.*, 2001), MreB-like proteins (Carballido-Lopez *et al.*, 2006) or ParM proteins (Møller-Jensen *et al.*, 2003), which could explain the lack of ms1/STARS homologues in prokaryotes.

Rational construct design was unsuitable for finding other possible domains as no other conserved areas were found. Instead combinatorial domain hunting (Reich *et al.*, 2006) was performed by DomainX. This method uses fragment libraries of the gene of interest and then expresses them to find soluble constructs; these are then rated on their level of expression. This has to be done as combinatorial domain hunting can produce up to 260,000 soluble fragments for a protein. Using the DomainX method, 26 soluble domains with high expression were produced (see appendix 4). Three were selected as they had high expression and covered the whole length of ms1/STARS (figure 3.4).

A soluble fragment covering the same area as ABD2 (residues 294-375) was found, but was not amongst the fragments given to us by DomainX as it expresses poorly. Those constructs that we were given agreed with the sizes of the fragments that ms1/STARS degraded into when full-length expression was attempted (figure 1.12). The ms1/STARS homologues show an interesting feature, in that they increase in size as the organisms in which they are expressed increase in complexity (figure 3.11).

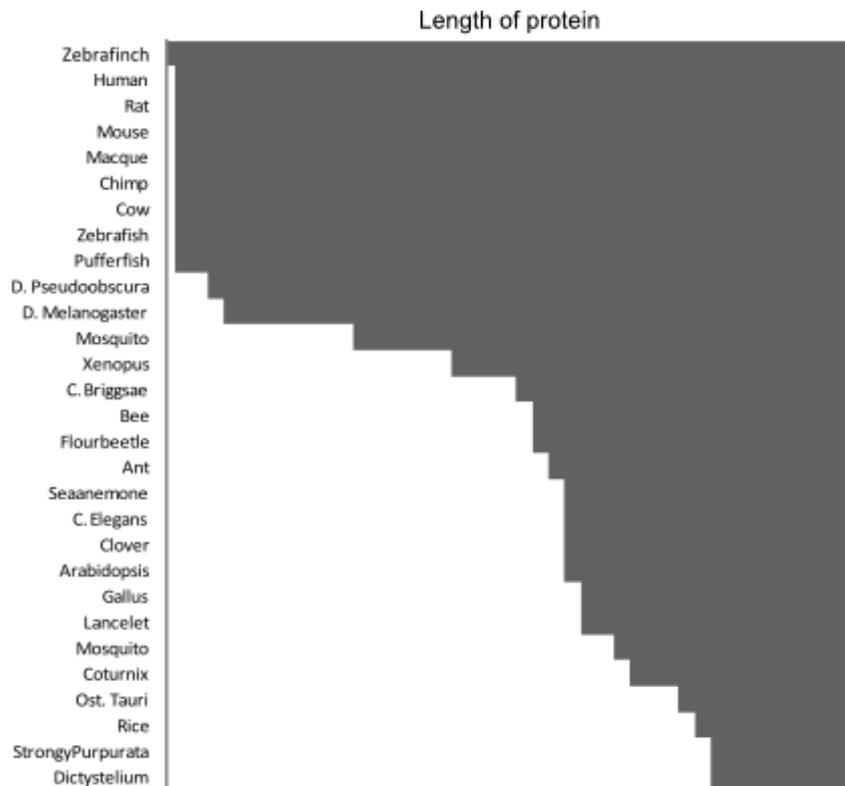


Figure 3.11 – Relative lengths of the homologues of ms1/STARS.

The increases in size happen in intervals similar in size to those of the putative domains. The putative domains were then examined using circular dichroism spectrometry (CD) and nuclear magnetic resonance spectrometry (NMR). Using both of these methods, only the C-terminal domain, ABD2, is folded. When it undergoes denaturation, it is a two-step process, forming a sigmoidal curve, which suggests cooperative unfolding. Cooperative unfolding is a sign of a folded domain. An unfolded domain would not undergo this change, and only a folded domain would go from folded to unfolded in one step.

The only folded domain, ABD2, was found by DomainX, but it was not classed as a highly expressing soluble domain by them. This suggests that there could have been other folded domains that were not found because of their low expression level. This is

unlikely due to the amino acid sequence of ms1/STARS which is predicted to be unfolded except for ABD2 (as predicted by RONN, figure 3.1). This may, however, be a problem with combinatorial domain hunting for other proteins. The sheer number of soluble fragments that it is possible to produce using the domain hunting method means that there has to be a form of selection, as full scale expression of all soluble fragments would be too much for a laboratory, even with a high-throughput methodology in place. Despite this possible problem, combinatorial domain hunting aids domain discovery when rational domain finding techniques such as looking for conserved regions cannot be used. The number of constructs makes organisation and labelling a vital part of the process, especially as the constructs undergo multiple tests of solubility and expression level.

# **Chapter 4 – Characterisation of ABD2**

## **Chapter 4 – Characterisation of ABD2**

### **4.1 ABD2 - Cloning, Expression, Purification, Biophysical Data and Structure Determination**

#### **4.1.1 Cloning of ABD2**

ABD2 was expressed in order to characterise it. Primers for ABD2 (see appendix 1) were used for the PCR reaction whose products were then used to clone the ABD2 construct into pETM11. The PCR products were run on an agarose gel (figure 4.1) to show that the PCR reaction was successful. Lane 1 contains a band of the right size to be the DNA of ABD2 construct 2. This can be seen by comparing it to the size of the bands in the lane containing the DNA marker ladder. Lane 2 contains a very faint band of the right size to be the DNA of ABD2 construct 3. Lane 3 contains a faint band of the size expected for ABD2 construct 4. Lane 4 contains a band of the same size as is expected for ABD2 construct 1. The negative control lane only contains a faint band of the expected size for ms1/STARS DNA. This shows that the primers are the reason there are bands of the right size in the other lanes, and not contamination of the DNA, as, if this were the case, these bands would appear in the negative control lane and the bands of construct DNA produced by the PCR reaction would all be the same size.

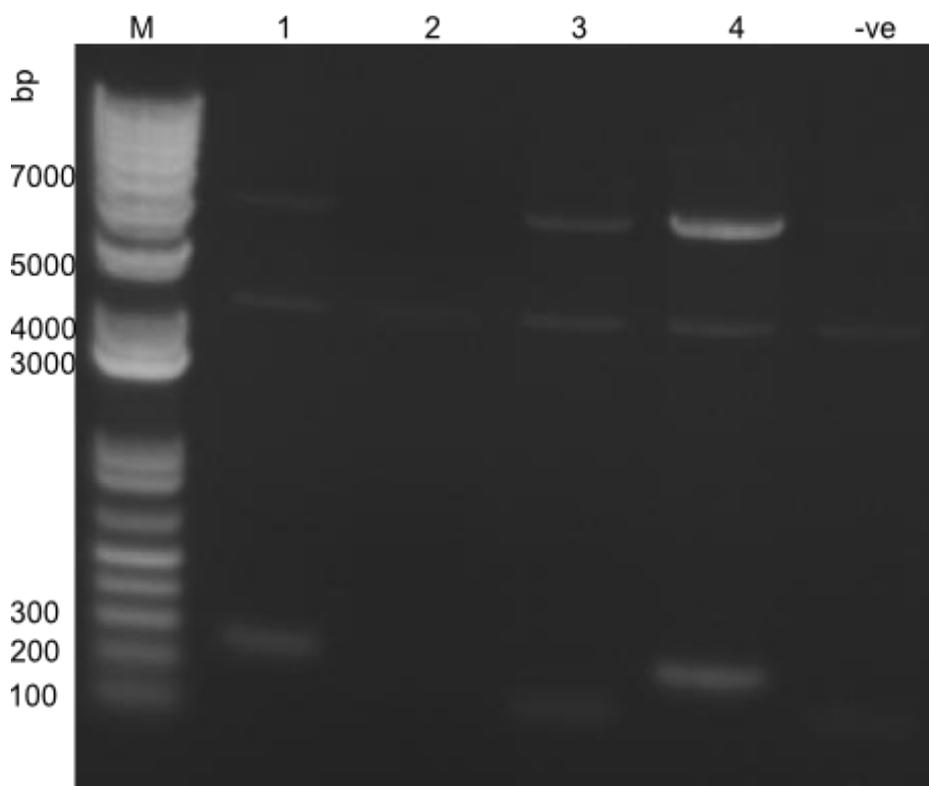


Figure 4.1 – Lane M contains the marker. Lane 1 contains the PCR product of ABD2 construct 2 (residues 297-375), lane 2 contains the PCR product of ABD2 construct 3 (residues 314-375), lane 3 contains the PCR product of ABD2 construct 4 (residues 317-375) and lane 4 contains the PCR product of ABD2 construct 1 (294-375). The lane marked –ve contains the negative control, a PCR reaction run with ms1/STARS alone. Lanes 1-4 contain bands of the correct size to be the desired constructs.

Construct 1 (residues 294-375) was chosen as the construct to be used for the characterisation of ABD2 as it covered the whole of the conserved region and was the construct with the highest yield of DNA after PCR. After cloning, DNA was transfected into Top10 One-Shot (Invitrogen) competent cells. DNA from these cells was sequenced and found to have been ligated in frame.

#### **4.1.2 Pilot Expression of ABD2**

To characterise ABD2, it must first be expressed. To do this, the DNA of ABD2 must be transfected into an expression vector and then pilot expressions must be performed to

find the best conditions to give maximum yield of soluble protein. ABD2 DNA was transfected into BL21 (STAR) DE3 cells and then pilot expression was performed. Figure 4.2 is an SDS PAGE gel showing the results of this. ABD2 with the His tag attached expresses as a protein of 12.76 kDa. A band of this size (shown by the arrow in figure 4.2) can be seen in all of the lanes. From this gel, it was concluded that incubation at 15°C for 16 hours (lane 4) is the best condition for the expression of soluble protein for this construct of ABD2.

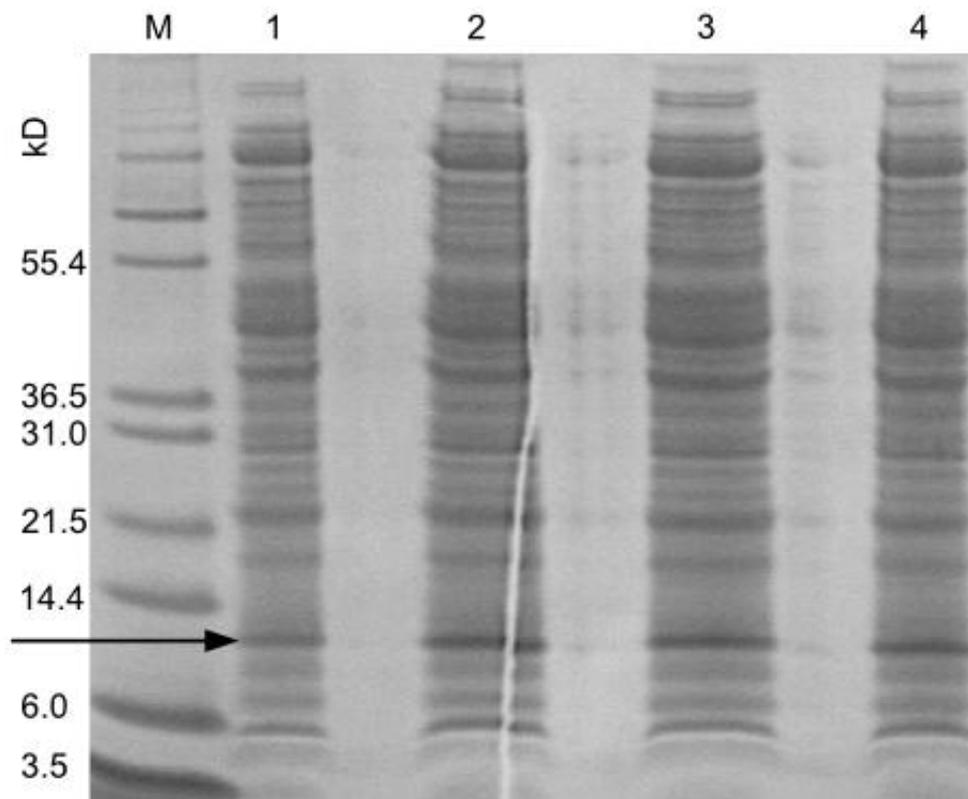


Figure 4.2 – Results of the pilot expression of ABD2, run on an SDS PAGE gel. M = marker, 1 = total protein after 4 hours incubation at 37°C, 2 = soluble protein after 4 hours incubation at 37°C, 3 = total protein after 16 hours incubation at 15°C, 4 = soluble protein after 16 hours incubation at 15°C. The black arrow indicates the expressed ABD2.

### **4.1.3 Purification of ABD2**

To be characterised, the protein must first be purified. For unlabelled ABD2, two nickel column purification steps, one before and one after the cleavage of the His-tag, were required for purification. The result of nickel column purification can be seen in the SDS PAGE gel in figure 4.3. Lane F1 contains the flow-through after the first nickel column purification step. This contains all the proteins with no nickel binding abilities. It does not contain ABD2, as ABD2 is expressed with a nickel-binding His-tag. Lane W1 contains the collected wash fraction. Washing removes any unbound proteins from the column. This fraction does not contain ABD2. Lane E1 contains the collected elution fraction from the first nickel column purification step. This fraction contains ABD2 with the His-tag, the band at 12.76 kDa. Following this step, the His-tag is cleaved from ABD2, and then the column purification step is repeated. Lane F2 contains the flow-through from this second step. This fraction contains ABD2 because ABD2 can no longer bind to the nickel column. The band of ABD2 is the only protein in this lane, as ABD2 alone has lost its ability to bind to nickel between the first and second column purification steps. ABD2 is the band at 10 kDa. Lane W2 contains the collected wash fraction, which also contains ABD2. Lane E2 contains the second collected elution fraction. This fraction does not contain ABD2 but contains uncleaved His-tagged ABD2 and various other impurities.

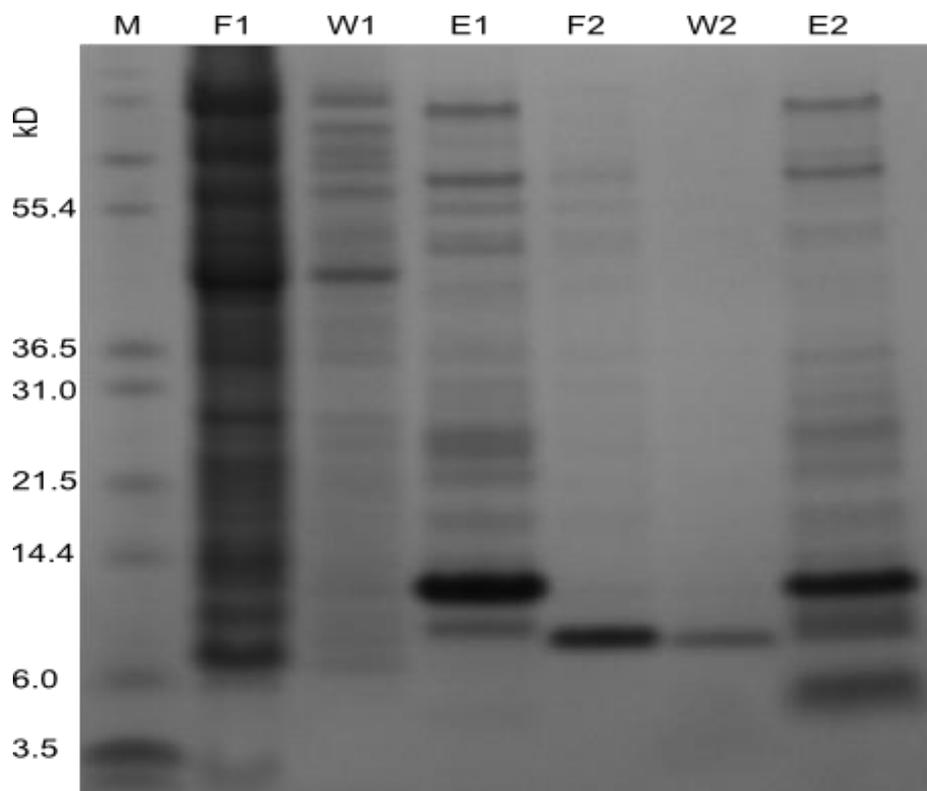


Figure 4.3 – The two stages of nickel column purification of ABD2 shown on an SDS PAGE gel. F = flow-through, W = wash and E = elution. 1 and 2 refer to the first and second stages.

However, nickel column purification was not sufficient to purify the ABD2 construct when it was labelled with  $^{15}\text{N}$  or  $^{13}\text{C}$  and  $^{15}\text{N}$ . For these samples, further purification was required. Gel filtration chromatography was used for this (figures 4.4 and 4.5). Figure 4.4 shows the gel filtration chromatograph. The y-axis shows the signal in mAu, while the x-axis shows the volume which has passed through the column. The red line is the absorption signal from the protein. There are three peaks in the signal, one at 44 - 56 ml, one at 56 - 72 ml and one at 86 - 106 ml. Samples from each peak are run on an SDS PAGE gel to find which peak contains ABD2.

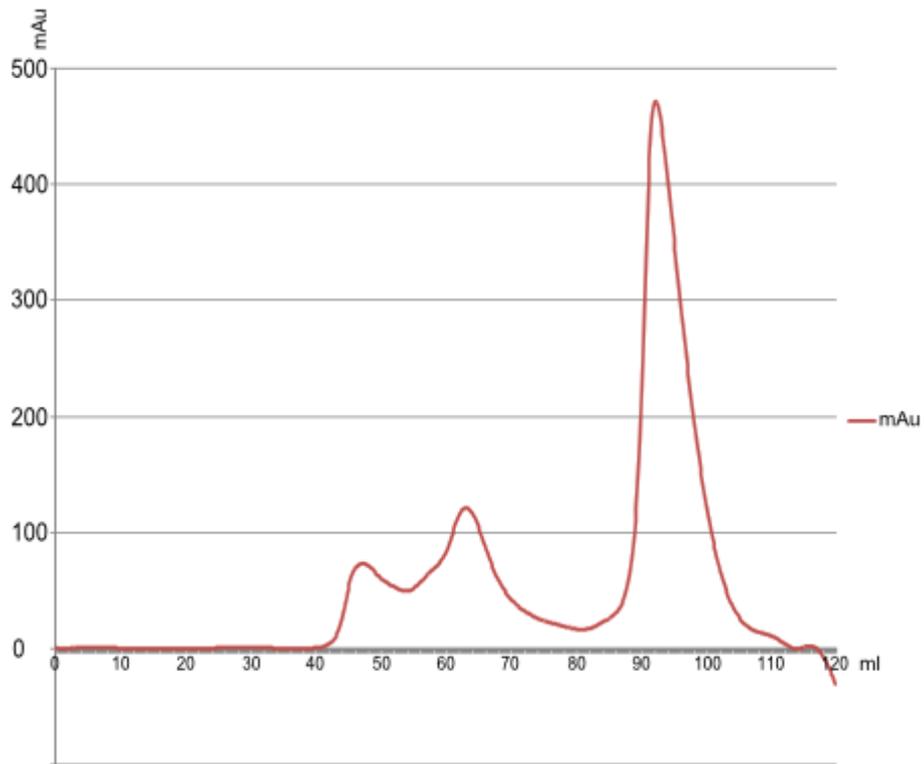


Figure 4.4 – Gel filtration elution profile of ABD2.

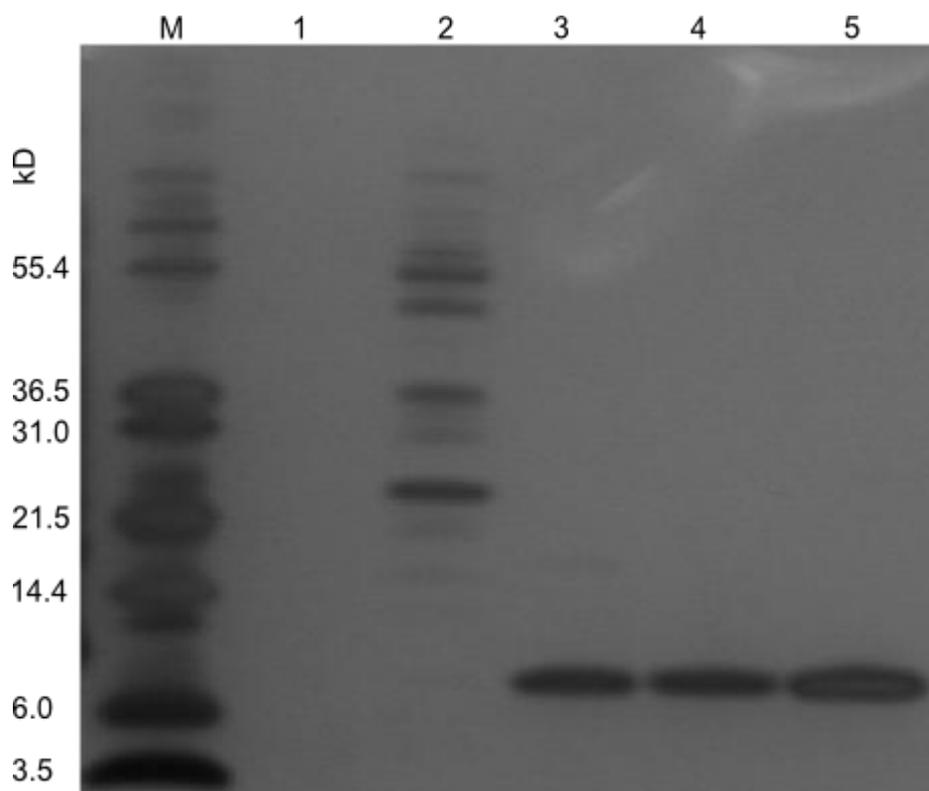


Figure 4.5 – SDS PAGE gel of the results of gel filtration purification. M = marker. Fraction 1 = 40-50 ml, fraction 2 = 60-70 ml, fraction 3 = 80-90 ml, fraction 4 = 90-100 ml and fraction 5 = 100-110 ml.

Figure 4.5 shows the SDS PAGE gel on which these samples were run. Lane M contains the marker. Lane 1 contains a sample from the fractions from 40-50 ml; this corresponds to the first peak in the elution profile (figure 4.4). This lane shows no protein material. Lane 2 contains a sample from the fractions from 60-70 ml, corresponding to the second peak in the elution profile. This lane contains various bands of higher molecular weight than ABD2 and represents contaminant material. Lane 3 contains fractions from 80-90 ml, lane 4 contains fractions from 90-100 ml and lane 5 contains fractions from 100-110 ml. These lanes correspond to the third peak of the elution profile and each contains a single band corresponding to pure ABD2.

#### **4.1.4 Mass Spectrometry**

Mass spectrometry was used for two reasons. Firstly, it would show that the purification strategy which was used works. It would also show the exact molecular weight of the purified protein, which will help to show that the correct protein has been purified. Figure 4.6 shows the results of the mass spectrometry. The y-axis is scaled so that the largest peak is at 100% intensity. The x-axis is mass in Daltons. There are two peaks, the larger, at 100% intensity, is at 10026 Daltons. The smaller is at 10% intensity, and weighs 10120 Daltons. The larger peak is the correct size to be ABD2, which, when the His-tag is removed, weighs 10025.7 Daltons. The smaller peak is an artefact of the mass spectrometry process as, when the protein is ionised, the matrix that the sample is attached to will also be ionised. The mass of the protein plus matrix can be calculated and any peaks due to this can be identified, and shown not to be a contaminant.

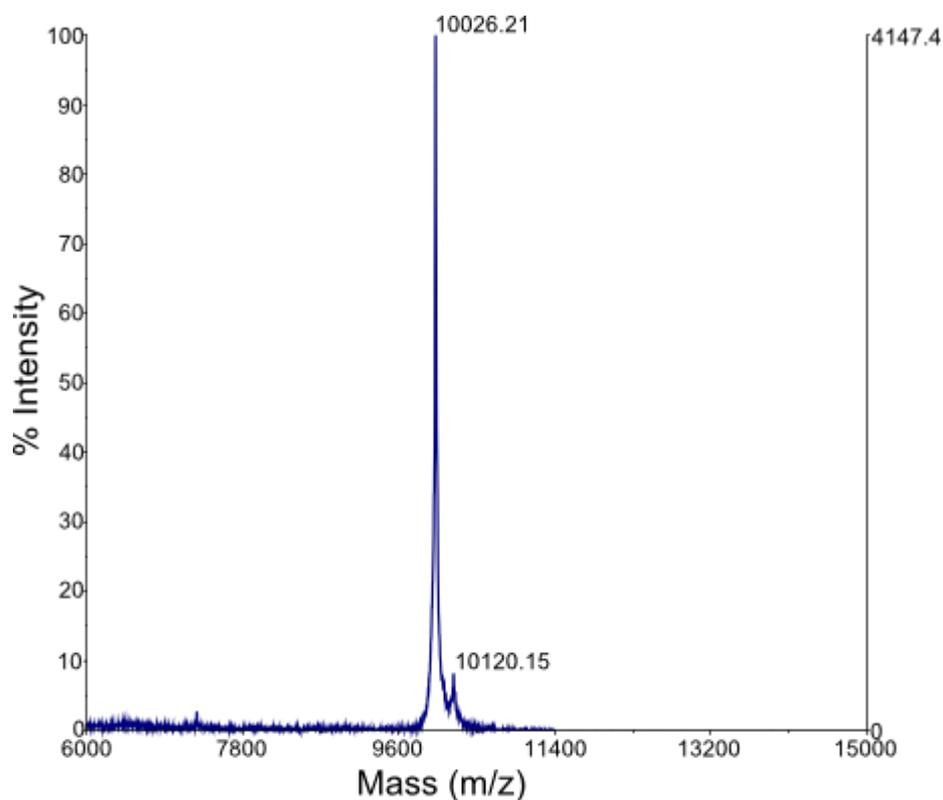


Figure 4.6 – Results of mass spectrometry. The largest peak is of the correct size to be ABD2. The smaller peak is that of the protein plus the matrix. Because of the methodology used, when the sample is ionised, the matrix to which the sample is attached can also be ionised. This leads to a second, smaller, peak (as there is less of protein plus matrix than of the protein on its own), but this will have the mass of the ionised protein plus ionised matrix and this value can be calculated so that the peak can be identified as this and not as a contaminant.

## **4.2 Characterisation of ABD2**

### **4.2.1 Characterisation of ABD2 Using Circular Dichroism**

Circular dichroism spectra of samples of ABD2 were recorded from 200 nm to 250 nm. The CD spectra of this domain show two peaks, one at 210 nm and one at 222 nm (figure 4.7). This is typical of  $\alpha$ -helical proteins. The fitted data shows that ABD2 is 93%  $\alpha$ -helical and 7% random coil.

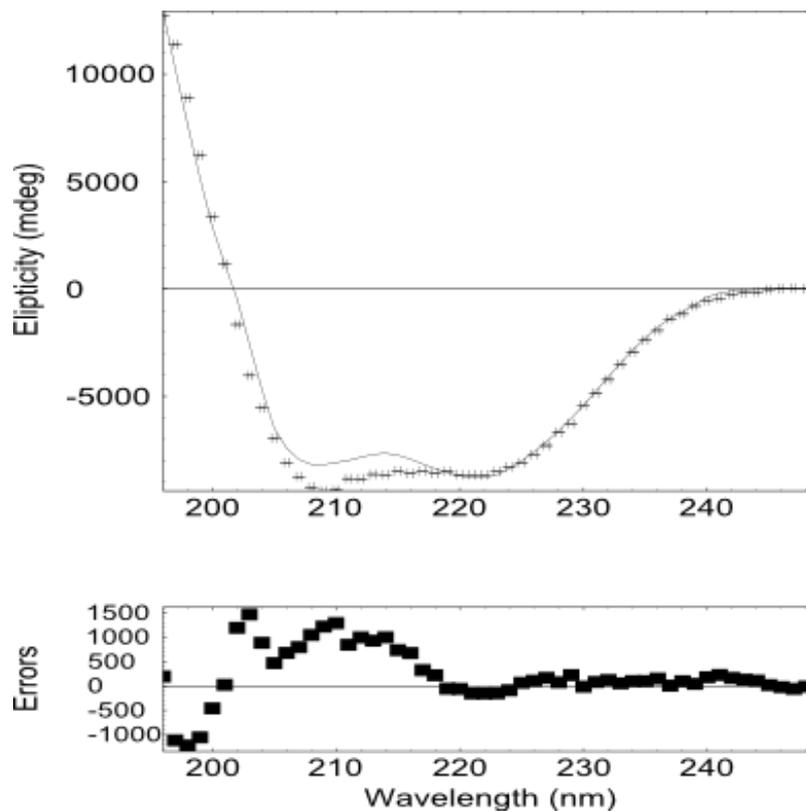


Figure 4.7 – CD spectrum of ABD2 (30  $\mu\text{M}$ ). The experimental data is the dotted line; the solid line is the fitted data. ABD2 is 93%  $\alpha$ -helical and 7% random coil.

Circular dichroism can also be used to evaluate the stability of a protein. This can be done in two ways, using either heat or a chemical such as urea as a denaturant (figures 4.8 and 4.9). Figure 4.8 shows the denaturation of ABD2 (30  $\mu\text{M}$ ) by increasing heat. The shape of the curve suggests that there is co-operative unfolding, a sign of a folded protein. The melting temperature of 56°C shows that ABD2 is stable. The box below the main figure shows the residual differences between the experimental data and the theoretical values used for data fitting. They are evenly spread out so the value for the melting temperature is not skewed by experimental error.

Figure 4.9 shows the denaturation of ABD2 (30  $\mu\text{M}$ ) using urea as a chemical denaturant. The concentration of urea is shown on the x-axis, with the molecular ellipticity along the y-axis. The graph shows a two-step curve suggesting that there is co-operative unfolding, something which only occurs if the domain is folded.

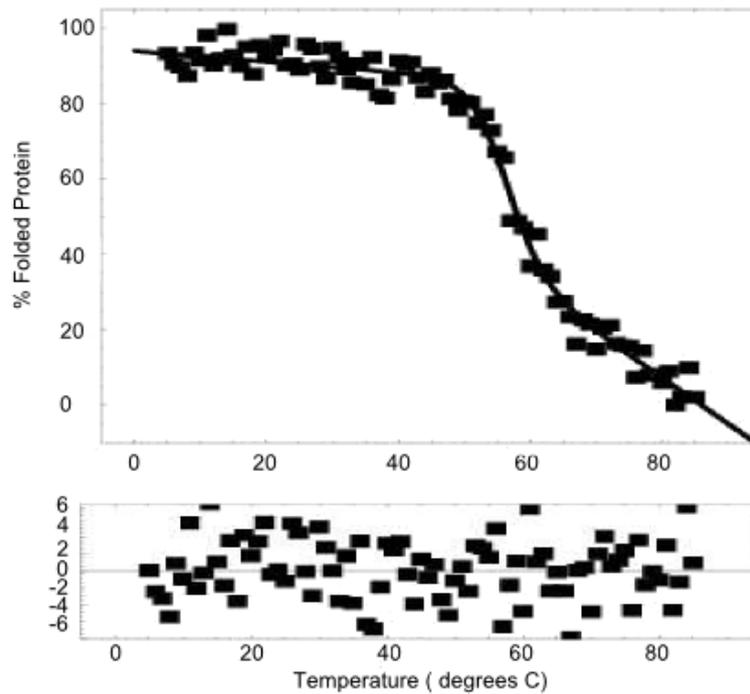


Figure 4.8 – Denaturation of ABD2 (30  $\mu\text{M}$ ) using increasing temperature. This gives a melting temperature of  $56^{\circ}\text{C} \pm 0.3^{\circ}\text{C}$ .

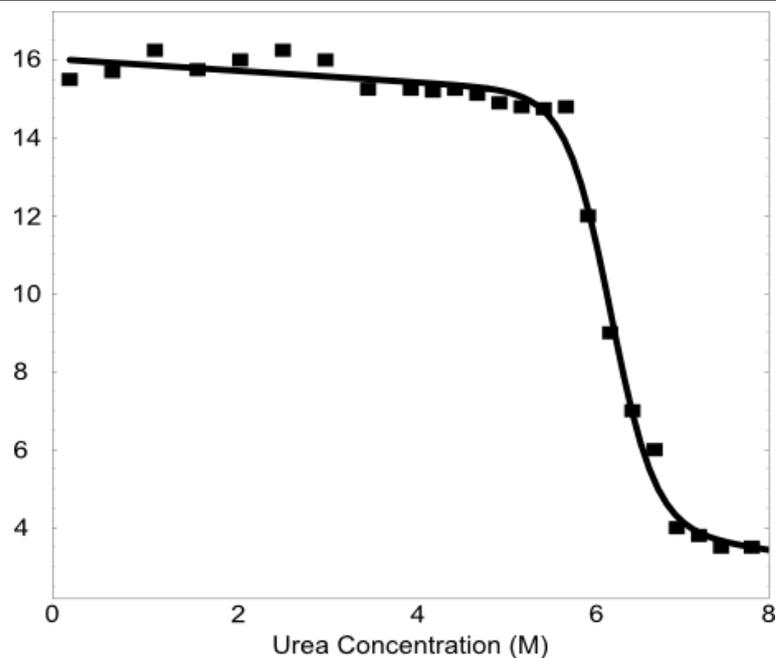


Figure 4.9 – Denaturation of ABD2 (30  $\mu\text{M}$ ) using increasing concentration of urea. Analysis of this data gave ABD2 a Gibbs free energy of  $13.41 \pm 1.38$  kcal per mol.

Both of these experiments showed that ABD2 is a stable protein with a melting temperature of 56°C, and denaturing at 6.2 M urea. Both experiments also show that ABD2 is a folded domain, as it undergoes co-operative unfolding, with a distinct transition between folded and unfolded states.

#### **4.2.2 Characterisation of ABD2 Using Fluorescence**

Another method that can be used to determine the folded nature of a protein is following the changes in tryptophan fluorescence when the protein is denatured using urea (figure 4.10). If a typical protein with a buried tryptophan unfolds, then the emission wavelength shifts from 330 nm to 360 nm (Jiskoot *et al.*, 1995). However, the peak for folded ABD2 (6 µM ABD2, 0 M urea) was at 353 nm, while the peak for urea treated ABD2 (6 µM ABD2, 8 M urea) is at 357 nm. Despite being folded, ABD2 shows very little shift in the emission wavelength when it becomes unfolded. This means unfolding of ABD2 cannot be followed using fluorescence spectroscopy.

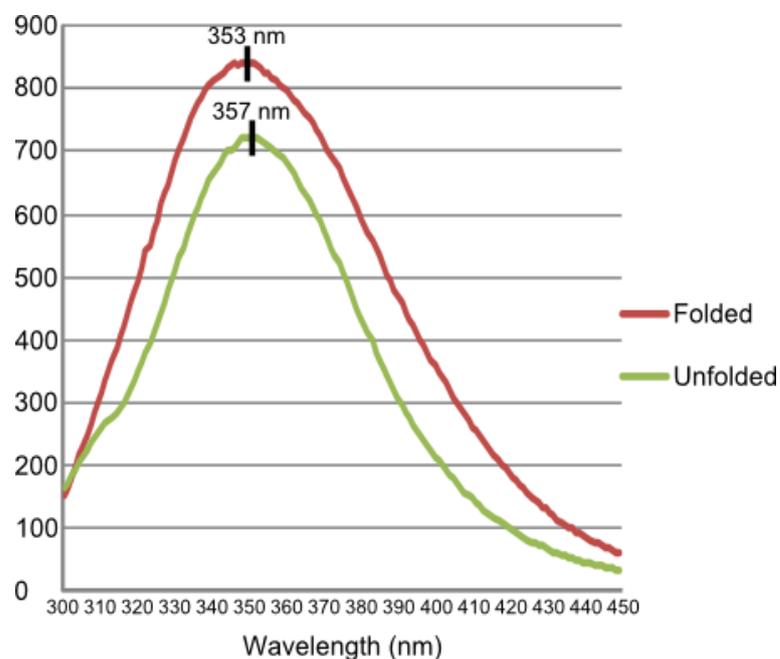


Figure 4.10 – Fluorescence spectrum of folded and unfolded ABD2 (both at 6 µM). The peak for folded ABD2 is 353 nm while the peak for unfolded ABD2 is 357 nm.

The emission value did provide useful information about the environment in which the sole tryptophan in ABD2 was found. That the starting value of the emission wavelength was already 353 nm suggests that the tryptophan is not in the hydrophobic core of ABD2. However, the value is also not as much as it would be if the tryptophan was in a completely water-accessible position, suggesting that it is instead on the edge of a pocket, where it is partially accessible to water.

### **4.2.3 Determination of the Multimeric State of ABD2**

#### **4.2.3.1 Results from Analytical Ultra-Centrifugation (AUC)**

The results of multimeric state determination by AUC were inconclusive, despite being repeated several times.

#### **4.2.3.2 Results From Analytical Gel Filtration**

Analytical gel filtration showed ABD2 (100  $\mu$ M) exists as a monomer. The elution profile of ABD2 compared to standards of known molecular weights can be seen in figure 4.11. The red line shows the elution profile of standards that had previously been run down the analytical gel filtration column. The blue line shows the elution profile of ABD2. The elution volume of ABD2 can be plotted against a standard curve produced by plotting the log of the elution volume of the standards against the log of their molecular weight. The calculated molecular weight is compared to the known molecular weight of the protein and it can be seen if the calculated molecular weight is equal to that of the protein acting as a monomer, dimer or multimer. This can be done because analytical gel filtration does not break non-covalent bonds between multimers.

The elution volume of ABD2 is 15 ml. This is the elution volume of a protein of 10 kDa. ABD2 has a molecular weight of 10.025 kDa. This suggests that ABD2 acts as a monomer.

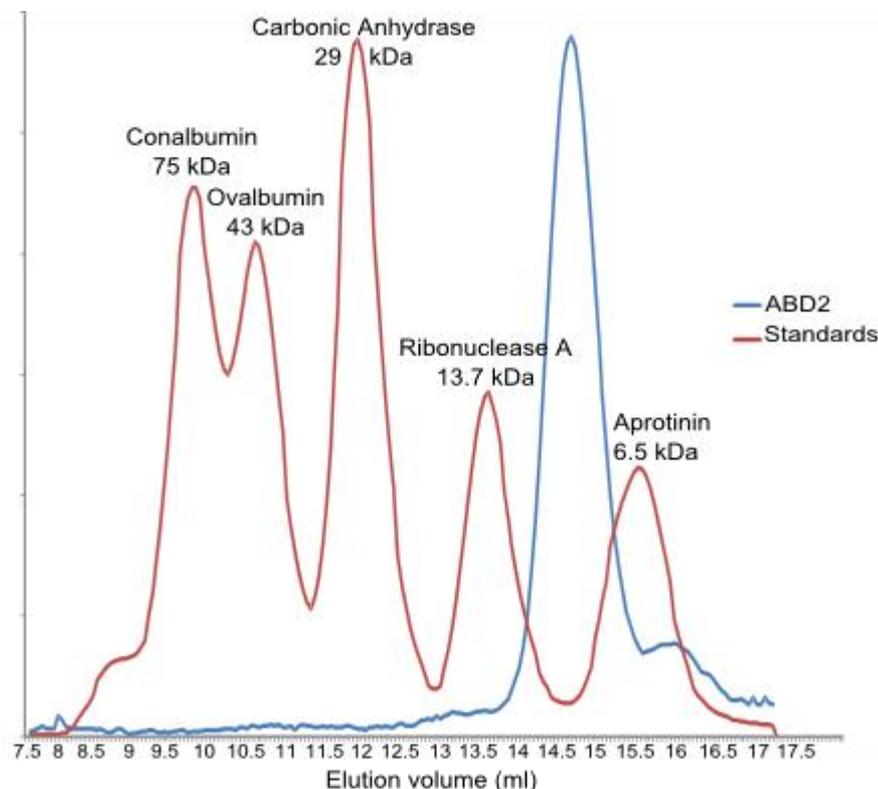


Figure 4.11 – Elution fraction of ABD2 (100  $\mu$ M) compared with the elution fraction of standards of known molecular weight when they are run on an analytical gel filtration column. ABD2 elutes at the fraction where a 10 kDa protein would elute. This is the size it would be if it is a monomer.

### **4.2.3.3 Results From NMR**

Combining the data from T1, T2 and hetero-NOE experiments performed on ABD2 allowed the calculation of its rotational correlation time ( $\tau_c$ ). The experiments were performed at 298 K on a 200  $\mu$ M sample. The results of this calculation can be seen in Figure 4.12. A Monte-Carlo simulation is used to calculate possible rotational correlation times which would produce the experimental results. A possible rotational correlation time which fits the data well will have a lower free energy. The 100 lowest free energy possible rotational correlation times are placed into bins in order to calculate

the most likely value for the rotational correlation time. The bin that contains the most of these lowest free energy possible rotational correlation times is that which is the most likely. The calculated rotational correlation time of ABD2 was 7.5 ns, which is consistent with the  $\tau_C$  of a protein of 10 kDa, the size of monomeric ABD2, when the rotation time is derived from the Stokes-Einstein equation. The derivation, which is the Stokes-Einstein rotational correlation time of a 10 kDa object plus  $\frac{1}{2}$  of that value again to account for non-spherical proteins and the hydration shell of the protein, gives a value of 8 ns for 10 kDa proteins. The theoretical value, again derived from the Stokes-Einstein equation, for a dimer would be 15 ns, and it is unlikely that the experimentally derived rotational time would be so much smaller if it were a dimer.

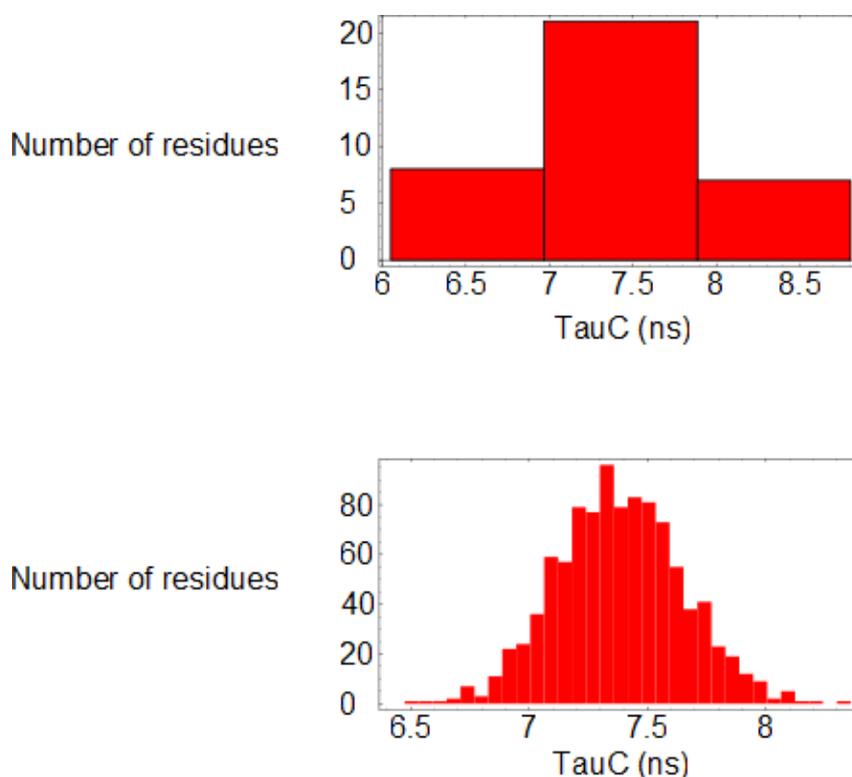


Figure 4.12 – A Monte-Carlo simulation is used to calculate possible rotation times from the experimental results. The 100 lowest energy possible rotation times are placed into bins in order to calculate the most likely value for the rotation time. For ABD2, this value is 7.5 ns.

## **4.3 Structure Determination Using NMR**

### **4.3.1 Residue Identification Using HNCACB and HNCOCACB Spectra**

Using a comparison of peaks in the HNCACB and the HNCOCACB spectra enabled the identification of the residue I and the residue I-1. This process was begun with the glycines which are recognisable from their chemical shifts in HSQC experiments (an HSQC of ABD2 can be seen in figure 4.13). Once a region of residues has been identified by residue type, the residues can be compared to the sequence of the protein and the region that they are from in the domain can be established. Figure 4.13 shows an HSQC spectrum of ABD2 (250  $\mu\text{M}$ ). The blue dots are the peaks produced by amide groups in ABD2. The x-axis shows their chemical shifts in ppm in the  $^1\text{H}$  axis, while the y-axis shows their chemical shift in the  $^{15}\text{N}$  axis. There is one peak for each residue in ABD2 and, in addition, for the tryptophan side chain and for the side chains of the glutamine residues.

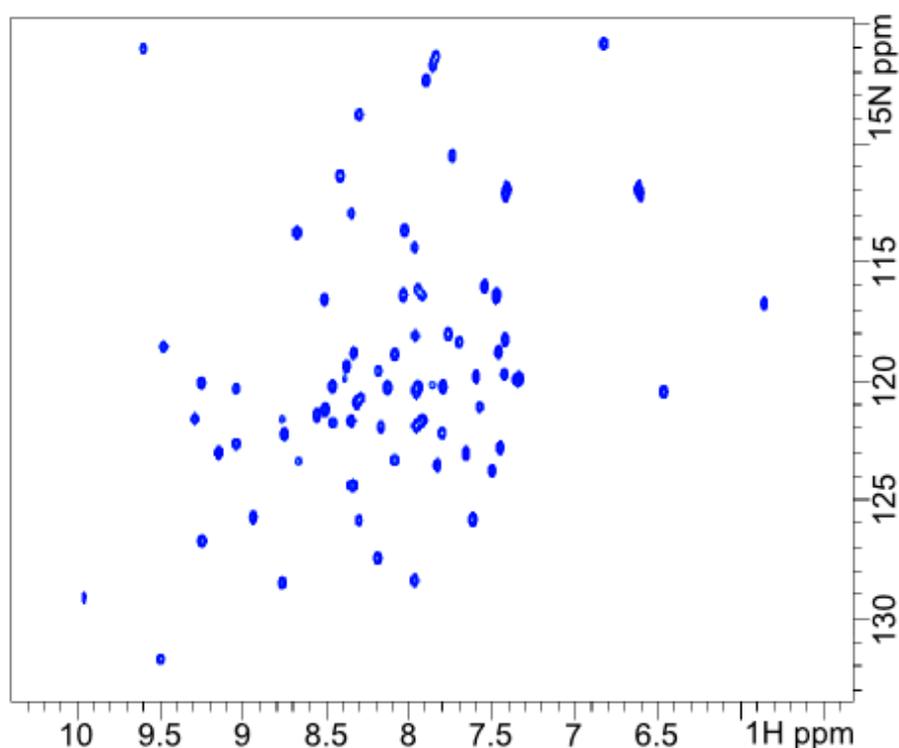


Figure 4.13 – HSQC of ABD2 (250  $\mu\text{M}$ ).

Secondary structure information can then be obtained by combining the identified residues with information from NOESY spectra of the domain. The distance between residues linked by NOESY cross peaks varies depending on secondary structure, with those residues in an  $\alpha$  helix having cross peaks to residues  $I \pm 4$ , while those in  $\beta$  strands have cross peaks to residues much further away. Two such regions are shown in figures 4.14 and 4.15.

In the figures, the purple dots are the peaks from the  $^{15}\text{N}$  NOESY spectrum (performed on a 500  $\mu\text{M}$  sample of ABD2), while the red peaks are the peaks from the HNCA spectrum (performed on a 300  $\mu\text{M}$  sample of ABD2). The light green peaks come from the HN(CO)CA spectrum (performed on a 188  $\mu\text{M}$  sample of ABD2). The blue peaks come from the HNCACB spectrum (performed on a 300  $\mu\text{M}$  sample of ABD2) and the dark green peaks come from the HN(CO)CACB spectrum (performed on a 300  $\mu\text{M}$  sample of ABD2). The peaks and the residues to which they belong are labelled in the figures.

The chemical shift values for  $\text{C}\alpha$  and  $\text{C}\beta$  can be used to identify the residues. The residues can then be linked to the peaks they produced in the  $^{15}\text{N}$  NOESY spectrum.  $^{15}\text{N}$  NOESY peaks have different patterns depending on the secondary structure of the residue. The contacts between different residues found in the  $^{15}\text{N}$  NOESY spectrum are shown by the black lines. Figure 4.14 shows a region of  $\beta$ -strand structure. The  $^{15}\text{N}$  NOESY peaks in this region only show contacts to the next residue, not to the residues  $I \pm 4$ . Figure 4.15, on the other hand, shows a region of  $\alpha$ -helical structure. In this region, the  $^{15}\text{N}$  NOESY peaks show contacts to the residues  $I \pm 4$ .

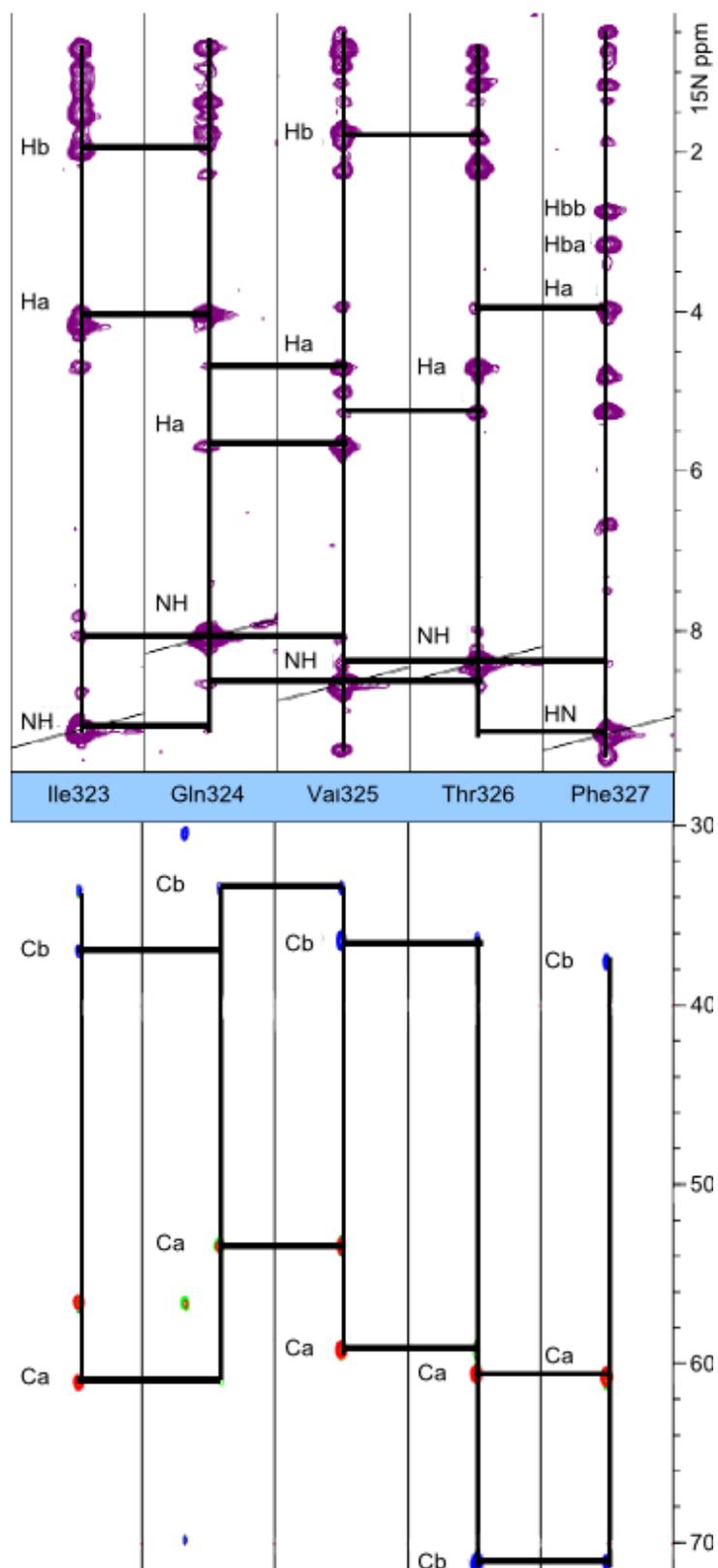


Figure 4.14 – Example of a region of  $\beta$ -strand structure, as seen in overlapped HNCA and HNCACB spectra (bottom half of the figure) and in the  $^{15}\text{N}$ -NOESY spectrum. The  $^{15}\text{N}$ -NOESY spectrum shows a distinct pattern in  $\beta$ -sheets, as there are long distance cross-peaks as opposed to the short distance  $I\pm 4$  residue cross-peaks found in  $\alpha$ -helical regions.

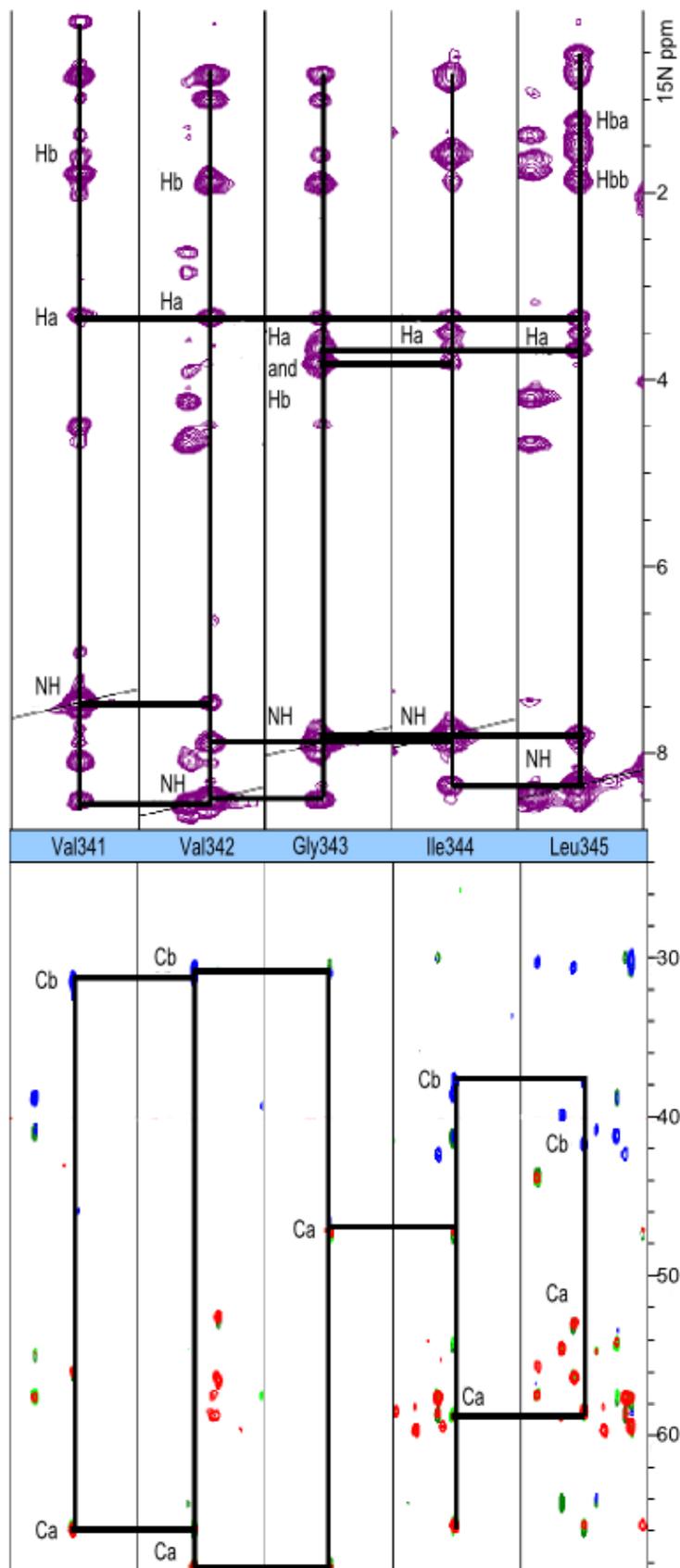


Figure 4.15 – Example of a region of  $\alpha$ -helical structure, as seen in overlapped HNCA and HNCACB spectra (bottom half of the figure) and in the  $^{15}\text{N}$ -NOESY spectrum. The  $^{15}\text{N}$ -NOESY spectrum shows a distinct pattern in  $\alpha$ -helices, where there are cross-peaks from residue  $i$  to residue  $i \pm 4$ . This is not present in  $\beta$ -strands, where cross-peaks are between residues further away in the sequence.

Once this procedure has been performed for all residues, a CSI chart can be made for the domain using the protocol described in “The Chemical Shift Index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy” (Wishart *et al.*, 1992).

The chemical shifts of  $H\alpha$ ,  $C\alpha$  and  $C\beta$  were compared to random coil values and used to plot a chemical shift index diagram (figure 4.16). The values given to each residue relate to the kind of secondary structure it is in, with +1 being assigned to residues with chemical shifts suggestive of being in a beta sheet and -1 being assigned to those with shifts suggesting an alpha helical conformation. A 0 is given to those residues whose chemical shifts are suggestive of random coil conformation. To produce the final diagram, those residues where all three chemical shift types had values suggestive of the same kind of secondary structure were given the value associated with that kind of secondary structure. Those residues where there was a majority of two chemical shifts suggesting a particular secondary structure were also given the value associated with that secondary structure. Where there was one chemical shift suggestive of that residue being in an  $\alpha$ -helix and one suggestive of that residue being in a  $\beta$ -strand and either the third chemical shift was missing, because of a lack of NMR data, or the third chemical shift suggested the residue was in a random coil, those residues were given a value of 0. Those residues without enough data to perform any of this analysis were also given a value of 0. The bars in the diagram represent each residue from 294 to 375.

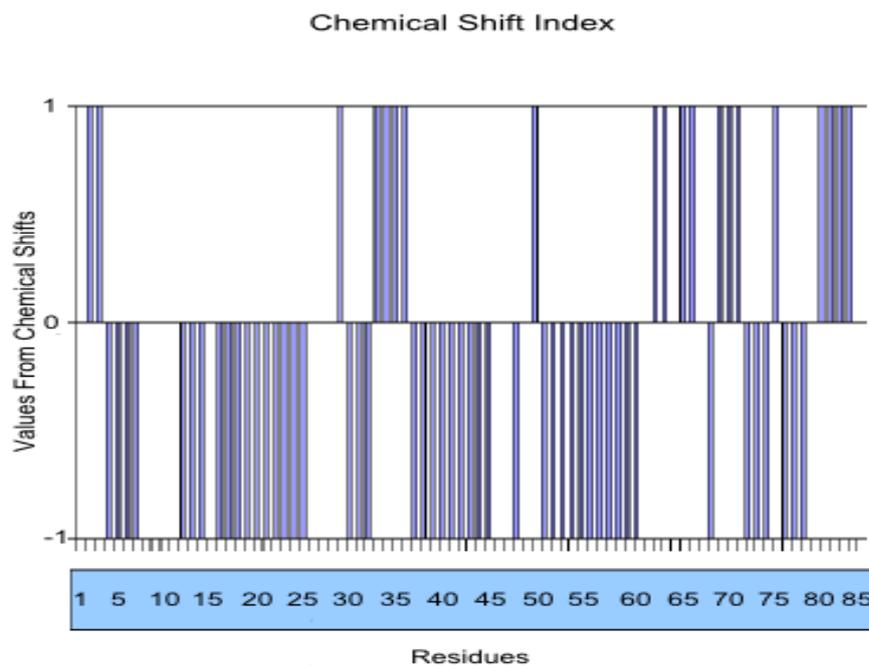


Figure 4.16 – Representation of the results of the chemical shift index of ABD2.

### **4.3.2 Side Chain Assignment**

Once  $C\alpha$  and  $C\beta$  values for each residue have been established, they can be used in combination with a carbon constant time HSQC (performed on a 300  $\mu\text{M}$  sample of ABD2), HCCH TOCSY (performed on a 300  $\mu\text{M}$  sample of ABD2) and  $^{13}\text{C}$ -edited NOESY (performed on a 200  $\mu\text{M}$  sample of ABD2) to identify the side chain carbons in each residue and to then identify the NOESY peaks and cross peaks that belong to that residue (figure 4.17 shows an example of this performed on leucine 330). In the figure, the light blue spots are the peaks from the  $^{13}\text{C}$ -edited NOESY spectrum. The dark blue peaks are from the HCCH TOCSY spectrum. The red spots are used to mark the peaks in the HCCH TOCSY that belong to leucine 330. These are also labelled in the figure.

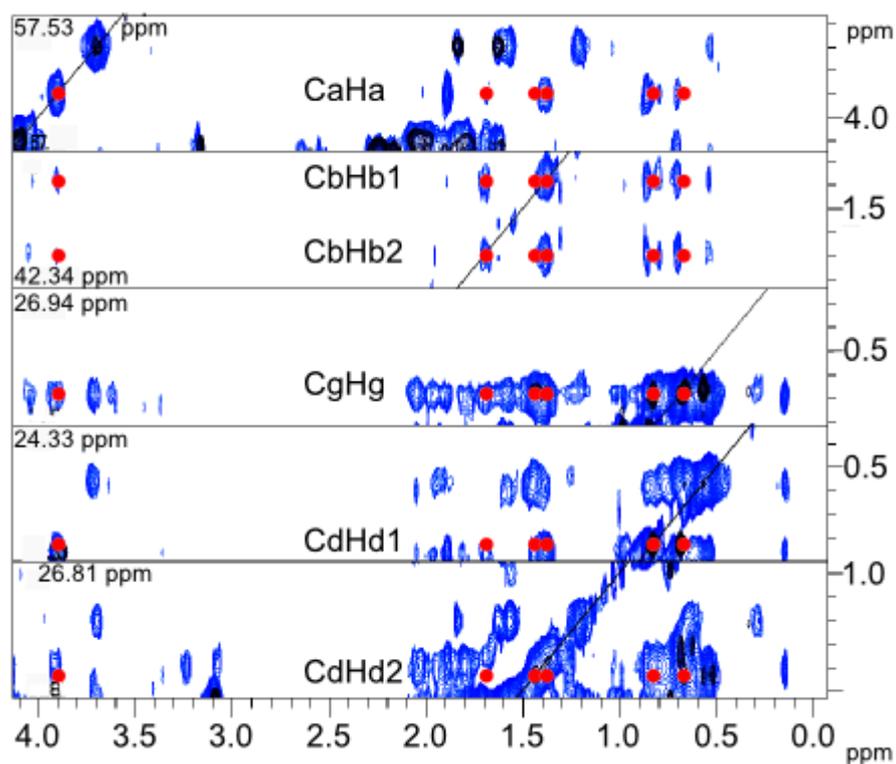


Figure 4.17 – An example of the combination of HCCH TOCSY (in dark blue) and  $^{13}\text{C}$ -NOESY (light blue) used to identify the carbons in the side chains of the residue. The residue in this example is leucine 330. The red spots are used to mark the peaks in the HCCH TOCSY that belong to the leucine. Not all HCCH TOCSY peaks are present in all of the spectra, meaning that more than one slice from the spectra may need to be compared in order to identify the side chain carbons. Peaks that are in both spectra are internal to the side chain of that residue. Those that are only the  $^{13}\text{C}$ -NOESY are peaks from other side chains that are within  $5\text{\AA}$  of the side chain. These inter-residue peaks provide some of the distance constraints for structure calculation.

### **4.3.3 TALOS output**

TALOS (Cornilescu *et al.*, 1999) predicts protein backbone dihedral torsion angles from NMR chemical shifts. This is done to help to ensure correct residue identification and to provide torsion angle restraints for structural calculation. To do this, TALOS takes the experimentally derived chemical shifts for the residues in the protein and compares them to similar chemical shifts in its database. It then compares the  $\phi$  and  $\psi$  angles of the shifts in the database, as discussed in materials and methods (section 2.18.5.1).

TALOS defines residues as being good, bad or ambiguous. The results of this analysis for ABD2 can be found below (figure 4.18). In this figure, residues that are within tolerance, and are called ‘good’ by TALOS are in green. Those which are uncertain according to TALOS are in yellow. The residues in grey are those for which TALOS did not have enough data to calculate a result.

G1	A2	M3	A4	R5	A6	E7	E8	H9	I10
Y11	R12	E13	I14	M15	E16	L17	C18	F19	V20
I21	R22	T23	M24	A25	R26	H27	R28	R29	D30
G31	K32	I33	Q34	V35	T36	F37	G38	E39	L40
F41	D42	R43	Y44	V45	R46	I47	S48	D49	K50
V51	V52	G53	I54	L55	M56	R57	A58	R59	K60
H61	G62	L63	V64	H65	F66	E67	G68	E69	M70
L71	W72	Q73	G74	K75	D76	D77	H78	V79	V80
I81	T82	L83	L84	E85					

Figure 4.18 – TALOS output for ABD2. Numbering is internal of ABD2 with G1-A4 being his-tag artefacts, R5 being R294 in ms1/STARS.

This output can also be plotted for each individual residue, to show where the ten best matches cluster in a Ramachandran plot. An example of this, for residue E306, is shown below (figure 4.19). The green squares represent the ten best matches in the TALOS database for that residue. They are close together in an island on the Ramachandran plot, showing the experimentally derived value for E306 is good according to TALOS.

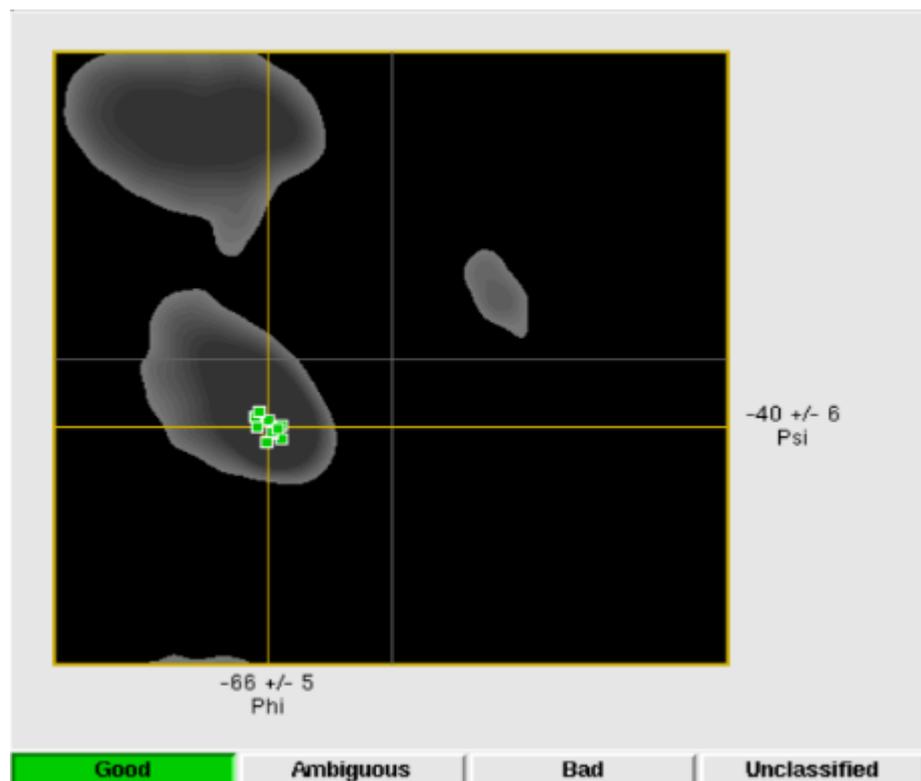


Figure 4.19 – TALOS output for ABD2 residue E306, showing the spread of the results in TALOS. Tight clustering shows that the values from TALOS were good.

As can be seen in figure 4.18, TALOS did not consider any of the assigned chemical shifts to be bad. Those that it considered to be uncertain were all in areas of beta sheet structure, where chemical shifts are more variable and more likely to be marked as uncertain by TALOS, or they were glycines at the end of  $\alpha$  helices, which are also more likely to be uncertain because of steric features of glycines that are caused by their lack of a side-chain.

#### **4.3.4 Structural Calculation**

The TALOS outputs, the sequence of ABD2, the peak list for the  $^{15}\text{N}$ -NOESY, the chemical shift list, the  $^{13}\text{C}$ -NOESY and the unlabelled homonuclear NOESY (performed on a 200  $\mu\text{M}$  sample of ABD2) were all used as inputs for CYANA (Güntert *et al.*, 1997).

CYANA calculates the most likely residue for a cross-peak to belong to, and repeats this for 7 cycles, taking the top twenty from the previous cycle as a guide. At the end of each cycle it produces a list of the cross-peaks, whether they have been discarded or not, and an overall summary of the process. Table 4.1 is the summary produced after the last cycle.

Certain conditions need to be met by the structure in order for it to be valid. Whether these conditions are met can be judged using data found in the output tables (see table 4.1).

Table 4.1

Peaks:		
selected	:	3650
in cnoenew.peaks	:	2338
in nnoenew.peaks	:	1238
in mslconlunoesy.peaks	:	74
assigned	:	2993
unassigned	:	657
in cnoenew.peaks	:	366
in nnoenew.peaks	:	286
in mslconlunoesy.peaks	:	5
with diagonal assignment	:	131
Cross peaks:		
with off-diagonal assignment	:	2862
with unique assignment	:	2258
with short-range assignment $ i-j  \leq 1$ :		1856
with medium-range assignment $1 <  i-j  < 5$ :		368
with long-range assignment $ i-j  \geq 5$ :		638

These conditions were met

1. The average target function for cycle 1 is below 250Å
2. The average target function for cycle 7 is below 10Å
3. There are less than 20% unassigned NOEs.
4. RMSD for cycle 1 is below 3Å
5. The change in the RMSD between cycles 1 and 7 is below 3Å

Condition 1 was met; the average target function after cycle 1 was 201.85Å. Condition 2 was also met, with the average target function after cycle 7 being 1.57Å. The percentage of unassigned NOEs is 18%, which is under the 20% threshold and means condition 3 is met. The average backbone RMSD for cycle 1 is 2.21Å; this is below 3Å so condition 4 is met. The RMSD between cycle 1 and cycle 7 is 1.67Å and therefore condition 5 is also met. The calculated structure is therefore a valid interpretation of the restraints provided.

These results were then used as the input files for REDAC (Güntert & Wüthrich, 1991) in order to further refine the structure, particularly with regard to the phi and psi angles. This resulted in the twenty lowest energy structures for ABD2, seen in grey in figure 4.20, and the lowest energy structure for ABD2, seen in figure 4.21. The lowest energy structure, in green, has been labelled with its major structural features. The sequence specific assignments for ABD2 can be found in appendix 5.

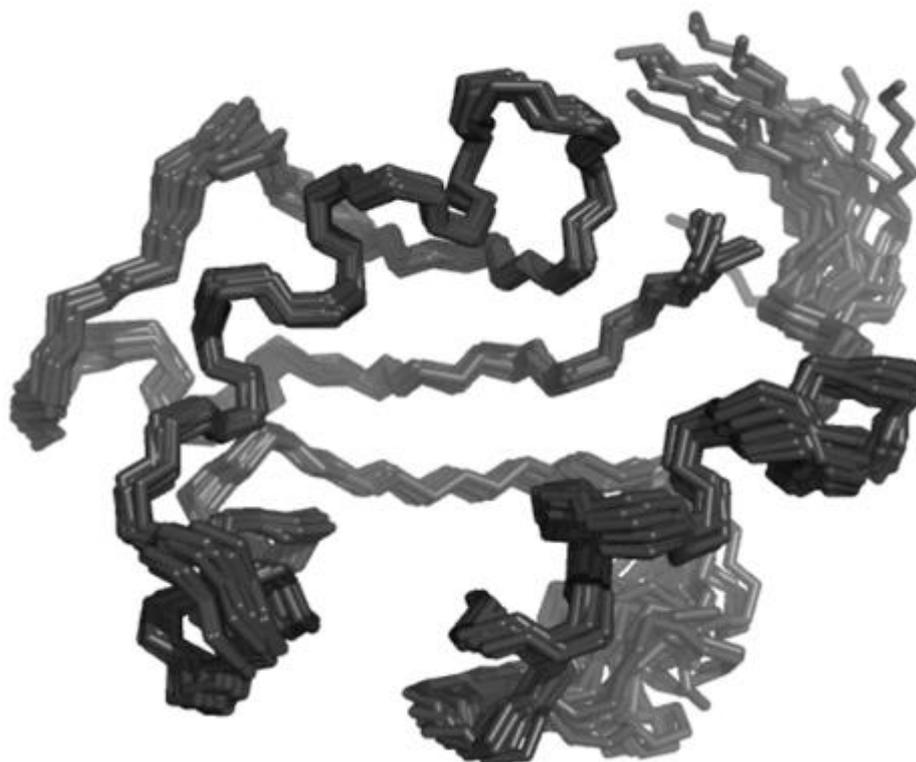


Figure 4.20 – The top 20 lowest energy structures of ABD2 produced by the structure calculation.

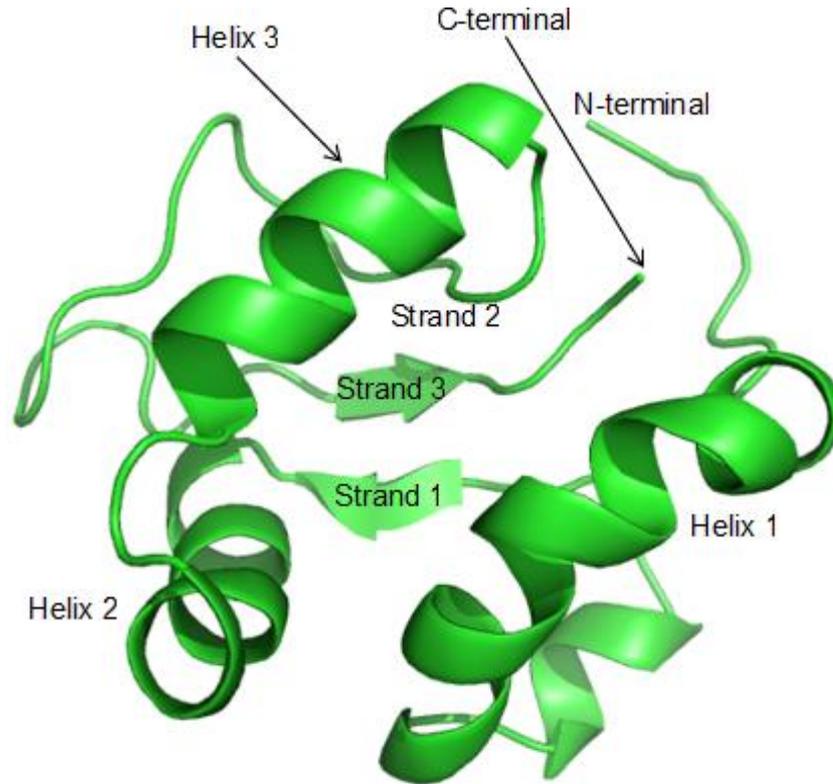


Figure 4.21 – The lowest free-energy structure of ABD2 from the structural calculation.

### **4.3.5 The Structure of ABD2**

ABD2, as seen in figure 4.21, is a mixed alpha-helical and beta-sheet protein. It consists of 3 alpha-helices and 3 beta-strands, which form 1 anti-parallel beta sheet. It contains two flexible regions, one from residue 294 to the start of the first helix and the second is a loop from residue 357 to 370. Flexible regions have faster dynamics and therefore relax more quickly. This means that any NMR peaks from them are broader, and sometimes they become so broad that they do not appear at all because they have been lowered to the same level as the noise in the spectrum. This explains the lack of information for these regions, including the lack of peaks for residues Y301 and R336 in the  $^{15}\text{N}$ -HSQC. There are also fewer cross-peaks in this region in the  $^{15}\text{N}$ -NOESY and the  $^{13}\text{C}$ -NOESY. This means there were fewer restraints in these areas, meaning

that there was a greater range of space they could inhabit in the structural calculations. This is why these regions vary more when the structure of ABD2 is calculated, and do not overlap as well on top of each other as the rest of ABD2 in the top 20 lowest energy structures seen in figure 4.20.

#### **4.3.6 Structure Validation by Residual Dipolar Coupling (RDC)**

There was an attempt to use RDCs as a separate method of validating the structure of ABD2, in a manner analogous to the  $R_{\text{free}}$  used by X-ray crystallographers when validating structures determined by X-ray crystallography. Unfortunately, results could only be obtained using one method of producing residual dipolar coupling, that method being alignment by filamentous phage (the aligned sample of ABD2 was at 120  $\mu\text{M}$ ). The results of this alignment can be seen in figure 4.22. The red bars indicate the residues of ABD2 for which there is RDC data. Their length indicates the shift of the HSQC peak caused by the alignment medium. For there to be confidence in RDCs, results from more than one alignment medium are required. The measured RDCs were compared to values back calculated from the structure of ABD2 using MODULE (Dosset *et al.*, 2001). The difference between the back calculated values and the experimental values are shown in figure 4.23. The experimental values are shown in orange while the back calculated values are shown in blue. There is a large difference between them, showing that the RDC values cannot be used to help to refine the structure as we cannot be sure that they are correct.

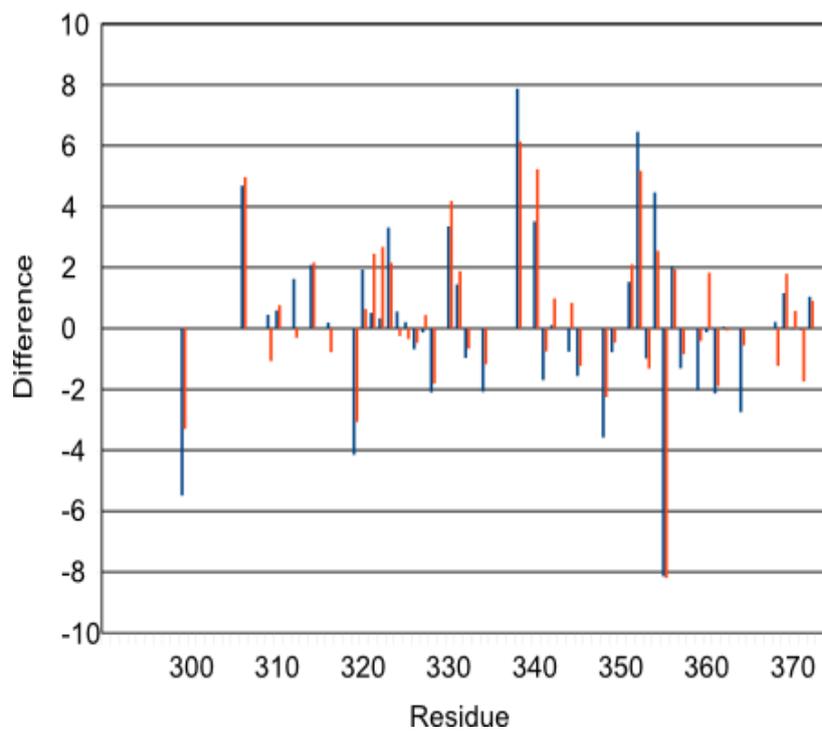


Figure 4.22 – Results of RDC alignment of ABD2 in filamentous phage.

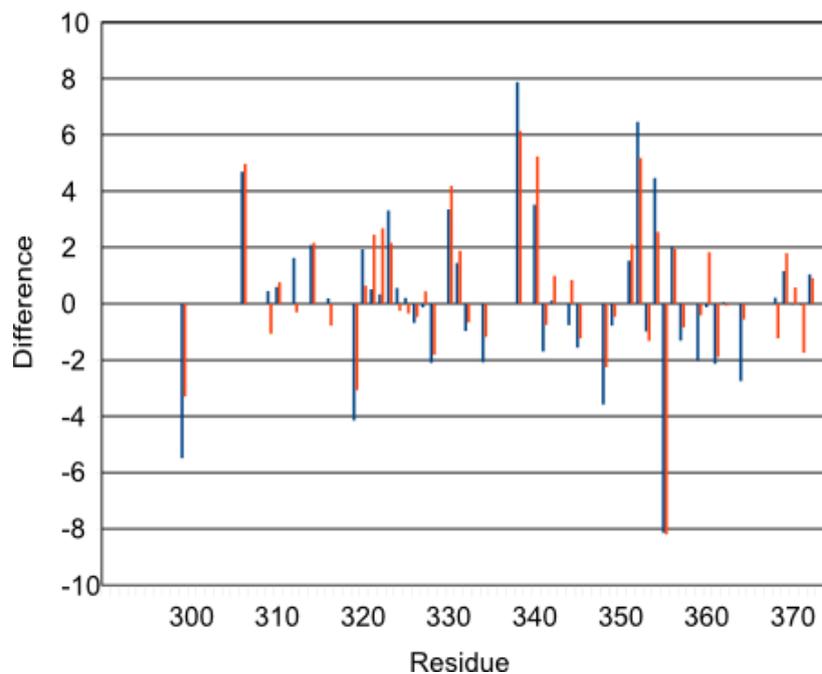


Figure 4.23 – Results of RDC alignment of ABD2 in filamentous phage (blue) and the difference between them and the back-calculated RDC values (orange).

### **4.3.7 PROCHECK Validation**

As discussed in section 2.21.5.4, PROCHECK was used to validate the structure produced. The results are presented in figure 4.24.

Part A of figure 4.24 shows the quality assessment of the Ramachandran plots produced for the residues in ABD2 in the 20 lowest free-energy structures calculated for it. The yellow area indicates disallowed values, while the purple area indicates allowed values. All of the 20 lowest free-energy structures have a high percentage of residues in the most favoured regions of the Ramachandran plot. These percentage values are within the allowed area of values in terms of percentage in the most favoured region versus resolution according to PROCHECK. This validates the structure in terms of psi and phi steric hindrance.

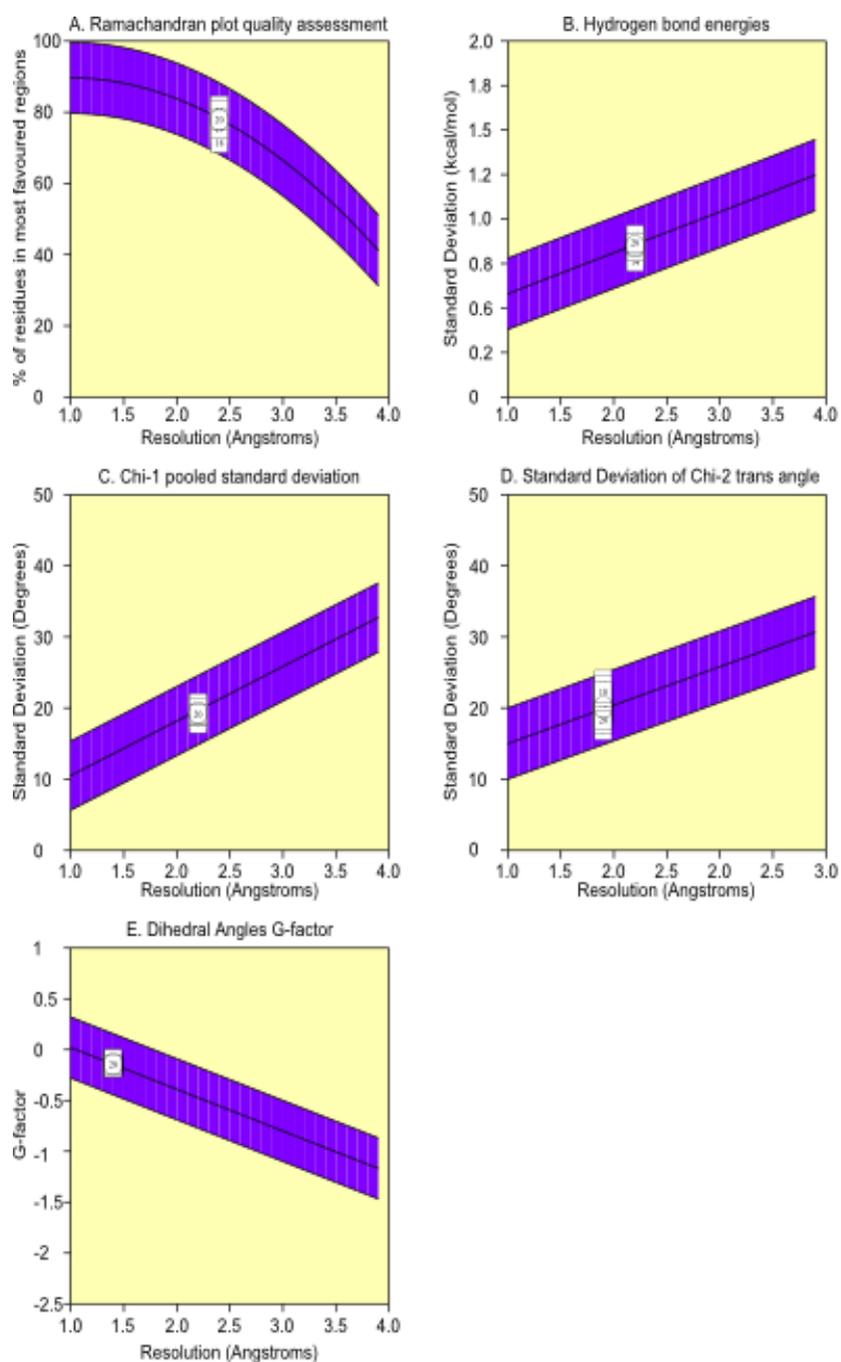
Part B of figure 4.24 shows the results of the quality assessment of the hydrogen bond energies in the 20 lowest free-energy structures calculated for ABD2. The yellow area indicates disallowed values for the hydrogen bond energies, while the purple area indicates allowed values for the hydrogen bond energies. All twenty lowest free-energy structures are within the allowed values for hydrogen bond energies. This validates the structure in terms of the hydrogen bond energies within it.

Part C of figure 4.24 shows the results of the quality assessment of the Chi-1 angled pooled deviation in the 20 lowest free-energy structures calculated for ABD2. The yellow area is the disallowed deviation, and the purple area is the allowed deviation. All twenty lowest free-energy structures are within the allowed deviation for Chi-1

angles in correct structures. This validates the structure in terms of the Chi-1 angles it contains.

Part D of figure 4.24 shows the results of the quality assessment of the Chi-2 trans angles in the 20 lowest free-energy structures calculated for ABD2. The yellow area is the disallowed region, and the purple area is the allowed deviation. All twenty lowest free-energy structures are within the allowed deviation for Chi-2 trans angles in correct structures. This validates the structure in terms of the Chi-2 trans angles it contains.

Part E of figure 4.24 shows the results of the quality assessment of the G-factor of the dihedral angles in the 20 lowest free-energy calculated for ABD2. The yellow area is the disallowed region, and the purple area is the allowed region. All twenty lowest free-energy structures are within the allowed G-factor for the dihedral angles. This validates the structure in terms of the dihedral angles it contains. The information that contributed to the plots can be seen in the table in figure 4.24 titled 'Plot statistics'.



#### Plot statistics

Stereochemical parameter	No. of data pts	Mean parameter value	Comparison values		No. of band widths from mean
			Typical value	Band width	
a. %-tage residues in A, B, L	20	77.9	37.8	10.0	-
b. H-bond energy st dev	20	0.8	1.3	0.2	-
c. Chi-1 pooled st dev	20	19.3	33.5	4.8	-
d. Chi-2 trans st dev	20	20.1	31.2	5.0	-

Figure 4.24 – PROCHECK assessment of the calculated ABD2 structure.

### **4.3.8 PDB Structure Submission**

The family of 20 lowest free-energy structures produced by the structure calculation was submitted to the PDB and run through the ADIT validation server. The ADIT validation server checks deviations from standards for close contacts, bond length, bond angle and torsion angle. It also looks for incorrectly positioned waters, chirality errors, sequence mismatches and missing or extra atoms or residues compared to what the submission claims to contain. The structure of ABD2 showed none of these and was submitted as PDB submission 2KRH.

### **4.3.9 Structure Comparison**

Once the structure was determined, it was compared to other structures in the PDB using DALI. The top 20 results can be seen below in table 4.2. The top 100 results can be found in appendix 6.

Table 4.2

Result	PDB ID	Z-score	rmsd	lali	No. Res.	% ID	Description
1	2hr3-b	4.7	2.7	55	135	16	Probable transcriptional regulator
2	1bi2-b	4.7	3.1	59	138	12	Diphtheria toxin transcription repressor
3	1q1h-a	4.6	3	64	85	9	Transcription factor E/II $\epsilon$ alpha
4	2p7c-b	4.6	3.7	62	82	19	BlaI monomeric form in complex with the blaP half-operator.
5	1g3y-a	4.6	3.2	59	212	12	Diphtheria toxin repressor
6	1bi2-a	4.6	3.1	58	214	12	Diphtheria toxin repressor
7	1bi3-b	4.6	3	58	137	12	Diphtheria toxin repressor
8	1bi3-a	4.6	3	57	212	12	Diphtheria toxin repressor
9	1g3t-a	4.6	3.2	59	215	12	Diphtheria toxin repressor
10	1fwz-a	4.6	3.1	59	211	14	Diphtheria toxin repressor
11	1pr6-a	4.5	3.1	58	82	19	Penicillinase repressor
12	1g3t-b	4.5	3.2	59	138	12	Diphtheria toxin repressor
13	2dtr-a	4.5	3.1	58	137	12	Diphtheria toxin repressor
14	2dtr	4.5	3.1	58	137	12	Diphtheria toxin repressor
15	3c18-c	4.4	3.5	60	290	15	Nucleotidyltransferase-like protein
16	1sax-b	4.4	3.2	58	120	10	Methicillin-resistance regulating transcriptional repressor meci in complex with 25-bp ds-DNA
17	1w7p-a	4.4	2.9	59	213	7	Endosomal complex ESCRT-II
18	1sd4-b	4.4	3.1	58	122	9	Penicillinase repressor
19	3bz6-a	4.3	3.9	64	168	9	Protein of unknown function from <i>Pseudomonas syringae</i>
20	2hr3-d	4.3	2.9	57	140	16	Probable transcription regulator

Of the top 500 results, all with a z score above 2 – z scores below 2 indicate spurious results - only one result was a known actin binding protein. The actin binding protein that was found was result 407 (see appendix 6). This protein is pdb entry 1sjj, a cryo-electron microscopy reconstruction of the structure domains which are part of the

chicken gizzard smooth muscle alpha-actinin (Liu *et al.*, 2004). The region of similarity is the closed conformation actin binding calponin homology domain. This region is entirely  $\alpha$ -helical, and contains no  $\beta$ -sheet structure, which suggests that at this z-score level the compared proteins are not very similar.

Of the top ten most similar proteins, all of those that had a known function were DNA binding proteins, and their structures were described as winged helix-turn-helix motifs. A comparison of six of these structures and ABD2 can be seen below (figure 4.25).

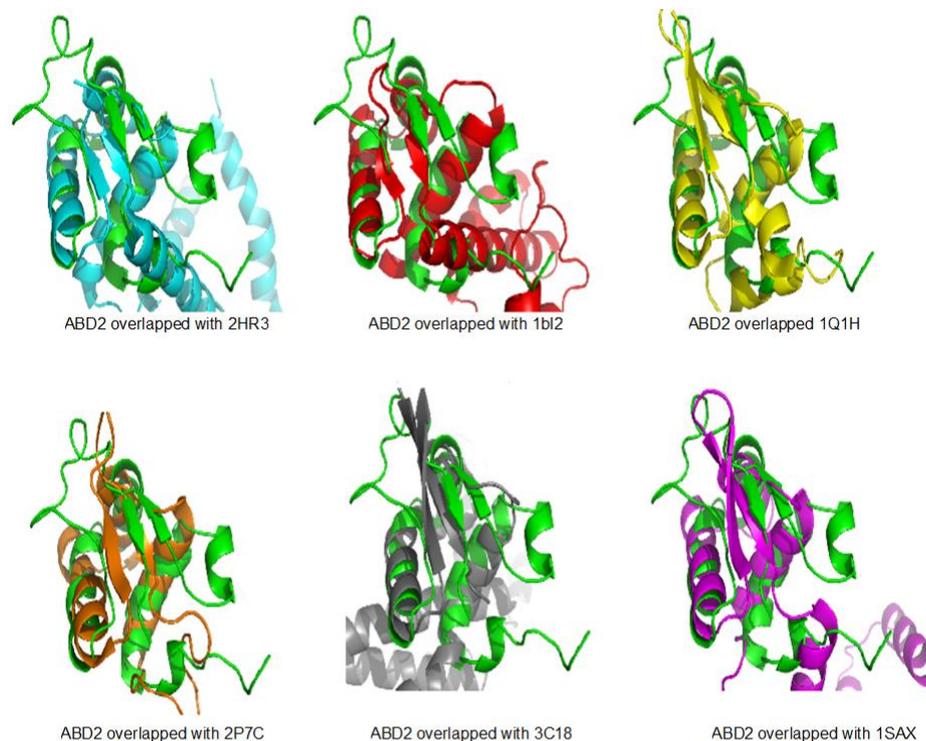


Figure 4.25 – ABD2 overlapped with the 6 most structurally similar proteins in the PDB, not including those where the structures have the same function as seen in table 3.2. The codes given to them are their entries in the PDB. 2HR3 = domain B of putative transcriptional regulator protein from *Pseudomonas aeruginosa* PA01, 1BI2 = domain B of diphtheria toxin repressor, 1Q1H = general transcription factor E/IIE alpha, 2P7C = domain B of BlaI monomeric form in complex with the blaP half-operator, 3C18 = domain C of nucleotidyltransferase-like protein and 1SAX = domain B of Methicillin-resistance regulating transcriptional repressor MecI.

#### **4.3.9.1 Winged Helix-Turn-Helix Domains**

After the structure of ABD2 had been calculated, it was compared to the other structures in the protein data bank using DALI (Holm *et al.*, 2008). DALI compares proteins by transforming their 3D structures into 2D matrices and then overlaying these.

Of the top 500 DALI results, only one was an actin-binding protein. This was result 470, 1sjj, the cyro-EM structure of alpha-actinin from chicken gizzard. The region of alpha-actinin which shows similarity is the CH domain. This is a helical domain, which has no beta-strand structure, which suggests that at this low level of results, any similarly sized helical protein will be picked up.

The top ten results were all winged helix-turn-helix domains.

Winged helix-turn-helix domains are a subset of the helix-turn-helix domain family. Helix-turn-helix domains are commonly found in transcription factors (Aravind *et al.*, 2005). The basic helix-turn-helix domain is a three helix core, which binds to DNA with its third helix (Aravind *et al.*, 2005). Along with DNA binding function, helix-turn-helix domains have been found to have many other functions including mediating protein-protein interactions (Wong *et al.*, 2000) or as part of larger enzymes (Zheng *et al.*, 2002).

The major distinguishing feature of a winged helix-turn-helix protein is the presence of a C-terminal  $\beta$ -strand hairpin unit (the wing in the name of the protein family) (Brennan, 1993). This wing sometimes packs into the shallow cleft in the open core of the domain (Clark *et al.*, 1993); however this packing is absent in ABD2.

The first winged helix-turn-helix domains to be fully investigated were Forkhead (Weigel *et al.*, 1989) and HNF-3 (Hepatocyte nuclear factor-3) (Lai *et al.*, 1990 and Lai *et al.*, 1991). They gave rise to the description of the typical winged helix-turn-helix motif (Kaufmann & Knöchel, 1996).

In a typical winged helix-turn-helix motif containing domain (as described in figure 4.26), the N-terminal contains three  $\alpha$ -helices (H1, H2 and H3) and three  $\beta$ -strands (S1, S2 and S3). In ABD2 these are found at the following residues, H1 = R302 to M314, H2 = F327 to S338, H3 = V341 to H351. Between H2 and H3, there is a short loop. The first  $\beta$ -strand is located between H1 and H2 (in ABD2 this is I323 to F327), while the remaining  $\beta$ -strands are found after the third helix. In ABD2 S2 = V354 to E357 and S3 = V370 to L372. S2 and S3 are separated by a loop which forms the wing. The three strands form a three-stranded anti-parallel  $\beta$ -sheet. In some winged helix-turn-helix domains, a second wing is found after the third  $\beta$ -strand, but this is not present in ABD2. This is not because of any artificially chosen domain boundaries, but is instead because the full length ms1/STARS protein ends at residue 375, which is the last residue of S3. The second wing of the winged helix-turn-helix domain is also not present in other, otherwise typical, winged helix-turn-helix motifs such as in E2F4, DP2, histone H5 and MarR (Brennan, 1993, Zheng *et al.*, 1999 and Alekshun *et al.*, 2001). The function of the second wing is to help in DNA recognition (Clark *et al.*, 1993)

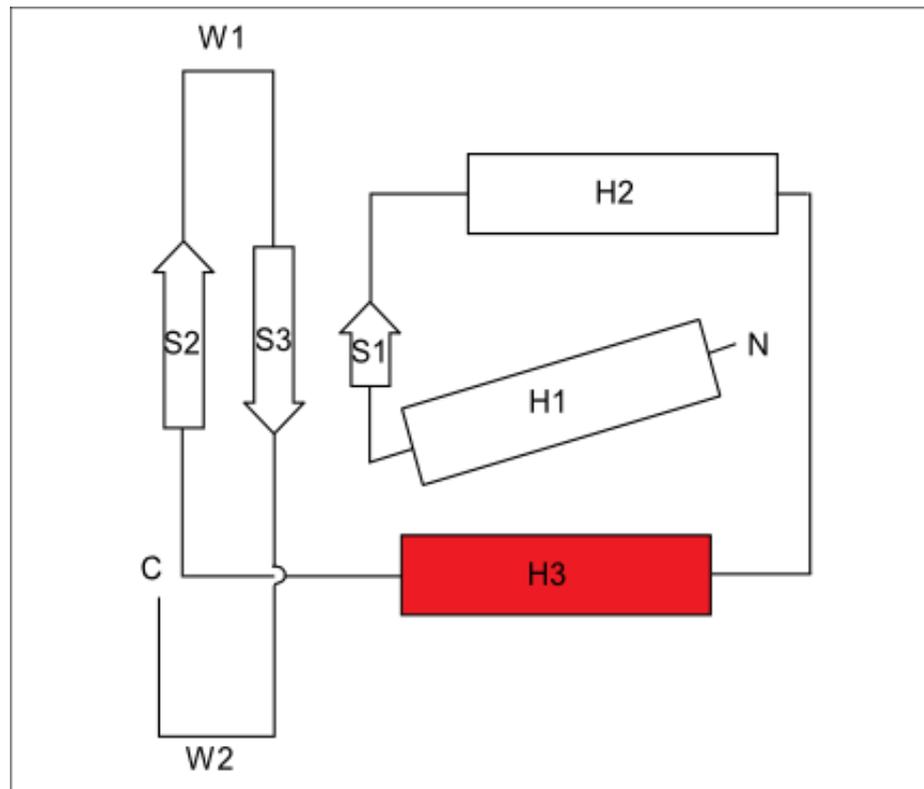


Figure 4.26 – Diagrammatic representation of the topology of the typical winged helix-turn-helix domain. Taken from “Winged Helix Proteins” Gajiwala K.S. & Burley S.K. (2000) *Curr. Opin. Struct. Biol.* 10, pp 110-116

Except for the wing, another major difference between winged helix-turn-helix domains and helix-turn-helix domains is the greater variety in the angle between helix-2 and helix-3 found in winged helix-turn-helix domains. In helix-turn-helix domains, the average angle between these two helices is  $120^\circ$  (Brennan *et al.*, 1986) while in winged helix-turn-helix domains it can vary from  $100^\circ$ , in biotin operator repressor protein BirA (Wilson *et al.*, 1992), to  $150^\circ$ , in transcription factor DB2 (Zheng *et al.*, 1999). In ABD2 the angle is  $116^\circ$ .

There are many variants within the winged helix-turn-helix domain family. In those examples with three  $\beta$ -strands, the loop between helix-1 and helix-2 can assume an extended configuration, and becomes the third strand, through hydrogen bonds with the C-terminal hairpin, while in the four-stranded version of the winged helix-turn-helix

motif it is the linker between helix-1 and helix-2 that forms the third strand while the C-terminal wing also extends to form a  $\beta$ -strand. These four strands then become an extended  $\beta$ -sheet.

Two- and three-stranded winged helix-turn-helix domains are used as DNA-binding domains by both prokaryotic and eukaryotic transcription factors. Binding is mediated by the third helix, highlighted in red in figure 4.26. In winged helix-turn-helix domains that bind nucleic acids, the wing provides an extra surface for interaction with the nucleic acids, normally with the minor groove of the DNA. The extra surface is produced by charged residues in the hairpin (Brennan, 1993, Clark *et al.*, 1993 and Swindells, 1995). Most winged helix-turn-helix domains bind to DNA using a large basic face. From the actin binding site mutation studies, this is the face where ABD2 binds to actin. How much the wing contributes to DNA-binding varies between proteins. In HNF3 $\gamma$ , the W1 takes no part in DNA binding (Brennan, 1993), in Foxo3a, W1's function is limited to making unspecific contacts with the phosphate backbone of the DNA it binds to (Tsai *et al.*, 2007), while in RFX1 the majority of the DNA contacts are made by W1, with only one contact being made by H3 (Gajiwala *et al.*, 2000). This suggests that ABD2 binds in a similar manner to RFX1, which binds to DNA in a different manner to the way HNF3 $\gamma$ /FKH forkhead domains bind to DNA. The difference is that in HNF3 $\gamma$ /FKH type domains, H3 fits into the major groove of the DNA while in RFX1-type domains H3 only interacts with the minor groove of the DNA, while the wing interacts with the major groove (Gajiwala *et al.*, 2000). It is suggested that SmtB (Cook *et al.*, 1998) and GH5 (Ramakrishnan *et al.*, 1993) also bind to DNA in a similar way (Gajiwala & Burley, 2000), as binding in both types occurs at

the most basic face. In HNF3 $\gamma$ /FKH type domains, this is H3; in RFX1, SmtB and GH5, this is W1.

ABD2 is not alone in being a winged helix-turn-helix domain that also has a non-DNA binding function. Another example is the version found on the N-termini of methionine aminopeptidases (Liu *et al.*, 1998). The PINT domain, which forms a scaffold for the lid of the proteasome, eIF3 and the signalsome, is another protein-protein interaction winged helix-turn-helix domain, but, unlike ABD2, it has lost its DNA binding activity (Hoffmann & Butcher, 1998 and Aravind & Ponting, 1998).

In many of the proteins that contain a winged helix-turn-helix domain, the winged helix-turn-helix domain exists as a separate domain within the protein, in the same way that ABD2 is independently folded and functional within ms1/STARS.

Many, but not all, helix-turn-helix and winged helix-turn-helix domains are flanked by dimerization domains or act as dimers to fulfil their function. Many of the proteins shown to be structurally similar to ABD2 act as dimers when they bind to DNA.

The dimerisation status of ABD2 had previously been determined in order to correctly calculate the structure, as structural calculation software must have this data in order to correctly interpret the restraints given to it. The data, analytical gel filtration and NMR dynamics experiments, pointed towards ABD2 being a monomer (section 4.1.8).

Monomeric winged helix-turn-helix domains include HNF3 $\gamma$ , one of the founder members of the winged helix-turn-helix domain family (Clark *et al.*, 1993). Other

monomeric winged helix-turn-helix domains include all the members of the Fox family except the FoxP subgroup (Wijchers *et al.*, 2006). In order to bind to DNA, winged helix-turn-helix domains require the second wing; these second wings are often rich in positively charged residues as the C-terminal of ABD1 is. This could mean that ABD1 acts as the second wing.

#### **4.4 Actin Binding of ABD2**

An actin co-sedimentation assay showed that ABD2 bound to actin (figure 4.27). ABD2, which remains in the supernatant when centrifuged in the absence of actin, is found in the pellet in the presence of actin. The results of this experiment can be seen in SDS-PAGE gel shown in figure 4.27. In this gel, S and P refer to the supernatant and pellet fractions after centrifugation. This notation will be used for all subsequent actin co-sedimentation experiment results. In lanes 1-5 all proteins were at 50  $\mu$ M. The actin binding region of calponin, residues 131-228, was used as a positive control in this experiment, as calponin is a known actin-binding protein. When centrifuged in the absence of actin, calponin 131-228 remains in the supernatant (lane 2). In the presence of actin, some calponin 131-228 is found in the pellet after centrifugation (lane 5). ABD2 behaves in the same way as calponin 131-228. In the absence of actin (lane 1), it remains in the supernatant when centrifuged, but it moves into the pellet when centrifuged with actin (lane 4). Lane 6 contains 70  $\mu$ M ABD2 and 50  $\mu$ M actin. This experiment was performed to see if the actin binding effect was reproducible at different ABD2 concentrations. ABD2 once again moves into the pellet in the presence of actin, but the amount that moves into the pellet does not appear to increase, although this method is not quantitative.

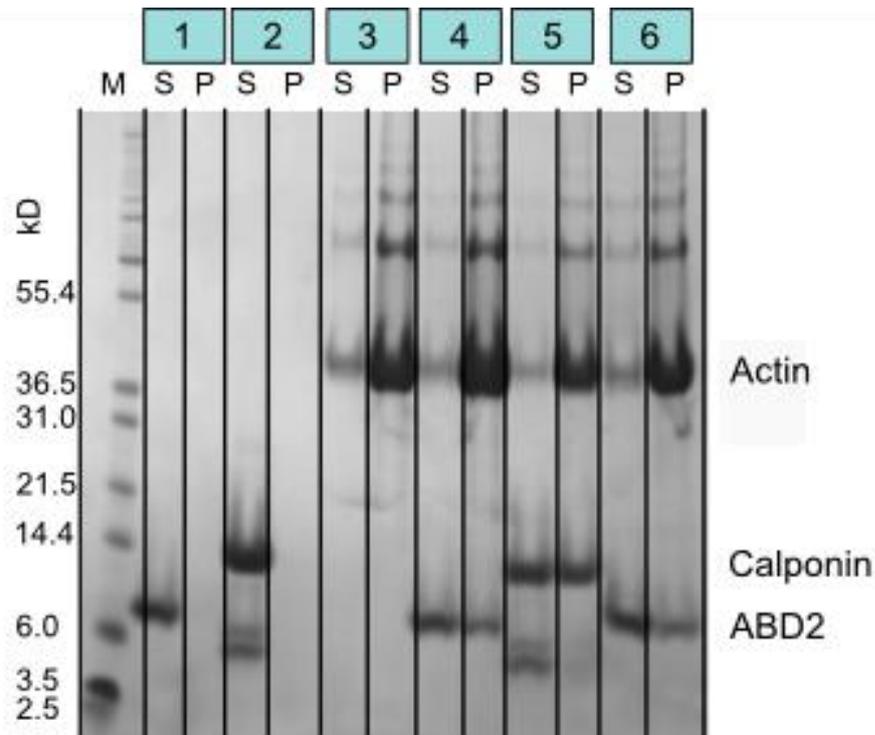


Figure 4.27 – S = supernatant and P = pellet after centrifugation. M = molecular weight marker. Lane 1 contains ABD2 (50  $\mu$ M) alone. Lane 2 contains calponin 131-228 (50  $\mu$ M) alone. Lane 3 contains actin (50  $\mu$ M) alone. Lane 4 contains ABD2 (50  $\mu$ M) and actin (50  $\mu$ M). Lane 5 contains calponin 131-228 (50  $\mu$ M) and actin (50  $\mu$ M). Lane 6 contains ABD2 (70  $\mu$ M) and actin (50  $\mu$ M). The actin binding assay shows that ABD2 behaves in the same way as a known actin binding protein when assayed for actin binding activity. This suggests that ABD2 is an actin binding domain.

An assay of fluorescence using pyrene-labelled actin was used to quantify the binding of ABD2 to actin. When the addition of further ABD2 did not change the fluorescence reading, the binding of actin was assumed to be saturated. This value was taken to be 100% saturation, and the other fluorescence readings were expressed as a percentage of this saturation. The percentage saturation was plotted against the amount of ABD2 added. The values from this graph were interpreted using a binding fit macro written for Mathematica by Dr. M. Pfuhl. The value for the  $K_D$  is iteratively improved until it fits the experimental data.

The graph of the data is shown in figure 4.28. The box below the graph shows the residual of the difference between the experimental data and the fitted data. The errors are evenly spread and are therefore unlikely to have biased the  $K_D$  value in a particular direction. The assay showed that the binding affinity of ABD2 to actin was  $\sim 11 \mu\text{M}$  (figure 4.28).

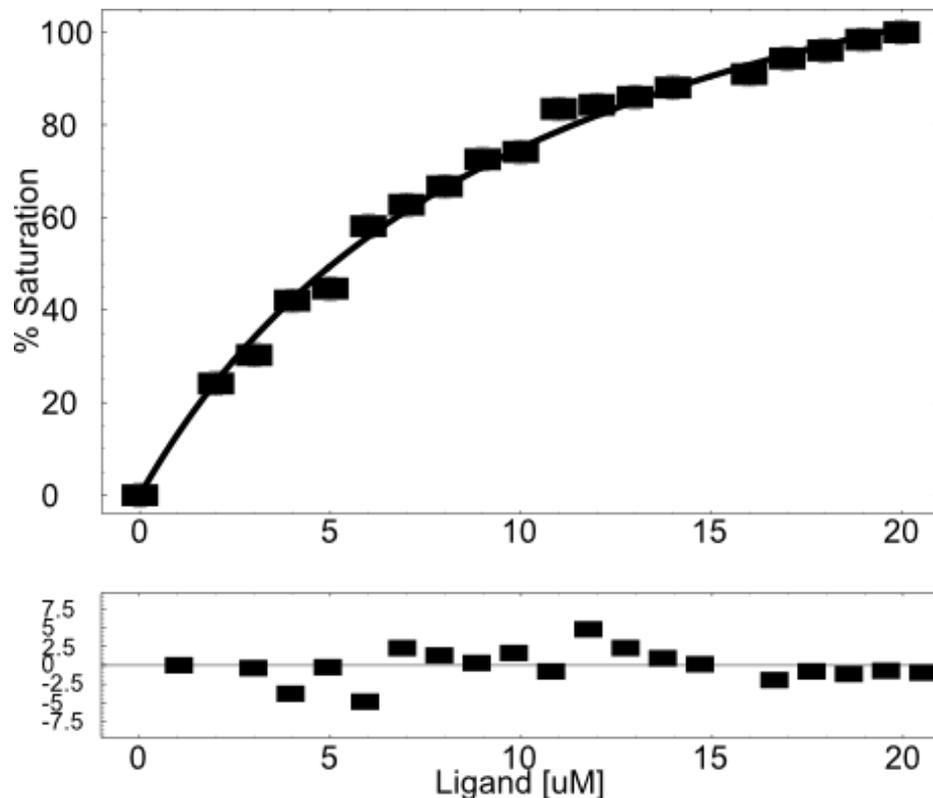


Figure 4.28 - Results of measuring actin binding of ABD2 using the fluorescence of pyrene-labelled actin. The  $k_D$  is  $10.61 \pm 0.7 \mu\text{M}$ .

#### **4.5 Characterisation of the Actin Binding Domain of ABD2**

ABD2 has been identified as an actin binding domain. We wished to identify the actin binding surface of ABD2. To do this, we used the structure that had been determined for ABD2 to provide information about where the actin binding surface could be. As actin is a negatively charged protein, binding by ABD2 will probably occur on a large positively charged surface. A representation of the surface electrostatic charge of

ABD2 (figure 4.29) shows two large positively charged patches. The positively charged patches are shown in blue. Negatively charged patches are shown in red. Areas with no charge are shown in white.

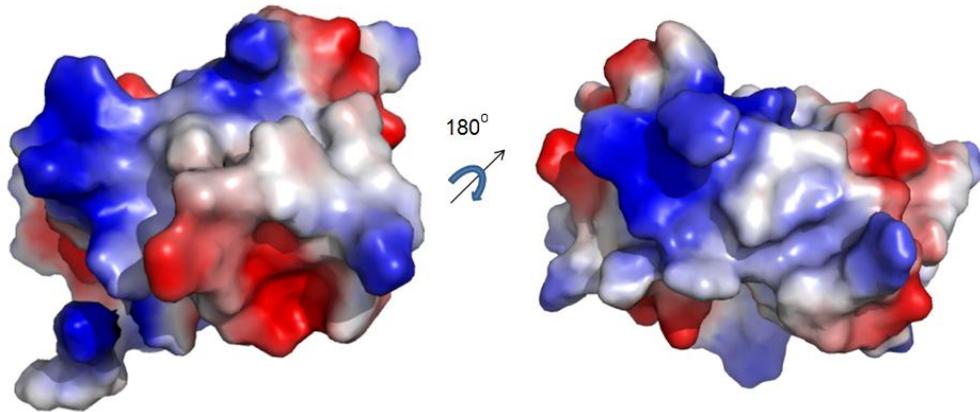


Figure 4.29 – Surface charge diagram of ABD2 showing the two large positively charged patches. The positive surfaces are in blue, negative surfaces are in red. White surface have no overall charge.

As actin binding appears to be a major function of ABD2, it would be expected that the residues involved would be conserved. Positive patches are created by the side chains of positively charged residues. These are arginine, lysine and sometimes histidine. Figure 4.30 shows the residues which are conserved in the positively charged patches. Part A shows where the conserved positively charged residues are on the structure of ABD2. Part B shows the electrostatic surface charge of ABD2. Part C is an overlay of the structure of ABD2 on the electrostatic surface charge of ABD2. It shows where the conserved positively charged residues are in terms of both the structure and the electrostatic surface charge. Part D shows the conservation of the residues of ABD2.

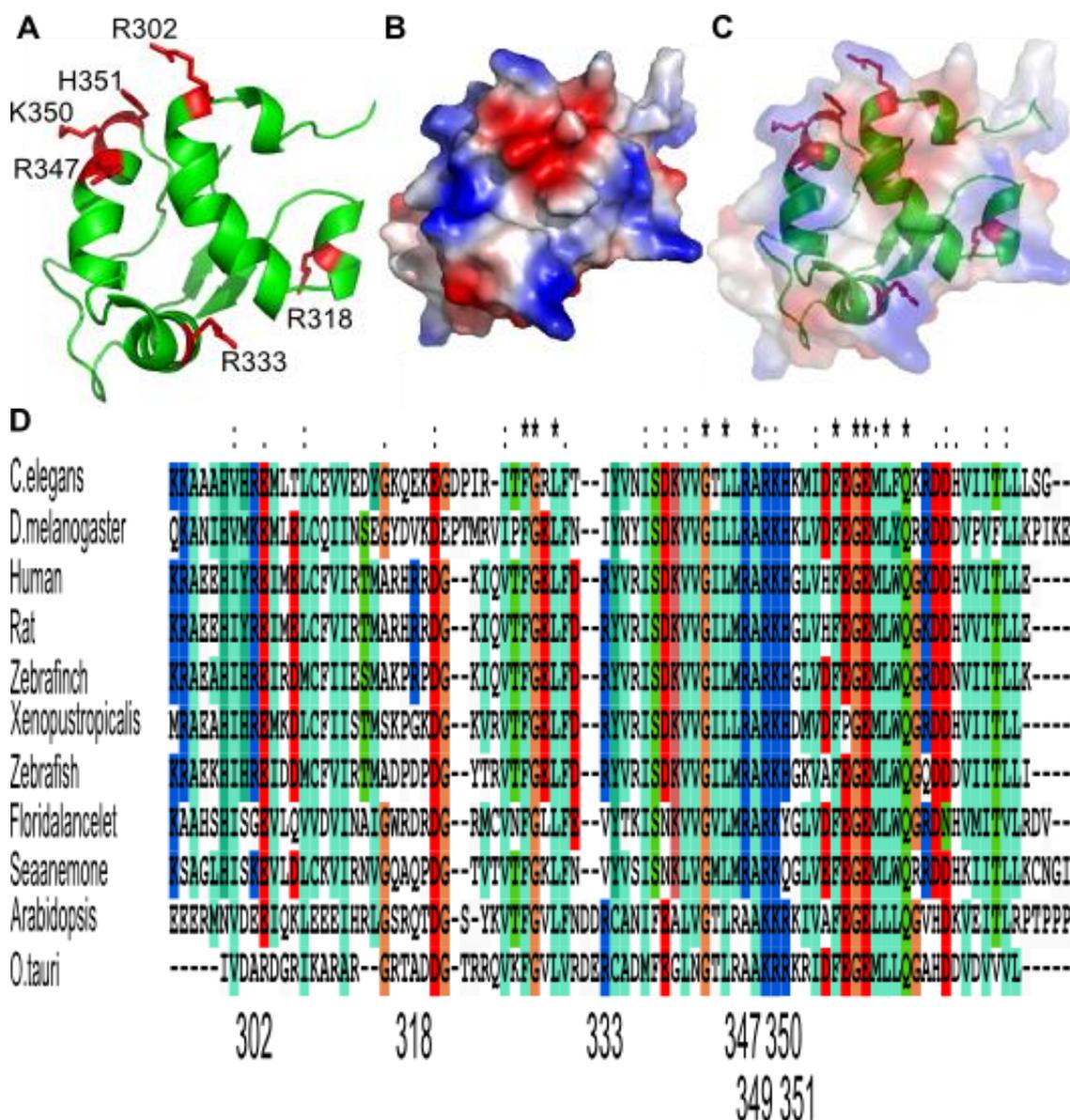


Figure 4.30 – A = Conserved positively charged residues that are also in the areas of positive charge seen in B. B = electrostatic surface charge of ABD2. C = A and B overlapped to show position of the conserved positively charged residues on the structure of ABD2 with respect to the surface charge. D = Aligned sequences of ABD2 and homologues, with conserved positively charged residues marked in blue.

If the conserved residues are important for actin binding, then if they are mutated, this will prevent the mutated ABD2 from binding to actin. The residues that were conserved and located in the positively charged patch on the surface of ABD2 (residues R302, R318, R347, R349, K350 and H351) were all separately mutated to alanines, and these mutated constructs were then expressed and characterised in terms of their folding and

their ability to bind to actin. Alanine was chosen as it is small and uncharged, and therefore, hopefully, the charge will be the only thing affected by the mutation. NMR was used to check if the mutant constructs were folded. The constructs produced by these mutations need to be folded as we believe that the actin-binding function of ABD2 is dependent on the structure of ABD2.

Figures 4.31 to 4.36 show the aliphatic region of the 1D NMR spectra of the mutant constructs. R302A (figure 4.31), R318A (figure 4.32), R347A (figure 4.33), K350A (figure 4.34) and H351A (figure 4.35) are folded. The presence of a peak below 0 ppm shows that these mutant constructs are folded, as peaks below 0 ppm indicate a methyl group which has been shifted by the presence of nearby aromatic side chains. This can only occur in a folded protein.

R349A (figure 4.36), on the other hand, is unfolded. This can be seen because there is no peak below 0 ppm in its 1D spectrum. If R349A were folded there would be methyl group side chains near aromatic side chains and this proximity would shift the methyl group NMR signal into the region below 0 ppm.

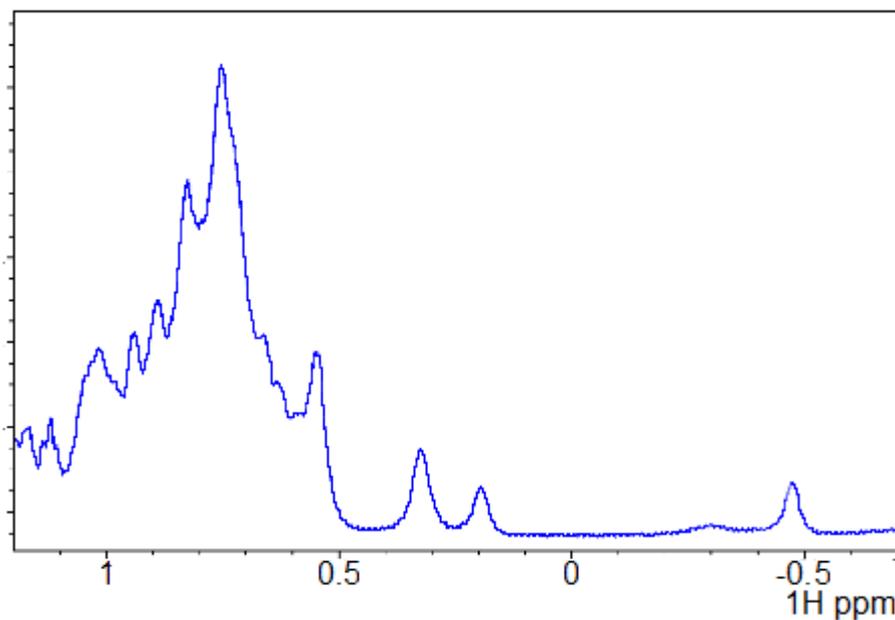


Figure 4.31 – 1D NMR spectrum of the R302A mutant, focusing on the aliphatic region. The R302A mutant is folded as can be seen by the peak at -0.45 ppm, which indicates a methyl group that has been shifted by the presence of a nearby aromatic side chain from another residue. This can only occur in a folded protein.

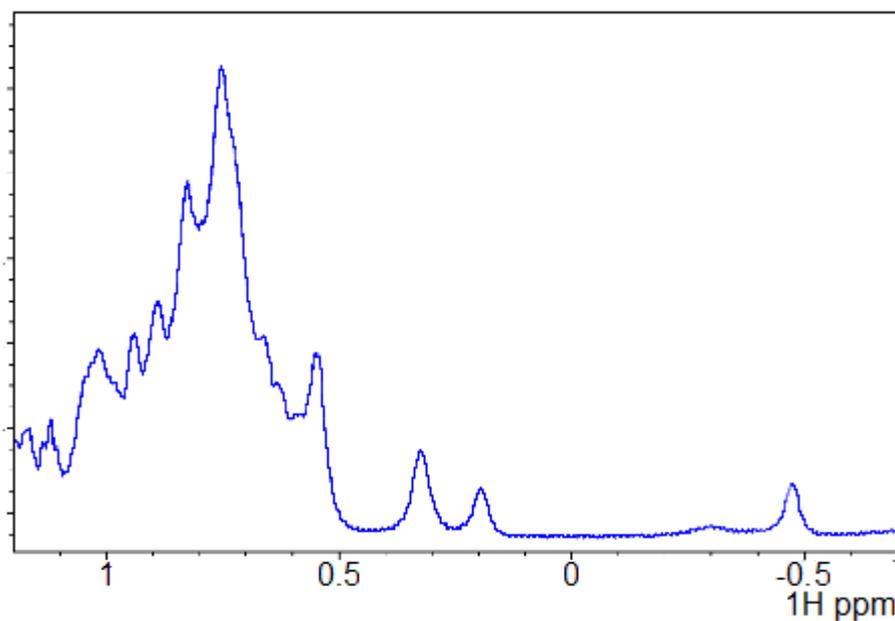


Figure 4.32 – 1D NMR spectrum of the R318A mutant, focusing on the aliphatic region. The R318A mutant is folded as can be seen by the peak at -0.46 ppm, which indicates a methyl group that has been shifted by the presence of a nearby aromatic side chain from another residue. This can only occur in a folded protein.

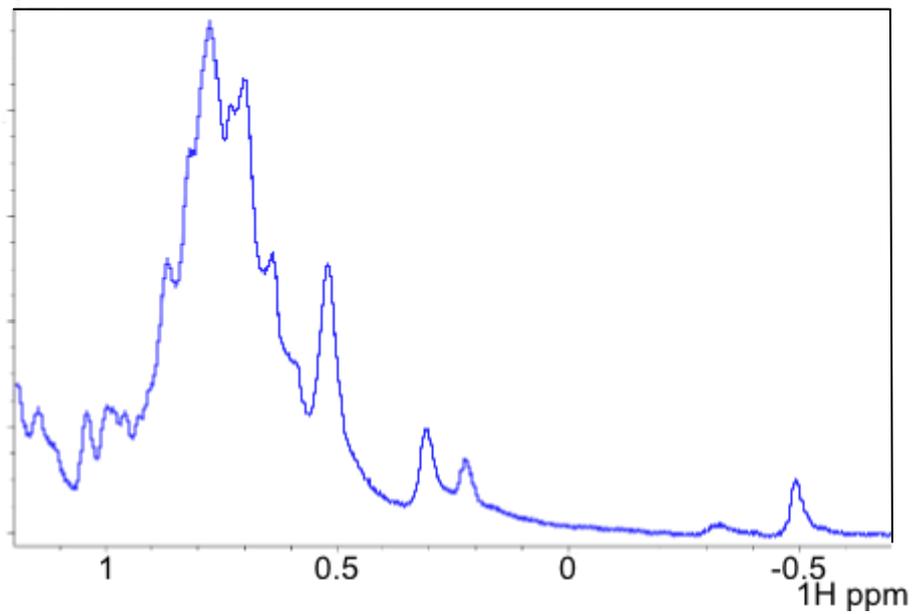


Figure 4.33 – 1D NMR spectrum of the R347A mutant, focusing on the aliphatic region. The R347A mutant is folded as can be seen by the peak at -0.5 ppm, which indicates a methyl group that has been shifted by the presence of a nearby aromatic side chain from another residue. This can only occur in a folded protein.

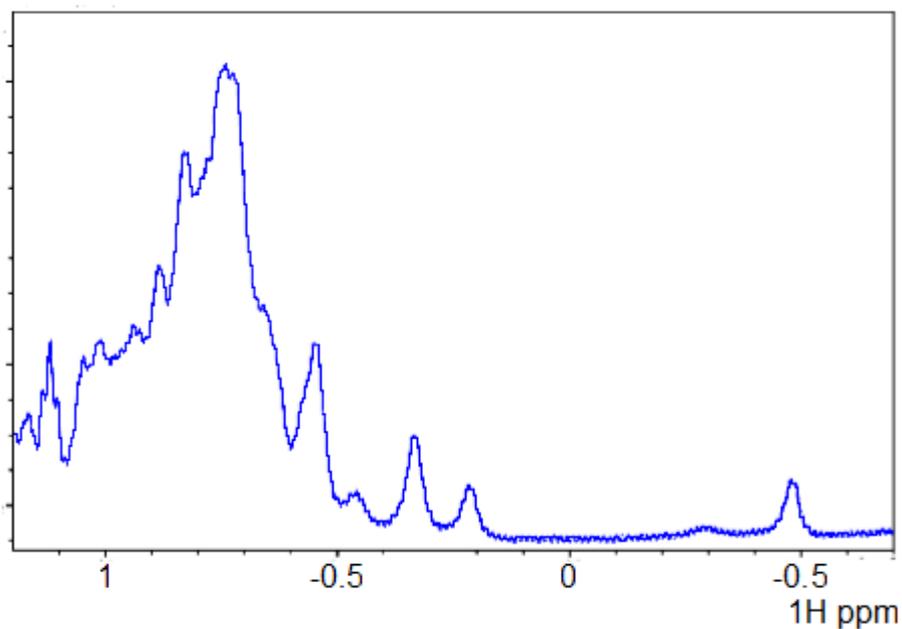


Figure 4.34 – 1D NMR spectrum of the K350A mutant, focusing on the aliphatic region. The K350A mutant is folded as can be seen by the peak at -0.47 ppm, which indicates a methyl group that has been shifted by the presence of a nearby aromatic side chain from another residue. This can only occur in a folded protein.

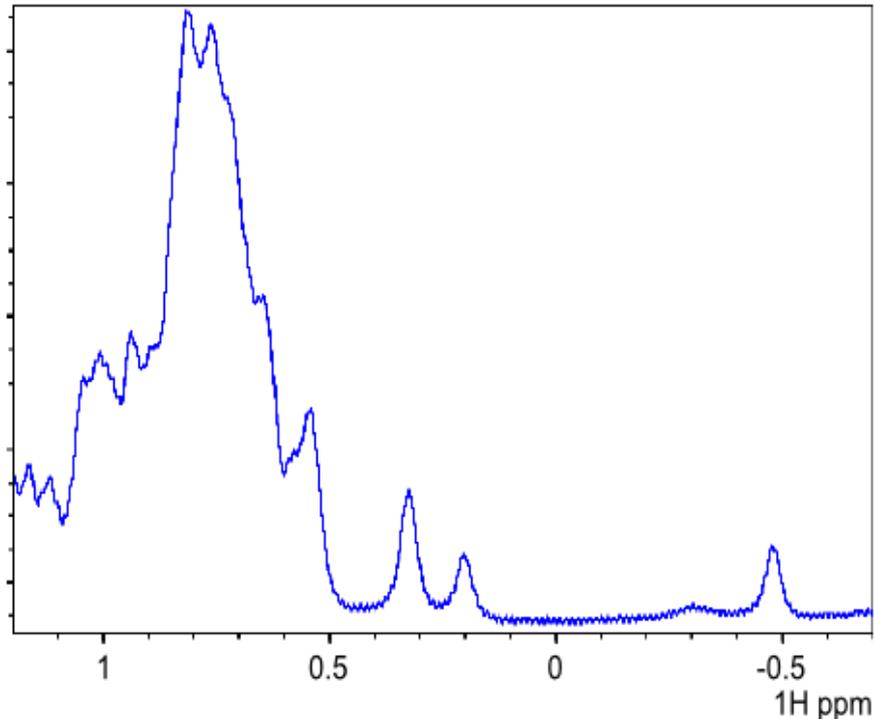


Figure 4.35 – The aliphatic region of the 1D NMR spectrum of the H351A mutant. The H351A mutant is folded as can be seen by the peak at -0.48 ppm, which indicates a methyl group that has been shifted by the presence of a nearby aromatic side chain from another residue. This can only occur in a folded protein.

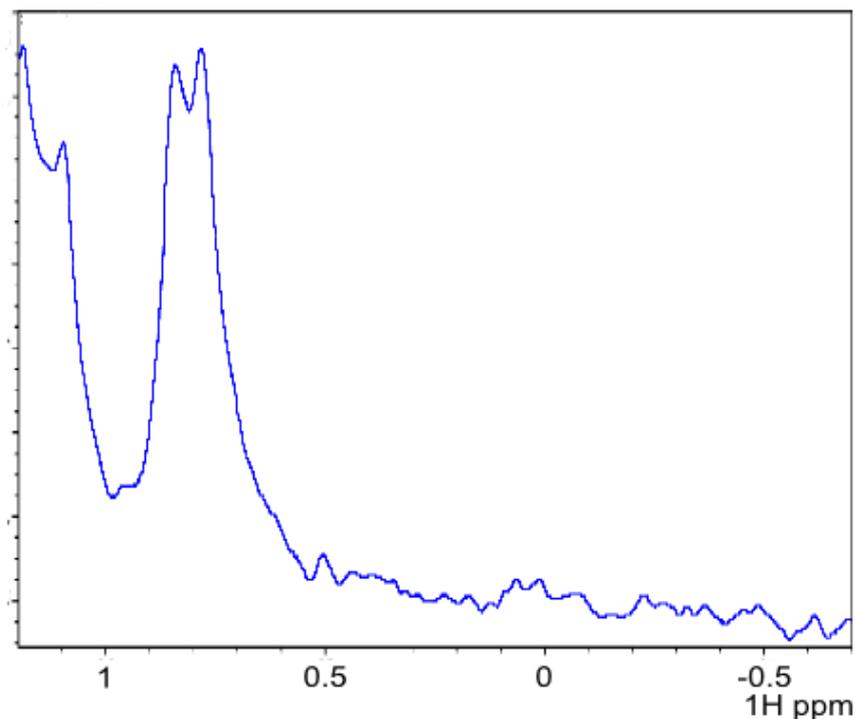


Figure 4.36 – Aliphatic region of the 1D NMR spectrum of the R349A mutant. The R349A mutant is unfolded, as can be seen by the lack of any peaks below 0 ppm. Peaks occur in this region if methyl groups in side chains are shifted by the presence of aromatic side chains from another residue. This will only occur in a folded protein. As there are no peaks below 0 ppm, it is unlikely that R349A is folded.

All the mutant constructs were folded, with the exception of R349A. As we believe the actin-binding activity of ABD2 to be a function of its structure, an unfolded construct will not bind or, if it does, it will be because exposed residues that should not be exposed are performing the binding. Because of this, no further experiments were performed on this mutant. Co-sedimentation assays for actin binding were performed on the remaining mutant constructs. If the mutant still binds to actin, it shows that the residue which has been mutated is not vital for actin binding. If the mutant does not bind to actin, it shows that the mutated residue was vital to actin binding.

Figure 4.37 shows the SDS PAGE gel run after the actin co-sedimentation assay was performed on WT ABD2, R302A ABD2 and R347A ABD2. S and P refer to the supernatant and pellet fractions after centrifugation. Lane M contains the molecular weight marker. All of the proteins were at 50  $\mu$ M. WT ABD2 on its own remains in the supernatant after centrifugation; this can be seen by the band at 10 kDa on the gel. When centrifuged in the presence of actin, WT ABD2 moves into the pellet. There is a band of the right size in both the pellet and supernatant lanes. When R302A and R347A are centrifuged in the absence of actin, they remain in the supernatant, as can be seen by the band on the gel in lanes 3 and 5. In the presence of actin, they remain in the supernatant after centrifugation, which can be seen in lanes 4 and 6. This shows that R302 and R347 are vital to actin binding, because when they are mutated, the protein can no longer bind to actin.

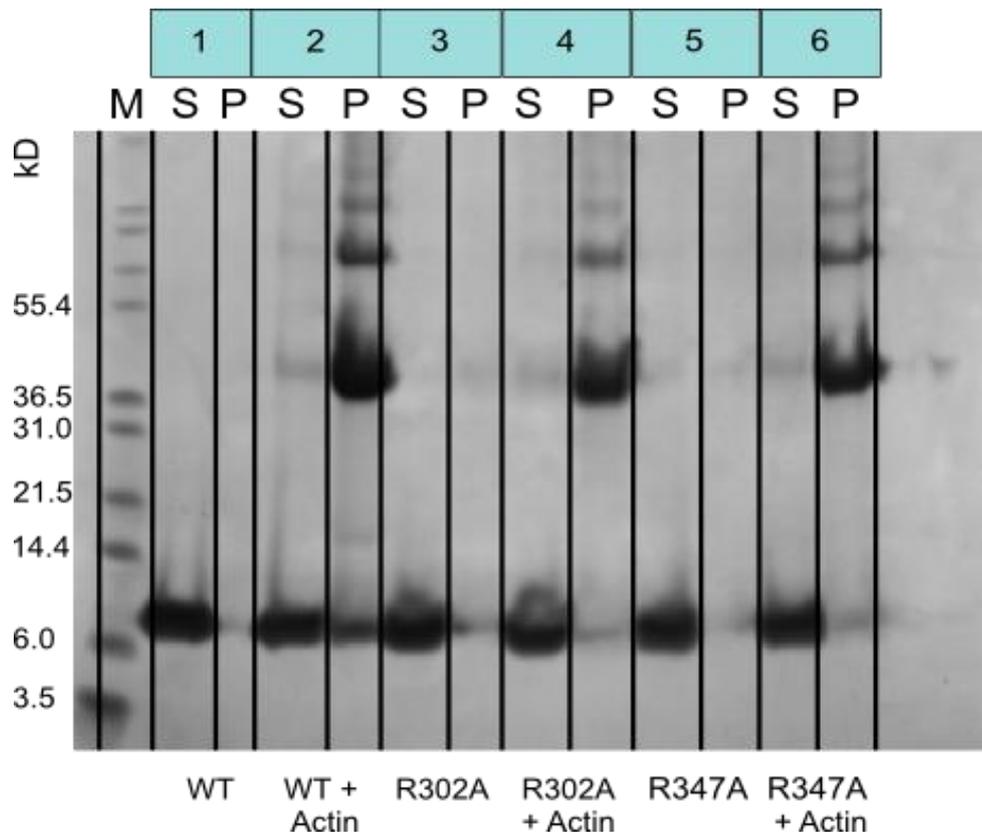


Figure 4.37 – S = supernatant and P = pellet after centrifugation. M = molecular weight marker. All proteins were at 50  $\mu$ M. Lane 1 contains WT ABD2 alone. Lane 2 contains WT ABD2 and actin. Lane 3 contains R302A alone. Lane 4 contains R302A and actin. Lane 5 contains R347A on its own. Lane 6 contains R347A and actin. The actin binding assay on the mutant constructs R302A and R347A shows that R302A and R347A do not bind to actin.

Figure 4.38 shows the SDS PAGE gel run after the actin co-sedimentation assay was performed on the mutant constructs, K350A ABD2 and H351A ABD2. S and P refer to the supernatant and pellet fractions after centrifugation. Lane M contains the molecular weight marker. All of the proteins were at 50  $\mu$ M. K350A and H351A on their own remain in the supernatant after centrifugation; this can be seen by the bands at 10 kDa on the gel in lanes 1 and 3. When centrifuged in the presence of actin, K350A remains in the supernatant. This shows that K350 is vital for actin binding as, when it is mutated, the resultant protein can no longer bind to actin. When H351A is centrifuged in the presence of actin, some of it moves into the pellet, although less of it than with

WT ABD2. This shows that H351 is not vital for actin binding as mutating this residue does not abolish the actin binding ability of ABD2.

The faint bands above the main actin band at 42 kDa could be actin dimers, although these are not supposed to occur in SDS PAGE gels. These can also be seen in the other lanes which contain actin in both this gel and the gels seen in figure 4.37 and 4.39.

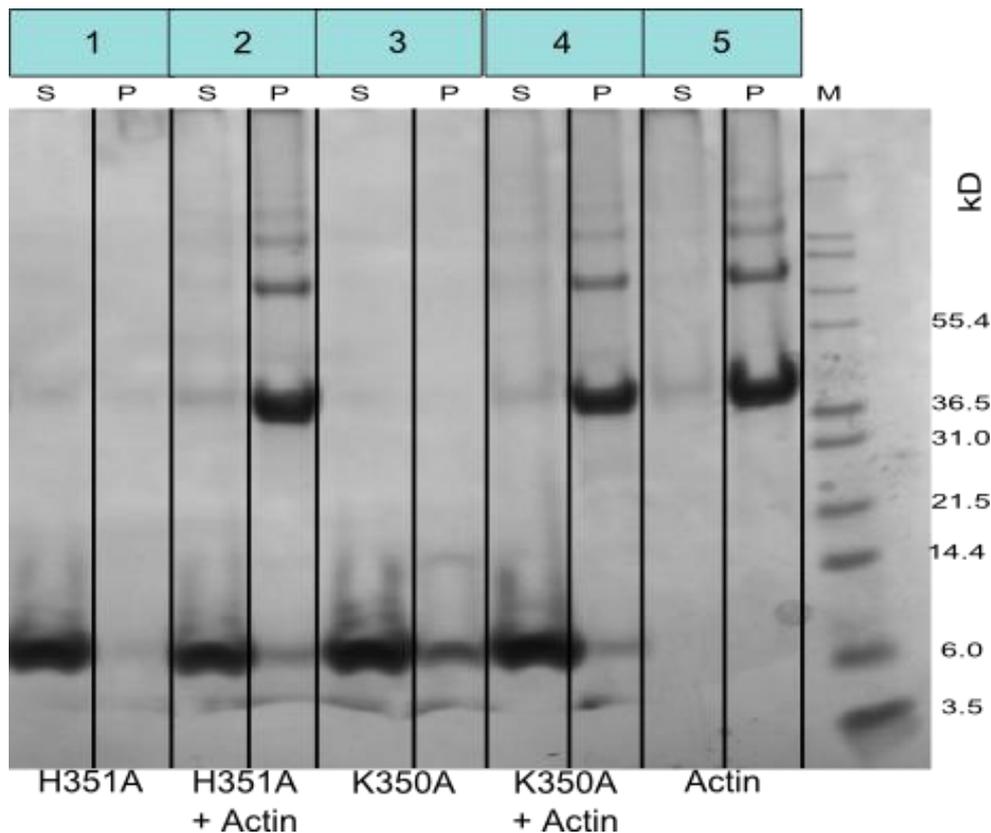


Figure 4.38 – S = supernatant and P = pellet after centrifugation. M = molecular weight marker. All proteins were at 50  $\mu$ M. Lane 1 contains H351A alone. Lane 2 contains H351A and actin. Lane 3 contains K350A alone. Lane 4 contains K350A and actin. Lane 5 contains actin. The actin binding assay on the mutant constructs K350A and H351A shows that H351A binds to actin, but K350A does not.

Figure 4.39 shows the SDS PAGE gel run after the actin co-sedimentation assay was performed on WT ABD2, K350A ABD2 and R318A ABD2. S and P refer to the supernatant and pellet fractions after centrifugation. Lane M contains the molecular

weight marker. All of the proteins were at 50  $\mu$ M. WT ABD2 on its own remains in the supernatant after centrifugation; this can be seen by the band at 10 kDa on the gel. When centrifuged in the presence of actin, WT ABD2 moves into the pellet. There is a band of the right size in both the pellet and supernatant lanes. When K350A and R318A are centrifuged in the absence of actin, they remain in the supernatant, as can be seen by the band on the gel in lanes 3 and 5. In the presence of actin, K350A remains in the supernatant after centrifugation, which can be seen in lane 4. R318A, on the other hand, moves into the pellet when centrifuged with actin (lane 6). This shows that K350 is vital to actin binding, because when it is mutated, the protein can no longer bind to actin. R318 is not vital to actin binding, as mutating this residue does not disturb the actin binding ability of ABD2.

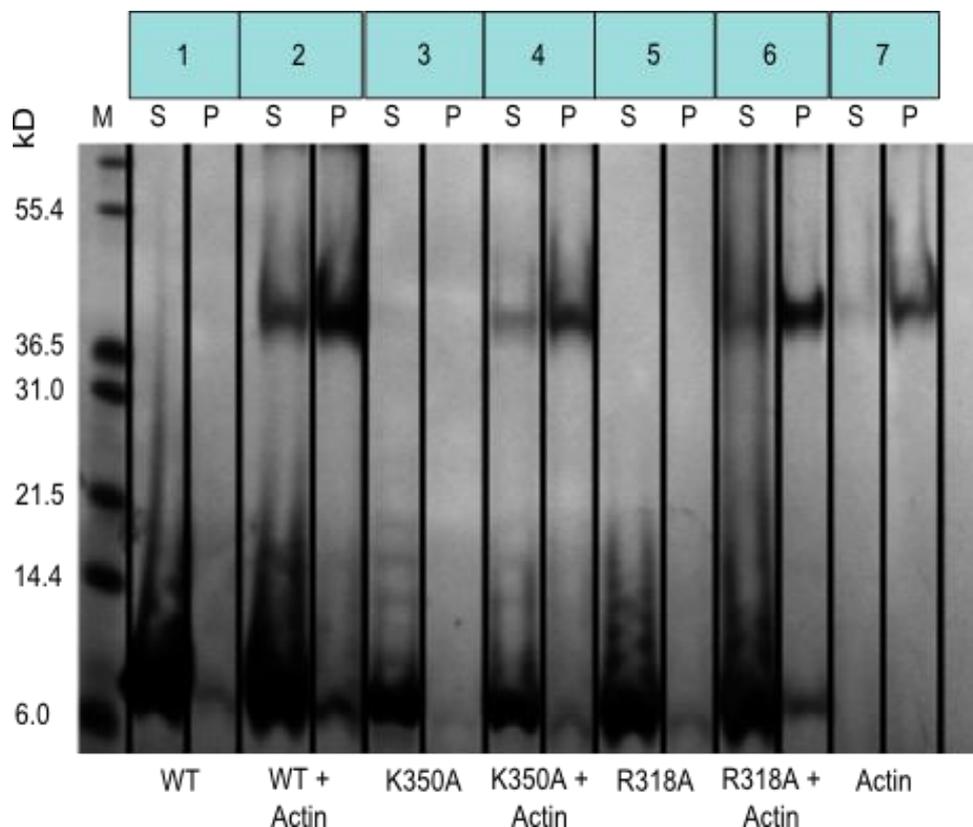


Figure 4.39 – S = supernatant and P = pellet after centrifugation. M = molecular weight marker. All proteins were at 50  $\mu$ M. Lane 1 contains WT ABD2 alone. Lane 2 contains WT ABD2 and actin. Lane 3 contains K350A alone. Lane 4 contains K350A and actin. Lane 5 contains R318A alone. Lane 6 contains R318A plus actin. Lane 7 contains actin. The actin binding assay on the mutant constructs K350A and R318A shows that R318A binds to actin, but K350A does not.

These experiments show that R302, R347 and K350 are vital to actin binding, while R318 and H351 are not. The position of these residues in the structure of ABD2 is shown in figure 4.40. It shows that the actin binding activity of ABD2 is concentrated in one of the areas of conserved positive charge, consisting of the N-terminal residue of helix 1 and the C-terminal of helix 3, suggesting that this is the actin binding site of ABD2.

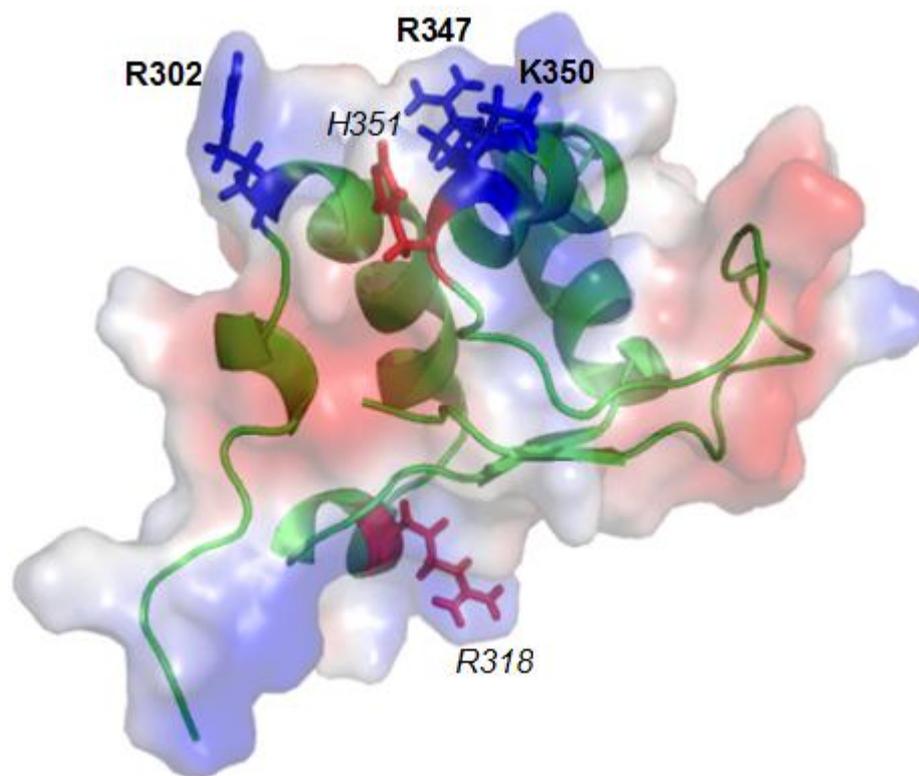


Figure 4.40 – Figure showing the location of the actin binding site. Those residues in bold were shown to be vital for actin binding when mutated, the residues in italics were shown not to be vital for actin binding.

#### **4.6 Discussion - Location of Actin Binding Surface of ABD2**

Determining the structure of ABD2 enabled the identification of areas of surface charge. ABD2 was shown to have two large positively charged patches (section 4.3). These positive patches were mapped using mutation studies and were shown to be produced

by two groups of residues, one at the N-terminal of helix 1 (residue R302) and the other by the residues at the C-terminal of helix 3 (R347 and K350) (figure 4.40).

There were two reasons why this method could not be used for ABD1. The first is that ABD1 shows no strong sequence homology with other proteins, so conserved positively charged regions and residues, which are the areas mostly like to be involved in actin binding, could not be identified in ABD1. When residue identification was attempted, there was difficulty in identifying much of ABD1 due to overlapping and weak signals. This poor signal quality is due to ABD1 being an unfolded protein. Unfolded proteins do not give dispersed spectral peaks such as those found in the NMR spectra of folded proteins. The reason for the lack of dispersion is that the side chains of the residues in unfolded proteins do not interact with each other so the chemical shifts in unfolded proteins remain close to their random coil value. The poor dispersion makes it more difficult to identify individual residues in the HSQC, HNCA, HNCACB, HN(CO)CA and HN(CO)CACB spectra used to identify residues in a protein. A few residues of ABD1 have been identified and information on these residues can be found in appendix 5. The co-operative behaviour of ABD1 seen upon actin binding when investigated using NMR, which caused most of the ABD1 peaks to disappear from the spectrum, meant that the residues involved in binding could not be identified.

**Chapter 5 – Actin Binding  
Properties of the Other  
Putative Domains of  
ms1/STARS**

## **Chapter 5 - Actin Binding Properties of the Other Putative Domains of ms1/STARS**

ms1/STARS binds to actin (Arai *et al.*, 2002). Deletion mutations have shown it binds to actin between residues 234 - 375 (Arai *et al.*, 2002) but did not show if any of the domains of ms1/STARS bind to actin independently. ABD2 was characterised, in terms of its actin binding ability, in chapter 4. The other domains of ms1/STARS, detailed in chapter 3, have their actin binding ability characterised in this chapter. This was done using an actin co-sedimentation assay to determine if the domains bound to actin independently, and then, if a domain did bind, the binding was quantified using a fluorescence based assay.

Figure 5.1 shows the SDS-PAGE gel which presents the results of the actin co-sedimentation assay upon msd1 and msd2. S and P refer to the supernatant and pellet after centrifugation. When centrifuged in the absence of actin, both msd1 (lane 2) and msd2 (lane 4) remain in the supernatant fraction. In the presence of actin, msd1 (lane 3) and msd2 (lane 5) again remain in the supernatant fraction. This result shows that neither msd1 nor msd2 binds to actin, as an actin binding protein would be found in the pellet with actin. The lane labelled M contains the molecular weight marker used in this gel.

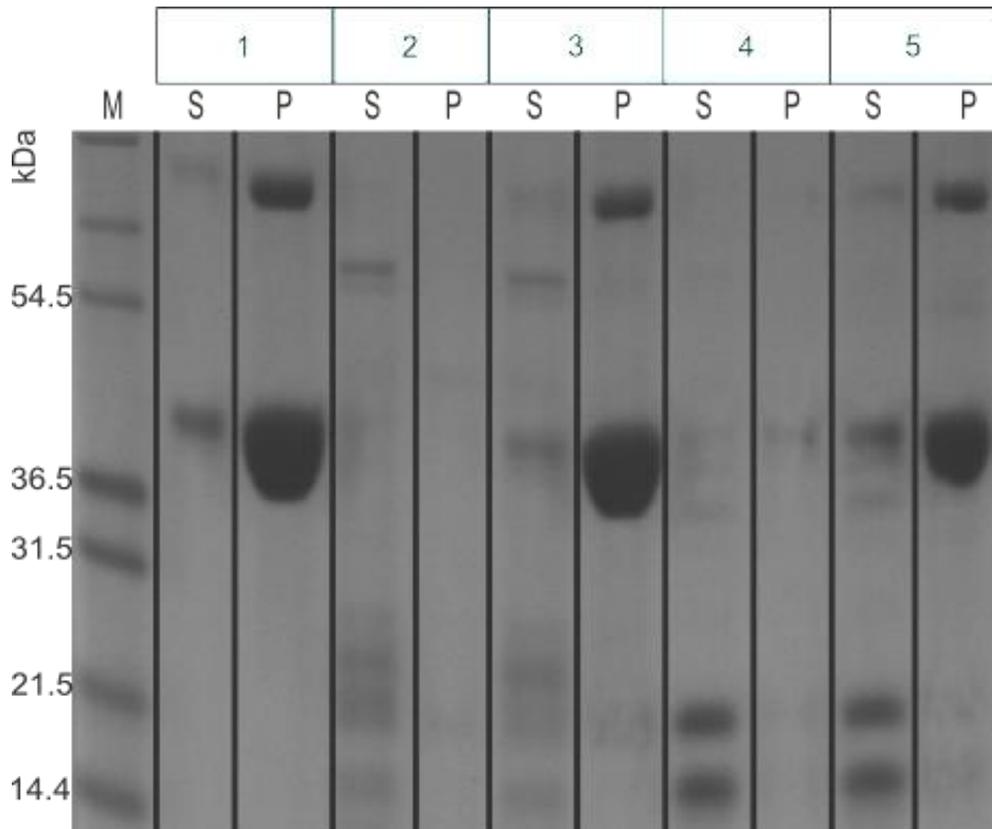


Figure 5.1 – SDS PAGE gel showing the results of the actin co-sedimentation assay performed on msd1 and msd2. All proteins are at 50  $\mu$ M. Lane M contains the marker. Lane 1 contains actin on its own. Lane 2 contains ms1d1 on its own. Lane 3 contains msd1 and actin. Lane 4 contains ms1d2 on its own. Lane 5 contains msd2 and actin. This experiment shows that neither msd1 or msd2 bind to actin.

Figure 5.2 shows the SDS-PAGE gel which shows the results of the actin co-sedimentation assay upon ABD1 and ABD2. S and P refer to the supernatant and pellet after centrifugation. All proteins are at 50  $\mu$ M. Lane M contains the marker. ABD1 (lane 1) and ABD2 (lane 3) both remain in the supernatant when they are centrifuged in the absence of actin. The fainter lower bands on the gel below ABD1 in lane 1 are degradation products of ABD1. They can also be seen in lane 2.

When centrifuged in the presence of actin, ABD1 (lane 2) and ABD2 (lane 4) move into the pellet, therefore both of them bind to actin.

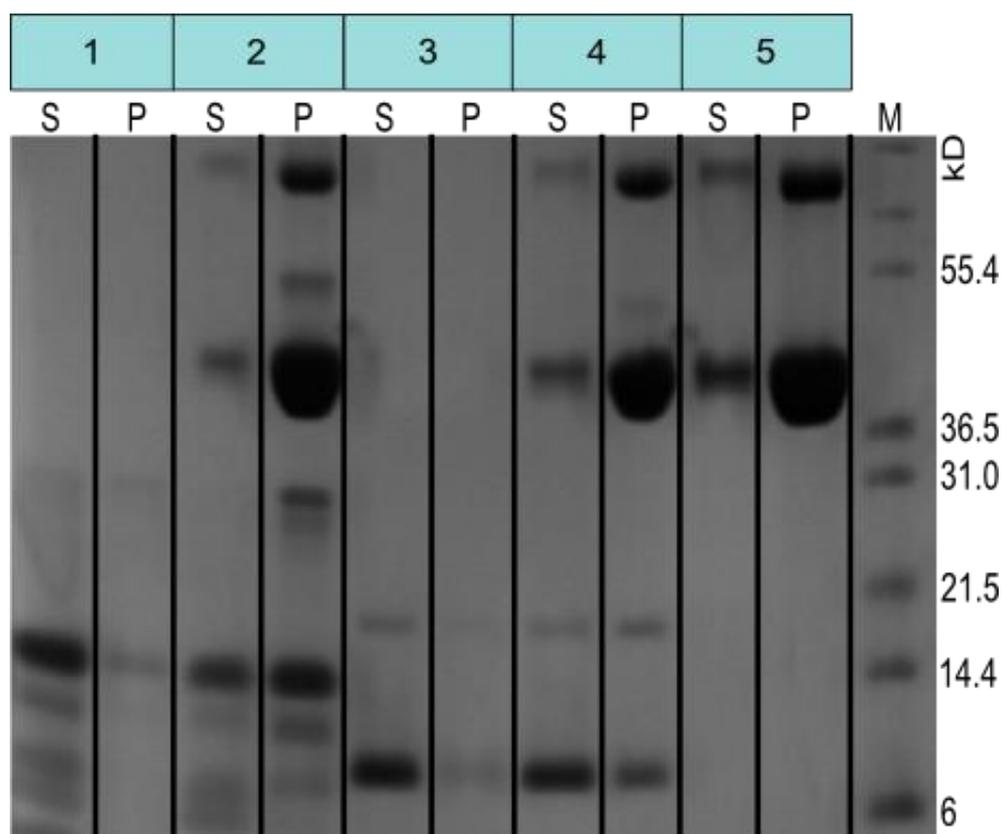


Figure 5.2 – SDS PAGE gel of the results of the actin co-sedimentation assay performed on ABD1 and ABD2. All proteins were at 50  $\mu$ M. Lane 1 contains ABD1 on its own. Lane 2 contains ABD1 and actin. ABD1 moves into the pellet after centrifugation in the presence of actin. Lane 3 contains ABD2 on its own. Lane 4 contains ABD2 and actin. ABD2 moves into the pellet after centrifugation in the presence of actin. Lane 5 contains actin on its own. Lane M contains the marker. This gel shows that ABD1 and ABD2 move into the pellet in the presence of actin, therefore both of them bind to actin.

Figures 5.1 and 5.2 show that ABD1 and ABD2 bind to actin, while msd1 and msd2 do not bind to it. Following this, the strength of the binding of ABD1 to actin was examined using the pyrene-labelled actin fluorescence assay (figure 5.3).

Increasing the amount of ABD1 led to an increase in fluorescence indicating an interaction of ABD1 with actin. When the addition of further ABD1 did not change the fluorescence reading, the binding of actin was assumed to be saturated. This value was taken to be 100% saturation, and the other fluorescence readings were expressed as a percentage of this saturation. The percentage saturation was plotted against the amount

of ABD1 added. The values from this graph were interpreted using a binding fit macro written for Mathematica by Dr. M. Pfuhl.

Figure 5.3 shows the results of this fitting. Saturation is shown as 100% binding, on the y-axis, while the concentration of ABD1 is shown on the x-axis. The box below the binding curve shows the residual error values between the experimental data and the fitted data. These are spread evenly throughout the measurement, so the value for the  $K_D$  is not biased by them.

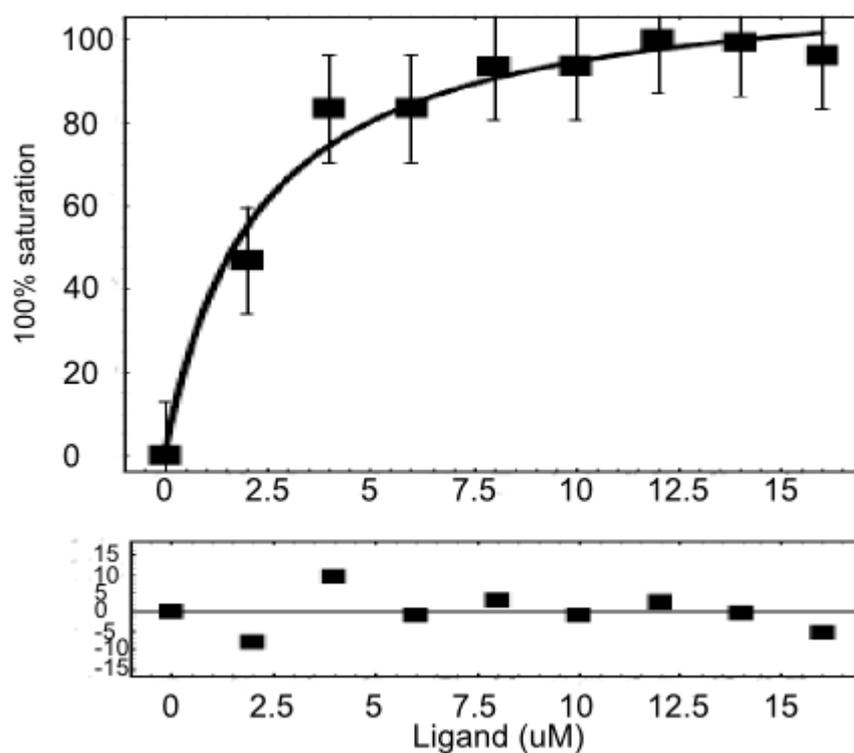


Figure 5.3 – Results of measuring actin binding of ABD1 using the fluorescence of pyrene-labelled actin. The  $K_D$  is  $2.21 \pm 0.47 \mu\text{M}$ .

This assay had previous been performed on ABD2 (see section 4.4).

The co-sedimentation assays showed that there were two independent actin-binding sites in the C-terminal of ms1/STARS; these two sites are found in ABD1 and ABD2. When these domains were expressed separately, they bound to actin. This suggests that they are functionally independent. As previous experiments had shown that deleting either of them stopped all actin binding (Arai *et al.*, 2002), we wished to see if the tandem of ABD1 and ABD2 bound more tightly than the domains on their own. The pyrene-labelled actin fluorescence assay was then used to determine this (figure 5.4). Saturation is shown as 100% binding, on the y-axis, while the concentration of the tandem of ABD1 and ABD2 is shown on the x-axis. The  $K_D$  is determined as described above. The box below the binding curve shows the error values. These are spread evenly throughout the measurement, so the value for the  $K_D$  was not biased by them.

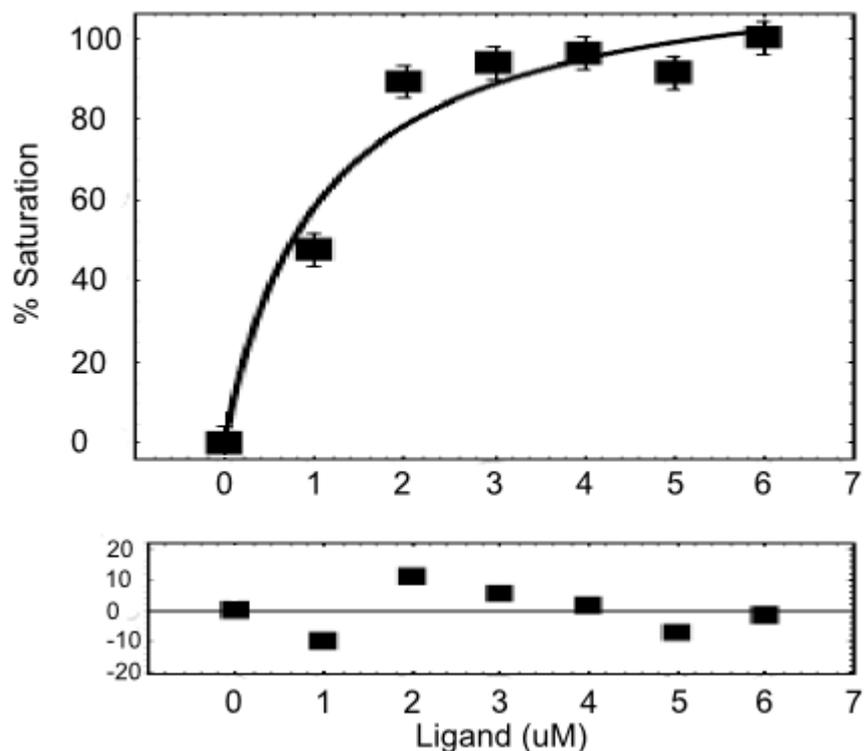


Figure 5.4 – Binding of the tandem of ABD1 and ABD2 to actin, measured using the fluorescence of pyrene-labelled actin. The  $K_D$  is  $1.07 \mu\text{M} (\pm 0.40 \mu\text{M})$ .

ABD1 binds more strongly to actin than ABD2 (a  $K_D$  of  $2.22 \pm 0.47 \mu\text{M}$  compared to  $10.61 \pm 0.7 \mu\text{M}$ ). The tandem of the two domains binds more tightly than either individually, binding with a  $K_D$  of  $1.07 \pm 0.40 \mu\text{M}$ .

### **5.1 Following the Binding of ABD1 to Actin Using NMR**

As ABD1 is unfolded, when its binding to actin is measured, only those residues which bind should disappear from the spectrum, as the rest will remain flexible enough to tumble at a rate that is fast enough to be seen using NMR spectroscopy. Figure 5.5 shows this. The black peaks in figure 5.5 are from the HSQC of ABD1 ( $50\mu\text{M}$ ) on its own. The red peaks are from the HSQC of  $50 \mu\text{M}$  ABD1 plus  $10 \mu\text{M}$  actin. The two HSQC spectra have been overlaid so that they can be compared. There are far fewer peaks in the ABD1 plus actin spectrum. This can be seen particularly clearly in the region that is at 8.5 ppm on the  $^1\text{H}$  axis and 110 ppm on the  $^{15}\text{N}$  axis. That this is not due to spectral quality being worse can be seen from some of the peaks, for example those peaks at 8.5 ppm on the  $^1\text{H}$  axis and 122 ppm on the  $^{15}\text{N}$  axis are the same size in both spectra.

More peaks disappear than expected in the ABD1 plus actin spectrum, suggesting that either more of the domain is involved in actin binding than was expected or that the domain becomes folded on binding to actin, which would mean that the structure would become rigid. That would mean that more of the peaks would tumble at the rate of actin, which is too slow to be measured by NMR.

The binding of ABD1 to actin could be studied further in future using solid state NMR, where the size of the protein is less important as the samples do not tumble.

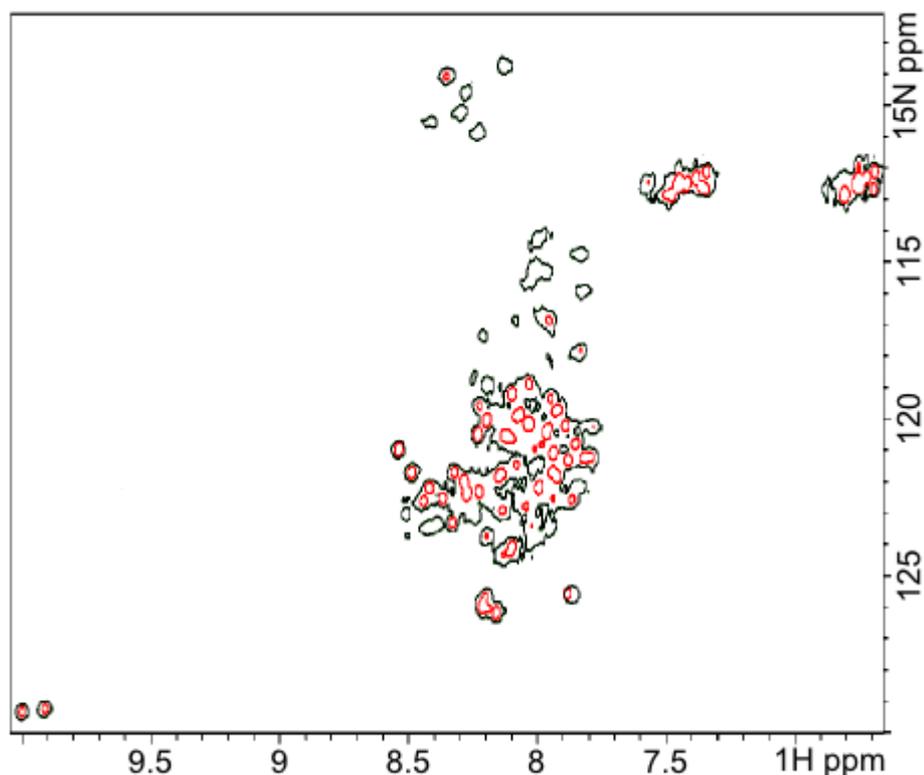


Figure 5.5 – HSQC of  $^{15}\text{N}$  labelled ABD1 without (black) and with (red) actin.  $10\ \mu\text{M}$  actin was added to  $50\ \mu\text{M}$  ABD1.

## **5.2 Residue Assignment of ABD1**

Assignment of ABD1 was attempted to try to identify the residues involved in actin binding. HNCA, HNCACB, HN(CO)CA and HN(CO)CACB spectra of ABD1 were taken and residue identification was attempted. Because ABD1 is unfolded, most of the residues have values close to the random coil values, making assignment difficult. Detailed sequence specific assignment for those residues of ABD1 that could be identified can be found in appendix 5.

## **5.3 Discussion of the Actin Binding of ms1/STARS**

The actin binding region of ms1/STARS was investigated by deletion mutation and the C-terminal of ms1/STARS (residues 234-375) was shown to be the region that binds to actin (Arai *et al.*, 2002). It was found that two separate deletions removed the actin

binding ability of ms1/STARS (Arai *et al.*, 2002). No rationale was given in the paper for the choice of deletion, and showing that two deletions both knock out actin binding does not mean that there are two actin binding sites as they could both be part of the same single actin binding site.

Previous work using negatively stained cryo-electron microscopy (Fock *et al.*, unpublished) (figure 1.7) on ms1/STARS showed that it must have more than one actin binding site as it causes actin bundling (see figure 5.6). This can only occur when the same protein molecule links more than one actin molecule.

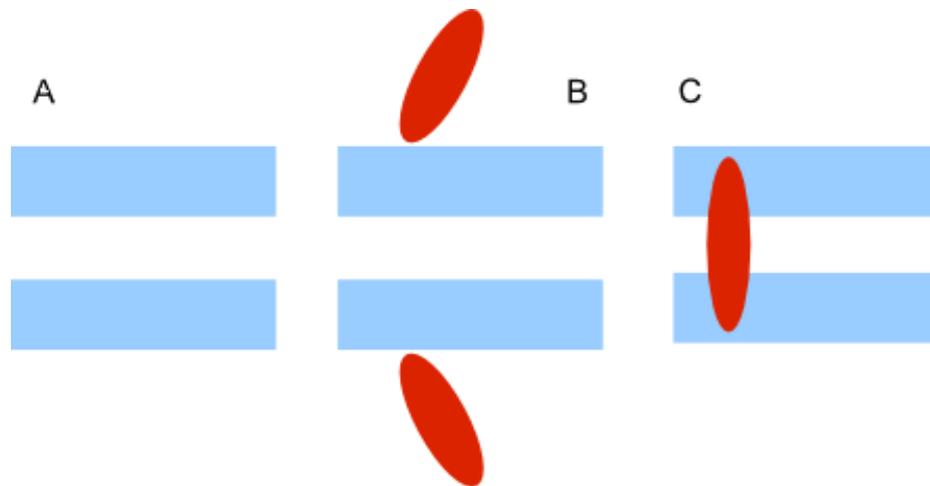


Figure 5.6 – A - Actin (blue) without any binding protein. B – actin plus binding protein (red) where the binding protein has only one actin binding site. C – actin becomes bundled when the actin binding protein has more than one actin binding site.

By characterising the actin binding activity of ABD1 and ABD2, this project has shown that ms1/STARS has two independent actin binding sites.

ABD1 binds to actin despite being unstructured. Unstructured actin binding proteins have been previously reported, for example, the regulatory domain of calponin, which binds to actin and is unfolded (El-Mezgueldi *et al.*, 1995 and Pfuhl *et al.*, 2011).

An attempt was made to see which residues of ABD1 bound to actin using NMR. This was attempted because ABD1 is unstructured. F-actin is too large to be seen using NMR, it tumbles too slowly. If the actin binding protein is folded, the complex with actin becomes too large to be observed by NMR and all of the peaks vanish. In an unstructured protein, only those residues that bind to actin will disappear as the rest of the protein remains tumbling at a rate that can be observed by NMR, enabling identification of the residues that bind to actin.

When this experiment was performed, most of the residues of ABD1 disappeared. This broadening, leading to the disappearance of peaks, suggests a co-operative effect occurs as so many disappear. This effect could be due to actin binding, if ABD1 becomes structured on binding. Other actin binding proteins which become more structured on binding include thymosin  $\beta$ 4, which has some structure when not bound to actin and then forms a more  $\alpha$ -helical structure on binding to actin (Dos Remedios *et al.*, 2003). However, thymosin  $\beta$ 4 is primarily a G-actin binding protein, not an F-actin binding protein.

This result means that solution state NMR is not suitable for studying the binding of ABD1 to actin. One way to further investigate the binding of ABD1 to actin would be to use solid state NMR, where the size of the protein is not an issue as the proteins do not tumble. This method could not be used on ABD2 as it required at least a 10 mg/ml yield of double-labelled protein, and this was not feasible for ABD2. It is more possible for ABD1 as it has a higher yield of 15 mg/ml. Experiments using this experimental strategy with another actin binding protein have already been performed by others in the lab showing this to be a feasible methodology.

ABD2 binds to actin with a  $K_D$  of  $10.61 \mu\text{M} \pm 0.7 \mu\text{M}$ . ABD1 binds more tightly to actin than ABD2, with a  $K_D$  of  $2.21 \mu\text{M} \pm 0.47 \mu\text{M}$ . The tandem of ABD1 and ABD2 binds more tightly than either individually with a  $K_D$  of  $1.07 \mu\text{M} \pm 0.40 \mu\text{M}$ .

The tandem binds more tightly than ABD2 but not significantly more tightly than ABD1. This suggests that the binding of two actin binding sites is not co-operative, as co-operative binding would cause the tandem to bind more tightly than either domain individually, since co-operative binding to one site would increase the affinity for actin of the other site. Even the tandem of ABD1 and ABD2 binds relatively weakly. Calponin, for instance, binds to actin with a  $k_D$  of  $0.046 \mu\text{M}$  (Sutherland *et al.*, 1990). Smaller proteins, highly homologous to ABD2, are found in organisms which express the longer ( $> 200$  amino acids) ms1/STARS homologues. These proteins exist separately (figure 5.7), and although no function is known for them, they would lack the second wing needed to bind tightly to DNA (section 7.5).

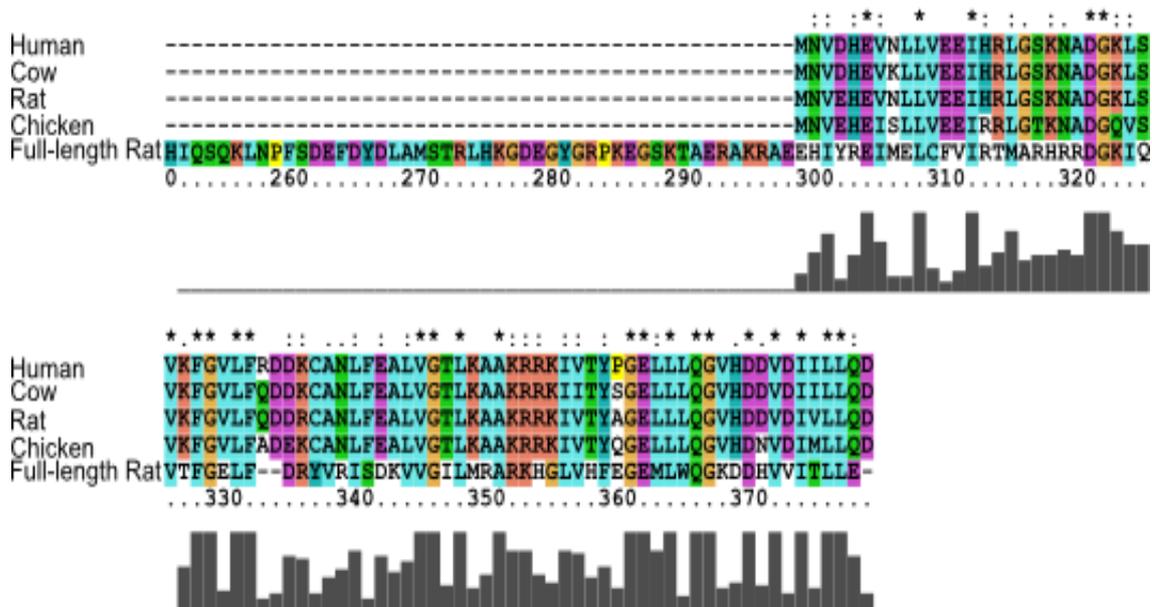


Figure 5.7 – Alignment of the short proteins found in organisms that express the long ms1/STARS homologue against full-length rat ms1/STARS. These cannot contain the second wing made by ABD1, as they only begin at the start of ABD2. The aligned proteins are human NP067066, cow NP001160033, rat NP001093117, chicken XP001233537 and full length rat ms1/STARS NP787038.

**Chapter 6 –  
The Binding of ms1/STARS  
to ABLIM-2**

## **Chapter 6 - ms1/STARS Binding to ABLIM-2**

Two other ms1/STARS binding proteins have been found, ABLIM-2 and ABLIM-3 (Barrientos *et al.*, 2007). They are members of the ABLIM (actin-binding LIM protein) protein family. Both contain four LIM domains, followed by a linker region and then a villin headpiece domain. ABLIM-2 and -3 were shown to bind to F-actin and to stimulate the activity of ms1/STARS (Barrientos *et al.*, 2007). ABLIM-2 is expressed in brain, spleen and skeletal muscle tissue but not in the heart while ABLIM-3 is expressed predominantly in the heart, along with some expression in the liver and lungs (Barrientos *et al.*, 2007). The interaction between ms1/STARS and ABLIM-2 and -3 was found using a yeast-2-hybrid screen of a skeletal muscle cDNA library. In the screen, the C-terminus of ms1/STARS (residues 234-375) was used as the bait (Barrientos *et al.*, 2007). Deletion mutation studies then showed that it is the C-terminal of ms1/STARS (residues 234-375) that binds to ABLIM-2, but it was not known which part of ABLIM-2 bound to ms1/STARS. It was not possible to express full-length ABLIM-2 in *E. coli*, and therefore expression of the domains of ABLIM-2 was attempted. The domains expressed were LIM domains 1, 2, 3 and 4 separately, and LIM 1 and 2, LIM 2 and 3 and LIM 3 and 4 as tandem domains. This was done as LIM domains often act in tandem *in vivo*. There is a long linker region between LIM4 and the VHD domain at the C-terminal of ABLIM-2 (residues 270 – 542), which does not express in a soluble form on its own. There are 7 isoforms of ABLIM-2; these isoforms are the product of differential splicing. The differences in the isoforms are in the linker region and can be seen in figure 6.1. The clone that was used to produce the constructs used in this project is isoform 2.

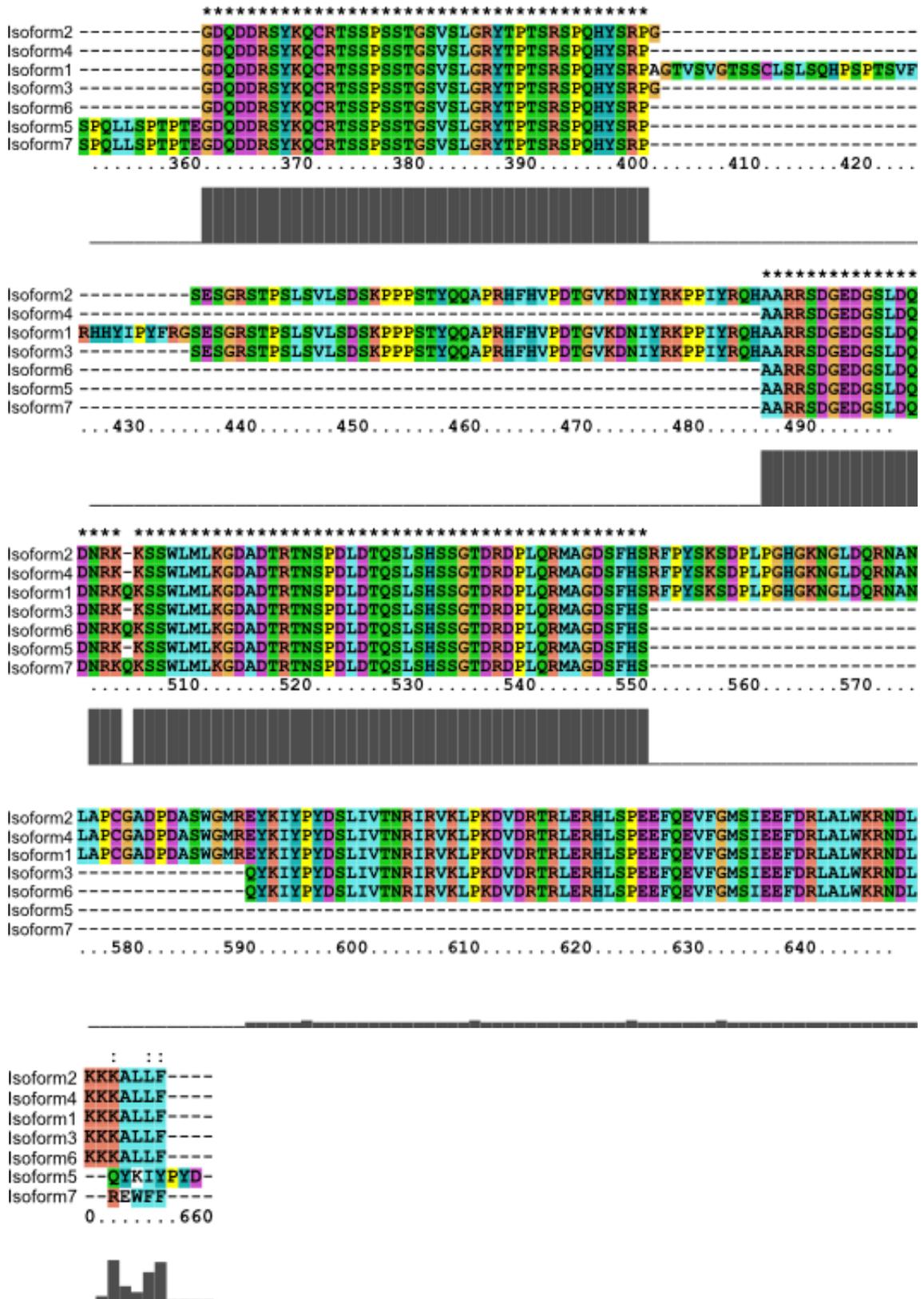


Figure 6.1 – Alignment of the 7 ABLIM-2 isoforms, showing the C-terminal that contains the differences between the isoforms.

## **6.1 Characterisation of the Structure of the Domains of ABLIM-2**

The domains of ABLIM-2 were characterised in terms of their folding using CD and NMR. It was impossible to do this for LIM domain 1 as it could not be stably expressed on its own.

LIM2 was characterised using CD (figure 6.2). The two peaks, one at 215 nm and the other at 224 nm, suggest it is at least partly  $\alpha$ -helical, although the values of the peaks are not typical of  $\alpha$ -helices. When the data was fitted, LIM2 was found to have 53%  $\alpha$ -helical and 47%  $\beta$ -sheet secondary structure. However, the experimental data and the fitted data are very dissimilar. This is possible due to precipitation of LIM2 while it was being studied.

LIM2 was then characterised using NMR. The 1D NMR spectrum can be seen in figure 6.3. The small peak at -0.65 ppm shows that LIM2 is folded as a peak below 0 ppm. This can only occur if a methyl side chain group is shifted due to its proximity to an aromatic side chain, which happens only in folded proteins.

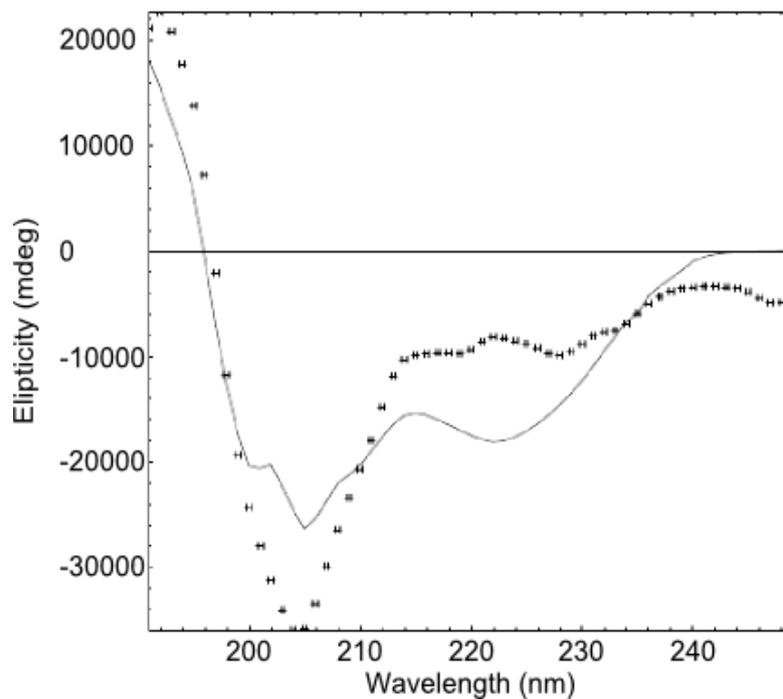


Figure 6.2 - CD spectrum of LIM domain 2. The sample was 100  $\mu\text{M}$ . LIM2 has 53%  $\alpha$ -helical structure and 47%  $\beta$ -sheet secondary structure. The dotted line is the raw data; the solid line is the fitted curve.

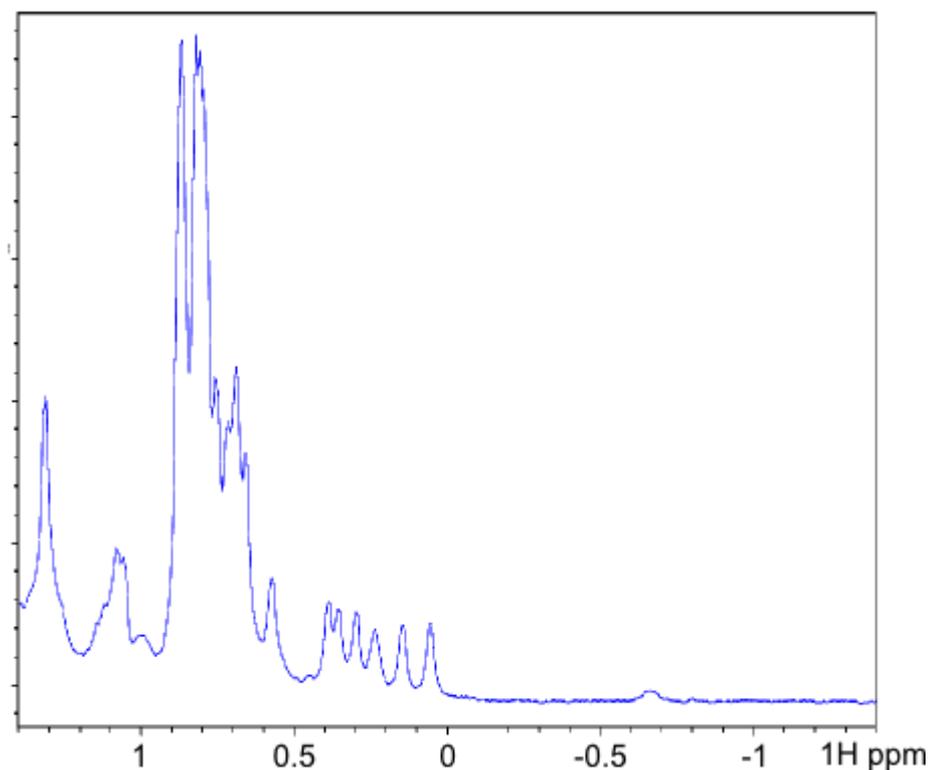


Figure 6.3 – Aliphatic region of 1D NMR performed on LIM2 (100  $\mu\text{M}$ ). The peak below 0 ppm suggests that it is folded.

The same techniques were used to investigate the folding of LIM3 (figures 6.4 and 6.5). The CD spectrum (figure 6.4) shows two peaks, one at 200 nm and one at 206 nm. When the structural content is calculated, LIM3 has 21%  $\alpha$ -helical, 22%  $\beta$ -sheet and 57% random coil secondary structure.

LIM3 was then characterised using NMR. The 1D NMR spectrum can be seen in figure 6.5. The peak at -0.4 ppm shows that LIM3 is folded, as a peak below 0 ppm can only occur if a methyl side chain group is shifted due to its proximity to an aromatic side chain. This shift only occurs in folded proteins.

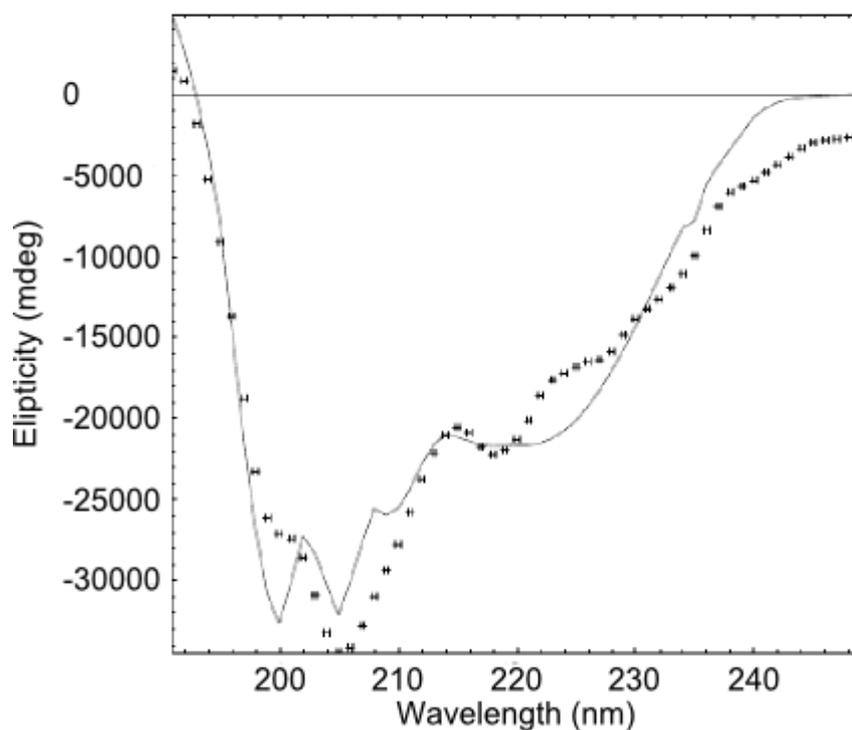


Figure 6.4 - CD spectrum of LIM domain 3. The sample was 100  $\mu$ M. LIM3 has 21%  $\alpha$ -helical, 22%  $\beta$ -sheet and 57% random coil secondary structure. The dotted line is the raw data; the solid line is the fitted curve.

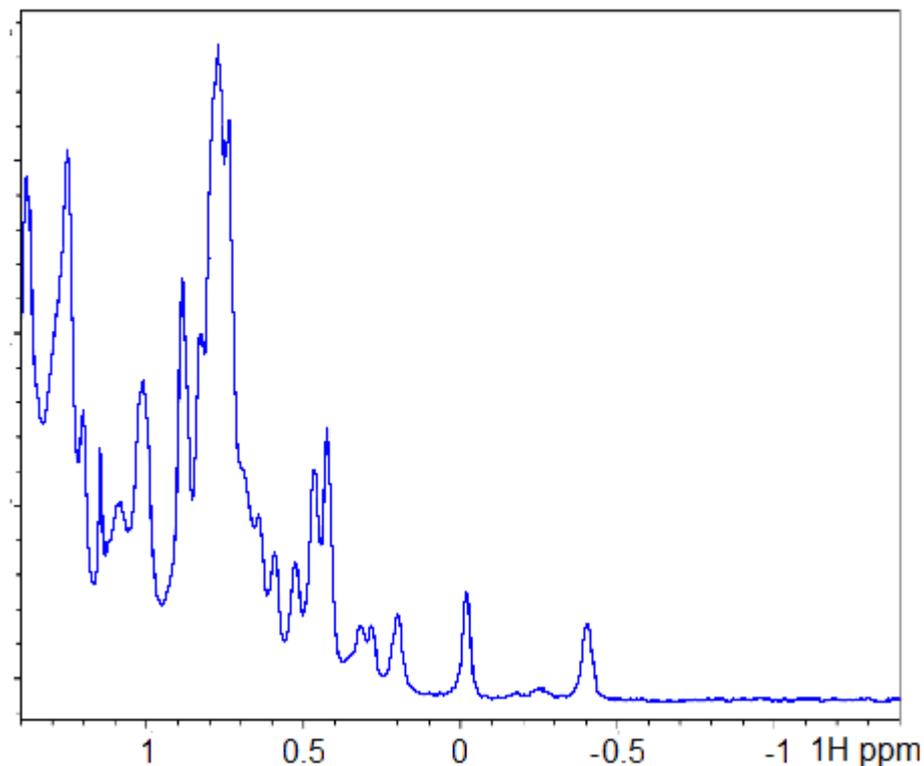


Figure 6.5 – Aliphatic region of 1D NMR performed on LIM3 (100  $\mu$ M). The peaks below 0 ppm suggest that it is folded.

LIM4 was also investigated using CD and NMR (figures 6.6 and 6.7). The CD spectrum (figure 6.6) shows two peaks, one at 200 nm and one at 224 nm. When the secondary structural content is calculated, LIM4 is 24%  $\alpha$ -helical, 10%  $\beta$ -sheet and 66% random coil.

LIM4 was then characterised using NMR. The 1D NMR spectrum can be seen in figure 6.7. The small peak at -0.78 ppm shows that LIM4 is folded, as a peak below 0 ppm can only occur if a methyl side chain group is shifted due to its proximity to an aromatic side chain, which only occurs in folded proteins.

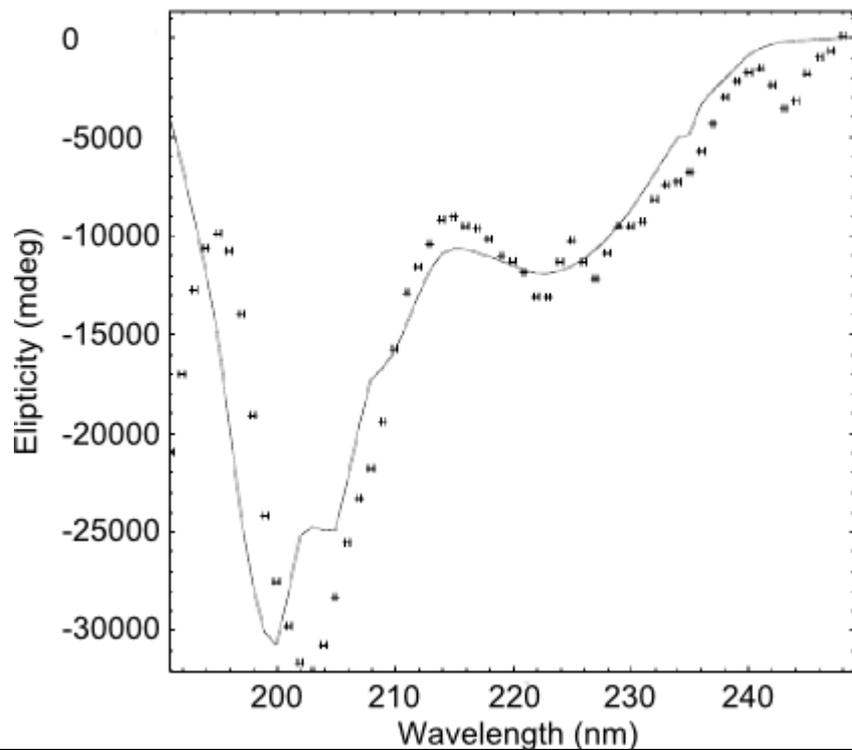


Figure 6.6 – CD spectrum of LIM domain 4 (100  $\mu$ M). It is 24% alpha-helical, 10% beta-sheet and 66% random coil. The dotted line is the raw data and the solid line is the fitted curve.

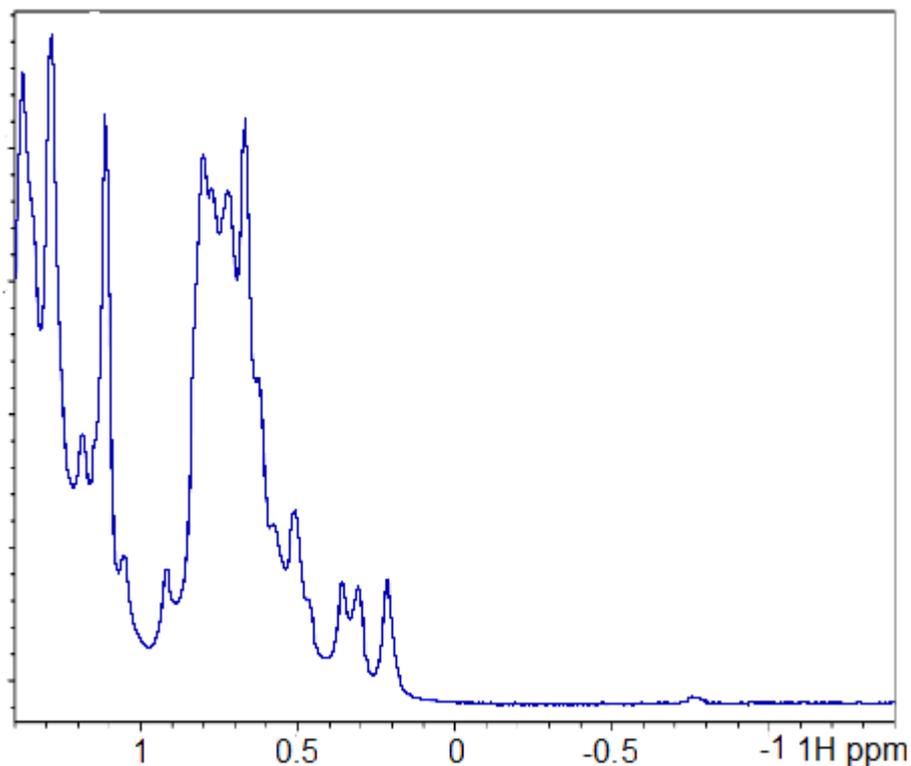


Figure 6.7 – Aliphatic region of 1D NMR performed on LIM4 (100  $\mu$ M). The peak below 0 ppm suggests that it is folded.

The first LIM domain was unstable when expressed, and could not be stabilised. This meant that CD and NMR could not be performed on LIM1. As the remaining LIM domains were folded, their melting temperature was studied using CD, to see how stable they were.

LIM2, LIM3 and LIM4 were then denatured using increasing temperature. This experiment was done for two reasons. Firstly it would confirm whether or not they were folded proteins as only folded proteins show co-operative unfolding. LIM2 (figure 6.8), LIM3 (figure 6.9) and LIM4 (figure 6.10) were shown to be folded as their denaturation is a two-step process, with a distinct transition between folded and unfolded states. Secondly, this experiment would determine the melting temperature of the protein domains. The melting temperature of LIM2 is  $26.1 \pm 0.3^\circ\text{C}$ , LIM3 is  $49.1 \pm 0.8^\circ\text{C}$  and LIM4 is  $60.1 \pm 0.3^\circ\text{C}$ .

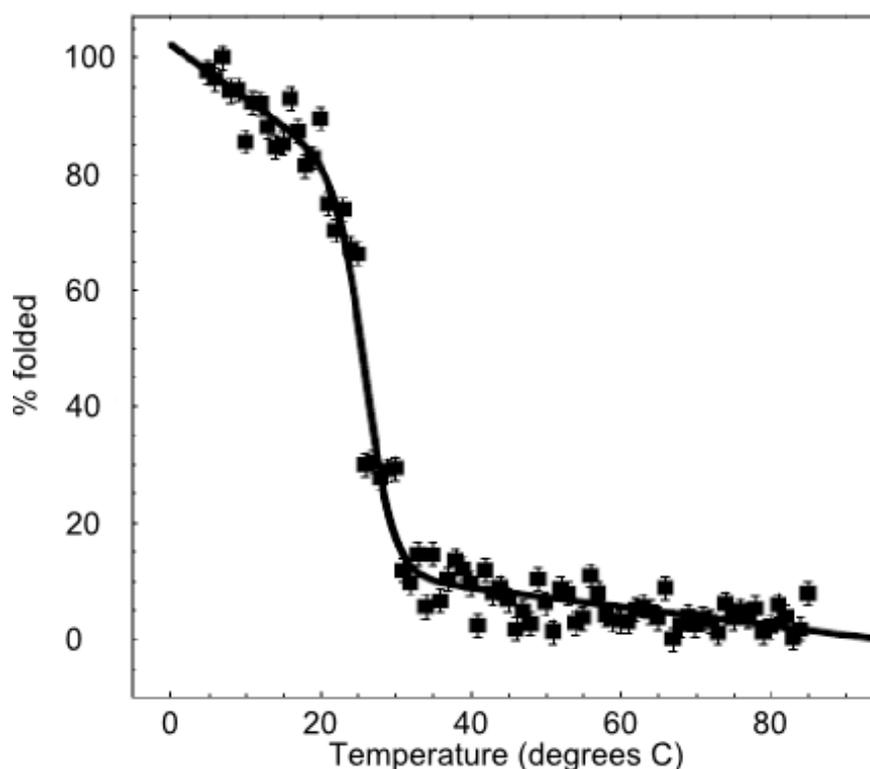


Figure 6.8 – Temperature denaturation of LIM domain 2 (100  $\mu\text{M}$ ). Melting temperature =  $26.1 \pm 0.3^\circ\text{C}$ .

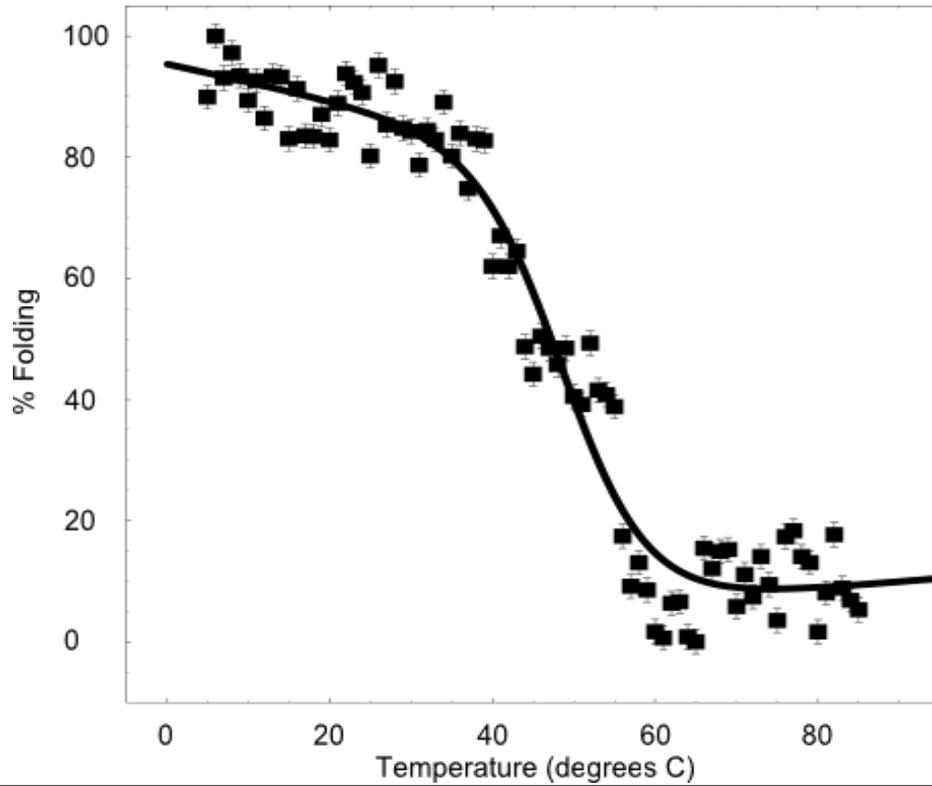


Figure 6.9 – Temperature denaturation of LIM domain 3 (100  $\mu$ M). Melting temperature =  $49.1 \pm 0.8^\circ\text{C}$ .

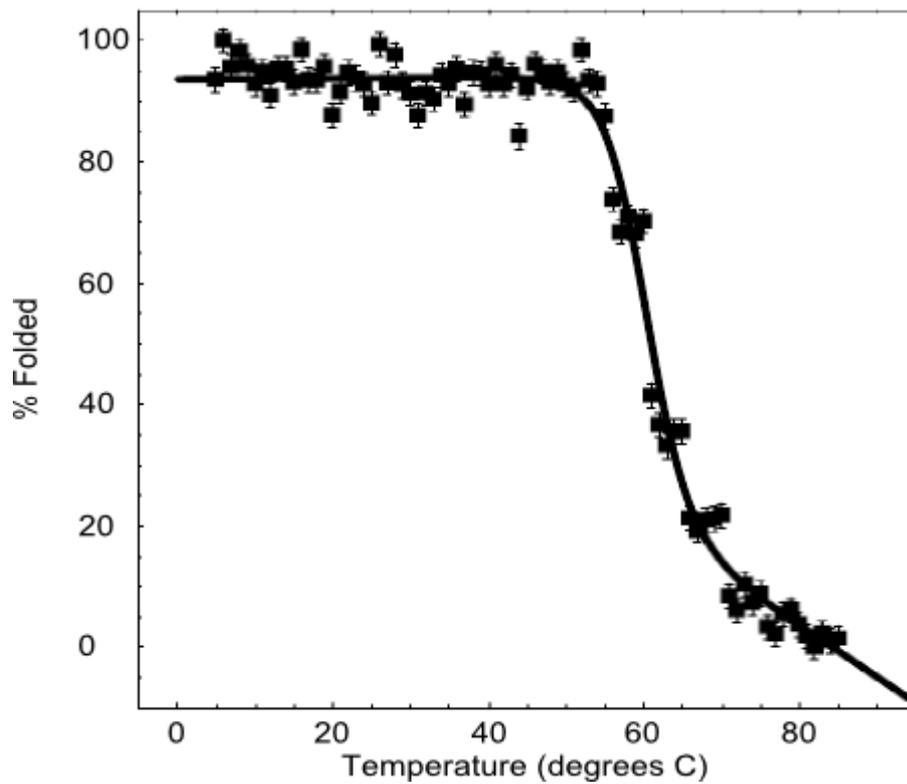


Figure 6.10 – Temperature denaturation of LIM domain 4 (100  $\mu$ M). Melting temperature =  $60.1 \pm 0.3^\circ\text{C}$ .

The stability of the LIM domains increases with their position along the protein with LIM1, which was not stable enough to study, being the least stable (melting temperature 26.1°C) and LIM4 the most stable (melting temperature 60.1°C).

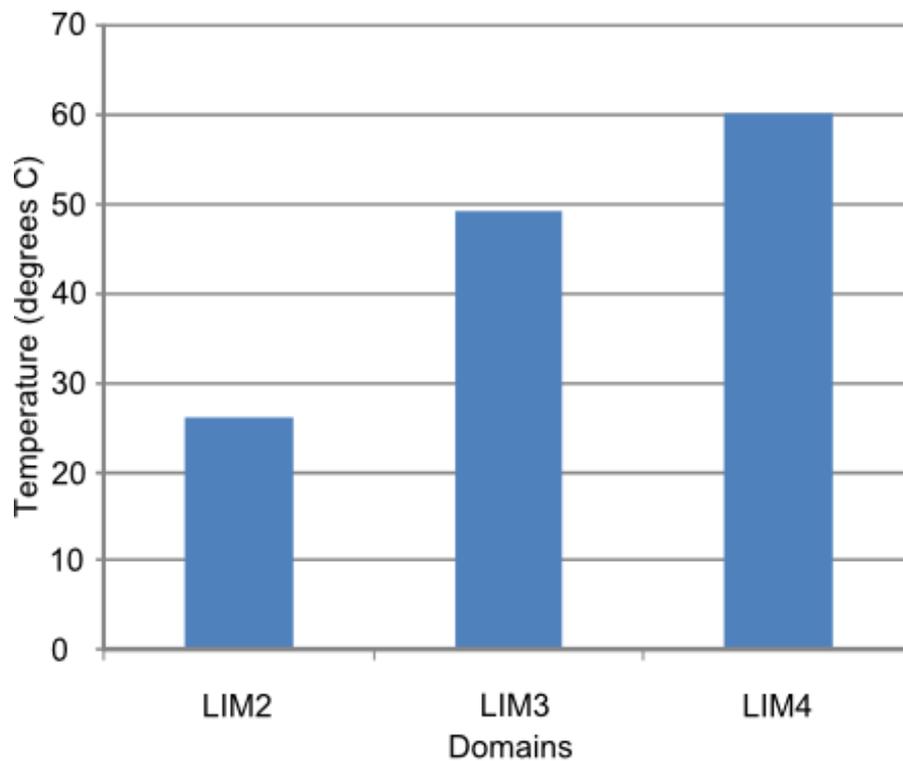


Figure 6.11 – Graph showing the increasing melting temperatures of the individual LIM domains of ABLIM-2.

In an attempt to test if the site of binding to ms1/STARS is in LIM1, it was expressed in tandem with LIM2, with the hope that this tandem construct would be soluble. Expression of the other possible LIM domain tandems was also attempted as LIM domains sometimes work in tandem when they bind to their ligands, for example LMO4 (LIM Only protein 4) (Deane *et al.*, 2004).

These tandem domains were characterised with respect to their secondary structure content and their folding. The CD spectrum of the tandem of LIM1 and LIM2 (figure

6.12) has two peaks, one at 209 nm and the other at 218 nm. When the secondary structure is calculated, it is 50%  $\alpha$ -helical, 10%  $\beta$ -sheet and 40% random coil.

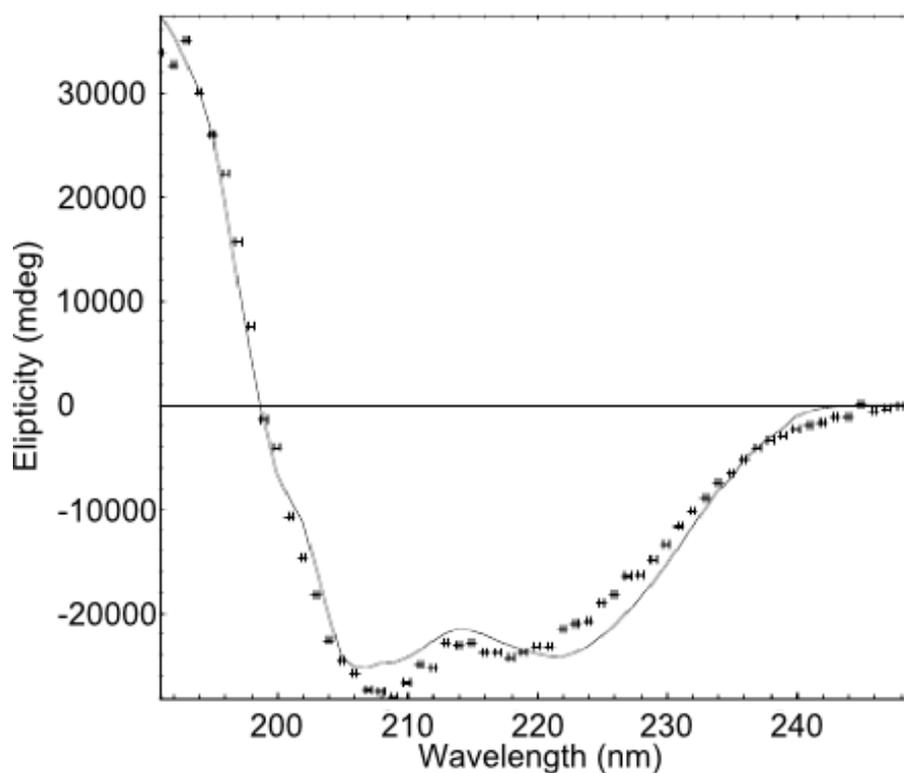


Figure 6.12 – CD spectrum of the tandem of LIM domains 1 and 2 (100  $\mu$ M). It has 50%  $\alpha$ -helical, 10 %  $\beta$ -sheet and 40 % random coil secondary structure. The dotted line is the raw data and the solid line is the fitted data.

The tandem of LIM1 and LIM2 was shown (figure 6.13) to be folded as its denaturation is a two-step process, with a distinct transition between folded and unfolded states. The melting temperature of the tandem of LIM1 and LIM2 is  $42.2 \pm 0.35^\circ\text{C}$ .

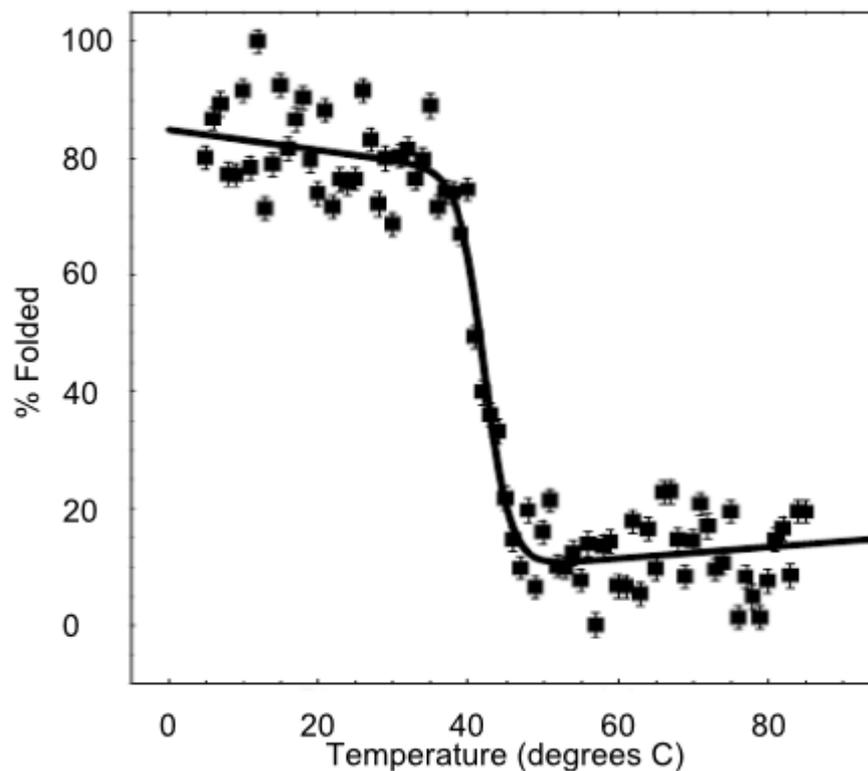


Figure 6.13 – Heat denaturation of the tandem of LIM domains 1 and 2 (100  $\mu$ M). The melting temperature is  $42.2 \pm 0.35^\circ\text{C}$ .

The CD spectrum of the tandem of LIM2 and LIM3 (figure 6.14) has two peaks, one at 208 nm and the other at 224 nm. When the secondary structure is calculated, it is 79%  $\alpha$ -helical, 1%  $\beta$ -sheet and 20% random coil.

The tandem of LIM2 and LIM3 was shown (figure 6.15) to be folded as there is cooperative unfolding when it unfolds. The melting temperature of the tandem of LIM2 and LIM3 is  $63.4 \pm 2.0^\circ\text{C}$ .

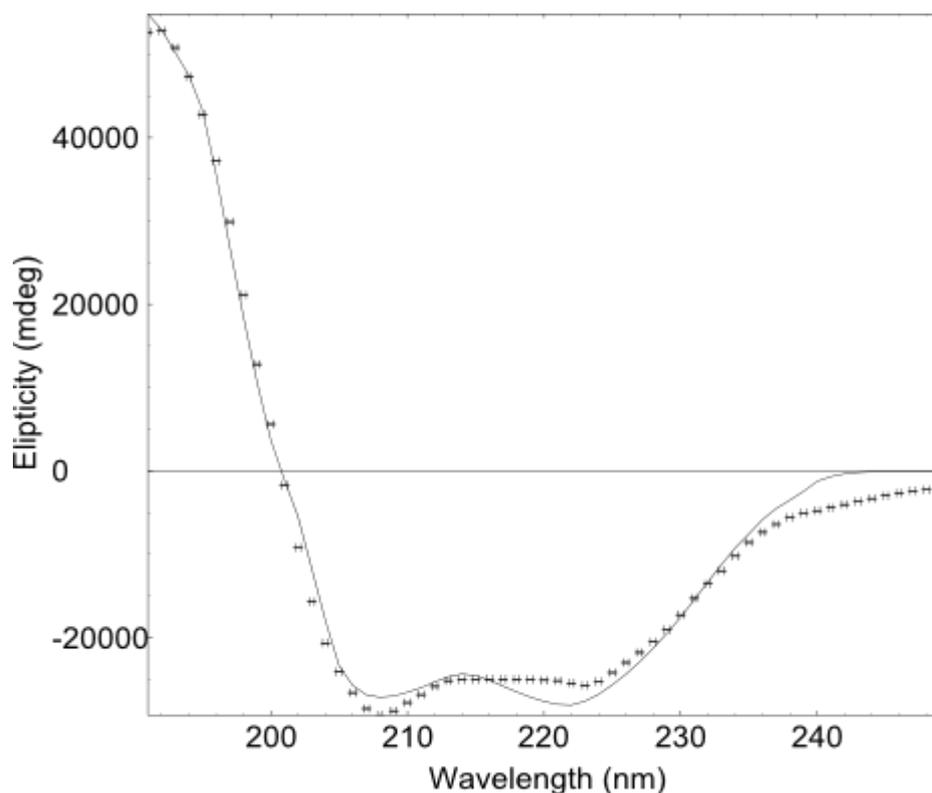


Figure 6.14 – CD spectrum of the tandem of LIM domains 2 and 3 (100  $\mu$ M). It has 79%  $\alpha$ -helical, 1%  $\beta$ -sheet and 20% random coil secondary structure. The dotted line is the raw data and the solid line is the fitted data.

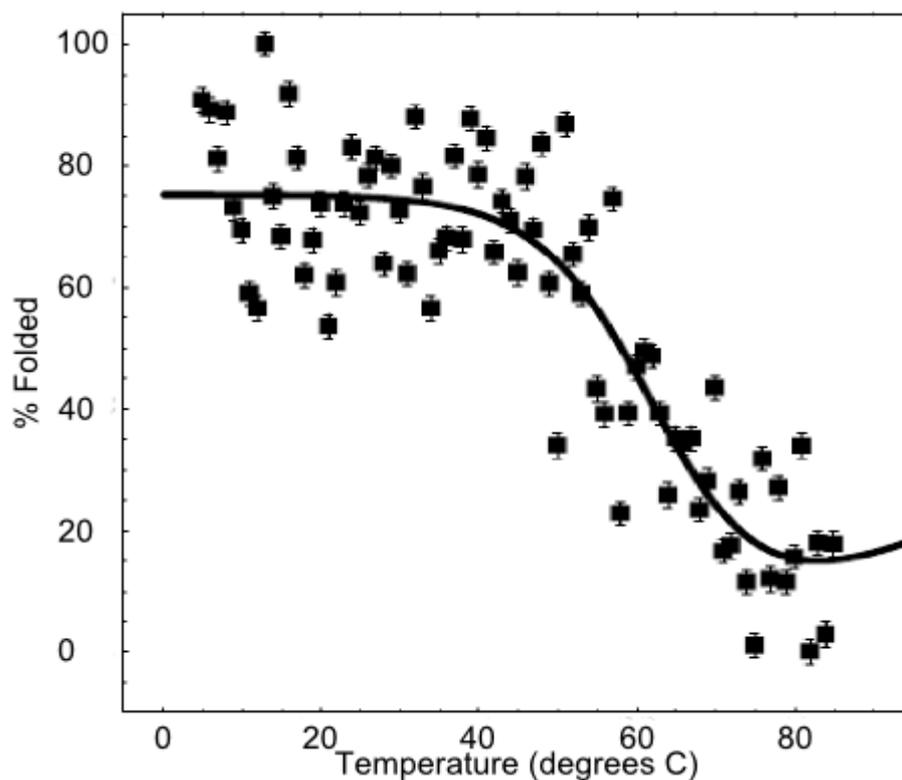


Figure 6.15 – Heat denaturation of the tandem of LIM domains 2 and 3 (100  $\mu$ M). The melting temperature is  $63.4 \pm 2.0^\circ\text{C}$ .

The CD spectrum of the tandem of LIM3 and LIM4 (figure 6.16) has two peaks, one at 206 nm and the other at 224 nm. When the secondary structure is calculated, it is 48%  $\alpha$ -helical, 1%  $\beta$ -sheet and 51% random coil.

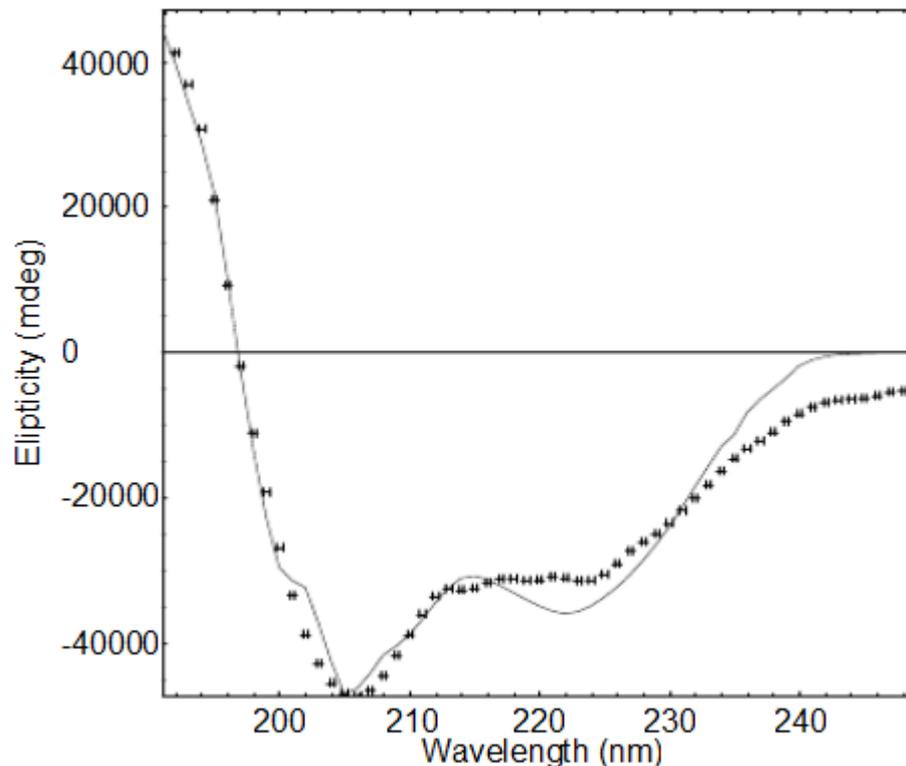


Figure 6.16 – CD spectrum of the tandem of LIM domains 3 and 4 (100  $\mu$ M). It has 48 %  $\alpha$ -helical, 1 %  $\beta$ -sheet and 51 % random coil secondary structure. The dotted line is the raw data; the solid line is the fitted data.

To complete the expression of the domains of ABLIM-2, the villin headpiece domain of ABLIM-2 (VHD) was also cloned and expressed, and then characterised in terms of its secondary structure and stability. It was also part of another separate project to examine its suitability as a protein to study actin using solid state NMR. Solid state NMR heats the sample more than solution state NMR does, which means that non-heat-stable proteins are not suitable for study using solid state NMR as they degrade too quickly. The ABLIM-2 VHD was shown to be stable enough to be studied by solid state NMR.

CD was used to determine the secondary structure of the villin headpiece domain (figure 6.17). The CD spectrum has two peaks, one at 207 nm and the second at 222 nm. When the secondary structure was calculated, it was 32%  $\alpha$ -helical, 24%  $\beta$ -sheet and 44% random coil.

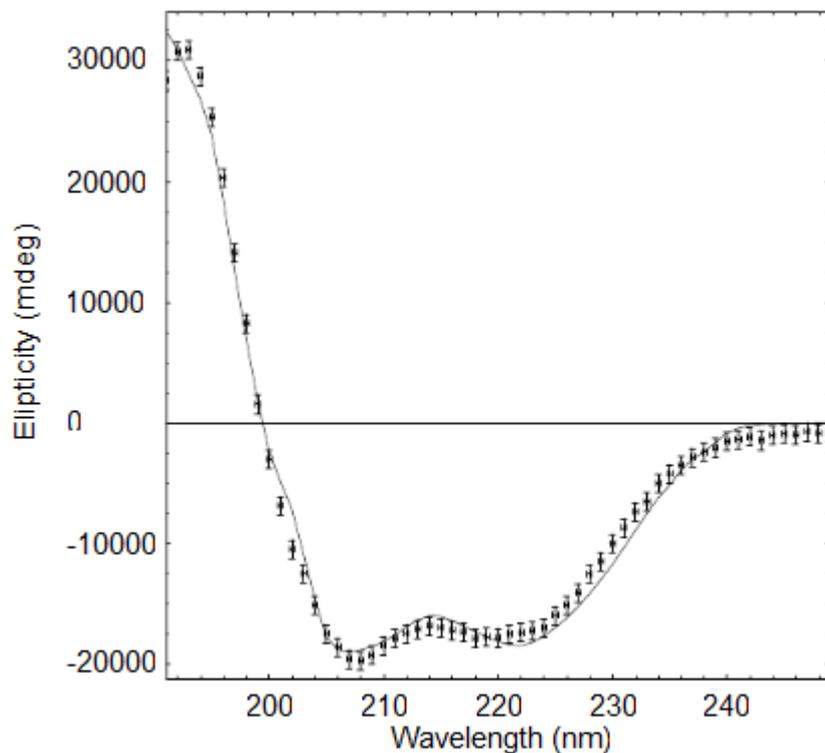


Figure 6.17 – The villin headpiece domain (100  $\mu$ M) has 32%  $\alpha$ -helical, 24%  $\beta$ -sheet and 44% random coil secondary structure. The dotted line is the raw data and the solid line is the fitted data.

CD was also used to determine the melting temperature of the VHD. However, there was very little change in the spectra between 5 $^{\circ}$  and 90 $^{\circ}$ C, as can be seen in figure 6.18. This shows that VHD is very stable.

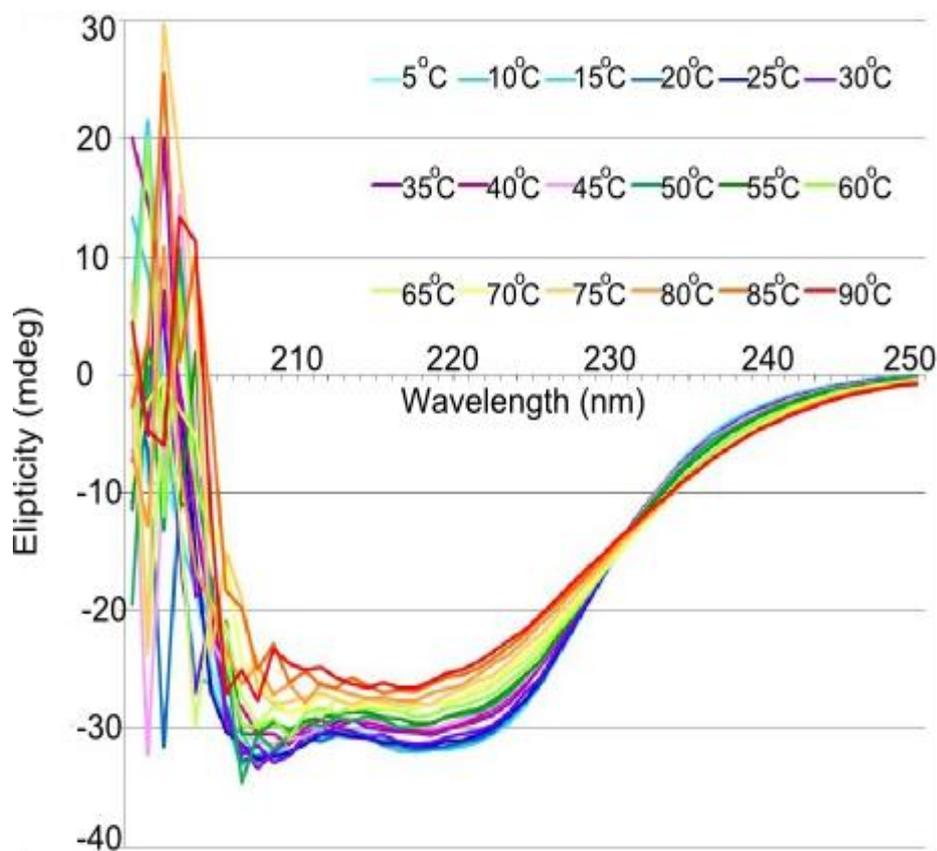


Figure 6.18 – CD spectra of the ABLIM-2 VHD at a range of temperatures from 5°C to 90°C. There is little change in the spectra indicating that it does not unfold even at 90°C.

## **6.2 Binding of the Domains of ABLIM-2 to ms1/STARS**

To discover which part of ABLIM-2 binds to ms1/STARS, the domains of ABLIM-2 were expressed and added to the C-terminal domains of ms1/STARS (ABD1, ABD2 and a tandem of ABD1 and ABD2). This region of ms1/STARS was implicated in binding to ABLIM-2 (Barrientos *et al.*, 2007). The ms1/STARS domains were <sup>15</sup>N labelled. Binding activity was followed by taking <sup>15</sup>N HSQC spectra of the ms1/STARS domains before and after the addition of the domains of ABLIM-2. If there is any binding, it will be seen by a shift in the peaks in the HSQC. Binding causes changes in the chemical environment of the protons of the residues involved in the binding. This change causes the peaks to shift.

There are no differences between the HSQC peaks of ABD2 without (green) and with (red) LIM2 (figure 6.19). This shows that there is no binding between ABD2 and LIM2. There are also no differences between the HSQC peaks of ABD2 without (green) and with (red) with LIM3 (figure 6.20). This shows that there is no binding between ABD2 and LIM3. Similarly, there are no differences between the HSQC peaks of ABD2 without (green) and with (red) LIM4 (figure 6.21). This shows that there is no binding between ABD2 and LIM4.

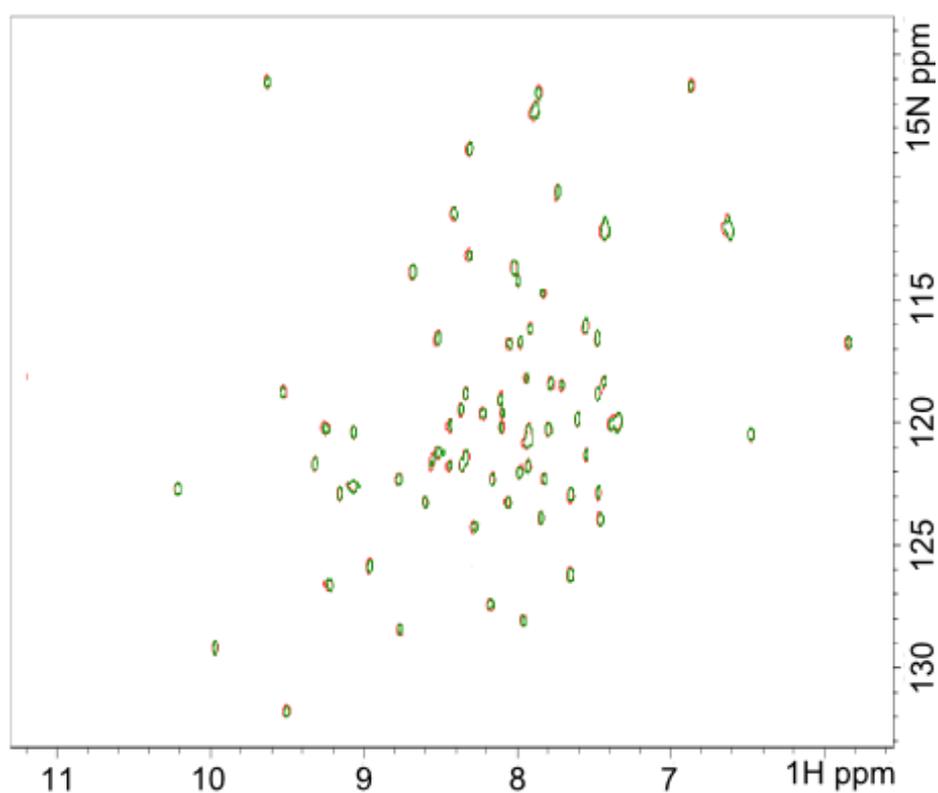


Figure 6.19 – 25 μM ABD2 without (green) and with (red) 50 μM LIM2.

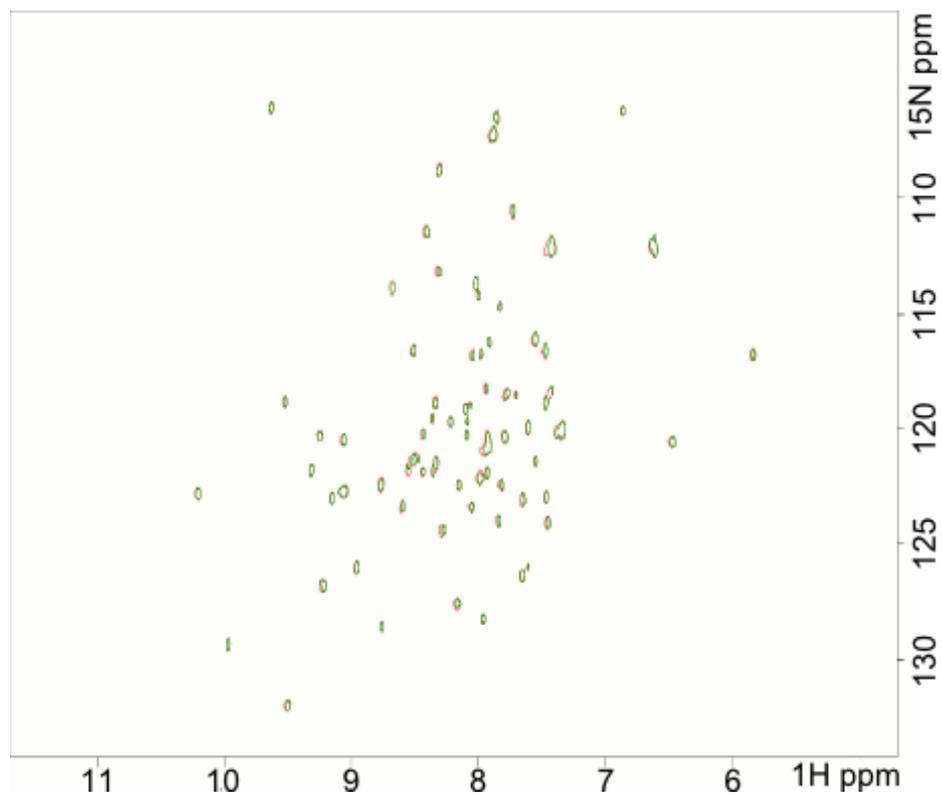


Figure 6.20 – 25 μM ABD2 without (green) and with (red) 50 μM LIM3.

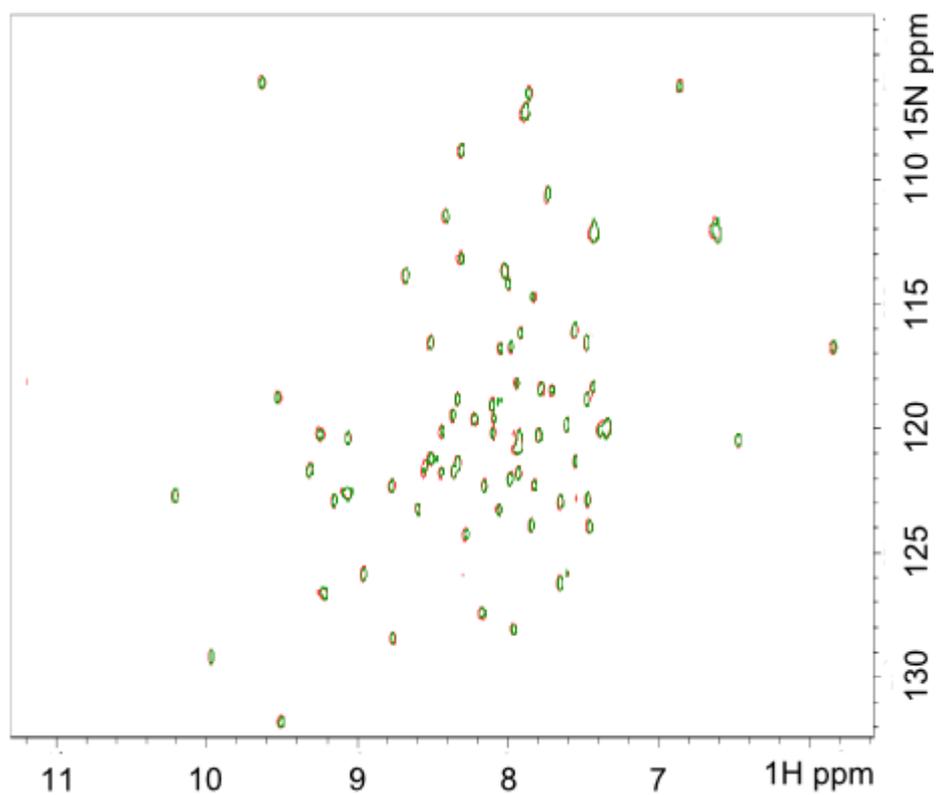


Figure 6.21 – 25 μM ABD2 without (green) and with (red) 50 μM LIM4.

None of the individual LIM domains bound to ABD2. The same experiments were then attempted with ABD1. Once again, no interaction was seen. The tandem LIM domain constructs were then tested against a tandem of ABD1 and ABD2 in the same way.

There are no differences between the HSQC peaks of the tandem of ABD1 and ABD2 without (green) and with (red) the tandem of LIM1 and LIM2 (figure 6.22). This shows that there is no binding between the tandem of ABD1 and ABD2 and the tandem of LIM1 and LIM2.

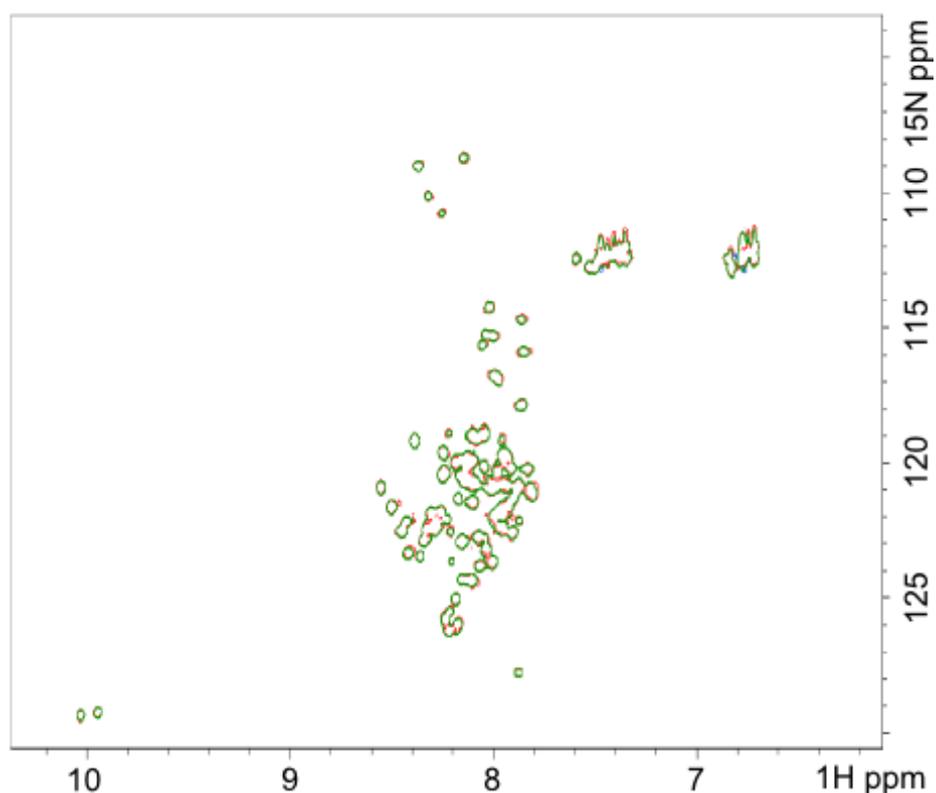


Figure 6.22 – 25  $\mu\text{M}$  tandem of ABD1 and ABD2 without (green) and with (red) 50  $\mu\text{M}$  tandem of LIM1 and LIM2.

There are no differences between the HSQC peaks of the tandem of ABD1 and ABD2 without (green) and with (red) the tandem of LIM2 and LIM3 (figure 6.23). This shows that there is no binding between the tandem of ABD1 and ABD2 and the tandem of LIM2 and LIM3.

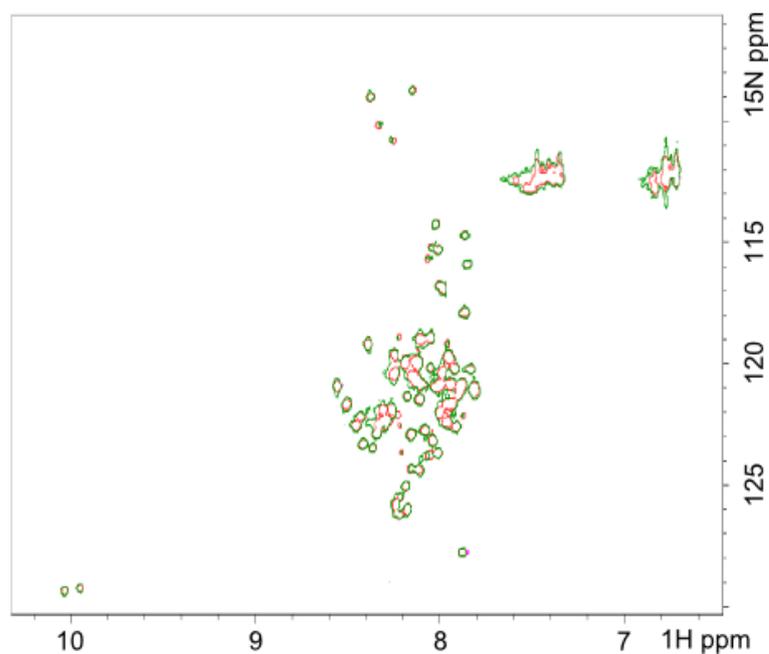


Figure 6.23 – 25  $\mu\text{M}$  tandem of ABD1 and ABD2 without (green) and with (red) 50  $\mu\text{M}$  tandem of LIM2 and LIM3.

There are no differences between the HSQC peaks of the tandem of ABD1 and ABD2 without (green) and with (red) the tandem of LIM3 and LIM4 (figure 6.24). This shows that there is no binding between the tandem of ABD1 and ABD2 and the tandem of LIM3 and LIM4.

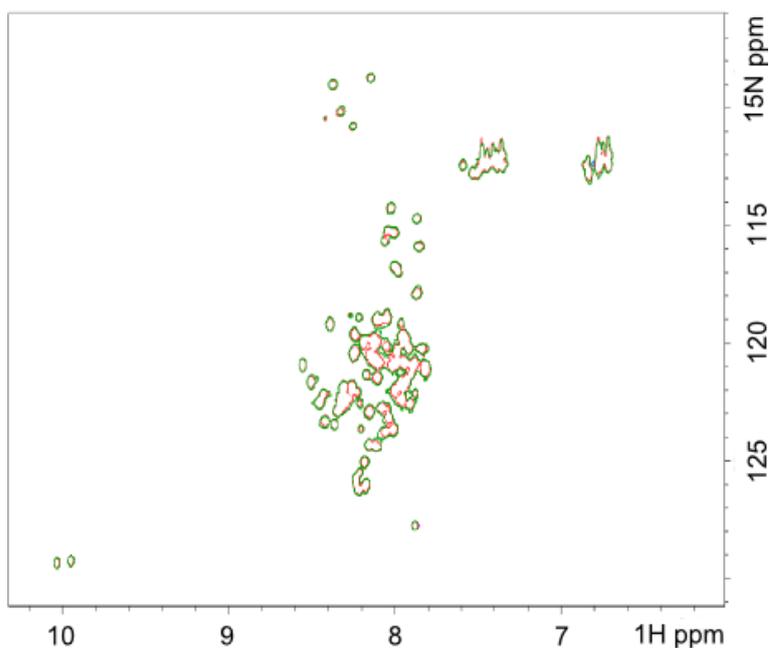


Figure 6.24 – 25  $\mu\text{M}$  tandem of ABD1 and ABD2 without (green) and with (red) 50  $\mu\text{M}$  tandem of LIM3 and LIM4.

None of the LIM tandem domains bind to the tandem of ABD1 and ABD2. The villin headpiece domain was then tested for binding to ABD1 and ABD2. This used the same technique but in this experiment, the  $^{15}\text{N}$  labelled protein was the villin headpiece domain, and the tandem of ABD1 and ABD2 was added to it to see if binding occurs.

There are no differences between the HSQC peaks of the tandem of VHD without (green) and with (red) the tandem of ABD1 and ABD2 (figure 6.25). This shows that there is no binding between the villin headpiece domain and the tandem of ABD1 and ABD2.

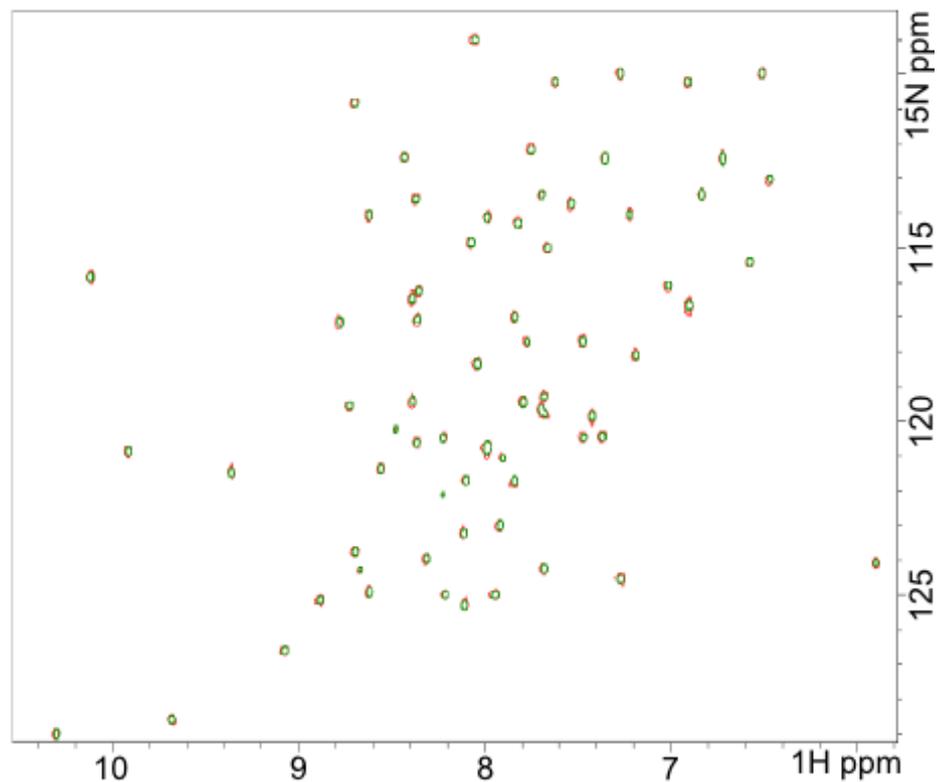


Figure 6.25 – 20  $\mu\text{M}$  villin headpiece domain without (green) and with (red) 40  $\mu\text{M}$  tandem of ABD1 and ABD2.

The villin headpiece domain of ABLIM-2 also shows no binding to the region of ms1/STARS identified as the ABLIM-2 binding region (ms1/STARS residues 234-375).

None of the domains of ABLIM-2 appeared to bind to any part of the C-terminal of ms1/STARS when tested by NMR. Further work on this is being performed by Dr. M. Dickens at the University of Leicester. Yeast-2-hybrid analysis screening is being used to try to identify the binding region in ABLIM2.

### **6.3 Discussion of ABLIM Binding to ms1/STARS**

Previous work showed that ABLIM-2 and -3 bound to ms1/STARS in the C-terminal of ms1/STARS (residues 234 – 375) (Barrientos *et al.*, 2007). This interaction was discovered by yeast-2-hybrid screening. They could not identify which parts of ABLIM-2 and -3 interacted with ms1/STARS.

ABLIM-2 and -3 contain 4 LIM domains. LIM domains are often protein-protein interaction domains and are therefore good candidates for being the site of the interaction between ABLIM-2 and the C-terminal of ms1/STARS.

The LIM domains were expressed; with 100  $\mu$ M zinc being added to the LB medium at the litre growth stage in order to encourage correct folding as LIM domains are zinc finger domains, which require zinc to maintain their structure.

When possible interaction between the individual LIM domains and ABD1, ABD2 and the tandem of ABD1 and ABD2 was investigated using NMR and pulldown assays, no interaction was seen.

The same experiments were then attempted with tandems of the LIM domains (LIM 1+2, LIM 2+3 and LIM3+4). Once again, no interaction was seen.

ABLIM-2 and -3 contain four LIM domains followed by a linker region, which links the LIM domains to the villin headpiece domain which is the C-terminus of ABLIM-2 and -3. The villin headpiece domain binds to actin. As well as the four LIM domains and the LIM domain tandems, the villin headpiece domain was also expressed, and interaction was assayed for this domain also. No interaction was found.

The lack of interaction in the experiments could be due to all commercially available ABLIM-2 cDNA being the cDNA of isoform 2 of ABLIM-2, in which residues 506 to 545 in the linker are missing and have been replaced by Q (figure 6.1). Binding could occur here. Another possibility is that a combination of domains that was not attempted was required for binding; however attempts at expressing longer constructs in *E. coli* were unsuccessful.

Further yeast-2-hybrid experiments are being performed by collaborators in order to further investigate the binding of ABLIM-2 and -3 with ms1/STARS. These have found that only full length ABLIM-2 binds to ABD1, ABD2 or a tandem of the two, and that separate domains of ABLIM-2 do not (Lewis *et al.*, unpublished).

# **Chapter 7 – DNA binding by ms1/STARS**

## **Chapter 7 - DNA Binding by ms1/STARS**

Once the structure of ms1/STARS ABD2 had been determined (see chapter 4), it was found that it shared a strong structural similarity to several winged helix-turn-helix domains (see chapter 4.3.8). The winged helix-turn-helix domains most similar to ABD2 have DNA-binding activity. As the structure of ABD2 is similar to these domains, it is possible that ABD2 also binds to DNA. For this to occur, the residues involved would need to be conserved.

### **7.1 Conservation of Residues Between the Recognition Helix of Structurally Similar Winged Helix-Turn-Helix Domains and Helix-3 of ABD2**

Helix-3 is the recognition helix in DNA-binding winged helix-turn-helix domains (an example of this can be seen in figure 7.1). The residues found in helix-3 of the most structurally similar winged helix-turn-helix domains and those in the third helix of ABD2 were compared. Figure 7.2 shows the conserved residues and the level of conservation. The position of these residues on the structure of ABD2 can be seen in figure 7.3. The residues highlighted are those which are conserved in both ABD2 and the proteins with which it shares a structural homology. All the numbers given are those of the residues in ABD2. S338 is conserved; this occurs because serines are often found at the start of  $\alpha$ -helices due to their structural properties. The two residues in blue, V342 and L345, are conserved as non-polar, hydrophilic residues. R347 is conserved as an arginine, while K350 and H351 are conserved as positively charged residues (K and R, and H in some conditions) in the similarly structured DNA binding proteins.

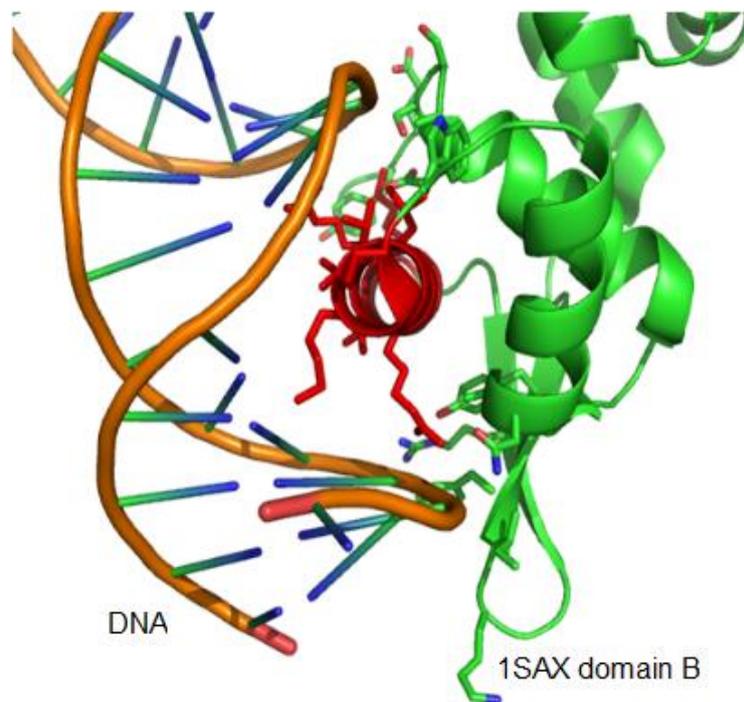


Figure 7.1 – 1SAX domain B, a protein domain structurally similar to ABD2, in complex with DNA. Helix-3, the DNA-binding helix, is highlighted in red.

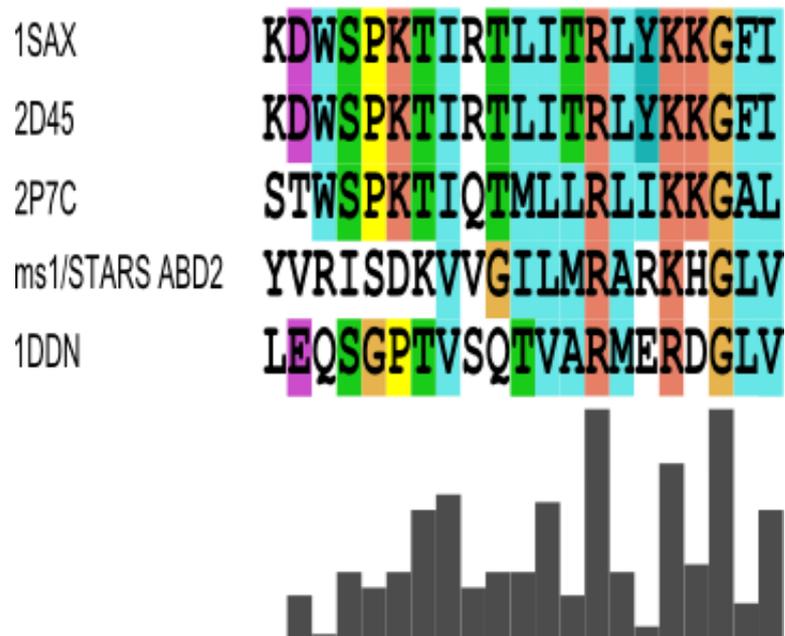


Figure 7.2 – Alignment of the DNA recognition helix of 4 winged helix-turn-helix domains with a similar structure to that of ABD2 of ms1/STARS. 1SAX = *S. Aureus* methicillin-resistance regulating transcriptional repressor MecI, 2D45 = the MecI-MecA repressor-operator complex, 2P7C = *B. Lichenformis* BlaI monomeric form in complex with the BlaP half-operator and 1DDN = Diphtheria toxin repressor.

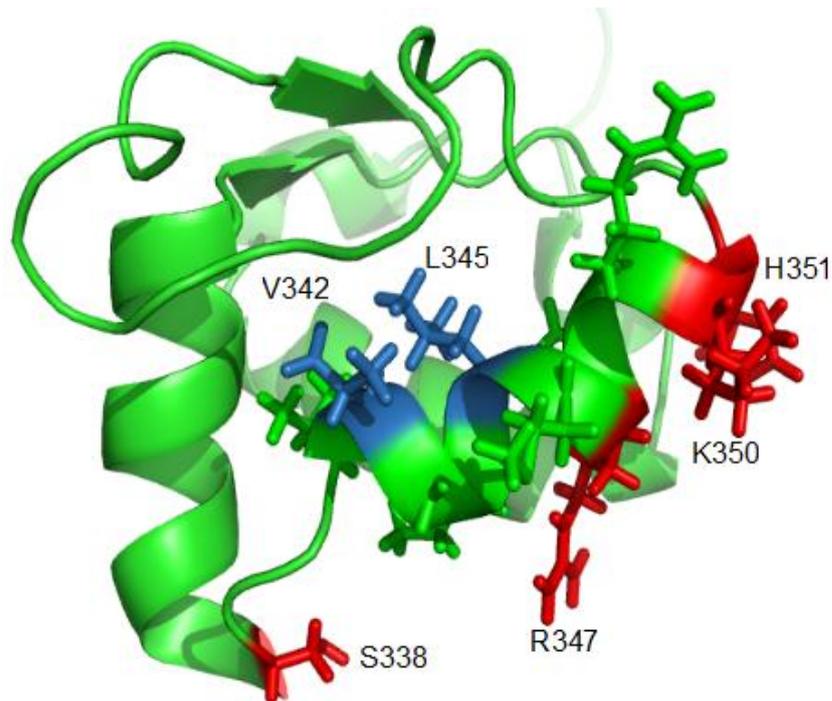


Figure 7.3 – Residues conserved in helix 3 of ABD2 when compared to DNA binding domains of similarly structured proteins. All numbers given are those of the residues in ABD2. S338 is conserved, as serines are often found at the start of helices. The two residues in blue, V342 and L345, are conserved as non-polar, hydrophilic residues. R347 is conserved as an arginine, while K350 and H351 are conserved as positively charged residues in the similarly structured DNA binding proteins.

## **7.2 DNA Binding to ABD2**

There is some conservation of the residues of helix-3 between ABD2 and the structurally similar DNA binding proteins. This means that ABD2 could bind to DNA. It was decided to test this experimentally. The DNA of the BACH1 (5'-3' = GTTGAGCACAGCAGTGCTGAGTCATGCTGAGTCATGCTGAG, 3'-5' = TCTCAGCATGACTCAGCATGACTCAGCACTGCTGTGCTCAA) promoter was used for this, as another group in the University of Leicester Biochemistry Department had used this in gel shift assays, which meant that there was enough BACH1 DNA available to enable this experiment to be performed. An agarose DNA gel shift assay

was used to test if ABD2 bound to DNA. In a DNA gel shift assay performed this way (see section 2.22.2) binding of the DNA by the protein is seen by the shifting of the DNA. If there is no binding, then the DNA remains as one band, as can be seen in lane 1 where there is 1  $\mu\text{M}$  DNA and 1  $\mu\text{M}$  ABD2. If there is binding, the band becomes shifted and smeared. The gel (figure 7.4) shows a small shift when the ratios of DNA : protein are 1  $\mu\text{M}$  : 100  $\mu\text{M}$ , and a smaller but more clearly seen shift at 3  $\mu\text{M}$  : 90  $\mu\text{M}$ .

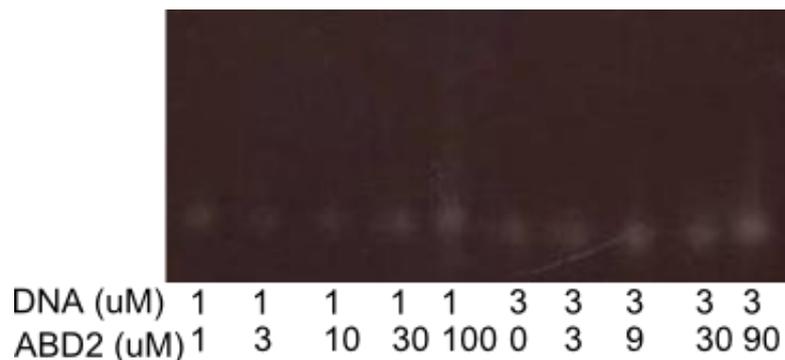


Figure 7.4 – Agarose gel showing the results of a DNA gel shift assay performed on ABD2. There is very weak binding at high protein concentrations.

There was very weak binding to DNA by ABD2. As ABD1 also bound to actin, we then attempted to find out if ABD1 bound to DNA. As the tandem of ABD1 and ABD2 bound more tightly to actin than the individual domains, we also tested its ability to bind to DNA.

### **7.3 DNA Binding to Tandem (ABD1 and ABD2), ABD1 and ABD2**

The same gel shift technique was used to study the binding of the tandem of ABD1 and ABD2, ABD1 and ABD2. The gel (figure 7.5) shows that the tandem binds more tightly than either domain individually. There is a strong shift in the lane which contains a ratio of 3  $\mu\text{M}$  DNA : 90  $\mu\text{M}$  tandem. There is a weaker shift in the lane which contains 3  $\mu\text{M}$  DNA : 90  $\mu\text{M}$  ABD1 and only a very weak shift in the lane which

contains 3  $\mu\text{M}$  DNA : 90  $\mu\text{M}$  ABD2. As shifts occur due to binding, this shows that the strongest binding is by the tandem.

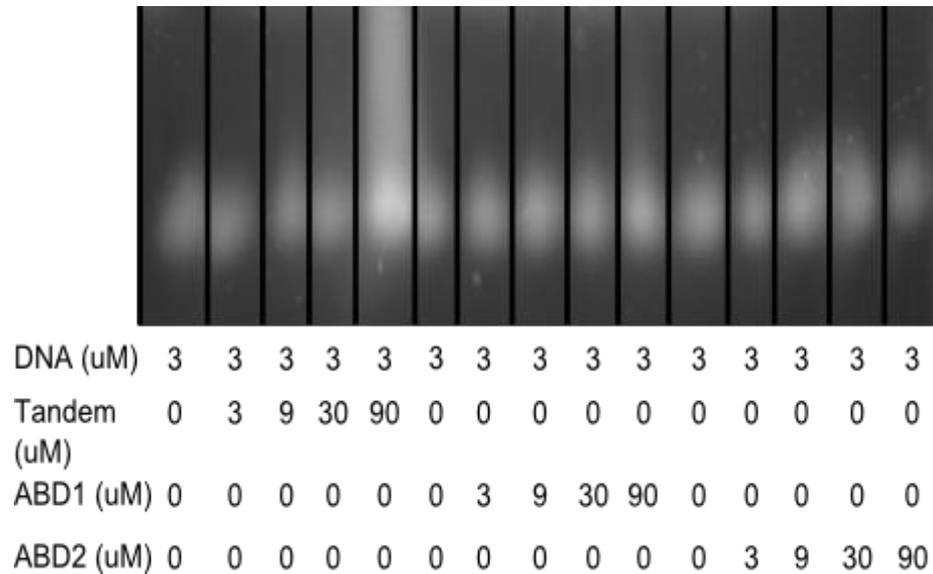


Figure 7.5 – Agarose gel showing the results of a DNA gel shift assay performed on ABD1, ABD2 and the tandem of ABD1 and ABD2.

As the tandem bound more strongly than the individual domains, the binding of the tandem to DNA was investigated further. The agarose gel of this experiment (figure 7.6) shows that the DNA begins to shift because of binding when the DNA : protein ratio is 1  $\mu\text{M}$  : 1  $\mu\text{M}$  and 3  $\mu\text{M}$  : 3  $\mu\text{M}$ .

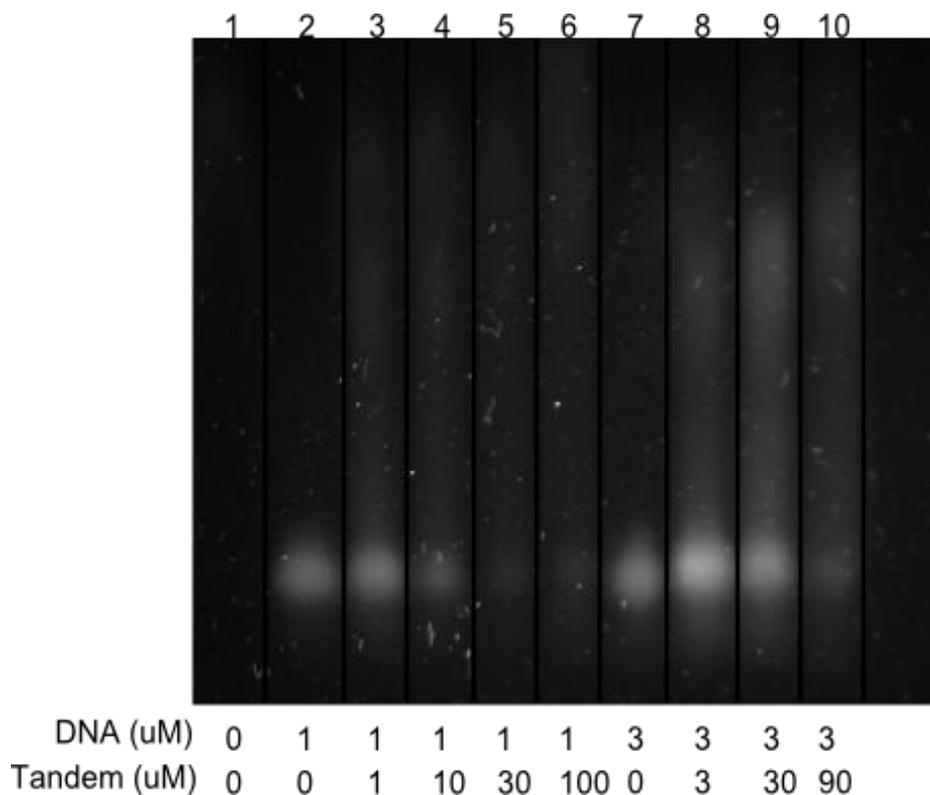


Figure 7.6 – Agarose gel showing the results of a DNA gel shift assay performed on the tandem of ABD1 and ABD2.

#### **7.4 Interesting Features of ABD1 Revealed By DNA Binding**

When an HSQC is performed on ABD1 on its own, the peaks are broad and are highly overlapped, as is typical for an unfolded protein. However, when HSQCs are performed on ABD1 bound to actin or DNA, the peaks of ABD1 become sharper and less broad (figure 7.7). A possible explanation for this behaviour could be that ABD1 is a molten globule.

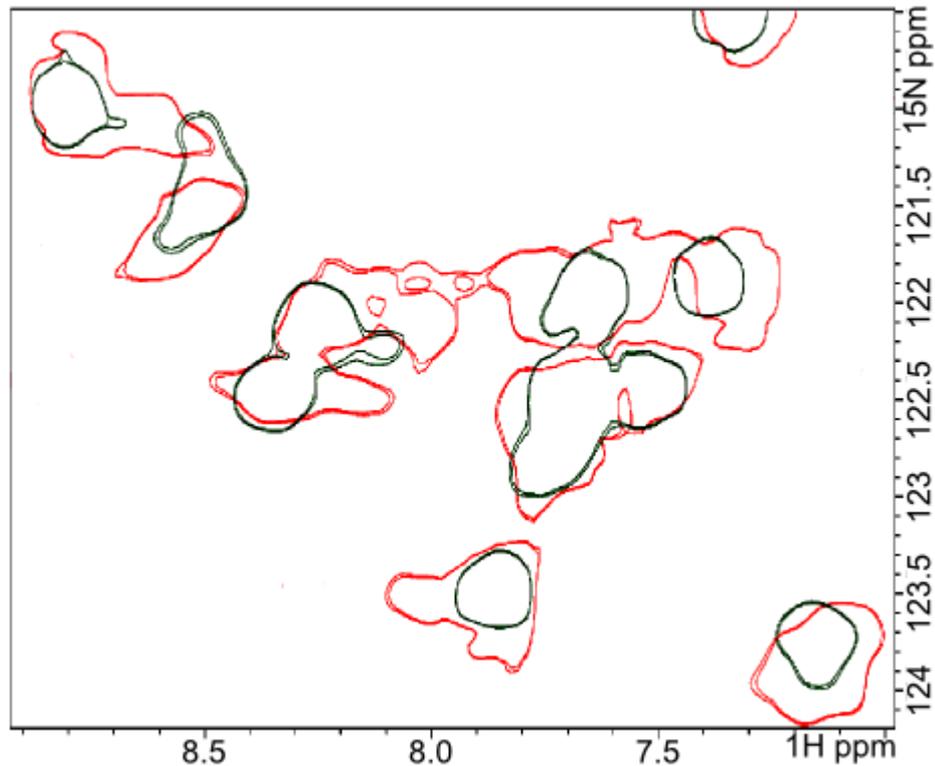


Figure 7.7 – 50  $\mu\text{M}$  ABD1 without (red) and with (black) 200  $\mu\text{M}$  DNA. The peaks of ABD1 are sharper when DNA is added.

Molten globules are proteins that have some residual native (folded) secondary protein structure but which do not have a fixed tertiary structure. They therefore appear to be unfolded by CD and NMR. This lack of fixed tertiary structure is due to an absence of the close packing of amino acid side chains that is found in a folded protein.

Many other unstructured DNA binding domains become folded upon binding to DNA (Spolar & Record, 1994). In the DNA binding region of leucine zipper GCN4 this folding increases when it binds to its specific DNA site (Weiss *et al.*, 1990). This suggests that, if the specific DNA that the tandem of ABD1 and ABD2 binds to is found and the experiment is repeated, ABD1 may become more folded on binding. If this occurs, the peaks of ABD1 may become more dispersed, aiding residue identification and structure determination. Something similar appears to occur when the binding of ABD1 to actin is studied (see chapter 5, section 1). ABD1 appears to become folded on

binding to actin, as the peaks of ABD1 disappear when actin is added, and this would only happen if a co-operative process such as binding occurred. If ABD1 remained unstructured on binding to actin no peaks would disappear.

As those ABD1 peaks that shift upon DNA binding and have been identified are all in the last 20 residues of ABD1, it may also be possible to produce a construct which features these last 20 residues and ABD2, which may give a better spectrum than the full tandem of ABD1 and ABD2 and enable a better understanding of the DNA and actin binding functions of the C-terminal of ms1/STARS.

### **7.5 The Binding of the Tandem (ABD1 and ABD2) to DNA Using NMR**

To discover which residues are involved in the binding of the tandem of ABD1 and ABD2, it was  $^{15}\text{N}$  labelled. An HSQC spectrum was taken of  $50\ \mu\text{M}$  of the tandem without (black) and with (red)  $200\ \mu\text{M}$  DNA. These HSQC spectra were overlaid and a section from them can be seen below in figure 7.8. Those residues of ABD2 which could be identified are labelled in the figure.

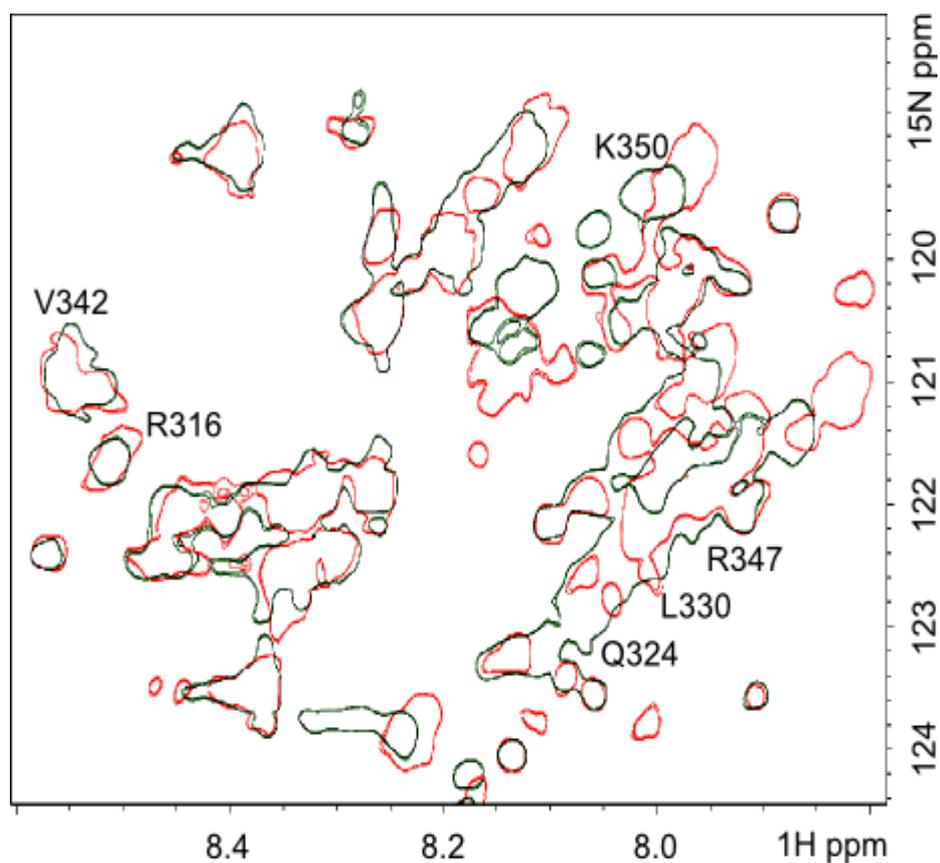


Figure 7.8 – NMR spectrum of the tandem of ABD1 and ABD2 (50 μM) without (black) and with (red) DNA (200 μM). There are significant changes to the position and strength of some peaks.

Figure 7.9 is a plot of the residues of ABD2 versus the change in their chemical shift when DNA is added. It shows that there are distinct regions where the shifts occur, showing that the binding is done by specific regions of ABD2. These residues are shown on the structure of ABD2 in figure 7.10. The two residues of ABD1 that could be identified and which shift upon DNA binding are also shown on the structure.

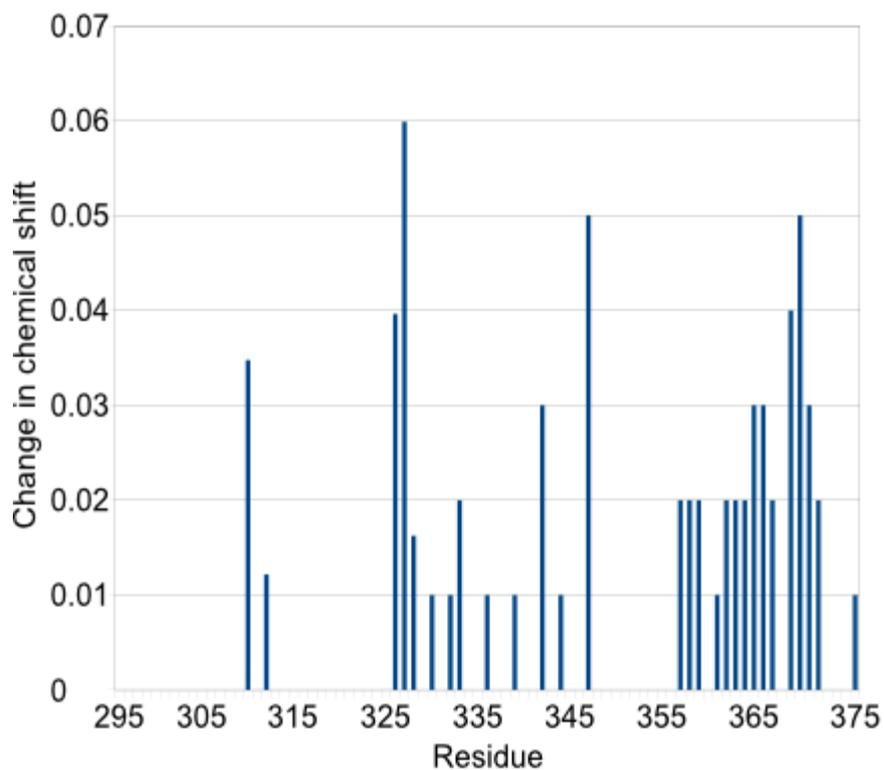


Figure 7.9 – Plot of change in chemical shift against residue number. Those residues where the peak could not be clearly identified in the tandem construct were given a value of 0.

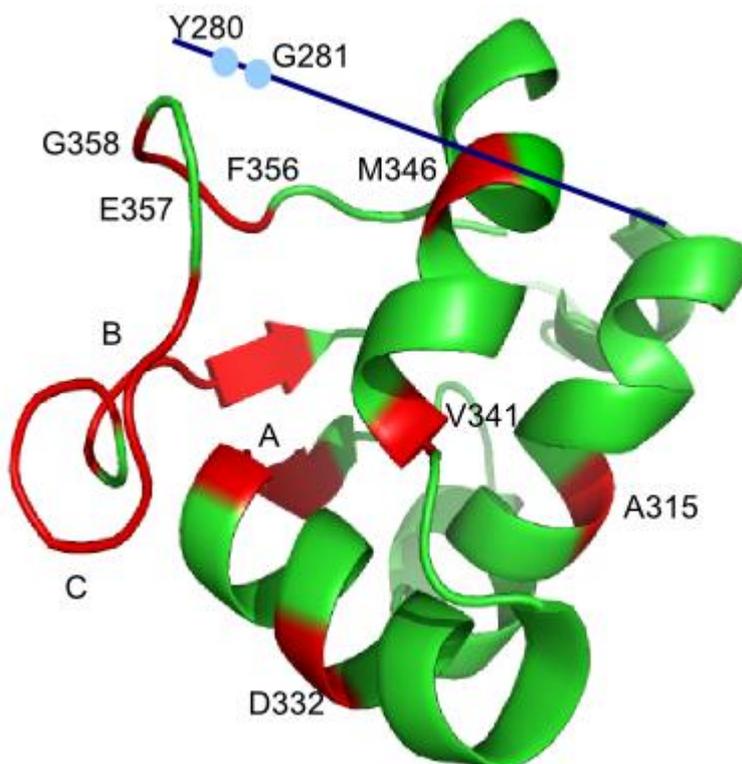


Figure 7.10 – The residues of ABD2 that are shifted on binding to DNA are shown in red. Group A is residues 325-327, group B is residues 361-366 and group C is residues 368-371. The C-terminal of ABD1 is shown in blue, and the blue dots are those residues of ABD1 which have been identified and shift on DNA binding.

The residues which shift are found in the wing and in helix 2 (residue F327 – S338). In winged helix-turn-helix domains that bind to DNA, the wing is known to make contacts with the DNA (Gajuwala and Burley, 2000). In the DNA binding of the heterodimer of E2F4 and DP2, helix 2 is involved in the dimerisation that enables the binding to DNA (Zheng *et al.*, 1999). Other winged helix-turn-helix domain proteins that bind without using the “recognition helix” (which would be helix 3 in ABD2) include FokI (Wah *et al.*, 1997). Another possible reason for the lack of shifts of the residues in helix 3 is that many of the residues of this helix are found in the centre of the spectrum and may therefore be hidden by ABD1 peaks in the tandem spectrum.

## **7.6 Discussion of the DNA Binding Activity of ABD1, ABD2 and the Tandem of ABD1 and ABD2**

As ABD2 has structural homology with winged helix-turn-helix domains which bind to DNA, if the DNA interaction helix (helix 3 in WHTH domains), was conserved, then there was a possibility that ABD2 could also bind to DNA. These residues were conserved (figure 7.2 and 7.3). An alignment of the residues of the 3<sup>rd</sup> helix in ABD2 and its homologues can be seen below (figure 7.11).

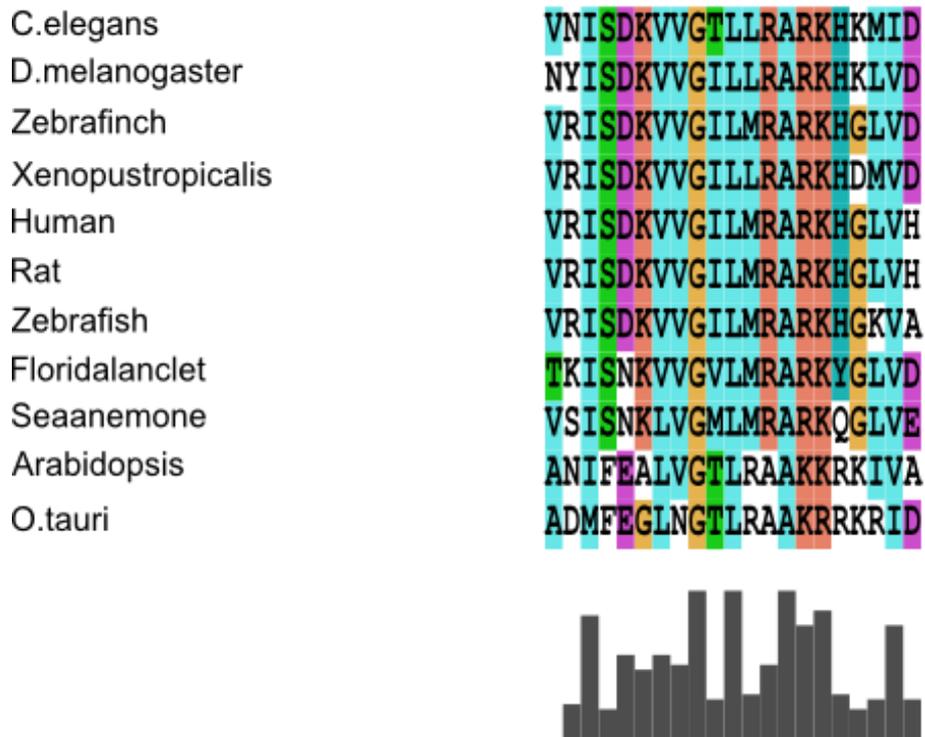


Figure 7.11 – Alignment of ms1/STARS and its homologues in the region which is the 3rd helix in ABD2.

The alignment of the last two homologues, which come from less complex organisms, appears shifted. These differences could help these homologues recognise different DNA sequences, as the 3<sup>rd</sup> helix is the recognition helix. Differences in this helix in other helix-turn-helix proteins cause them to bind to different DNA sequences (Gajiwala & Burley, 2000).

The conserved residues in ABD2 were compared to the residues of the third helix of winged helix-turn-helix domains with similar structure (figure 7.2). There is conservation of the positively-charged residues that are the equivalent to those that have been shown to interact with DNA in similar WHTH domains whose structures have been solved in the presence of DNA (figure 7.1).

Because of the conservation of these residues, it was felt that there was a possibility that ABD2 bound to DNA. The experiment was performed with double stranded DNA from BACH1. This DNA is not expected to be specifically bound by ABD2, and therefore binding will be weak.

When the experiment was performed, there was indeed very weak binding, only when ABD2 is 100 fold in excess.

Monomeric WHTH domains lose their ability to bind to DNA when the second wing is cleaved (Kitano *et al.*, 2010). ABD2 appears to have no second wing, certainly not in the C-terminal. However, the last 20 residues of ABD1 contain a large number of arginines and lysines, similar to the second wing of WHTH domains.

When the DNA binding assay was performed on ABD1 it showed that ABD1 binds to DNA more strongly than ABD2, at around a 30:1 protein to DNA ratio. The tandem of ABD1 and ABD2 was also tested for DNA binding and was shown to bind much more tightly than either ABD1 or ABD2, starting at around 10:1 protein to DNA ratio. This suggests that while ABD2 exists as a monomer, its endogenous binding to DNA may be mediated by ABD1, the C-terminal of which acts as a second wing. DNA binding activity may require ABD2 to identify the specific DNA to which ms1/STARS binds and then ABD1 performs the actual binding. The unspecific nature of the binding may explain why the residues involved in binding are not those of helix 3, which makes base specific contacts, but those of wing 1, which binds non-specifically to the phosphate backbone of the DNA. Identifying the consensus sequence to which ABD2 binds to could lead to a greater understanding of the function of ms1/STARS. One method of

identifying this sequence would be to use SELEX. SELEX is the Systematic Evolution of Ligands by Exponential enrichment (Tuerk & Gold, 1990). SELEX, when used to find the specific DNA to which a DNA binding protein binds, uses randomised DNA. DNA is selected if it binds to the protein. The DNA is then amplified. This is repeated until the specific DNA is reached. The point where the specific DNA is identified is measured by changes in the binding co-efficient. Once there are no further changes in the binding co-efficient, it is confirmed that the specific DNA has been found (figure 7.12).

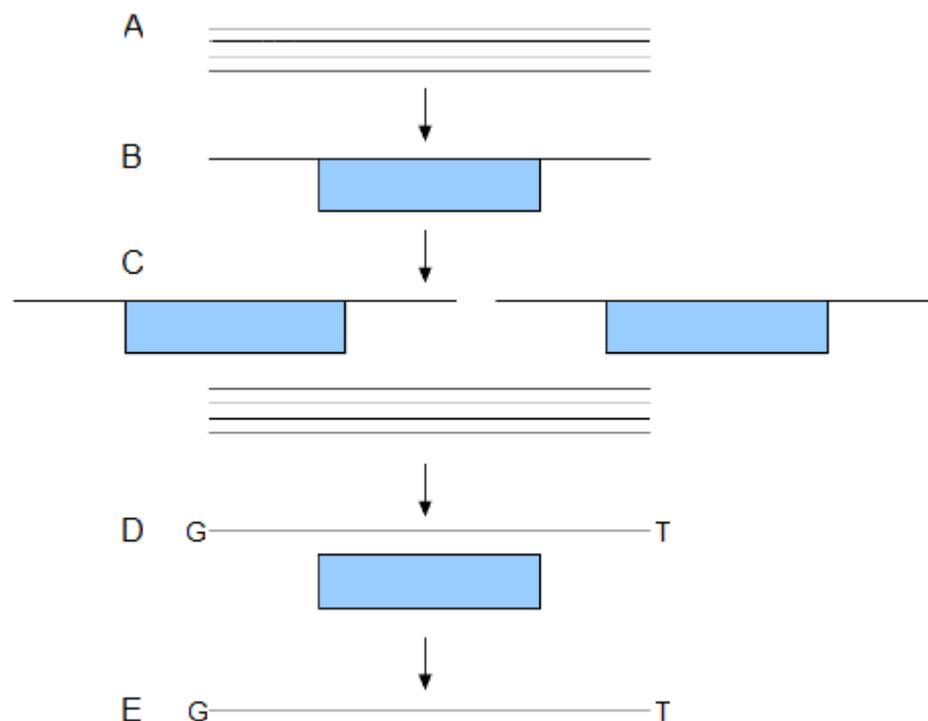


Figure 7.12 – How SELEX works. A = random oligonucleotides produced. B = oligonucleotides are mixed with DNA binding protein and allowed to bind. C = unbound oligonucleotides removed. D = protein and bound DNA are separated. E = the oligomers that bound to the protein are used as the templates for the next round of SELEX, until no increase in binding affinity is found.

An interesting feature of ABD1 is its arginine-rich C-terminal (the last 30 residues of ABD1 are SEEFD YELAMSTR**RLH** **KGDEGYGRPK** EGT**KT**), and it is also conserved, but not as conserved as ABD2 (figure 7.13).

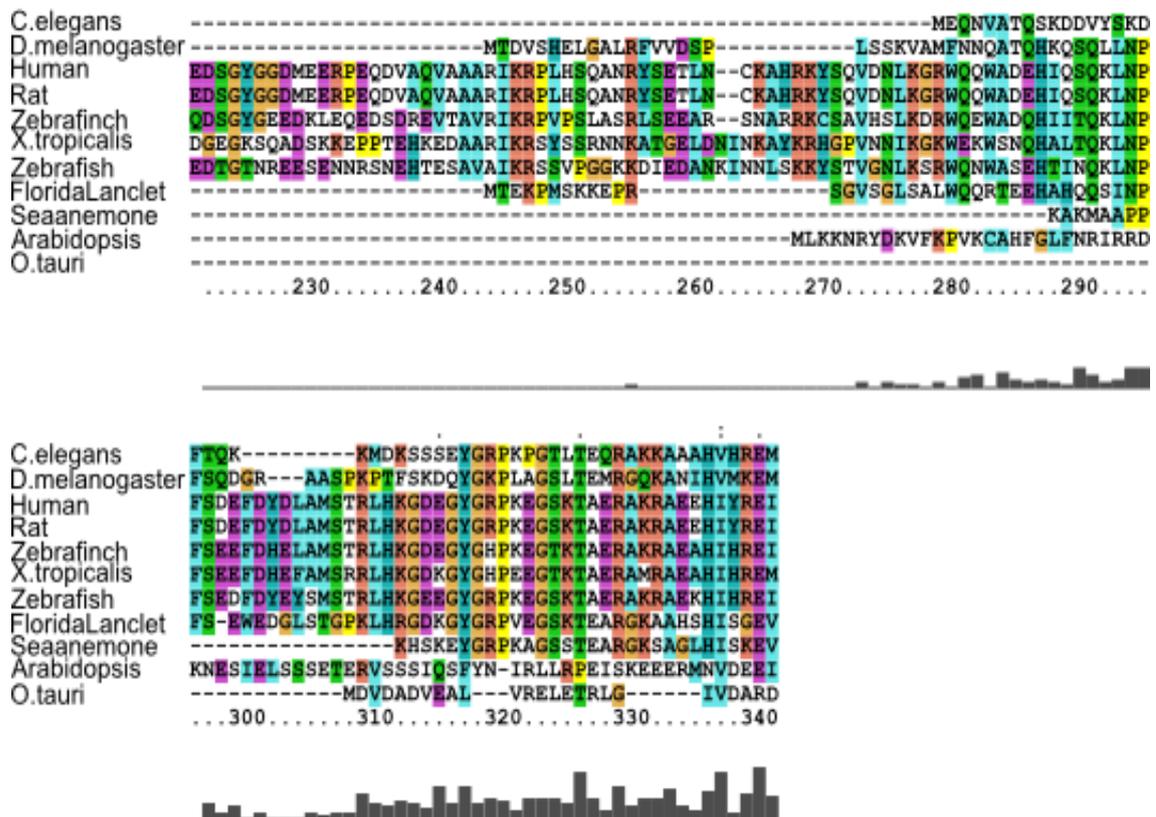


Figure 7.13 – An alignment of the C-terminal of proteins with sequence homology to ms1/STARS, focusing on the ABD1 region of ms1/STARS. The aligned proteins are C. elegans - NP001129772, D. melanogaster - NP569983, Human – NP631905, Rat NP787038, Dog XP539120, Zebrafinch XP002199130, Xenopus tropicalis NP001106590, Zebrafish CAK03875, Florida lancllet XP002605450, Sea anemone XP001619805, Arabidopsis CAA20577 and O. tauri CAL57253.

Arginine is a positively charged amino acid, which means it could bind to the negatively charged phosphate backbone of DNA. As ABD1 is an unstructured domain, the C-terminal of ABD1 could act as a second wing. Although most W2s in winged helix-turn-helix domains are C-terminal to the rest of the domain, some WTH proteins have their second wings at their N-terminal. This group includes the C-terminal accessory domain of methionine aminopeptidase 2 (Liu *et al.*, 1998). Those residues of ABD1 that have been identified in the tandem spectrum as shifting when DNA binding occurs are all found in the C-terminal of ABD1. Similar folding on DNA binding occurs in wing 2 of the RecQ C-terminal domain of human WRN protein (Kitano *et al.*, 2010), a non-specific DNA binding winged helix-turn-helix domain.

# **Chapter 8 - Discussion**

## **Chapter 8 - Discussion**

### **8.1 - Introduction**

The purpose of this research project was to determine the domain organisation of ms1/STARS and to elucidate the actin binding function of the domains.

ms1/STARS was implicated in the development of the heart condition left ventricular hypertrophy (LVH) (Mahadeva *et al.*, 2002). Left ventricular hypertrophy develops in some cases of high blood pressure. Individuals with LVH have an increased risk of heart attack and stroke. ms1/STARS is upregulated in rat models of LVH an hour after aortic banding (Mahadeva *et al.*, 2002). Aortic banding was used to mimic the increased pressure which results in LVH. Following the upregulation of ms1/STARS, early response genes and then foetal isoforms of cytoskeletal proteins such as alpha-actin and beta-myosin heavy chains are expressed. This leads to remodelling of the heart and the development of left ventricular hypertrophy.

### **8.2 The Domain Organisation of ms1/STARS**

Before the start of this project, nothing was known about the domain organisation of ms1/STARS other than deletion of the residues 234-375 caused the abolition of actin binding (Arai *et al.*, 2002). However, it was not known if this region alone was sufficient for actin binding.

It was believed that ms1/STARS consisted of several domains, since, when the full-length protein is expressed in *E. coli.*, it degrades into distinct bands, suggestive of

domains, rather than the smear that would be produced by the degradation of a protein without domain organisation.

One putative domain, ABD2, was found using rational domain design. The other three putative domains, msd1, msd2 and ABD1, were found by DomainX, a company who use combinatorial domain hunting to look for solubly expressed putative protein domains.

The only conserved region, ABD2, has homologues throughout eukaryotic organisms, from mammals, reptiles, birds and fish, through to nematode worms, flies and yeast. There are no known prokaryotic homologues. This could be because prokaryotes do not use actin as a cytoskeletal protein.

The putative domains were characterised in terms of their folding using NMR and CD. The first three putative domains, msd1, msd2 and ABD1, were unfolded. ABD2 was folded. The putative domains were then characterised in terms of their actin binding ability using a co-sedimentation assay. msd1 and msd2 did not bind to actin, while ABD1 and ABD2 did. ABD2 is therefore independently folded and independently active, suggesting it is indeed a domain. Although ABD1 is not independently folded, its independent activity suggests that it too is a domain. The two N-terminal domains, msd1 and msd2, show neither feature and therefore remain merely putative domains, as no function is known for them.

### **8.3 The Structure of ABD2**

The structure of ABD2 was determined using nuclear magnetic resonance spectroscopy (NMR). 2700 constraints were used for this, which is greater than the number of residues x 30 minimum required. The structure was validated, by both the conditions CYANA suggests (Guentert *et al.*, 1997) and the ADIT validation server used by the PDB to validate all new protein structures that are submitted. The structure of ABD2 is mixed  $\alpha$ -helical and  $\beta$ -sheet. It contains three  $\alpha$ -helices and three  $\beta$ -strands, which together form an anti-parallel  $\beta$ -sheet. It also contains a long flexible loop, which lies between strand 2 and strand 3. The obtained structure was not found in any actin binding protein with a known structure. Therefore a search for proteins with a similar structure to that of ABD2 was performed using DALI (Holm *et al.*, 2008). The proteins with the most similar structure to ABD2 were all winged helix-turn-helix domains. The structure of ABD2 has the structural features of a canonical winged helix-turn-helix domain. A typical winged helix-turn-helix domain contains 3  $\alpha$ -helices in the N-terminal. This is followed by 3  $\beta$ -strands which form an anti-parallel  $\beta$ -sheet. Between strand 2 and strand 3 there is a loop, which is the wing which gives this family of domains its name. Many winged helix-turn-helix domains are found in DNA binding proteins as DNA binding domains.

Determining the structure of ABD2, and its part in the domain organisation of ms1/STARS, enabled the evaluation of three possible binding partners of ms1/STARS.

## **8.4 Actin binding of ms1/STARS**

Of the four putative domains of ms1/STARS - msd1, msd2, ABD1 and ABD2 - only ABD1 and ABD2 bound to actin. Previous work (Arai *et al.*, 2002) had shown that deleting either residues 234 to 279 or 346 to 375 abolished actin binding. The research performed during this project has shown that both ABD1 (residues 193 to 295) and ABD2 (residues 294 to 375) have independent actin binding activity.

Determining the structure of ABD2 led to insights into the possible location of the actin binding site of ABD2, as two large positively-charged surfaces were found on the electrostatic surface of ABD2. Mutation studies were performed on ABD2 to determine if one or both of these positively charged patches was the actin binding site. Positively-charged residues which were conserved in ms1/STARS homologues and were also found in one of the two positively-charged patches were mutated to alanines. These mutant constructs were tested using NMR to see if they were folded. One of them, R349A, was not folded. It was believed that it was the folded structure of ABD2 which produced its actin-binding ability; therefore no further studies were carried out on this unfolded mutant.

The rest of the mutant constructs were tested for actin binding. If a mutant construct bound to actin, then the original, unmutated residue was not vital for actin binding. Residues R302, R347 and K350 were shown to be vital for actin binding. These experiments showed that the actin binding site of ABD2 is found in the positively-charged patch produced by the residues in the N-terminus of helix 1 and the C-terminus of helix 3.

As ABD1 is unfolded and it produces heavily overlapped spectra, its structure could not be determined. This meant that the technique used to determine the actin binding site of ABD2 could not be used to ascertain the actin-binding site of ABD1. Even if structure determination had been possible for ABD1, because its sequence is not conserved it would not be known which structural elements to test to see if they are involved in actin binding.

The actin binding ability of ABD1, ABD2 and a tandem of ABD1 and ABD2 were quantitatively measured using a fluorescence-based assay. ABD2 had the weakest actin binding ability ( $k_D = 10.61 \mu\text{M} \pm 0.7 \mu\text{M}$ ), while ABD1 ( $k_D = 2.21 \mu\text{M} \pm 0.47 \mu\text{M}$ ) and the tandem of ABD1 and ABD2 ( $k_D = 1.07 \mu\text{M} \pm 0.40 \mu\text{M}$ ) showed strong actin binding. This is interesting and could suggest that ABD2 has another, non-actin-binding function.

### **8.5 The Binding of ms1/STARS to ABLIM2**

The C-terminal region (residues 234-375) of ms1/STARS had been shown to bind to isoform 2 of ABLIM2 by yeast-2-hybrid screening (Barrientos *et al.*, 2007). Combinations of ABLIM2 domains and the two C-terminal ms1/STARS domains (ABD1 and ABD2) were tested for interaction using NMR. No interaction was found.

### **8.6 The Binding of ms1/STARS to DNA**

When the structure of ABD2 was determined it was found to be a winged helix-turn-helix domain. These mixed  $\alpha$ -helical and  $\beta$ -sheet domains are often found as DNA-

binding domains. In particular, of the top twenty proteins with the most similar structure to ABD2, those with a known function are DNA-binding domains. There is a high level of conservation between the residues involved in DNA-binding in these winged helix-turn-helix domains and ABD2. Because of this, it was decided to experimentally test whether ABD2 bound to DNA.

ABD2 bound weakly to DNA.

The DNA-binding ability of ABD1 and the tandem of ABD1 and ABD2 were also tested. It was shown that ABD1 binds to DNA more strongly than ABD2, and the tandem binds more tightly than either domain on its own.

The DNA-binding ability of the two C-terminal domains of ms1/STARS was unknown before this project.

### **8.7 How These Results Aid Our Understanding of ms1/STARS, its Role and What This Could Mean for the Rho Signalling Pathway**

Before the start of this project little was known about the function, domain organisation and structure of ms1/STARS. It had been shown that deletion of the C-terminal (residues 234-375) of ms1/STARS abolished actin-binding (Arai *et al.*, 2002). Nothing was known about the function or organisation of the rest of the protein (residues 1-233). The biophysical characterisation data collected during this project show that ms1/STARS contains four possible domains, and that only the last of these, ABD2 (residues 294-375), is folded. Previous work (Fock *et al.*, unpublished data) showed that the C-terminal of ms1/STARS (residues 234-375) caused actin bundling through

the cross-linking of F-actin, and the actin binding experiments performed in this project showed that ABD1 (residues 193-295) and ABD2 bound to actin independently.

When the structure of ABD2 was determined, it was found to be a winged helix-turn-helix domain. Domains with this structure are often found in DNA binding domains. When ABD2, ABD1 and a tandem of the two domains were tested to see if they bound to DNA, it was found that they did.

The DNA binding function of ABD1 and ABD2 may explain the varying results found by other groups studying ms1/STARS, in terms of the cellular location of ms1/STARS (Arai *et al.*, 2002), the effect of over-expression of ms1/STARS on cells (Mahadeva *et al.*, 2002 and Troidl *et al.*, 2009) and how stressed cells respond to stimuli (Mahadeva *et al.*, 2002, Lamon *et al.*, 2009 and Troidl *et al.*, 2009).

The RhoA signalling pathway is a stress response pathway, which could explain why ms1/STARS responds to all of these different stresses (increased blood pressure (Mahadeva *et al.*, 2002), resistance training (Lamon *et al.*, 2009) and fluid shear stress (Troidl *et al.*, 2009)), and the differences in response could be due to either ms1/STARS acting as a transcription factor itself, or as a co-factor to other transcription factors. Identifying the DNA sequence that ms1/STARS binds to specifically, if it binds to a specific sequence, may help to explain this variety of responses.

The response by ms1/STARS to stresses could be vital for healthy muscle function. A morphelino knockdown of ms1/STARS in zebrafish caused musculoskeletal deformities and curvature and shortening of the longitudinal axis of the muscle (Mahadeva *et al.*, 2007), which showed that ms1/STARS is required in the adult. This suggests

ms1/STARS has a maintenance role in adult hearts and only causes pathological symptoms when the pathway is constantly rather than intermittently active. This can be seen in exercise, as modelled using resistance training. In exercise, there is only occasional use of this pathway to repair damage done during exercise. This leads to the development of thicker, stronger skeletal muscles (Lamon *et al.*, 2009). In high blood pressure, the pathway is constantly active and focused on the cardiac tissue. This means that ms1/STARS is overexpressed for a longer time period, and leads to LVH. If the pressure increase is only for a short period of time it leads to a non-pathological increase in the thickness of the left ventricular wall. This is seen in athletes in physiological LVH, which shows that it not merely overexpression, but also the length of time that the overexpression occurs that produces the effects of ms1/STARS overexpression.

The importance of the length of time ms1/STARS is expressed could be that it allows the stress signal to be recognised, for ms1/STARS to be activated, for it to translocate to the nucleus and for gene expression to occur. There is evidence that ms1/STARS could translocate into the nucleus as MSD2 contains a nuclear localisation signal (NLS) from residue 153 to residue 179 (figure 8.1). The putative NLS has only been tested in silico (Cokol *et al.*, 2000), but of the proteins in the database, 97.6% that contain this NLS motif are found in the nucleus.

	1	11	21	31	41	51	
1	MAPGETVREA	GPAKSALQKV	RRATLVINLA	RGWQQWANEN	STRQAQEPAG	WLPGATQDLP	60
61	HTPKPEGPRQ	HAPKPPSPKP	DGDREGRGSE	EATEVSHIKR	KEVTRTVVSK	AYERGGDVNY	120
121	LSHRYEHDGG	VSEAVQPDND	IDRILLSHDS	PT <b>RRRKCTNL</b>	<b>VSKLTKGWKV</b>	<b>MEQEPPKWS</b>	180
181	DSIDTEDSGY	GGDMEERPEQ	DVAQVAAARI	KRPLHSQANR	YSETLNCKAH	RKYSQVDNLK	240
241	GRWQQWADEH	IQSQKLNPF	DEFDYDLAMS	TRLHKGDEGY	GRPKEGSKTA	ERAKRAEEHI	300
301	YREIMELCFV	IRTMARHRRD	GKIQVTFGEL	FDRYVRISDK	VVGILMRARK	HGLVHFEGEM	360
361	LWQ GKDDHV	ITLLE					

Figure 8.1 – Location and sequence of the putative NLS found in MSD2.

This means there is a possibility that, instead of removing inhibition of serum response factor (SRF) by reducing the available pool of G-actin as suggested by previous research (Kuwahara *et al.*, 2005), activation of ms1/STARS by RhoA could cause ms1/STARS to translocate into the nucleus using the NLS where it either directly interacts with SRF to activate it or works with other proteins in order to activate SRF. The possibility of this mechanism of action is supported by unpublished data from Koekemoer (figure 8.2) showing that ms1/STARS is found in the nucleus of H9c2 cell when over-expressed. When the effects of fluid shear stress on endothelial cells was examined, ms1/STARS was also found in the nucleus after stress induction (Troidl *et al.*, 2009) (figure 8.3).

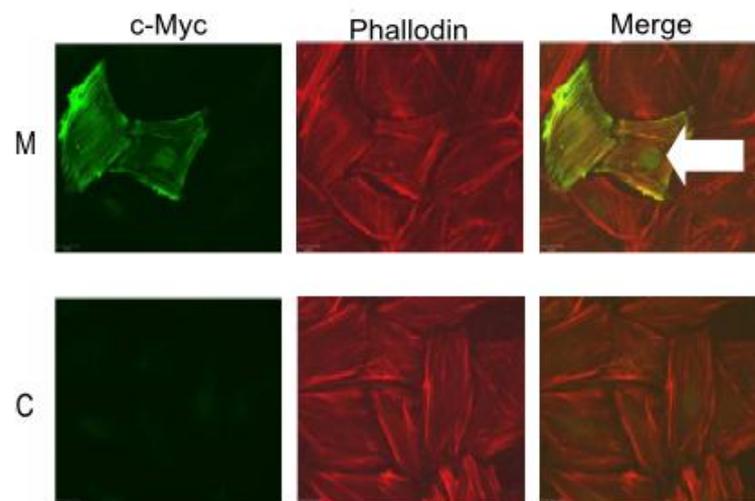


Figure 8.2 – Unpublished data from A. Koekemoer showing ms1/STARS over-expression in H9c2 cells. M = ms1/STARS expression vector, C = empty vector control. Cells were stained with c-Myc antibody and phalloidin, and then the images were merged. Only the cells with the ms1/STARS expression vector show the nucleus, showing that ms1/STARS is found in the nucleus when over-expressed.

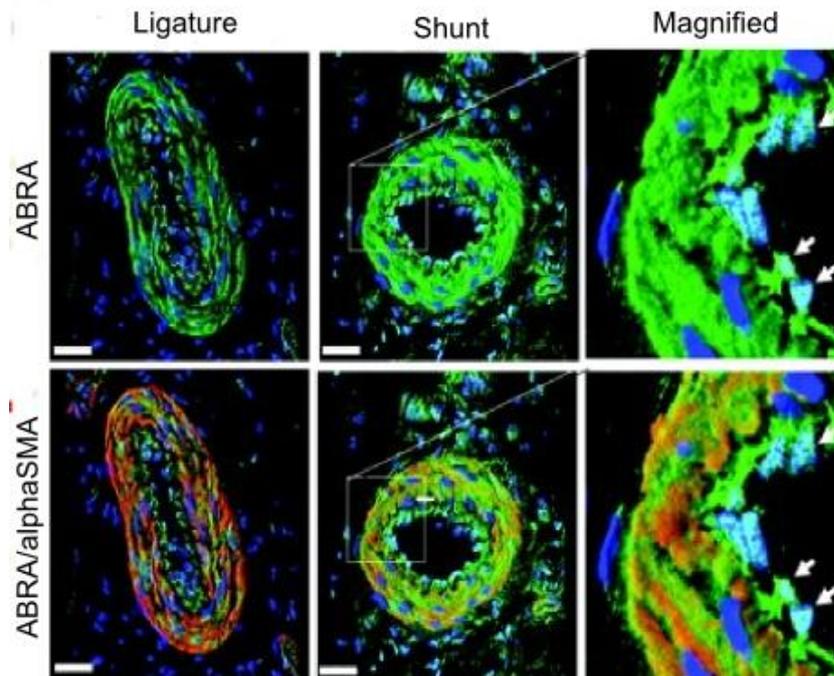


Figure 8.3 – Taken from Troidl *et al.*, “Actin-binding Rho Activating Protein (ABRA) is essential for fluid shear stress-induced arteriogenesis” (2009) *Arterioscler. Thromb. Vasc. Biol.* vol. 29 pp 2093-2101. ABRA, stained green, is found in the nucleus of fluid-shear stressed endothelial cells.  $\alpha$ -SMA, stained red, shows the cytoskeleton. The nucleus is stained blue. The merged image shows that although some of the ABRA co-localises with the cytoskeleton, some of it co-localises with the nuclear staining, as can be seen in the cyan dots, marked with the white arrow.

There exists also the possibility of secondary method of control of ms1/STARS signalling, as it has been shown that attenuation of Rho signalling does not fully attenuate STARS signalling (Arai *et al.*, 2002).

ms1/STARS contains a MyoD binding site in the promoter, and also contains an SRF binding site (Ounzain *et al.*, 2008). This means that continuous activation of ms1/STARS could create a positive feedback loop, as SRF is activated by ms1/STARS (Arai *et al.*, 2002). This may explain why hypertrophy becomes pathological with continuous exposure to increased blood pressure because the pathway is sent into overdrive.

A possible mechanism is shown below in figure 8.4. In it, stress activates RhoA. RhoA signalling then causes the disassociation of ms1/STARS from F-actin. ms1/STARS then translocates to the nucleus, using the NLS found in MSD2, where it may either bind to DNA or form part of a complex that binds to DNA. Through an unknown mechanism this then leads to SRF activation and heart remodelling.

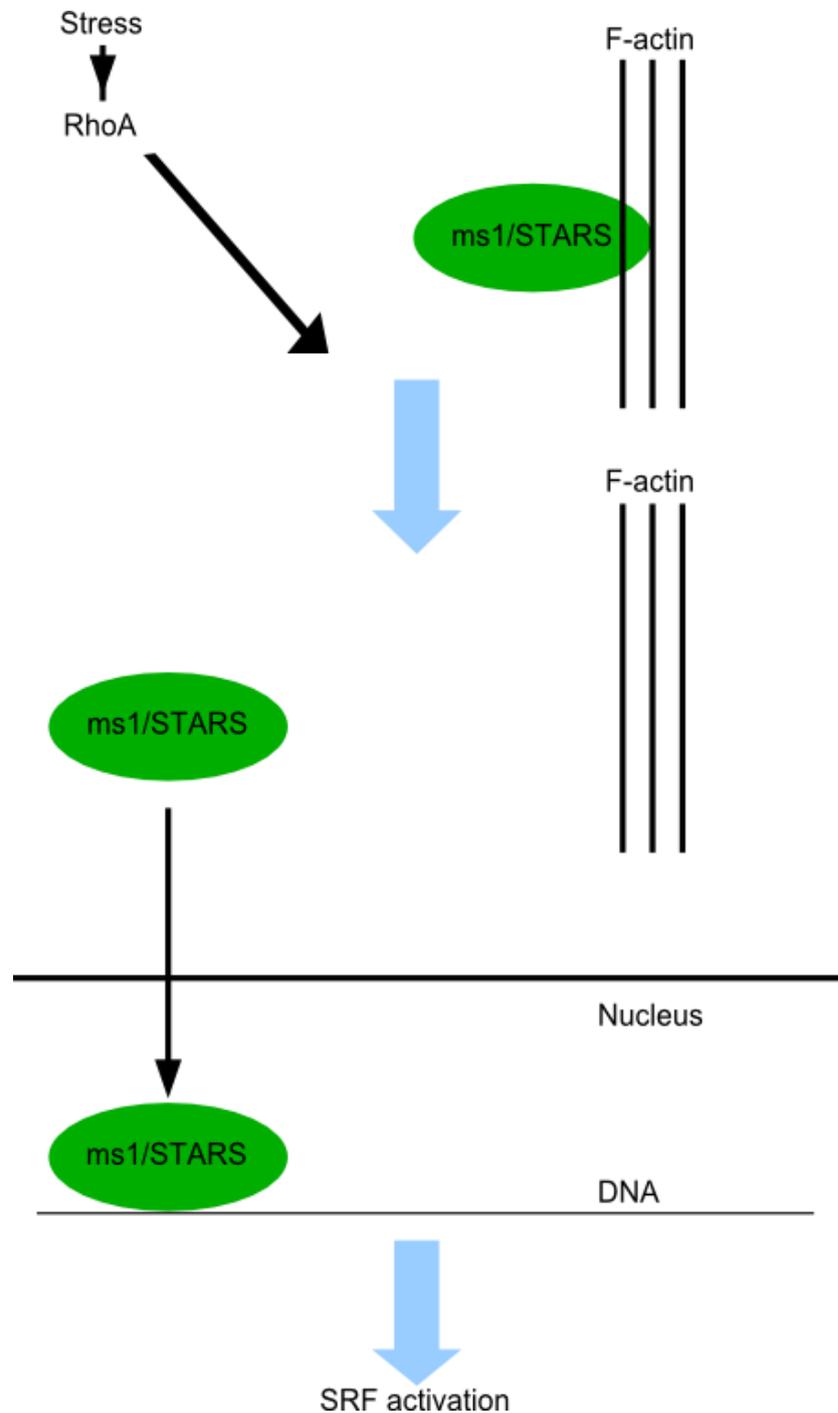


Figure 8.4 – Proposed pathway of ms1/STARS action.

This pathway is suggested as it explains why, in unstressed rat neomyocytes, ms1/STARS is located on the I band and M line of the sarcomere (Arai *et al.*, 2002) while, in stressed rat endothelial cells, ms1/STARS is found in the nucleus (Troidl *et al.*, 2009). It would explain both binding partners of ms1/STARS found during this project, the reason for the presence of an NLS in MSD2, and may explain why the tandem of ABD1 and ABD2 binds more strongly than the individual domains to DNA, but not to actin. The previously suggested role of ms1/STARS, that of actin treadmilling (Arai *et al.*, 2002), does not explain the data we have produced or that of other groups. It could be that ms1/STARS acts, with RhoA, as part of a stress sensor that enables the cell to respond to a variety of stresses which require hypertrophic growth as a response. It is part of the signalling mechanism involved in the expression of the necessary genes for hypertrophy to occur after the stress signals increase over a certain threshold.

As, in several of the cell types studied, expression of ms1/STARS is only seen when the cell is stressed, it is possible that expression of ms1/STARS may occur in other cell types and has yet not been found because these cell types have not been stressed in a way that causes ms1/STARS to respond. The effect of ms1/STARS may be tissue-specific, and it may not move into the nucleus in all cell types.

In this project, the domain organisation of ms1/STARS was established including those domains which bind to actin. The two actin binding domains are called ABD1 and ABD2. The structure of the second actin binding domain, ABD2, was determined. The structure of ABD2 is that of a winged helix-turn-helix domain. These domains frequently bind to DNA and it was discovered that ABD1 and ABD2 also bind to DNA.

# Appendices

## Appendix 1 – Sequences and Primers Used to Produce Constructs

### Sequences

#### Full length ms1/STARS

MAPGETVREA GPAKSALQKV RRATLVINLA RGWQQWANEN STRQAQEPAG WLPGATQDLP  
HTPKPEGPRQ HAPKPPSPKP DGDREGRGSE EATEVSHIKR KEVTRTVVSK AYERGGDVNY  
LSHRYEHDGG VSEAVQPDND IDRILLSHDS PTRRRKCTNL VSKLTGWKWKV MEQEPPKWK  
DSIDTEDSGY GGDMEERPEQ DVAQVAAARI KRPLHSQANR YSETLNCKAH RKYSQVDNLK  
GRWQQWADEH IQSQKLNPFV DEFYDLAMS TRHLKGDGEGY GRPKEGSKTA ERAKRAEEHI  
YREIMELCFV IRTMARHRRD GKIQVTFGEL FDRYVRISDK VVGILMRARK HGLVHFEGEM  
LWQGKDDHVV ITLLE

#### MSD1

PGEKESGEGP AKSALRKIRT ATLVISLARG WQQWANENST RQAQEPAGWL PGATQDLPHT  
PKPEGPRQHA PKPPSPKPDG DREGRGSEEA TEVSHIKRKE VTRTVVSKAY ERGGD

#### MSD2

STRQAQEPAG WLPGATQDLP HTPKEPGPRQ HAPKPPSPKP DGDREGRGSE EATEVSHIKR  
KEVTRTVVSK AYERGGDVNY LSHRYEHDGG VSEAVQPDND IDRILLSHDS PTRRRKCTNL  
VSKLTGWKWKV MEQEPPKWKV DSIDTEDSGY GGDMEE

#### ABD1

DMEERPEQDV AQVAAARIKR PLHSQANRYS ETLNCKAHRK YSQVDNLKGR WQQWADEHIQ  
SQKLNPFVDE FDYDLAMSTR LHKGDGEGYGR PKEGSKTAER AKR

#### ABD2

RAEEHIYREI MELCFVIRTM ARHRRDGGKIQ VTFGELFDRY VRISDKVVGI LMRARKHGLV  
HFEGEMLWQG KDDHVVITLL E

#### Full length ABLIM2

MSAVSQPQAA PSPLEKSPST AILCNYCGNV CKGEVLRVQD KYFHIKCFVC KACQCDAEG  
GFFVRQGEYI CTLDYQRLYG TRCFSCDQFI EGEVVSALGK TYHPDCFVCA VCRLPFPPGD  
RVTFNGKECM CQKCSLPVSV GSSAHLSQLG RSCGGCGTEI KNGQALVALD KHWHLGCFKC  
KSCGKLLNAE YISKDGLPYC EADYHAKFGI RCDSCEKYIT GRVLEAGEKH YHPSCALCVR  
CGQMFAEGEE MYLQGSSIWH PACRQAARTE DRNKETRTSS ESIISVPASS TSGSPSRVIY  
AKLGGEILDY RDLAALPKSK AIYDIDRPDM ISYSPYISHS AGDRQSYGEG DQFFESYKQC  
RTSSPSSTGS VSLGRYTPTS RSPQHYSRPG SESGRSTPSL SVLSDSKPPP STYQQAPRHF  
HVPDTGVKDN IYRKPPIYRQ HAARRSDGED GSLDQDNRKK SSWLMLKGDA DTRTNSPDL  
TQSLSHSSGT DRDPLQRMAG DSFHSRFPYS KSDPLPGHGK NGLDQRNANL APCGADPDAS  
WGMREYKIYP YDSLIVTNRI RVKLPKDVDR TRLERHLSPE EFQEVFGMSI EEFDRALWK  
RNDLKKKALL F

#### LIM1

ILCNTCGNVC KGEVLRVQDK YFHIKCFVCK ACGCDLAEGG FFVRQGEYIC TLDYQRLYGT

## LIM2

LDYQRLYGTR CFSCDQFI EG EVVSALGKTY HPDCFVCAVC RLPFPPGDRV TFNGKECMCQ  
KCSLPVSV

## LIM3

RSCGGCGTEI KNGQALVALD KHWHLGCFKC KSCGKLLNAE YISKDGLPYC EADYHAKFGI

## LIM4

CDSCEKYITG RVLEAGEKHY HPSCALCVRG GQMFAEGEEM YLQGSSIWHP ACRQAARTED

## LIM1+2

ILCNTCGNVC KGEVLRVQDK YFHIKCFVCK ACGCDLAEGG FFVRQGEYIC TLDYQRLYGT  
RCFSCDQFIE GEVVSALGKT YHPDCFVCAV CRLPFPPGDR VTFNGKECMC QKCSLPVSV

## LIM2+3

LDYQRLYGTR CFSCDQFIEG EVVSALGKTY HPDCFVCAVC RLPFPPGDRV TFNGKECMCQ  
KCSLPVSVGS SAHLSQGLRS CGGCGTEIKN GQALVALDKH WHLGCFKCKS CGKLLNAEYI  
SKDGLPYCEA DYHAKFGI

## LIM3+4

RSCGGCGTEI KNGQALVALD KHWHLGCFKC KSCGKLLNAE YISKDGLPYC EADYHAKFGI  
RCDSCEKYIT GRVLEAGEKH YHPSCALCVR CGQMFAEGEE MYLQGSSIWH PACRQAARTE D

## VHD

QYKIYPYDSL IVTNRIRVKL PKDVDRTRLE RHLSPEEFQE VFGMSIEEFD RLALWKRNDL  
KKKALLF

## **Primers Used To Produce Constructs**

### Primers used to clone MSD1 into pLeics03

Forward = TAC TTC CAA TCC ATG CCG GGC GAA AAG GAA AGC

Reverse = TAT CCA CCT TTA CTG TCA ACG TCC CCT CCT CTC

### Primers used to clone MSD2 into pLeics03

Forward = GTA TTT TCA GGG CGC CAA CAG TAC CAG ACA GGC CCA

Reverse = GAC GGA GCT CGA ATT TCA CTC CTC CAT GTC CCC T

Primers used to clone ABD1 into pLeics03

Forward = GTA TTT TCA GGG CGC CGA CAT GGA GGA GAG C

Reverse = GAC GGA GCT CGA ATT TCA TCT CTT GGC CCT TTC

Primers used to clone ABD1+ABD2 tandem into pLeics03

Forward = GTA TTT TCA GGG CGC CGA CAT GGA GGA GAG G

Reverse = GAC GGA GCT CGA ATT TCA CTC AAG GAG AGT AAT CAC

Primers used to clone ABD2 constructs into pETM-11

Construct 1 (residues 294-375)

Forward: GCC TAG CGC CCA TGG CGA GAG CCG AGG AGC ACA TC

Reverse: GCG CGG ATC CTT ACT CAA GGA GAG TAA TCA CAA CAT GGT CAT CTT TGC C

Construct 2 (residues 297-375)

Forward: GCC TAG CGC CCA TGG CGG AGC ACA TCT ATC GGG AA

Reverse: GCG CGG ATC CTT ACT CAA GGA GAG TAA TCA CAA CAT GGT CAT CTT TGC C

Construct 3 (residues 314-375)

Forward: GCC TAG CGC CCA TGG CGC TCT TTG ATA GAT ATG TT

Reverse: GCG CGG ATC CTT ACT CAA GGA GAG TAA TCA CAA CAT G

Construct 4 (residues 317-375)

Forward: GCC TAG CGC CCA TGG CGG ATA GAT ATG TTC GAA TT

Reverse: GCG CGG ATC CTT ACT CAA GGA GAG TAA TCA CAA CAT G

Primers Used to Clone the ABLIM2 domains into pLeics07

LIM1

Forward: GTA TTT TCA GGG CGC CAT CCT GTG CAA CAC G

Reverse: GAC GGA GCT CGA ATT TCA GGT GCC GTA GAG CCT

LIM2

Forward: GTA TTT TCA GGG CGC CCT GGA CTA CCA GAG G

Reverse: GAC GGA GCT CGA ATT TCA CAC CGA TAC GGG CAG

LIM3

Forward: GTA TTT TCA GGG CGC CCG AAG TTG TGG GGG C

Reverse: GAC GGA GCT CGA ATT TCA GAT GCC GAA CTT GGC

LIM4

Forward: GTA TTT TCA GGG CGC CTG TGA CAG CTG TGA G

Reverse: GAC GGA GCT CGA ATT TCA GTC TTC AGT TCT GGC

VHD

Forward: TAC TTC CAA TCC ATG CAA TAC AAG ATC TAT CCG TAT GAC

Reverse: GTC ATA CGG ATA GAT CTT GTA TTG CAT GGA TTG GAA GTA

Primers Used To Produce Mutant ABD2 Constructs For Expression in pLeics03

R302A

Forward: GCC GAG GAG CAC ATC TAT GCT GAA ATT ATG GAA TTG TGC

Reverse: GCA CAA TTC CAT AAT TTC AGC ATA GAT GTG CTC CTC GGC

R318A

Forward: CGC ACA ATG GCT CGC CAT GCA CGA GAT GGC AAG ATC CAG

Reverse: CTG GAT CTT GCC ATC TGC TGC ATG GCG AGC CAT TGT GCG

R347A

Forward: GTG GTG GGC ATC CTC ATG GCT GCC AGG AAA CAC GGA CTG

Reverse: CAG TCC GTG TTT CCT GGC AGC CAT GAG GAT GCC CAC CAC

R349A

Forward: GGC ATC CTC ATG CGT GCC GCT AAA CAC GGA CTG GTG CAC

Reverse: GTG CAC CAG TCC GTG TTT AGC GGC ACG CAT GAG GAT GCC

K350A

Forward: ATC CTC ATG CGT GCC AGG GCT CAC GGA CTG GTG CAC TTT

Reverse: AAA GTG CAC CAG TCC GTG AGC CCT GGC ACG CAT GAG GAT

H351A

Forward: CTC ATG CGT GCC AGG AAA GCT GGA CTG GTG CAC TTT GAA

Reverse: TTC AAA GTG CAC CAG TCC AGC TTT CCT GGC ACG CAT GAG

## **Appendix 2 – Vectors and Cell Strains**

Table 1: Antibiotic resistance of expression vectors used

Expression Vector	Antibiotic Resistance
pETM11	Kanamycin
pLeics07	Kanamycin
pLeics03	Kanamycin
pDXV	Carbenicillin

Table 2: Antibiotic resistance of expression cell lines used

Expression Cell	Antibiotic Resistance
BL21 STAR (DE3)	None
Top10	None
Arctic Express (RP)	Gentamycin
Arctic Express (RIL)	Gentamycin
Codon Plus	Chloramphenicol

## **Appendix 3 – Buffer Recipes**

### **Luria Broth (LB) Agar**

5 g yeast extract (Melford Laboratories)

10 g tryptone (Melford Laboratories)

10g NaCl (Fisher Scientific)

15 g Agar (Acros)

per litre of LB agar produced

### **1% Agarose Gel**

3g agarose powder (Melford Laboratories) in 300 ml 1 x TAE buffer

### **Luria Broth (LB):**

5 g yeast extract (Melford Laboratories)

10g tryptone (Melford Laboratories)

10 g NaCl (Fisher Scientific)

in 1 litre of water

### **FF6 Wash Buffer**

50 mM NaPO<sub>4</sub> (Fisher Scientific)

10 mM NaCl (Fisher Scientific)

10 mM Imidazole (Sigma Aldrich)

0.02% NaN<sub>3</sub> (Fisher Scientific)

100 ul β-mercaptoethanol (Fisher Scientific)

### FF6 Elution buffer

50 mM NaPO<sub>4</sub> (Fisher Scientific)

500 mM NaCl (Fisher Scientific)

500 mM Imidazole (Sigma Aldrich)

0.02% NaN<sub>3</sub> (Fisher Scientific)

100 ul β-mercaptoethanol (Fisher Scientific)

### Gel Filtration Buffer

20 mM PO<sub>4</sub> (Fisher Scientific)

50 mM NaCl (Fisher Scientific)

2 mM DTT (Melford)

0.02% NaN<sub>3</sub> (Fisher Scientific)

pH 7.5

### Measuring Buffer

20 mM PO<sub>4</sub> (Fisher Scientific)

50 mM NaCl (Fisher Scientific)

2 mM DTT (Melford)

0.02% NaN<sub>3</sub> (Fisher Scientific)

### G-Actin Buffer

2mM Tris-HCl (pH 8) (Fisher Scientific)

0.2 mM ATP (Melford)

0.1 mM CaCl<sub>2</sub> (Melford)

0.5 mM DTT (Melford)

1 mM NaN<sub>3</sub> (Fisher Scientific)

### 10 x KME Buffer

100 mM Tris HCl (pH 8)

500 mM KCl (Melford)

25 mM MgCl<sub>2</sub> (Fisher Scientific)

10 mM EGTA (Melford)

10mM NaN<sub>3</sub>

### Pyrene-Labeling Buffer

25 mM Tris

100 mM KCl

2 mM MgCl<sub>2</sub>

0.3 mM ATP

0.02% NaN<sub>3</sub>

pH 7.5

### Actin Binding Buffer

10 mM Hepes (Melford)

50 mM KCl

3.5 mM MgCl<sub>2</sub>

1 mM DTT

0.02% NaN<sub>3</sub>

pH 7.2

Bicelle Buffer

25 mM PO<sub>4</sub>

100 mM NaCl

2 mM MgCl<sub>2</sub>

pH 6.5

Filamentous Phage Working Buffer

25 mM PO<sub>4</sub>

100 mM NaCl

2 mM MgCl<sub>2</sub>

0.01% NaN<sub>3</sub>

pH 7

½ TB Buffer

0.04 M Tris base

0.04 M boric acid (Fisher Scientific)

## **Appendix 4 – Domains Produced by DomainX**

Name	Size (kD)	Start DNA (bp)	End DNA (bp)	Expression Level	Comments
E12	17.5	5	450	Medium	Possibly folded by thermofluor assay. Degrades over time into smaller fragments.
G09	15.3	5	389	Low	
B7	14.0	5	345	Medium	Strong peak in thermofluor. Most likely folded, T <sub>m</sub> around 45 and 60 degrees (two peaks). No degradation by ESI-MS.
G12	14.0	5	345	Medium	
C10	10.1	6	248	Medium	
G11	10.1	6	248	Medium	
B2	10.1	7	247	Medium	
B12	10.1	7	247	Medium	
HR1	11.6	36	318	Low	
HR2	22.1	42	612	Low	
D11	12.5	47	356	Medium	Elutes from S200 column in the void volume.
F05	11.9	68	360	Medium	
G10	11.9	68	359	High	Precipitated overnight following Ni purification.
F03	9.8	100	333	Low	
C2	18.4	119	588	High	Three peaks in gel filtration trace, the first 2 give strong peaks in thermofluor assay, T <sub>m</sub> around 55°C. Significant PMF in MASCOT (all others are insignificant by MASCOT but still from MS-1). Best grown in LB then induced with 1mM for 5 hours at 22°C.
E8	18.4	119	589	Low	
F5	18.4	119	589	High	
C1	8.9	181	391	High	Very small peak in thermofluor trace, likely unfolded, best grown in autoinduction media, single band after Ni+S200 columns. Has lost start MET by ESI-MS.
C4	20.9	215	752	Medium	Significant peak in thermofluor trace, T <sub>m</sub> around 45-50°C, sample degrades.
F2	8.5	274	472	Low	
C9	13.3	285	616	Medium	
E10	10.6	308	563	Low	
G1	12.5	579	886	Medium	Significant peak in thermofluor trace, T <sub>m</sub> around 50°C, best grown in autoinduction media.
H7	8.4	589	786	Low	
HR3	16.7	696	1119	Low	
HR4	9.6	897	1125	Low	

## Appendix 5 – Sequence Specific Assignments

### Sequence Specific Assignments for ABD2

<u>Residue</u>	<u>Atom</u>	<u>Chemical Shift(ppm)</u>
[2] ALA	C	179.443
[2] ALA	CA	52.670
[2] ALA	CB	19.394
[2] ALA	H	8.351
[2] ALA	HA	4.228
[2] ALA	HB1	1.311
[2] ALA	HB2	1.311
[2] ALA	HB3	1.311
[2] ALA	N	125.955
[3] MET	C	181.369
[3] MET	CA	55.541
[3] MET	CB	33.034
[3] MET	CG	31.985
[3] MET	H	8.382
[3] MET	HA	4.331
[3] MET	HB2	1.968
[3] MET	HB3	1.935
[3] MET	HG2	2.513
[3] MET	HG3	2.474
[3] MET	N	119.901
[4] ALA	C	179.541
[4] ALA	CA	52.555
[4] ALA	CB	19.067
[4] ALA	H	8.297
[4] ALA	HA	4.197
[4] ALA	HB1	1.296
[4] ALA	HB2	1.296
[4] ALA	HB3	1.296
[4] ALA	N	125.905
<u>Residue</u>	<u>Atom</u>	<u>Chemical Shift(ppm)</u>
295 ARG	C	180.613
295 ARG	CA	56.586
295 ARG	CB	30.548
295 ARG	CD	43.189
295 ARG	CG	27.159
295 ARG	H	8.317
295 ARG	HA	4.137
295 ARG	HB2	1.843
295 ARG	HB3	1.752
295 ARG	HD2	3.080
295 ARG	HD3	3.164
295 ARG	HG2	1.619
295 ARG	HG3	1.563
295 ARG	N	120.637

<u>Residue</u>	<u>Atom</u>	<u>Chemical Shift(ppm)</u>
296 ALA	C	179.039
296 ALA	CA	53.234
296 ALA	CB	18.907
296 ALA	H	8.232
296 ALA	HA	4.145
296 ALA	HB1	1.297
296 ALA	HB2	1.297
296 ALA	HB3	1.297
296 ALA	N	124.144
297 GLU	C	179.913
297 GLU	CA	57.497
297 GLU	CB	29.796
297 GLU	CG	36.073
297 GLU	H	8.334
297 GLU	HA	4.112
297 GLU	HB2	1.904
297 GLU	HB3	2.050
297 GLU	HG2	2.225
297 GLU	HG3	2.180
297 GLU	N	119.206
298 GLU	CA	58.741
298 GLU	CB	30.025
298 GLU	CG	36.226
298 GLU	H	8.326
298 GLU	HA	4.131
298 GLU	HB2	1.816
298 GLU	HB3	1.978
298 GLU	HG2	2.199
298 GLU	HG3	2.159
298 GLU	N	120.902
299 HIS	C	181.025
299 HIS	CA	56.859
299 HIS	CB	29.694
299 HIS	HA	4.088
299 HIS	HB3	3.089

<u>Residue</u>	<u>Atom</u>	<u>Chemical Shift(ppm)</u>	<u>Residue</u>	<u>Atom</u>	<u>Chemical Shift(ppm)</u>
300 ILE	CA	63.219	303 GLU	C	177.435
300 ILE	CB	39.003	303 GLU	CA	59.464
300 ILE	CD1	12.753	303 GLU	CB	29.721
300 ILE	CG1	28.006	303 GLU	CG	34.339
300 ILE	CG2	17.308	303 GLU	H	7.927
300 ILE	H	7.786	303 GLU	HA	4.016
300 ILE	HA	3.803	303 GLU	HB2	1.930
300 ILE	HB	1.818	303 GLU	HB3	2.144
300 ILE	HD11	0.773	303 GLU	HG2	2.161
300 ILE	HD12	0.773	303 GLU	HG3	2.187
300 ILE	HD13	0.773	303 GLU	N	118.557
300 ILE	HG12	1.197			
300 ILE	HG13	0.916	304 ILE	C	179.818
300 ILE	HG21	0.627	304 ILE	CA	65.301
300 ILE	HG22	0.627	304 ILE	CB	38.009
300 ILE	HG23	0.627	304 ILE	CD1	14.179
300 ILE	N	117.928	304 ILE	CG1	28.218
			304 ILE	CG2	18.037
301 TYR	CA	57.462	304 ILE	H	7.693
301 TYR	CB	38.748	304 ILE	HA	3.571
301 TYR	CD1	132.975	304 ILE	HB	1.819
301 TYR	CD2	132.975	304 ILE	HD11	0.489
301 TYR	CE1	118.812	304 ILE	HD12	0.489
301 TYR	CE2	118.812	304 ILE	HD13	0.489
301 TYR	HA	4.428	304 ILE	HG12	0.996
301 TYR	HB2	2.866	304 ILE	HG13	0.967
301 TYR	HB3	2.738	304 ILE	HG21	0.753
301 TYR	HD1	6.924	304 ILE	HG22	0.753
301 TYR	HD2	6.924	304 ILE	HG23	0.753
301 TYR	HE1	6.606	304 ILE	N	118.824
301 TYR	HE2	6.606			
			305 MET	C	177.828
302 ARG	C	178.217	305 MET	CA	58.172
302 ARG	CA	60.220	305 MET	CB	30.606
302 ARG	CB	29.854	305 MET	H	8.094
302 ARG	CD	42.549	305 MET	HA	4.344
302 ARG	CG	27.747	305 MET	HB2	3.081
302 ARG	H	8.087	305 MET	HB3	3.081
302 ARG	HA	3.809	305 MET	N	120.610
302 ARG	HB2	1.872			
302 ARG	HB3	1.850	306 GLU	C	179.683
302 ARG	HD2	3.088	306 GLU	CA	59.662
302 ARG	HD3	3.169	306 GLU	CB	29.383
302 ARG	HG2	1.592	306 GLU	CG	36.986
302 ARG	HG3	1.649	306 GLU	H	8.151
302 ARG	N	119.827	306 GLU	HA	4.030
			306 GLU	HB2	1.871
			306 GLU	HB3	1.581
			306 GLU	HG2	2.427
			306 GLU	HG3	2.186
			306 GLU	N	122.247

<u>Residue</u>	<u>Atom</u>	<u>Chemical Shift(ppm)</u>	<u>Residue</u>	<u>Atom</u>	<u>Chemical Shift(ppm)</u>
307 LEU	C	178.418	311 ILE	C	180.508
307 LEU	CA	58.679	311 ILE	CA	66.891
307 LEU	CB	41.270	311 ILE	CB	37.751
307 LEU	CD1	24.856	311 ILE	CD1	13.883
307 LEU	CD2	26.928	311 ILE	CG1	28.255
307 LEU	H	7.562	311 ILE	CG2	17.240
307 LEU	HA	4.314	311 ILE	H	8.373
307 LEU	HB2	2.039	311 ILE	HA	3.307
307 LEU	HB3	1.546	311 ILE	HB	1.764
307 LEU	HD11	0.610	311 ILE	HD11	0.673
307 LEU	HD12	0.610	311 ILE	HD12	0.673
307 LEU	HD13	0.610	311 ILE	HD13	0.673
307 LEU	HD21	0.700	311 ILE	HG12	0.912
307 LEU	HD22	0.700	311 ILE	HG13	1.931
307 LEU	HD23	0.700	311 ILE	HG21	0.709
307 LEU	HG	1.451	311 ILE	HG22	0.709
307 LEU	N	121.205	311 ILE	HG23	0.709
			311 ILE	N	119.837
308 CYS	CA	64.677	312 ARG	C	178.651
308 CYS	CB	26.902	312 ARG	CA	59.990
308 CYS	H	7.906	312 ARG	CB	30.398
308 CYS	HA	3.833	312 ARG	H	7.953
308 CYS	HB2	3.146	312 ARG	HA	3.624
308 CYS	HB3	2.723	312 ARG	HB2	1.707
308 CYS	N	116.467	312 ARG	HB3	1.550
			312 ARG	N	116.869
309 PHE	C	179.059	313 THR	C	181.240
309 PHE	CA	61.163	313 THR	CA	64.991
309 PHE	CB	38.524	313 THR	CB	69.454
309 PHE	CD1	132.112	313 THR	CG2	21.500
309 PHE	CD2	132.112	313 THR	H	7.750
309 PHE	H	7.912	313 THR	HA	3.824
309 PHE	HA	4.173	313 THR	HB	4.019
309 PHE	HB2	3.285	313 THR	HG21	0.926
309 PHE	HB3	3.285	313 THR	HG22	0.926
309 PHE	HD1	7.161	313 THR	HG23	0.926
309 PHE	HD2	7.161	313 THR	N	110.574
309 PHE	N	120.587			
310 VAL	C	178.650	314 MET	C	180.534
310 VAL	CA	67.588	314 MET	CA	57.284
310 VAL	CB	31.292	314 MET	CB	34.801
310 VAL	CG1	23.261	314 MET	CG	31.764
310 VAL	CG2	22.086	314 MET	H	7.924
310 VAL	H	8.628	314 MET	HA	4.180
310 VAL	HA	3.276	314 MET	HB2	1.926
310 VAL	HB	2.360	314 MET	HB3	2.094
310 VAL	HG11	0.998	314 MET	HG2	2.583
310 VAL	HG12	0.998	314 MET	HG3	2.391
310 VAL	HG13	0.998	314 MET	N	120.735
310 VAL	HG21	0.798			
310 VAL	HG22	0.798			
310 VAL	HG23	0.798			
310 VAL	N	123.504			

<u>Residue</u>	<u>Atom</u>	<u>Chemical Shift(ppm)</u>	<u>Residue</u>	<u>Atom</u>	<u>Chemical Shift(ppm)</u>
315 ALA	C	179.496	321 GLY	C	182.936
315 ALA	CA	52.968	321 GLY	CA	45.530
315 ALA	CB	20.236	321 GLY	H	7.852
315 ALA	H	7.453	321 GLY	HA2	3.689
315 ALA	HA	4.205	321 GLY	HA3	4.017
315 ALA	HB1	1.386	321 GLY	N	106.262
315 ALA	HB2	1.386			
315 ALA	HB3	1.386	322 LYS	C	180.777
315 ALA	N	123.294	322 LYS	CA	56.567
			322 LYS	CB	33.668
316 ARG	CA	56.349	322 LYS	CD	27.010
316 ARG	CB	30.675	322 LYS	CE	42.124
316 ARG	CD	43.033	322 LYS	CG	25.494
316 ARG	H	8.420	322 LYS	H	7.797
316 ARG	HA	4.341	322 LYS	HA	4.188
316 ARG	HB2	1.760	322 LYS	HB2	1.442
316 ARG	HB3	1.760	322 LYS	HB3	1.569
316 ARG	HD2	3.170	322 LYS	HD2	1.417
316 ARG	HD3	3.280	322 LYS	HD3	1.781
316 ARG	N	121.949	322 LYS	HE2	3.091
			322 LYS	HE3	3.171
317 HIS	CA	63.270	322 LYS	HG2	1.282
317 HIS	CB	32.076	322 LYS	HG3	1.242
317 HIS	CD2	118.675	322 LYS	N	120.739
317 HIS	CE1	139.075			
317 HIS	H	7.439	323 ILE	C	183.393
317 HIS	HA	4.433	323 ILE	CA	61.031
317 HIS	HB2	2.253	323 ILE	CB	37.025
317 HIS	HB3	1.788	323 ILE	CD1	13.544
317 HIS	HD2	6.990	323 ILE	CG1	26.527
317 HIS	HE1	7.587	323 ILE	CG2	19.169
317 HIS	N	119.310	323 ILE	H	9.219
			323 ILE	HA	4.049
318 ARG	CA	61.823	323 ILE	HB	1.999
318 ARG	CB	34.695	323 ILE	HD11	0.741
318 ARG	H	8.222	323 ILE	HD12	0.741
318 ARG	HA	4.208	323 ILE	HD13	0.741
318 ARG	HB2	1.114	323 ILE	HG12	0.979
318 ARG	N	114.750	323 ILE	HG13	1.551
			323 ILE	HG21	0.699
319 ARG	HA	5.147	323 ILE	HG22	0.699
319 ARG	HB2	2.365	323 ILE	HG23	0.699
319 ARG	HB3	2.849	323 ILE	N	126.643
320 ASP	C	180.228			
320 ASP	CA	53.613			
320 ASP	CB	40.342			
320 ASP	H	7.598			
320 ASP	HA	4.362			
320 ASP	HB2	2.910			
320 ASP	HB3	2.583			
320 ASP	N	120.217			

<u>Residue</u>	<u>Atom</u>	<u>Chemical Shift(ppm)</u>	<u>Residue</u>	<u>Atom</u>	<u>Chemical Shift(ppm)</u>
324 GLN	C	182.797	327 PHE	C	180.598
324 GLN	CA	53.423	327 PHE	CA	60.825
324 GLN	CB	33.447	327 PHE	CB	37.608
324 GLN	CG	33.847	327 PHE	CD1	132.265
324 GLN	H	8.040	327 PHE	CD2	132.265
324 GLN	HA	5.700	327 PHE	CE1	130.349
324 GLN	HB2	1.783	327 PHE	CE2	130.349
324 GLN	HB3	1.783	327 PHE	H	9.320
324 GLN	HG2	2.284	327 PHE	HA	3.991
324 GLN	HG3	1.932	327 PHE	HB2	3.170
324 GLN	N	123.768	327 PHE	HB3	2.749
324 GLN	NE2	112.031 0.200	327 PHE	HD1	6.718
			327 PHE	HD2	6.718
325 VAL	C	182.823	327 PHE	HE1	6.602
325 VAL	CA	59.210	327 PHE	HE2	6.602
325 VAL	CB	36.475	327 PHE	N	121.524
325 VAL	CG1	22.762			
325 VAL	CG2	19.921	328 GLY	C	181.127
325 VAL	H	8.692	328 GLY	CA	46.772
325 VAL	HA	4.700	328 GLY	H	9.636
325 VAL	HB	2.229	328 GLY	HA2	3.353
325 VAL	HG11	0.935	328 GLY	HA3	3.635
325 VAL	HG12	0.935	328 GLY	N	106.176
325 VAL	HG13	0.935			
325 VAL	HG21	0.742	329 GLU	C	177.920
325 VAL	HG22	0.742	329 GLU	CA	58.643
325 VAL	HG23	0.742	329 GLU	CB	30.022
325 VAL	N	113.895	329 GLU	CG	37.186
			329 GLU	H	7.350
326 THR	C	180.827	329 GLU	HA	3.999
326 THR	CA	60.532	329 GLU	HB2	1.848
326 THR	CB	71.126	329 GLU	HB3	2.142
326 THR	CG2	22.311	329 GLU	HG2	2.279
326 THR	H	8.419	329 GLU	HG3	2.212
326 THR	HA	5.263	329 GLU	N	119.770
326 THR	HB	4.835			
326 THR	HG21	1.160	330 LEU	C	179.587
326 THR	HG22	1.160	330 LEU	CA	57.613
326 THR	HG23	1.160	330 LEU	CB	42.412
326 THR	N	111.421	330 LEU	CD1	27.064
			330 LEU	CD2	24.291
			330 LEU	CG	26.998
			330 LEU	H	7.971
			330 LEU	HA	3.863
			330 LEU	HB2	1.645
			330 LEU	HB3	1.353
			330 LEU	HD11	0.701
			330 LEU	HD12	0.701
			330 LEU	HD13	0.701
			330 LEU	HD21	0.775
			330 LEU	HD22	0.775
			330 LEU	HD23	0.775
			330 LEU	HG	1.343
			330 LEU	N	122.016

<u>Residue</u>	<u>Atom</u>	<u>Chemical Shift(ppm)</u>	<u>Residue</u>	<u>Atom</u>	<u>Chemical Shift(ppm)</u>
331 PHE	CA	62.446	334 TYR	C	180.054
331 PHE	CB	38.503	334 TYR	CA	60.185
331 PHE	CD1	132.470	334 TYR	CB	38.743
331 PHE	CD2	132.470	334 TYR	CD1	132.418
331 PHE	CE1	131.687	334 TYR	CD2	132.418
331 PHE	CE2	131.687	334 TYR	CE1	118.254
331 PHE	CZ	130.426	334 TYR	CE2	118.254
331 PHE	H	9.059	334 TYR	H	8.197
331 PHE	HA	3.429	334 TYR	HA	3.955
331 PHE	HB2	2.664	334 TYR	HB2	2.920
331 PHE	HB3	1.953	334 TYR	HB3	2.633
331 PHE	HD1	6.830	334 TYR	HD1	6.956
331 PHE	HD2	6.830	334 TYR	HD2	6.956
331 PHE	HE1	7.199	334 TYR	HE1	6.531
331 PHE	HE2	7.199	334 TYR	HE2	6.531
331 PHE	HZ	7.188	334 TYR	N	119.995
331 PHE	N	120.779			
			335 VAL	CA	65.487
332 ASP	C	178.832	335 VAL	CB	31.307
332 ASP	CA	57.281	335 VAL	CG1	22.054
332 ASP	CB	41.073	335 VAL	CG2	20.583
332 ASP	H	7.480	335 VAL	H	7.357
332 ASP	HA	4.291	335 VAL	HA	3.644
332 ASP	HB2	2.668	335 VAL	HB	1.871
332 ASP	HB3	2.540	335 VAL	HG11	0.940
332 ASP	N	116.574	335 VAL	HG12	0.940
			335 VAL	HG13	0.940
333 ARG	C	179.621	335 VAL	HG21	0.825
333 ARG	CA	56.928	335 VAL	HG22	0.825
333 ARG	CB	29.808	335 VAL	HG23	0.825
333 ARG	CD	42.986	335 VAL	N	120.498
333 ARG	CG	26.997			
333 ARG	H	7.441	337 ILE	C	181.079
333 ARG	HA	4.092	337 ILE	CA	62.440
333 ARG	HB2	1.814	337 ILE	CB	38.797
333 ARG	HB3	1.877	337 ILE	CD1	13.228
333 ARG	HD2	3.158	337 ILE	CG1	26.753
333 ARG	HD3	3.070	337 ILE	CG2	17.541
333 ARG	HG2	1.621	337 ILE	HA	4.077
333 ARG	HG3	1.530	337 ILE	HB	1.366
333 ARG	N	118.691	337 ILE	HD11	0.337
			337 ILE	HD12	0.337
			337 ILE	HD13	0.337
			337 ILE	HG12	1.223
			337 ILE	HG13	0.675
			337 ILE	HG21	0.623
			337 ILE	HG22	0.623
			337 ILE	HG23	0.623

<u>Residue</u>	<u>Atom</u>	<u>Chemical Shift(ppm)</u>	<u>Residue</u>	<u>Atom</u>	<u>Chemical Shift(ppm)</u>
338 SER	C	183.302	342 VAL	C	179.941
338 SER	CA	58.868	342 VAL	CA	68.282
338 SER	CB	64.083	342 VAL	CB	31.027
338 SER	H	7.977	342 VAL	CG1	24.258
338 SER	HA	4.242	342 VAL	CG2	21.975
338 SER	HB2	3.909	342 VAL	H	8.506
338 SER	HB3	3.596	342 VAL	HA	3.333
338 SER	N	114.528	342 VAL	HB	1.904
			342 VAL	HG11	1.005
339 ASP	C	179.562	342 VAL	HG12	1.005
339 ASP	CA	54.554	342 VAL	HG13	1.005
339 ASP	CB	39.933	342 VAL	HG21	0.745
339 ASP	H	8.561	342 VAL	HG22	0.745
339 ASP	HA	4.688	342 VAL	HG23	0.745
339 ASP	HB2	2.869	342 VAL	N	121.665
339 ASP	HB3	2.641			
339 ASP	N	121.243	343 GLY	C	180.310
			343 GLY	CA	47.143
340 LYS	C	180.185	343 GLY	H	7.895
340 LYS	CA	55.994	343 GLY	HA2	3.658
340 LYS	CB	32.566	343 GLY	HA3	3.670
340 LYS	CD	28.840	343 GLY	N	107.446
340 LYS	CE	41.749			
340 LYS	CG	24.716	344 ILE	C	179.069
340 LYS	H	8.088	344 ILE	CA	65.619
340 LYS	HA	4.491	344 ILE	CB	37.948
340 LYS	HB2	1.604	344 ILE	CD1	13.860
340 LYS	HB3	2.013	344 ILE	CG1	28.221
340 LYS	HD2	1.568	344 ILE	CG2	19.264
340 LYS	HD3	1.648	344 ILE	H	7.809
340 LYS	HE2	2.944	344 ILE	HA	3.493
340 LYS	HG2	1.465	344 ILE	HB	1.548
340 LYS	HG3	1.395	344 ILE	HD11	0.720
340 LYS	N	119.427	344 ILE	HD12	0.720
			344 ILE	HD13	0.720
341 VAL	C	182.048	344 ILE	HG12	0.870
341 VAL	CA	65.864	344 ILE	HG13	1.654
341 VAL	CB	31.431	344 ILE	HG21	0.787
341 VAL	CG1	20.709	344 ILE	HG22	0.787
341 VAL	CG2	21.793	344 ILE	HG23	0.787
341 VAL	H	7.487	344 ILE	N	122.291
341 VAL	HA	3.312			
341 VAL	HB	1.793			
341 VAL	HG11	0.746			
341 VAL	HG12	0.746			
341 VAL	HG13	0.746			
341 VAL	HG21	0.172			
341 VAL	HG22	0.172			
341 VAL	HG23	0.172			
341 VAL	N	118.702			

<u>Residue</u>	<u>Atom</u>	<u>Chemical Shift(ppm)</u>	<u>Residue</u>	<u>Atom</u>	<u>Chemical Shift(ppm)</u>
345 LEU	C	178.127	349 ARG	C	177.353
345 LEU	CA	58.492	349 ARG	CA	59.270
345 LEU	CB	41.714	349 ARG	CB	30.352
345 LEU	CD1	25.702	349 ARG	H	8.448
345 LEU	CD2	24.903	349 ARG	HA	4.385
345 LEU	CG	26.542	349 ARG	HB2	1.416
345 LEU	H	8.356	349 ARG	HB3	1.891
345 LEU	HA	3.678	349 ARG	N	120.536
345 LEU	HB2	1.890			
345 LEU	HB3	1.241	350 LYS	CA	59.332
345 LEU	HD11	0.502	350 LYS	CB	32.098
345 LEU	HD12	0.502	350 LYS	CD	29.180
345 LEU	HD13	0.502	350 LYS	CE	41.688
345 LEU	HD21	0.551	350 LYS	CG	24.770
345 LEU	HD22	0.551	350 LYS	H	7.848
345 LEU	HD23	0.551	350 LYS	HA	3.996
345 LEU	HG	1.408	350 LYS	HB2	1.883
345 LEU	N	122.202	350 LYS	HB3	1.883
			350 LYS	HD2	1.469
346 MET	C	176.580	350 LYS	HD3	1.552
346 MET	CA	57.682	350 LYS	HE2	2.755
346 MET	CB	30.938	350 LYS	HG2	1.363
346 MET	CG	32.710	350 LYS	HG3	1.170
346 MET	H	8.495	350 LYS	N	123.506
346 MET	HA	4.286			
346 MET	HB2	1.940	351 HIS	C	182.389
346 MET	HB3	2.275	351 HIS	CA	57.269
346 MET	HG2	2.751	351 HIS	CB	31.008
346 MET	HG3	2.810	351 HIS	CD2	119.65
346 MET	N	116.579	351 HIS	CE1	139.493
			351 HIS	H	7.680
347 ARG	C	178.902	351 HIS	HA	4.434
347 ARG	CA	59.449	351 HIS	HB2	2.854
347 ARG	CB	30.189	351 HIS	HB3	2.989
347 ARG	CD	43.160	351 HIS	HD2	6.912
347 ARG	H	7.911	351 HIS	HE1	7.677
347 ARG	HA	4.035	351 HIS	N	113.931
347 ARG	HB2	1.740			
347 ARG	HB3	1.716	352 GLY	C	182.883
347 ARG	HD2	3.205	352 GLY	CA	46.166
347 ARG	HD3	3.019	352 GLY	H	7.861
347 ARG	N	121.772	352 GLY	HA2	3.886
			352 GLY	HA3	4.045
348 ALA	C	177.968	352 GLY	N	107.102
348 ALA	CA	55.490			
348 ALA	CB	18.794			
348 ALA	H	9.143			
348 ALA	HA	3.904			
348 ALA	HB1	1.413			
348 ALA	HB2	1.413			
348 ALA	HB3	1.413			
348 ALA	N	122.853			

<u>Residue</u>	<u>Atom</u>	<u>Chemical Shift(ppm)</u>	<u>Residue</u>	<u>Atom</u>	<u>Chemical Shift(ppm)</u>
353 LEU	C	179.556	356 PHE	C	183.769
353 LEU	CA	55.881	356 PHE	CA	56.646
353 LEU	CB	43.370	356 PHE	CB	40.724
353 LEU	CD1	25.844	356 PHE	CD1	132.974
353 LEU	CD2	23.297	356 PHE	CD2	132.974
353 LEU	CG	27.799	356 PHE	CE1	131.866
353 LEU	H	8.044	356 PHE	CE2	131.866
353 LEU	HA	4.471	356 PHE	CZ	129.481
353 LEU	HB2	1.516	356 PHE	H	5.874
353 LEU	HB3	1.703	356 PHE	HA	4.326
353 LEU	HD11	0.807	356 PHE	HB2	3.424
353 LEU	HD12	0.807	356 PHE	HB3	2.578
353 LEU	HD13	0.807	356 PHE	HD1	6.912
353 LEU	HD21	0.891	356 PHE	HD2	6.912
353 LEU	HD22	0.891	356 PHE	HE1	6.764
353 LEU	HD23	0.891	356 PHE	HE2	6.764
353 LEU	HG	1.576	356 PHE	HZ	6.905
353 LEU	N	116.829	356 PHE	N	116.789
354 VAL	C	183.243	357 GLU	C	181.209
354 VAL	CA	57.442	357 GLU	CA	55.682
354 VAL	CB	36.041	357 GLU	CB	31.207
354 VAL	CG1	21.376	357 GLU	CG	36.307
354 VAL	CG2	19.289	357 GLU	H	8.774
354 VAL	H	6.858	357 GLU	HA	4.212
354 VAL	HA	5.400	357 GLU	HB2	1.957
354 VAL	HB	1.939	357 GLU	HB3	1.814
354 VAL	HG11	0.713	357 GLU	HG2	2.219
354 VAL	HG12	0.713	357 GLU	HG3	2.273
354 VAL	HG13	0.713	357 GLU	N	122.825
354 VAL	HG21	0.639	358 GLY	C	182.880
354 VAL	HG22	0.639	358 GLY	CA	43.862
354 VAL	HG23	0.639	358 GLY	H	8.307
354 VAL	N	106.151	358 GLY	HA2	3.656
355 HIS	C	183.175	358 GLY	HA3	4.446
355 HIS	CA	56.825	358 GLY	N	108.928
355 HIS	CB	35.294	359 GLU	C	182.295
355 HIS	H	9.252	359 GLU	CA	57.482
355 HIS	HA	4.543	359 GLU	CB	30.202
355 HIS	HB2	3.025	359 GLU	CG	36.258
355 HIS	HB3	2.686	359 GLU	H	8.764
355 HIS	N	120.073	359 GLU	HA	4.142
			359 GLU	HB2	1.819
			359 GLU	HB3	1.997
			359 GLU	HG2	2.233
			359 GLU	HG3	2.173
			359 GLU	N	121.640

<u>Residue</u>	<u>Atom</u>	<u>Chemical Shift(ppm)</u>	<u>Residue</u>	<u>Atom</u>	<u>Chemical Shift(ppm)</u>
360 MET	C	184.552	363 GLN	C	180.783
360 MET	CA	53.548	363 GLN	CA	57.721
360 MET	CB	33.853	363 GLN	CB	27.475
360 MET	CG	31.985	363 GLN	CG	32.733
360 MET	H	7.565	363 GLN	H	8.957
360 MET	HA	4.147	363 GLN	HA	3.659
360 MET	HB2	1.813	363 GLN	HB2	1.828
360 MET	HB3	1.528	363 GLN	HB3	1.562
360 MET	HG2	2.220	363 GLN	HG2	1.814
360 MET	HG3	1.533	363 GLN	HG3	1.652
360 MET	N	116.071	363 GLN	N	125.743
			363 GLN	NE2	111.769
361 LEU	C	182.724			
361 LEU	CA	53.194	364 GLY	C	182.254
361 LEU	CB	44.698	364 GLY	CA	45.983
361 LEU	CD1	22.402	364 GLY	H	9.522
361 LEU	CD2	25.639	364 GLY	HA2	3.372
361 LEU	CG	25.889	364 GLY	HA3	4.263
361 LEU	H	6.456	364 GLY	N	118.427
361 LEU	HA	4.476			
361 LEU	HB2	0.532	365 LYS	C	180.029
361 LEU	HB3	-0.348	365 LYS	CA	58.666
361 LEU	HD11	0.293	365 LYS	CB	33.682
361 LEU	HD12	0.293	365 LYS	CG	24.141
361 LEU	HD13	0.293	365 LYS	H	7.652
361 LEU	HD21	-0.519	365 LYS	HA	4.117
361 LEU	HD22	-0.519	365 LYS	HB2	1.680
361 LEU	HD23	-0.519	365 LYS	HB3	1.772
361 LEU	HG	0.703	365 LYS	HG2	1.324
361 LEU	N	120.367	365 LYS	HG3	1.525
			365 LYS	N	123.414
362 TRP	C	181.475			
362 TRP	CA	55.702	366 ASP	C	180.619
362 TRP	CB	31.405	366 ASP	CA	54.508
362 TRP	CD1	128.051	366 ASP	CB	41.704
362 TRP	CE3	121.523	366 ASP	H	8.004
362 TRP	CH2	125.043	366 ASP	HA	5.107
362 TRP	CZ2	115.121	366 ASP	HB2	2.808
362 TRP	CZ3	122.422	366 ASP	HB3	2.320
362 TRP	H	10.215	366 ASP	N	113.753
362 TRP	HA	4.570			
362 TRP	HB2	3.120	367 ASP	C	180.556
362 TRP	HB3	2.780	367 ASP	CA	57.642
362 TRP	HD1	7.254	367 ASP	CB	40.270
362 TRP	HE3	7.753	367 ASP	H	7.598
362 TRP	HH2	7.149	367 ASP	HA	4.132
362 TRP	HZ2	7.383	367 ASP	HB2	2.597
362 TRP	HZ3	7.091	367 ASP	HB3	2.533
362 TRP	N	122.504	367 ASP	N	120.169

<u>Residue</u>	<u>Atom</u>	<u>Chemical Shift(ppm)</u>	<u>Residue</u>	<u>Atom</u>	<u>Chemical Shift(ppm)</u>
368 HIS	C	181.706	371 ILE	C	182.546
368 HIS	CA	56.222	371 ILE	CA	60.147
368 HIS	CB	30.669	371 ILE	CB	39.319
368 HIS	CD2	118.665	371 ILE	CD1	14.846
368 HIS	CE1	140.343	371 ILE	CG1	27.317
368 HIS	H	8.281	371 ILE	CG2	17.879
368 HIS	HA	4.500	371 ILE	H	9.507
368 HIS	HB2	2.985	371 ILE	HA	4.724
368 HIS	HB3	3.118	371 ILE	HB	1.755
368 HIS	HD2	6.993	371 ILE	HD11	0.522
368 HIS	HE1	8.014	371 ILE	HD12	0.522
368 HIS	N	113.359	371 ILE	HD13	0.522
			371 ILE	HG12	0.742
369 VAL	C	181.757	371 ILE	HG13	1.357
369 VAL	CA	64.898	371 ILE	HG21	0.706
369 VAL	CB	31.738	371 ILE	HG22	0.706
369 VAL	CG1	20.964	371 ILE	HG23	0.706
369 VAL	CG2	23.586	371 ILE	N	132.068
369 VAL	H	7.564			
369 VAL	HA	3.476	372 THR	C	182.846
369 VAL	HB	1.880	372 THR	CA	61.662
369 VAL	HG11	0.898	372 THR	CB	70.883
369 VAL	HG12	0.898	372 THR	CG2	21.538
369 VAL	HG13	0.898	372 THR	H	9.070
369 VAL	HG21	0.746	372 THR	HA	5.018
369 VAL	HG22	0.746	372 THR	HB	3.919
369 VAL	HG23	0.746	372 THR	HG21	1.138
369 VAL	N	123.639	372 THR	HG22	1.138
			372 THR	HG23	1.138
370 VAL	C	181.312	372 THR	N	123.150
370 VAL	CA	63.639			
370 VAL	CB	32.848	373 LEU	C	182.188
370 VAL	CG1	22.267	373 LEU	CA	55.005
370 VAL	CG2	21.620	373 LEU	CB	43.336
370 VAL	H	8.156	373 LEU	CD1	25.264
370 VAL	HA	3.940	373 LEU	CD2	26.993
370 VAL	HB	1.879	373 LEU	CG	27.327
370 VAL	HG11	0.747	373 LEU	H	8.749
370 VAL	HG12	0.747	373 LEU	HA	4.512
370 VAL	HG13	0.747	373 LEU	HB2	1.908
370 VAL	HG21	0.955	373 LEU	HB3	1.149
370 VAL	HG22	0.955	373 LEU	HD11	0.996
370 VAL	HG23	0.955	373 LEU	HD12	0.996
370 VAL	N	127.710	373 LEU	HD13	0.996
			373 LEU	HD21	0.861
			373 LEU	HD22	0.861
			373 LEU	HD23	0.861
			373 LEU	HG	1.621
			373 LEU	N	128.414

<u>Residue</u>	<u>Atom</u>	<u>Chemical Shift(ppm)</u>
374 LEU	C	181.567
374 LEU	CA	54.743
374 LEU	CB	40.629
374 LEU	CD1	25.307
374 LEU	CD2	25.291
374 LEU	CG	27.088
374 LEU	H	7.949
374 LEU	HA	4.397
374 LEU	HB2	1.594
374 LEU	HB3	1.292
374 LEU	HD11	0.646
374 LEU	HD12	0.646
374 LEU	HD13	0.646
374 LEU	HD21	0.685
374 LEU	HD22	0.685
374 LEU	HD23	0.685
374 LEU	HG	1.299
374 LEU	N	128.328
375 GLU	CA	57.790
375 GLU	CB	32.138
375 GLU	CG	32.135
375 GLU	H	7.636
375 GLU	HA	4.077
375 GLU	HB2	2.200
375 GLU	HB3	2.047
375 GLU	HG2	1.788
375 GLU	HG3	1.789
375 GLU	N	126.278

### **Sequence Specific Assignment of**

#### **ABD1**

<u>Residue</u>	<u>Atom</u>	<u>Chemical Shift(ppm)</u>	<u>Residue</u>	<u>Atom</u>	<u>Chemical Shift(ppm)</u>
277 ASP	CA	55.292	280 TYR	CA	63.140
277 ASP	CB	42.309	280 TYR	CB	33.535
278 GLU	CA	58.315	280 TYR	H	7.982
278 GLU	CB	30.851	280 TYR	N	120.134
278 GLU	H	8.510	289 THR	CA	63.030
278 GLU	N	121.577	289 THR	CB	70.584
279 GLY	CA	46.378	290 ALA	CA	53.513
279 GLY	H	8.387	290 ALA	CB	20.30
279 GLY	N	109.038	290 ALA	H	8.227
			290 ALA	N	126.144

## **Appendix 6 – Top 100 DALI Results for ABD2**

Number	Name	Z-score	Description
1	2HR3-B	4.7	Probable transcriptional regulator
2	1BI2-B	4.7	Diphtheria toxin repressor
3	1Q1H-A	4.6	Transcription factor E
4	2P7C-B	4.6	Bacillus Licheniformis Blai Monomeric Form In Complex With The Blap Half-Operator.
5	1G3Y-A	4.6	Diphtheria toxin repressor
6	1G3T-A	4.6	Diphtheria toxin repressor
7	1BI3-B	4.6	Diphtheria toxin repressor
8	1FWZ-A	4.6	Diphtheria toxin repressor
9	1BI3-A	4.6	Diphtheria toxin repressor
10	1P6R-A	4.5	Penicillinase repressor
11	1G3T-B	4.5	Diphtheria toxin repressor
12	2DTR	4.5	Diphtheria toxin repressor
13	3C18-C	4.4	Nucleotidyltransferase-like protein
14	1SAX-B	4.4	Methicillin resistance regulatory protein MECI
15	1W7P-A	4.4	Endosomal complex Escrt-li
16	1SD4-B	4.4	Penicillinase repressor
17	3BZ6-A	4.3	Conserved protein of unknown function from Pseudomonas syringae
18	2HR3-D	4.3	Probable transcriptional regulator
19	3C18-A	4.3	Nucleotidyltransferase-like protein
20	2NR3-A	4.3	Protein unknown function
21	2RGV-B	4.3	Peroxide operon regulator
22	2QQA-A	4.3	Diphtheria toxin repressor
23	2G9W-A	4.2	Rv1846c, a putative transcriptional regulatory protein of Mycobacterium tuberculosis
24	2FE3-B	4.2	Peroxide operon regulator
25	2FU4-A	4.2	Ferric uptake regulation protein
26	1OKR-B	4.2	Methicillin resistance regulatory protein MECI
27	3CTA-A	4.2	Riboflavin kinase
28	3C18-B	4.2	Nucleotidyltransferase-like protein
29	1OKR-A	4.2	Methicillin resistance regulatory protein MECI
30	2FU4-B	4.2	Ferric uptake regulation protein
31	1YQA-A	4.2	Histone H1
32	1DPR-B	4.2	Diphtheria toxin repressor
33	1MZB-A	4.1	Ferric uptake regulation protein
34	3DV8-A	4.1	Transcriptional regulator, CRP/FNR family
35	1W1W-E	4.1	Sister chromatic cohesion protein 1
36	2RGV-A	4.1	Peroxide operon regulator
37	2FE3-A	4.1	Peroxide operon regulator
38	2D45-C	4.1	Methicillin resistance regulatory protein MECI
39	1DDN-D	4.1	Diphtheria toxin repressor
40	1U6G-A	4.1	Cullin homolog 1
41	2PFB-B	4.1	Transcriptional regulator OHRR
42	2IT0-B	4.1	Iron-dependent repressor IDER

43	2ISZ-A	4.1	Iron-dependent repressor IDER
44	1U8R-I	4.1	Iron-dependent repressor IDER
45	1U8R-B	4.1	Iron-dependent repressor IDER
46	2IT0-D	4.1	Iron-dependent repressor IDER
47	1U8R-H	4.1	Iron-dependent repressor IDER
48	1U8R-C	4.1	Iron-dependent repressor IDER
49	2QWW-A	4.0	Transcriptional regulator, MARR family
50	1DPU-A	4.0	Replication protein A (RPA32) C-terminal domain
51	2HR3-A	4.0	Probable transcriptional regulator
52	1W1W-F	4.0	Sister chromatic cohesion protein 1
53	1JHH-A	4.0	LexA repressor
54	1SD4-A	4.0	Penicillinase repressor
55	2QWW-G	4.0	Transcriptional regulator, MARR family
56	2D45-B	4.0	Methicillin resistance regulator protein MECI
57	1W1W-H	4.0	Sister chromatic cohesion protein 1
58	1W1W-G	4.0	Sister chromatic cohesion protein 1
59	2ZME-A	4.0	Vacuolar-sorting protein SNF8
60	1F5T-D	4.0	Diphtheria toxin repressor
61	3BPV-A	4.0	Transcriptional regulator
62	2ISZ-B	4.0	Iron-dependent repressor IDER
63	2ISZ-C	4.0	Iron-dependent repressor IDER
64	2IT0-C	4.0	Iron-dependent repressor IDER
65	1U8R-D	4.0	Iron-dependent repressor IDER
66	2ISY-A	4.0	Iron-dependent repressor IDER
67	1U8R-G	4.0	Iron-dependent repressor IDER
68	IU8R-J	4.0	Iron-dependent repressor IDER
69	IU8R-A	4.0	Iron-dependent repressor IDER
70	1C0W-B	3.9	Diphtheria toxin repressor
71	2JT1-A	3.9	PEFI protein
72	1U5T-A	3.9	Appears to be functionally related to SNF7
73	2HR3-C	3.9	Probable transcriptional regulator
74	11EA	3.9	LEXA repressor DNA binding domain
75	1Z1D-A	3.9	Replication protein A – 32 KDa subunit
76	1SD6-B	3.9	Methicillin resistance regulatory protein MECI
77	1B1B-A	3.9	Iron dependent regulator
78	1SD7-A	3.9	Methicillin resistance regulatory protein MECI
79	2QWW-F	3.9	Transcriptional regulator, MARR family
80	2PFB-A	3.9	Transcriptional regulator OHRR
81	2NNN-F	3.9	Probable transcriptional regulator
82	2NNN-H	3.9	Probable transcriptional regulator
83	2NNN-D	3.9	Probable transcriptional regulator
84	2NNN-G	3.9	Probable transcriptional regulator
85	2IT0-A	3.9	Iron-dependent repressor IDER
86	2GXB-B	3.9	Double-stranded RNA-specific adenosine deaminase
87	2ACJ-D	3.9	Double-stranded RNA-specific adenosine deaminase
88	2GXB-A	3.9	Double-stranded RNA-specific adenosine deaminase
89	2ISZ-D	3.9	Iron-dependent repressor IDER
90	1FX7-B	3.9	Iron-dependent repressor IDER
91	2ISY-B	3.9	Iron-dependent repressor IDER
92	1WQ2-B	3.9	Dissimilatory Sulfite Reductase D

93	1XSD-A	3.8	Penicillinase repressor
94	11NW-A	3.8	Multidrug resistance operon repressor
95	1HST-A	3.8	Histone H5 (globular domain)
96	2CO5-A	3.8	Viral protein F93
97	1SD7-B	3.8	Methicillin resistance regulatory protein MECI
98	2G9W-B	3.8	Rv1846c, a putative transcriptional regulatory protein of Mycobacterium tuberculosis
99	1SAX-A	3.8	Methicillin resistance regulatory protein MECI
100	1COW-C	3.8	Diphtheria toxin repressor

# Bibliography

## **Bibliography**

Alekshun M. N., Levy S. B., Mealy T. R., Seaton B. A. & Head J. F. "The crystal structure of MarR, a regulator of multiple antibiotic resistance, at 2.3 Å resolution" (2001) *Nature Struct. Biol.* vol. 8 pp 710-714

Altschul S. F., Gish W., Miller W., Myers E. W. & Lipman D. J. "Basic local alignment search tool" (1990) *J. Mol. Biol.* vol. 215 pp 403-410

Anversa P., Beghi C., Kikkawa Y. & Olivetti G. "Myocardial infarction in rats. Infarct size, myocyte hypertrophy and capillary growth" (1986) *Circ. Res.* vol. 58 pp 26-37

Arai A., Spencer J. A. & Olson E. N. "STARS, a striated muscle activator of Rho signalling and serum response factor-dependent transcription" (2002) *J. Biol. Chem.* vol. 277 pp 24453-24459

Aravind L., Anantharaman V., Balaji S., Mohan Babu M. & Iyer L. M. "The many faces of the helix-turn-helix domain: Transcription regulation and beyond" (2005) *FEMS Microbiology Review* vol. 29 pp 231-262

Aravind L. & Pontin C. P. "Homologues of the 26S proteasome subunits are regulators of transcription and translation" (1998) *Protein Sci.* vol. 7 pp 1250-1254

Arber S. & Caroni P. "Specificity of single LIM motifs in targeting and LIM/LIM interactions in situ" (1996) *Genes Dev.* vol. 10 pp 289-300

Arsenian S., Weinhold B., Oelgeschlager M., Ruther U. & Nordheim A. "Serum response factor is essential for mesoderm formation during mouse embryogenesis" (1998) *EMBO J.* vol. 17 pp 6289-6299

Ashley E. A. & Niebauer J. (2004) "Cardiology Explained" (Remedica Explained Series) (1<sup>st</sup> Edition), Remedica

Ashley E. A., Raxwal V. K. & Froelicher V. F. "The prevalence and prognostic significance of electrocardiographic abnormalities" (2002) *Curr. Probl. Cardiol.* vol. 1 pp 1-72

Barrientos T., Frank D., Kuwahara K., Bezprozvannaya S., Pipes G. C., Bassel-Duby R., Richardson J. A., Katus H. A., Olsen E. N. & Frey N. "Two novel members of the ABLIM protein family, ABLIM-2 and -3, associate with STARS and directly bind F-Actin" (2007) *J. Biol. Chem.* vol. 282 pp 8393-8403

Bashour A. M., Fullerton A. T., Hart M. J. & Bloom G. S. "IQGAP1, a Rac- and Cdc42-binding protein, directly binds and cross-links microfilaments" (1997) *J. Cell. Biol.* vol. 137 pp 1555-1566

Belaguli N. S., Sepulveda J. L., Nigam V., Charron F., Nemer M. & Schwartz R. J. "Cardiac tissue enriched factors serum response factor and GATA-4 are mutual coregulators" (2000) *Mol. Cell. Biol.* vol. 20 pp 7550-7558

- Black B. L. & Olson E. N. "Transcriptional control of muscle development by myocyte enhancer factor-2 (MEF2) proteins" (1998) *Annu. Rev. Cell Dev. Biol.* vol. 14 pp 167-196
- Blais A. & Dynlacht B. D. "E2F-associated chromatin modifiers and cell cycle control" (2005) *Curr. Opin. Cell Biol.* vol. 19 pp 658-662
- Bodenhausen G. & Ruben D. J. "Natural Abundance Nitrogen-15 NMR by enhanced heteronuclear spectroscopy" (1980) *Chemical Physics Letters* vol. 69 pp 185-189
- Brennan R. G. "The winged-helix DNA-binding motif: another helix-turn-helix takeoff" (1993) *Cell* vol. 74 pp 773-776
- Brennan R. G., Takeda Y., Kim J., Anderson W. F. & Matthews B. W. "Crystallisation of a complex of cro repressor with a 17 base-pair operator" (1986) *J. Mol. Biol.* vol. 188 pp 115-118
- Brill S., Li S., Lyman C.W., Church D. M., Wasmuth J. J., Weissbach L., Bernards A. & Snijder A. J. "The Ras GTPase-activating-protein-related human protein IQGAP2 harbours a potential actin binding domain and interacts with calmodulin and Rho family GTPases" (1996) *Mol Cell Biol.* vol. 16 pp 4869-78
- Buchwalter G., Gross C. & Waslyk B. "Ets ternary complex transcription factors" (2004) *Gene* vol. 324 pp 1-14
- Burridge K. & Chrzanowska-Wodnicka M. "Focal adhesions, contractility and signalling" (1996) *Annu. Rev. Cell Dev. Biol.* vol 12 pp 463-518
- Carballido-Lopez R., Formstone A., Li Y., Ehrlich S. D., Noirot P. & Errington J. "Actin homolog MreBH governs cell morphogenesis by localization of the cell wall hydrolase LytE" (2006) *Dev. Cell* vol. 11 pp 399-409
- Chang D. F., Belguli N. S., Iyer D., Roberts W. B., Wu S. P., Dong X. R., Marx J. G., Moore M. S., Beckerle M. C., Majesky M. W. & Schwartz R. J. "Cysteine-rich LIM-only proteins CRP1 and CRP2 are potent smooth muscle differentiation cofactors" (2003) *Dev. Cell.* vol. 4 pp 107-118
- Cen B., Selvaraj A. & Prywes R. "Myocardin/MKL family of SRF co-activators: key regulators of immediate early and muscle specific gene expression" (2004) *J. Cell Biochem.* vol 93 pp 74-82
- Chen C. Y. & Schwartz R. J. "Recruitment of the tinman homolog Nkx-2.5 by serum response factor activates cardiac  $\alpha$ -actin gene transcription" (1996) *Mol. Cell. Biol.* vol. 16 pp 6372-6384
- Clark K. L., Halay E. D., Lai E. & Burley S. K. "Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5" (1993) *Nature* vol. 364 pp 412-420

Clore G. M., Driscoll P. C., Wingfield P. T. & Gronenborn A. M. "Analysis of the backbone dynamics of interleukin-1 beta using two-dimensional inverse detected heteronuclear  $^{15}\text{N}$ - $^1\text{H}$  NMR spectroscopy" (1990) *Biochemistry* vol. 29 pp 7387-7401

Clore G. M., Starich M. R. & Gronenborn A. M. "Measurement of residual dipolar couplings of macromolecules aligned in the nematic phase of a colloidal suspension of rod-shaped viruses" (1998) *J. Am. Chem. Soc.* Vol. 120 pp 10571-10572

Cook W. J., Kar S. R., Taylor K. B. & Hall L. M. "Crystal structure of the cyanobacterial metallothionein repressor SmtB: a model for metalloregulatory proteins" (1998) *J. Mol. Biol.* vol 275 pp 337-346

Cokol M., Nair R. & Rost B. "Finding nuclear localisation signals" (2000) *EMBO Reports* vol 1 pp 411-415

Cornilescu G., Delaglio F. & Bax A. "Protein backbone angle restraints from searching a database for chemical shift and sequence homology" (1999) *Journal of Biomolecular NMR* vol 13 pp 289-302

Croissant J. D., Kim J. H., Eichele G., Goering L., Lough J., Prywes R. & Schwartz R. J. "Avian serum response factor expression restricted primarily to muscle cell lineages is required for alpha-actin gene transcription" (1996) *Dev. Biol.* vol. 177 pp 250-264

Dalton S. & Treisman R. "Characterisation of SAP-1, a protein recruited by serum response factor to the c-Fos serum response element" (1992) *Cell* vol. 68 pp 597-612

Davis F. J., Gupta M., Camorerri-Mercado B., Schwartz R. J. & Gupta M. P. "Calcium/calmodulin-dependent kinase activates serum response factor transcription activity by its dissociation from histone deacetylase, HDAC4. Implications in cardiac muscle gene regulation during hypertrophy" (2003) *J. Biol. Chem.* vol. 278 pp 20047-20058

Deane J. E., Ryan D. P., Sunde M., Maher M. J., Guss J. M., Visvader J. E. & Matthews J. M. "Tandem LIM domains provide synergistic binding in the LMO4:Ldb1 complex" (2004) *EMBO J.* vol. 23 pp 3589-3598

Dos Remedios C. G., Chhabra D., Kekic M., Dedova I. V., Tsubakihara M., Berry D. A. & Nosworthy N. J. "Actin binding proteins: regulation of cytoskeletal microfilaments" (2003) *Physiol. Rev.* vol. 83 pp 433-473

Dosset P., Hus J.C., Marion D. & Blackledge M. "A novel interactive tool for rigid-body modelling of multi-domain macromolecules using residual dipolar couplings" (2001) *J. Biomol. NMR* vol. 20 pp 223-231

El-Mezgueldi M., Mendre C., Calas B., Kassab R. and Fattoun A., "Characterisation of the regulatory domain of gizzard calponin: interaction of 145-163 region with F-actin, calcium-binding proteins and tropomyosin" (1995) *J. Biol. Chem.* vol. 270 pp 8867-8876

Feuerstein R., Wang X., Song D., Cooke N. E. & Liebhaber S. A. "The LIM/double zinc-finger motif functions as a protein dimerization domain" (1994) *Proc. Natl. Acad. Sci. USA* vol. 91 pp 10655-10659

Fogh R., Ionides J., Ulrich E., Boucher W., Vranken V., Linge J. P., Habeck M., Rieping W., Bhat T. N., Westbrook J., Henrick K., Gilliland G., Berman H., Thornton J., Nilges M., Markley J. & Laue E. "The CCPN project: An Interim Report on a data model for the NMR community" (2002) *Nat. Struct. Biol.* vol. 9 pp 416-418

Freyd G., Kim S. K. & Horvitz H. R. "Novel cysteine-rich motif and homeodomain in the product of the *Caenorhabditis elegans* cell lineage gene *lin-11*" (1990) *Nature* vol. 344 pp 876-879

Fukata M., Kuroda S., Fujii K., Nakamura T., Shoji I., Matsuura Y., Okawa K., Iwamatsu A., Kikuchi A. & Kaibuchi K. "Regulation of cross-linking of actin filament by IQGAP1, a target for Cdc42" (1997) *J. Biol. Chem.* vol. 272 pp 29579-29583

Gajiwala K. S. & Burley S. K. "Winged Helix Proteins" (2000) *Current Opinion in Structural Biology* vol. 10 pp 110-116

Glennon P.E., Sugden P. H. & Poole-Wilson P. A. "Cellular mechanisms of cardiac hypertrophy" (1995) *Br. Heart J.* vol. 73 pp 496-499

Gronenborn A. M. "The Importance of being ordered: improving NMR structures using residual dipolar couplings" (2002) *Biophysics* vol. 325 pp 957-966

Grzesiek S. "Correlating backbone amide and side-chain resonances in larger proteins by multiple relayed triple resonance NMR" (1992) *Journal of the American Chemical Society* vol. 114 pp 6291-6293

Güntert P., 2004 "Automated NMR Structure Calculation With CYANA" in A. K. Downing ed. *Protein NMR Techniques (Methods in Molecular Biology)* (2<sup>nd</sup> Edition), Humana Press, pp 353 -378

Güntert P., Mumenthaler C. & Wüthrich K. "Torsion Angle Dynamics for NMR Structure Calculation with the New Program DYANA" (1997) *J. Mol. Biol.* vol. 273 pp 283-288

Güntert P. & Wüthrich K. "Improved efficiency of protein structure calculations from NMR data using the program DIANA with redundant dihedral angle constraints" (1991) *J. Biomol. NMR* vol. 1 pp 447-456

Guettler S., Vartianinen M. K., Miralles F., Larijani B. & Treisman R. "RPEL motifs link the serum response factor cofactor MAL but not myocardin to Rho signalling via actin binding" (2008) *Mol. Cell. Biol.* vol. 28 pp 732-742

Hansen M. R., Mueller L., & Pardi A. "Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions" (1998) *Nat. Struct. Biol.* vol. 5 pp 1065-1074

- Hart M. J., Callow M. G., Souza B. & Polakis P. "IQGAP1, a calmodulin-binding protein with a rasGAP-related domain, is a potential effector for cdc42Hs" (1996) *Embo J.* vol. 15 pp 2997-3005
- Haun R. S., Serventi I. M. & Moss J. "Rapid, reliable ligation-independent cloning of PCR products using modified plasmid vectors" (1992) *Biotechniques* vol. 13 pp 515-518
- Herrmann T., Güntert P. & Wüthrich K. "Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA." (2002) *J. Mol. Biol.* vol. 319 pp 209-227
- Hipskind R. A., Rao V. N., Mueller C. G., Reddy E. S. & Nordheim A. "Ets-related protein Elk-1 is homologous to the c-Fos regulatory factor p63TCF" (1991) *Nature* vol. 354 pp 531-534
- Hoffmann K. & Butcher P. "The PCI domain: a common theme in three multiprotein complexes" (1998) *Trends Biochem. Sci.* vol. 23 pp 204-205
- Holm L., Kaariainen S., Rosenstrom P. & Shenkel A. "Searching Protein Structure Databases With Dali Lite v3" (2008) *Bioinformatics* vol. 24 pp 2780-2781
- Hooft R. W., Sander C. & Vriend G. "Objectively judging the quality of a protein structure from a Ramachandran plot" (1997) *Comput. Appl. Biosci.* vol. 13 pp 425-430
- Hudlicka O., Brown M. & Egginton S. "Angiogenesis in Skeletal and Cardiac Muscle" (1992) *Physiol Rev.* vol. 72 pp 369-417
- Ikura M., Kay L. E. & Bax A. "A Novel Approach for Sequential Assignment of <sup>1</sup>H, <sup>13</sup>C and <sup>15</sup>N Spectra of Larger Proteins: Heteronuclear Triple-Resonance Three-Dimensional NMR Spectroscopy. Application to Calmodulin" (1990) *Biochemistry* vol. 29 pp 4659-4667
- Jardetzky O. "Protein dynamics and conformational transitions in allosteric proteins" (1996) *Prog. Biophys. Mol. Biol.* vol. 65 pp 171-219
- Jeffries C. M., Graham S. C., Stokes P. H., Collyer C. A., Guss J. M. & Matthews J. M. "Stabilization of a binary protein complex by intein-mediated cyclization" (2006) *Protein Sci.* vol. 15 pp 2612-2618
- Jiskoot W., Hlady V., Naleway J. J. & Herron J. N. "Application of Fluorescence Spectroscopy for Determining the Structure and Function of Proteins" (1995) *Pharm. Biotechnol.* vol. 7 pp 1-63
- Jones L. J. F., Carballido-Lopez R. & Errington J. "Control of Cell Shape in Bacteria: helical, actin-like filaments in *Bacillus subtilis*" (2001) *Cell* vol. 104 pp 913-922
- Kannel W. B. "Prevalence and natural history of electrocardiographic left ventricular hypertrophy" (1983) *American Journal of Medicine* vol. 75 pp 4-11

- Karlsson O., Thor S., Norberg T., Ohlsson H. & Edlund T. "Insulin gene enhancer binding protein Isl-1 is a member of a novel class proteins containing both a homeo- and a Cys-His domain" (1990) *Nature* vol. 344 pp 879-882
- Kaufmann E. & Knöchel W. "Five years on the wings of fork head" (1996) *Mechanisms of Development* vol. 57 pp 3 - 20
- Kibbe W. A. "OligoCalc: an online oligonucleotide properties calculator" (2007) *Nucleic Acid Research* vol. 35 (web server issue) W43-46
- Kirkpatrick S., Gelatt C. D. Jr. & Vecchi M. P. "Optimization by simulated annealing" (1983) *Science* vol. 220 pp 671 – 680
- Kitano K., Kim S. Y. & Hakoshima T. "Structural basis for DNA strand separation by the unconventional winged-helix domain of RecQ helicase WRN" (2010) *Structure* vol. 10 pp 177-187
- Koenig B. W., Hu J. S., Ottiger M., Bose S., Hendler R. W. & Bax A. "NMR measurement of dipolar coupling in proteins aligned by transient binding to purple membrane fragments" (1999) *J. Am. Chem. Soc.* Vol. 121 pp 1385-1386
- Kosa J. L., Michelsen J. W., Louis H. A., Olsen J. I., Davis D. R., Beckerle M. C. & Winge D. R. "Common metal ion coordination in LIM domain proteins" (1994) *Biochemistry* vol. 33 pp 468-477
- Kuwahara K., Barrientos T., Pipes G. C., Li S. & Olson E. N. "Muscle-specific signalling mechanism that links actin dynamics to serum response factor" (2005) *Mol. Cell Biol.* vol 25 pp 3173-3181
- Kuwahara K., Teg Pipes G. C., McAnally J., Richardson J. A., Hill J. A., Bassel-Duby R. & Olson E. N. "Modulation of adverse cardiac remodelling by STARS, a mediator of MEF2 signalling and SRF activity" (2007) *J. Clin. Invest.* Vol 117 pp 1324-1334
- Lai E., Prezioso V. R., Rao W. F., Chen W. S. & Darnell jr J. E. "Hepatocyte nuclear factor 3 alpha belongs to a gene family in mammals that is homologous to the *Drosophila* homeotic gene fork head" (1991) *Genes Dev.* Vol 5 pp 416-427
- Lai E., Prezioso V. R., Smith E., Litvin O., Costa R. H. & Darnell jr J. E. "HNF-3A, a hepatocyte-enriched transcription factor of novel structure is regulated transcriptionally" (1990) *Genes Dev.* Vol 4 pp 1427-1436
- Lamon S., Wallace M. A., Léger B. & Russel A.P. "Regulation of STARS and its downstream targets suggests a novel pathway involved in human skeletal muscle hypertrophy and atrophy" (2009) *J. Physiol* vol 587 pp 1795-1803
- Larkin M. A., Blackshields G., Brown N. P., Chenna R., McGettigan P. A., McWilliam H., Valentin F., Wallace I. M., Wilm A., Lopez R., Thompson J. D., Gibson T. J. & Higgins D. G. "Clustal W and Clustal X version 2.0" (2007) *Bioinformatics* vol. 23 pp 2947-2948

Laskowski R. A., MacArthur M. W., Moss D. S. & Thornton J. M. "PROCHECK – a programme to check the stereochemical quality of protein structures." (1993) *J. App. Cryst.* vol 26 pp 283-291

Li J., Zhu X., Chen M., Cheng L., Zhou D., Lu M. M., Du K., Epstein J. A. & Parmacek M. S. "Myocardin-related transcription factor B is required in cardiac neural crest for smooth muscle differentiation and cardiovascular development" (2005) *Proc. Natl. Acad. Sci. USA* vol. 102 pp 8916-8921

Lipari G. & Szabo A. "Model-free approach to the interpretation of nuclear magnetic-resonance relaxation in macromolecules. 1. Theory and range of validity" (1982) *Journal of the American Chemical Society* vol. 104 pp 4546-4559

Lipari G. & Szabo A. "Model-free approach to the interpretation of nuclear magnetic-resonance relaxation in macromolecules. 2. Analysis of experimental results" (1982) *Journal of the American Chemical Society* vol. 104 pp 4559-4570

Liu Y., Sinha S., McDonald O. G., Shang Y., Hoofnagle M. H. & Owens G. K. "Kruppel-like factor 4 abrogates myocardin-induced activation of smooth muscle gene expression" (2005) *J. Biol. Chem.* vol. 280 pp 9719-9727

Liu J., Taylor D. W. & Taylor K. A. "A 3-D reconstruction of smooth muscle alpha-actinin by cryo-EM reveals two different conformations at the actin binding region" (2004) *J. Mol. Biology* vol. 338 pp 115-125

Liu S., Widom J., Kemp C. W., Crews C. M. & Clardy J. "Structure of human methionine aminopeptidase-2 complexed with fumafillin" (1998) *Science* vol. 282 pp 1324-1327

Lundquist E. A., Herman R. K., Shaw J. E. & Bargmann C. I. "UNC-115, a conserved protein with predicted LIM and actin-binding domains, mediates axon guidance in *C. elegans*" (1998) *Neuron*. vol. 21 pp 385-392

Mack C.P., Somlyo A. V., Hautmann M., Somlyo A. P. & Owens G. K. "Smooth muscle differentiation marker gene expression is regulated by RhoA-mediated actin polymerisation" (2001) *J. Biol. Chem.* vol. 276 pp 341-347

Mahadeva H., Brooks G., Lodwick D., Chong N. W. & Samani N. J. "ms1, a novel stress-responsive, muscle-specific gene that is up-regulated in the early stages of pressure overload-induced left ventricular hypertrophy" (2002) *FEBS Lett.* vol. 521 pp 100-104

Medvedeva S., Simorre J-P., Brutscher B., Guerlesquin F. & Marion D. "Extensive <sup>1</sup>H NMR resonance assignment of proteins using natural abundance gradient-enhanced <sup>13</sup>C-<sup>1</sup>H correlation spectroscopy" (1993) *FEBS Lett.* vol. 333 pp 251-256

Meissner A. & Sorensen O. W. "Optimisation of three-dimensional TROSY-type HCCH correlation of aromatic (1)H-(13)C groups in proteins" (1999) *J. Magn. Reson.* vol. 139 pp 447-450

- Miano J. M. "Serum response factor: toggling between disparate programs of gene expression" (2003) *J. Mol. Cell Cardiol.* vol. 35 pp 577-593
- Michelsen J. W., Sewell A. K., Louis H. A., Olsen J. I., Davis D. R., Winge D. R. & Beckerle M. C. "Mutational analysis of the metal sites in an LIM domain" (1994) *J. Biol. Chem.* vol. 269 pp 11108-11113
- Miralles F., Posern G., Zaromytidou A. I. & Treisman R. "Actin dynamics control SRF activity by regulation of its coactivator MAL" (2003) *Cell* vol. 113 pp 329-342
- Molkentin J. D., Black B. L., Martin J. F. & Olson E. N. "Cooperative activation of muscle gene expression by MEF2 and myogenic bHLH proteins" (1995) *Cell* vol. 83 pp 1125-1136
- Molkentin J. D. & Olson E. N. "Combinatorial control of muscle development by basic helix-loop-helix and MADS-box transcription factors" (1996) *Proc. Natl. Acad. Sci. USA* vol. 93 pp 9366-9373
- Møller-Jensen J., Borch J., Dam M., Jensen R. B. Roepstorff P., Gerdes K. and Löwe J. "Bacterial mitosis: ParM of plasmid R1 moves plasmid DNA by an actin-like insertional polymerisation mechanism" *Mol Cell* (2003) vol 12 pp 1477-87
- Neuman N. A., Ma S., Schnitzler G. R., Zhu Y., Lagna G. & Hata A. "The four-and-a-half LIM domain protein 2 regulates vascular smooth muscle phenotype and vascular tone" *J. Biol. Chem.* (2009) vol 284 pp 13202-13212
- Ohkawa Y., Marfella C. G. & Imbalzano A. N. "Skeletal muscle specification by myogenin and Mef2D via the SWI/SNF ATPase Brg1" (2006) *EMBO J.* vol. 25 pp 490-501
- Ohkawa Y., Yoshimura S., Higashi C., Marfella C. G., Dacwag C. S., Tachibana T. & Imbalzano A. N. "Myogenin and the SWI/SNF ATPase Brg1 maintain myogenic gene expression at different stages of skeletal myogenesis" (2007) *J. Biol. Chem.* vol. 282 pp 6564-6570
- Ottiger M., Delaglio F. & Bax A. "Measurement of J and Dipolar Couplings from Simplified Two-Dimensional NMR Spectra" (1998) *J. Magn. Resonance* vol. 131 pp 373-378
- Ounzain S., Dacwag C. S., Samani N. J., Imbalzano A. N. & Chong N. W. "Comparative in silico analysis identifies bona fide MyoD binding sites within the Myocyte Stress 1 gene promoter" (2008) *B.M.C Molecular Biology* vol. 9 pp 50
- Owens G. K., Kumar M. S. & Wamhoff B. R. "Molecular regulation of vascular smooth muscle cell differentiation in development and disease" (2004) *Physiol. Rev.* vol. 84 pp 767-801
- Palmer A. G. "Probing molecular motion by NMR" (1997) *Curr. Opin. Struct. Biol.* vol. 7 pp 732-737

- Pantaloni D., Le Clainche C. & Carlier M. F. "Mechanism of actin-based motility" (2001) *Science* vol. 292 pp 1502-1506
- Parmacek M. S. "Myocardin-related transcription factors: critical co-activators regulating cardiovascular development and adaptation" (2007) *Circ. Res.* vol. 100 pp 633-644
- Pellegrini L., Tan S. & Richmond T. J. "Structure of serum response factor core bound to DNA" (1995) *Nature* vol. 376 pp 490-498
- Perez-Alvarado G. C., Miles C., Michelsen J. W., Louis H. A., Winge D. R., Beckerle M. C. & Summers M. F. "Structure of the carboxy-terminal LIM domain from the cysteine rich protein CRP" (1994) *Nat. Struct. Biol.* vol 1 pp 388-398
- Pfuhl M., Al-Sarayreh S. & El-Mezgueldi M. "The Calponin Regulatory Region is Intrinsically Unstructured: Novel Insight Into Actin-Calponin and Calmodulin-Calponin Interfaces Using NMR Spectroscopy" (2011) *Biophys. J.* vol. 100 pp 1718-1728
- Philippar U., Schrott G., Dieterich C., Müller J. M., Galgóczy O., Engel F. B., Keating M. T., Gertler F., Schüle R., Vingron M. & Nordheim A., "The SRF target gene Fhl2 antagonises RhoA/MAL-dependent activation of SRF" (2004) *Mol. Cell.* vol. 16 pp 867-880
- Pipes G. C., Creemers E. E. & Olson E. N. "The myocardin family of transcriptional coactivators: versatile regulators of cell growth, migration and myogenesis" (2006) *Genes Dev.* vol. 20 pp 1545-1556
- Ramachandran G. N., Ramakrishnan C. & Sasisekharan V. "Stereochemistry of polypeptide chain configurations" (1963) *J. Mol. Biol.* vol. 7 pp 95-99
- Ramakrishnan V., Finch J. T., Graziano V., Lee P. L. & Sweet R. M. "Crystal structure of globular domain of histone H5 and its implications for nucleosome binding" (1993) *Nature* vol. 362 pp 219-223
- Reich S., Puckey L. H., Cheetham C. L., Harris R., Ali A. A., Bhattacharyya U., Maclagan K., Powell K. A., Prodromou C., Pearl L. H., Driscoll P. C. & Savva R. "Combinatorial Domain Hunting: An effective approach for the identification of soluble protein domains adaptable to high-throughput applications" (2006) *Protein Sci.* vol. 15 pp 2356-2365
- Richey P. A. & Brown S. P. "Pathological versus Physiological Left Ventricular Hypertrophy: A Review" (1998) *Journal of Sports Sciences* vol. 16 pp 121-141
- Ridley A. J. & Hall A. "The small GTP-binding protein rho regulates the assembly of focal adhesions and actin stress fibers in response to growth factors" (1992) *Cell* vol. 70 pp 389-399
- Roof D. J., Hayes A., Adamian M., Chishti A. H. & Li T. "Molecular characterization of abLIM, a novel actin-binding and double zinc finger protein" (1997) *J. Cell. Biol.* vol. 138 pp 575-588

- Rückert M. & Otting G. "Alignment of biological macromolecules in novel non-ionic liquid crystalline media for NMR experiments" (2000) *J. Am. Chem. Soc.* Vol. 122 pp 7793-7797
- Sadoshima J., Takahashi T., Jahn L. & Izumo S. "Roles of mechano-sensitive ion channels, cytoskeleton, and contractile activity in stretch-induced immediate-early gene expression and hypertrophy of cardiac myocytes" (1992) *Proc. Natl. Acad. Sci. USA* vol 89 pp 9905-9909
- Sass H. J., Cordier F., Hoffmann A., Rogowski M., Cousin A., Omichinski J. G., Lowen H., & Grzesiek S. "Purple membrane induced alignment of biological macromolecules in the magnetic field" (1999) *J. Am. Chem. Soc.* vol. 121 pp 2047-2055
- Sass H. J., Musco G., Stahl S. J., Wingfield P. T. & Grzesiek S. "Solution NMR of proteins within polyacrylamide gels: diffusional properties and residual alignment by mechanical stress or embedding of orientated purple membranes" (2000) *J. Biomol. NMR* vol. 18 pp 303-309
- Schmeichel K. L. & Beckerle M. C. "The LIM domain is a modular protein-binding interface" (1994) *Cell* vol. 79 pp 211-219
- Schmidt A. & Hall M. N. "Signalling to the actin cytoskeleton" (1998) *Annu. Rev. Cell Dev. Biol.* vol. 14 pp 305-338
- Sepulveda J. L., Vlahopoulos S., Iyer D., Belaguli N. & Schwartz R. J. "Combinatorial expression of GATA4, Nkx2-5 and serum response factor directs early cardiac gene activity" (2002) *J. Biol. Chem* vol. 277 pp 25775-25782
- Sokolow M. & Lyon T. P. "The Ventricular Complex in Left Ventricular Hypertrophy as Obtained by Unipolar Precordial and Limb Leads" (1949) *Am. Heart J.* vol. 37 pp 161-186
- Sotiropoulos A., Gineitis D., Copeland J. & Treisman R. "Signal-regulated activation of serum response factor is mediated by changes in actin dynamics" (1999) *Cell* vol. 98 pp 159-169
- Spolar R. S. & Record M. T. Jr. "Coupling of local folding to site-specific binding of proteins to DNA" (1994) *Science* vol. 263 pp 777-784
- Struckhoff E. C. & Lundquist E. A. "The actin-binding protein UNC-115 is an effector of Rac signalling during axon pathfinding in *C. elegans*" (2003) *Development* vol. 130 pp 693-704
- Swindells M. B. "Identification of a common fold in the replication terminator protein suggests a possible mode for Dna binding" (1995) *Trends Biochem. Sci.* vol. 20 pp 300-302
- Takano H., Komuro I., Oka T., Shiojima I., Mizuno T. & Yazaki Y. "The Rho family G proteins play a critical role in muscle differentiation" (1998) *Mol. Cell. Biol.* vol. 18 pp 1580-1589

- Thompson J. D., Higgins D. G. & Gibson T. J. "CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice" (1994) *Nucleic Acids Res.* vol. 22 pp 4673-4680
- Tjandra N. & Bax A. "Direct Measurement of Distances and Angles in Biomolecules by NMR in a Dilute Liquid Crystalline Medium" (1997) *Science* vol. 278 pp 1111-1114
- Treisman R. "Identification of a protein-binding site that mediates transcriptional response of the *c-fos* gene to serum factors" (1986) *Cell* vol. 46 pp 567-574
- Troidl K., Rüdiger I., Wei-Jun C., Mücke Y., Grossekkettler L., Piotrowska I., Apfelbeck H., Schierling W., Volger O.L., Horrevoets A. J., Grote K., Schmitz-Rixen T., Schaper W. & Troidl C. "Actin-binding Rho Activating Protein (ABRA) is essential for fluid shear stress-induced arteriogenesis" (2009) *Arterioscler. Thromb. Vasc. Biol.* vol. 29 pp 2093-2101
- Tsai K-L., Sun Y-J., Huang C-Y., Yang J-Y., Hung M-C. & Hsiao C-D. "Crystal structure of the human FOXO3a-DBD/DNA complex suggests the effects of post-translational modification." (2007) *Nucleic Acids Research* vol. 35 pp 6984-6994
- Tuerk C. & Gold L. "Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase" (1990) *Science* vol. 249 pp 505-510
- Tycko R., Blanco F. J. & Ishii Y. "Alignment of biopolymers in strained gels: a new way to create detectable dipole-dipole couplings in high resolution biomolecular NMR" (2000) *J. Am. Chem. Soc.* vol. 122 pp 9340-9341
- Vandromme M., Gauthier-Rouviere C., Carnac G., Lamb N. & Fernandez A. "Serum response factor p67SRF is expressed and required during myogenic differentiation of both mouse C2 and rat L6 muscle cell lines" (1992) *J. Cell Biol.* vol. 118 pp 1489-1500
- Wagner G. "NMR relaxation and protein mobility" (1993) *Curr. Opin. Struct. Biol.* vol. 3 pp 748-754
- Wah D. A., Hirsch J. A., Dorner L. F., Schildkraut I. & Aggarwal A. K. "Structure of the multimodular endonuclease FokI bound to DNA" (1997) *Nature* vol. 388 pp 97-100
- Wang D., Chang P. S., Wang Z., Sutherland L., Richardson J. A., Small E., Krieg P. A. & Olson E. N. "Activation of cardiac gene expression by myocardin, a transcriptional cofactor for serum response factor" (2001) *Cell* vol. 105 pp 851-62
- Wang D., Li S., Hockemayer D., Sutherland L., Wang Z., Schrott G., Richardson J. A., Nordheim A. & Olson E. N. "Potentiation of serum response factor activity by a family of myocardin-related transcription factors" (2002) *Proc. Natl. Acad. Sci. USA* vol. 99 pp 14855-14860
- Wang Z., Wang D. Z., Hockemeyer D., McAnally J., Nordheim A. & Olson E. N. "Myocardin and ternary complex factors compete for SRF to control smooth muscle gene expression" (2004) *Nature* vol. 428 pp 185-189

- Wang Z., Wang D.Z., Pipes G. C. & Olson E. N. "Myocardin is a master regulator of smooth muscle gene expression" (2003) *Proc. Natl. Acad. Sci. USA* vol. 106 pp 18734-18739
- Way J. C. & Chalfie M. "The *mec-3* gene of *Caenorhabditis elegans* requires its own product for maintained expression and is expressed in three neuronal cell types" (1989) *Genes Dev.* vol. 3 pp 1823-1833
- Wei L., Zhou W., Croissant J. D., Johansen F. E., Prywes R., Balasubramanyam A. & Schwartz R. J. "RhoA signalling via serum response factor plays an obligatory role in myogenic differentiation" (1998) *J. Biol. Chem.* vol. 273 pp 30287-30294
- Weigel D., Jürgens G., Küttner F., Seifert E. & Jäckle H. "The homeotic gene *fork head* encodes a nuclear protein and is expressed in the terminal regions of the *Drosophila* embryo" (1989) *Cell* vol. 57 pp 645-658
- Weiss M. A., Ellenberger T., Wobbe C. R., Lee J. P., Harrison S. C. & Struhl K. "Folding transition in the DNA-binding domain of GCN4 on specific binding to DNA" (1990) *Nature* vol. 347 pp 575-578
- Wijchers P. J. E. C., Burbach J. P. H. & Smidt M. P. "In control of biology: of mice, men and Foxes" (2006) *Biochem J.* vol. 397 pp 233-246
- Wilson K. P., Shewchuck L. M., Brennan R. G., Otsuka A. J. & Matthews B. W. "Escherichia coli biotin holoenzyme synthetase/bio repressor crystal structure delineates the biotin- and DNA-binding domains" (1992) *Proc. Natl. Acad. Sci. USA* vol. 89 pp 9257-9261
- Wittekind M. & Mueller L. "HNCACB, a High-Sensitivity 3D NMR Experiment to Correlate Amide-Proton and Nitrogen Resonances with the Alpha- and Beta-Carbon Resonances in Proteins" (1993) *JOURNAL OF MAGNETIC RESONANCE SERIES B* vol. 101 pp 201-205
- Wong H. C., Mao J., Nguyen J. T., Srinivas S., Zhang W., Liu B., Li L., Wu D. & Zheng J. "Structural basis of the recognition of the dishevelled DEP domain in the Wnt signalling pathway" (2000) *Nat. Struct. Biol.* vol. 7 pp 1178-1184
- Yang Z. R., Thomson R., McMeil P. & Esnouf R. M. "RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins" (2005) *Bioinformatics* vol. 21 pp 3369-3376
- Zaromytidou A. I., Miralles F. & Treisman R. "MAL and ternary complex factor use different mechanisms to contact a common surface on the serum response factor DNA binding domain" (2006) *Mol. Cell. Biol.* vol. 26 pp 4134-4148
- Zhang X., Azhar G., Chai J., Sheridan P., Nagano K., Brown T., Yang J., Khrapk K., Borrás A. M., Lawitts J., Misra R. P. & Wei J. Y. "Cardiomyopathy in transgenic mice with cardiac-specific overexpression of serum response factor" (2001) *Am. J. Physiol. Heart Circ. Physiol.* vol. 280 pp 1782-1792

Zhang Z. R., Thomson R., McNeil P. & Esnouf R. M. "RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins" (2005) *Bioinformatics* vol. 21 pp 3369-3376

Zheng N., Fraenkel E., Pabo C.O. & Pavletich N. P. "Structural basis of DNA recognition by the heterodimeric cell cycle transcription factor E2F-DP" (1999) *Genes Dev.* vol. 13 pp 666-674

Zheng N., Schulman B. A., Song L., Miller J. J., Jeffrey P. D., Wang P., Chu C., Koepf D. M., Elledge S. J., Pagano M., Conaway R. C., Conaway J. W. Harper J. W. & Pavletich N. P., "Structure of the Cull1-Rbx1-Skp1-FboxSkp2 SCF ubiquitin ligase complex" (2002) *Nature* vol. 416 pp 703-709