INSIGHTS INTO THE STRUCTURES AND DYNAMICS OF THE PATHOGEN SECRETED EFFECTORS AVR3A11 AND TARP THROUGH THE APPLICATION OF NMR SPECTROSCOPY

James Tolchard

A thesis submitted to the School of Chemistry, University of East Anglia in the fulfilment of the requirement for the degree of Doctor of Philosophy.

December 2013

Declaration

I declare that the work contained in this thesis, submitted by me for the degree of Doctor of Philosophy, is my own original work, except where due reference has been made to other authors or co-workers and has not been previously submitted by me for a degree at this or any other university.

James Tolchard
December 2013

Abstract

The lifecycles of obligate pathogenic and parasitic microorganisms depend on a myriad of interactions with their hosts at the molecular level. The class of bacterial proteins directly responsible for these inter-organism interactions have been termed effector proteins and can function in either an extracellular or, once secreted into the host cell, an intracellular environment. Primarily through the use of nuclear magnetic resonance spectroscopy (NMR), I have investigated the biophysical properties of two such bacterial effector proteins.

TARP (translocated actin recruiting protein) is a largely disordered 100 kDa effector, common to all chlamydial species, which functions to remodel the host actin cytoskeleton to facilitate the internalisation of the chlamydial cell. Using constructs of TARP comprising an expected actin binding domain, I have shown through NMR chemical shift indexing and 15N relaxation that although the unbound domain is intrinsically disordered a short region, which aligns to other helical actin binding domains, maintains some helical propensity. Furthermore, these residues map to chemical shift variations in the bound state and the K_d for the interaction has also been determined using isothermal titration calorimetry.

AVR3a11 is an 8 kDa effector from the pepper pathogen Phytophthora capsici that has been shown to inhibit plant programmed cell death. Using a combination of 2D and 3D NMR experiments I have assigned the majority of the backbone and side-chain resonances from the structured regions of AVR3a11. Through the acquisition and analysis of ^{13}C and ^{15}N edited HSQC-NOESY spectra I have also calculated a water refined, structural ensemble for AVR3a11. Additionally, analysis of the slow (H:D exchange) and fast (T_1, T_2 and heteronuclear NOE) dynamic regimes, describes AVR3a11 as a relatively tightly folded helical bundle which also exhibits a significant degree of conformational exchange.

Contents

List of Figures

List of Tables

Abbreviations

Å	Angstroms (10^{-10} m)
ABP	Actin binding protein
ABD	Actin binding domain
Abs, A	Absorbance
ADP	Adenosine diphosphate
AIM	Autoinduction medium
APSY	Automated Projection Spectroscopy
ATP	Adenosine triphosphate
AVR	Avirulence resistance
BLAST	Basic Local Alignment Search Tool
BMRB	Biological Magnetic Resonance Bank
D_2O	Deuterium oxide
CCPN	Collaborative Computing Project for NMR
CSA	Chemical shift anisotropy
CSI	Chemical shift indexing
CV	Column volume
DNA	Deoxyribonucleic Acid
DLS	Dynamic light scattering
DSC	Differential scanning calorimetry
DSS	4, 4-dimethyl-4-silapentane-1-sulfonic acid
DTT	Dithioltheritol
ε	Extinction coefficient ($M^{-1}cm^{-1}$)
EB	Elementary body
EDTA	Ethylenediaminetetraacetic acid
EM	Electro-magnetic
EPR	Electron Paramagnetic Resonance
ESI-MS	Electrospray ionisation mass spectroscopy
ETI	Effector-triggered immunity
ETS	Effector-triggered susceptibility
FID	Free Induction Decay
FT	Fourier Transform
FTIR	Fourier transform infrared spectroscopy
GST	Glutathione-S-transferase
hetNOE	Heteronuclear Nuclear Overhauser enhancement
HR	Hypersensitive response

HSQC	Heteronuclear Single Quantum Coherence
IDP	Intrinsically Disordered Protein
IMAC	Immobilised-Metal Affinity Chromatography
INEPT	Insensitive Nuclei Enhanced by Polarisation Transfer
IPTG	Isopropyl thiogalactoside
ITC	Isothermal Titration Calorimetry
JIC	John Innes Centre
kD	Dissociation constant
kDa	Kilo (10^3) Daltons
LB	Lysogeny-Broth
LPS	Lipopolysaccharide
Mbp	Mega (10^6) base pairs
MEM	Minimal Essential Medium
MOPS	3-(N-morpholino)propanesulfonic acid
MRI	Magnetic Resonance Imaging
MWCO	Molecular Weight Cut-Off
NaCl	Sodium chloride
NaN3	Sodium Azide
NMR	Nuclear Magnetic Resonance
NOESY	Nuclear Overhauser Effect Spectroscopy
OD_{600nm}	Optical density (600nm)
PAGE	Polyacrylamide gel electrophoresis
PAMP	Pathogen-associated molecular patter
PBS	Phosphate buffered saline
PCD	Programmed Cell Death
PDB	Protein Data Bank
PEP	Preservation of equivalent pathways
PFG	Pulsed field gradient
PID	Pelvic inflammatory disease
ppb	Parts per billion
ppm	Parts per million
PSVS	Protein Structure Validation Suite
PTI	Pathogen-triggered immunity
PWP	Protein waiting for a partner
RB	Reticulate body
rcf	Relative Centrifugal Force
RF	Radio frequency

rpm	Revolutions per minute
ROS	Reactive Oxygen Species
RSDM	Reduced spectral density mapping
SAXS	Small angle X-ray scattering
SDS	Sodium Dodecyl Sulphate
SRCD	Synchrotron Radiation Circular Dichroism
STD	Sexually Transmitted Disease
TARP	Translocated actin recruiting protein
TEMED	N,N,N,N tetramethylethylenediamine
TMS	Tetramethylsilane
TOCSY	Total Correlation Spectroscopy
Tris	Tris - (hydroxymethyl)-aminomethane
TROSY	Transverse relaxation optimised spectroscopy
UEA	University of East Anglia
UV	Ultraviolet
VUV	Vacuum ultraviolet
WASP	Wiskott-Aldrich syndrome protein
WH2	Wiskott-Aldrich syndrome protein homology domain 2

Acknowledgements

Firstly, I would like to thank my supervisor Dr Tharin Blumenschein. Her invaluable experience, guidance, patience and optimism have had a comprehensive impact on my character and scientific approach and I am forever indebted to her excellent supervisory skills.Acknowledgements must also go to all of the past (Olli, Angelo, Krissy, Matt and Morgan) and present (Liz, Alex and Jess) members of the NMR and EPR groups for their advice, discussions on magnetic resonance, sanity checks, coffees, beers and failed pub quiz attempts. Your quest to beat The Bar will unfortunately have to go on without my knowledge of horrendous 90s music.

I would also like to give thanks to the technicians Nick Cull (for helping me in the wet lab when I first started) and Dr Colin MacDonald (for help me throughout my PhD with our temperamental spectrometers) as well as my secondary supervisors Professor Nick Le Brun and Professor Geoff Moore for their unparalleled knowledge and input.

Special thanks must also go to my fiance Ellie, for her continued love and support and for putting up with my endless repetition of nonsense when practicing for scientific talks and to her parents Mike, and the omnipresent Janice, for their support, friendship and donation of a place to live. However, final thanks must go to my parents Lynn and Chris, for without their unwavering financial and often tested emotional support, I would not be in the position I am in today.

And, Yes..... I can FINALLY start earning actual money now!

Chapter 1

Pathogen Secreted Effectors

1.1 Introduction

Since the emergence of single celled life, the act of molecular communication amongst, and even between, species has been a fundamentally important driving force behind the evolution of life on our planet. The evolutionary advantages that communication can bestow with respect to survival and genetic propagation are clear, whether defensive, symbiotic or predatory in nature. Indeed, the obvious worth of cellular interactions can be evidenced from the fact that cellular adhesion (1), and arguably endosymbiosis (2), which have shaped multicellular life as we know it, were relatively early evolutionary adaptions.

However, perhaps because these have been such lengthy and convoluted evolutionary paths, the mechanisms that underpin cellular interactions commonly involve multifaceted networks of molecular signals. Understanding the complexities of these interactions has therefore become a significant motivation for the majority of the biological sciences due to their far reaching implications regarding disease (3), immunity (4), medicine (5), and even wider environmental concerns such as bioremediation (6), conservation (7) and the synthesis of renewable fuels (8) and sustainable foodstuffs (9).

The incredible scope allowed to intercellular interactions through chemistry and biology has also been equally exploited by pathogenic species in order to hijack, damage and kill other species for their own benefit. To this effect, many microbial pathogens have evolved sets of proteins that are not immedi-

ately damaging in themselves, but which are able to interact with and modulate the cellular processes and signalling networks of their hosts. Termed "effectors", these proteins have evolved so much so that they usually share similar structural elements or binding sites to those of the native system. Therefore, even if the host interaction has previously been studied, a thorough understanding of how pathogens are able to subtly manipulate these mechanisms will most likely require a comprehensive interdisciplinary understanding of the underlying biological, chemical and physical phenomena supporting them.

There are many known examples of pathogenic species throughout the tree of life that directly employ effector proteins to their advantage. Interestingly, whilst great diversity exists between the mechanisms targeted by pathogens, there are also many examples of common molecular targets that are particularly vulnerable to subversion. Typically, common targets are molecules that are either overly abundant or overly networked, as these characteristics can tend to back an organism into an exploitable "evolutionary corner". Having total reliance on a particular form or arrangement of a molecular system can significantly reduce the probability of mutational events prescribing defensive advantages whilst simultaneously not disrupting any required characteristic.

One such molecular constituent, which several unrelated parasites have convergently evolved to manipulate, is that of eukaryotic ubiquitin(10). Ubiquitin is a highly conserved protein, found in all eukaryotes, which functions primarily as a molecular identifier for other proteins. As a tag, ubiquitin is perhaps most synonymous with protein degradation and recycling, but as a broader cellular signal it also has implications in DNA- repair, transcriptional regulation, apoptosis and the cell-cycle(11). By interacting with ubiquitin and fine-tuning some of these interactions, pathogens therefore have potential access to a significant range of core cellular functions.

For example, a secreted effector (OspI) from the pathogen *Shigella flexneri* has been shown to interfere with host ubiquitination to inhibit a pertinent host cell immune response. Once secreted into the host cell, OspI functions to modify a host E2 enzyme (UBC13) which is then unable to polyuniquitinate an E3 enzyme (TRAF6) required for the transduction of the NF-KB signalling cascade involved in the inflammation-related immune response(10). Furthermore, *Legionella pneumophila*, another intracellular pathogen, has been shown to utilise both host ubiquitin and the host proteasome to aid in the temporal

regulation of its own pathogenesis. Upon infection, *L. pneumophila* secretes a number of protein effectors into host cells, two of which are LubX and SidH.

Through homology to E3 ubiquitin ligases, LubX is able to polyubiquitinate SidH, which is then degraded by the host proteasome. Although the function of SidH itself has not yet been identified, this post-infection degradation of SidH has been shown to be required for optimal disease progression (12) relative to the initial time of infection.

Another common pathogenic target is that of the actin cytoskeleton, an equally abundant cellular system. By interacting with the actin cytoskeleton pathogens are able to exploit its structural properties to aid processes such as cellular adherence (13) and host cell internalisation (14). For instance, the highly pathogenic enteropathogenic (EPEC) and enterohemorraghic (EHEC) *E. coli* species secrete a small actin binding protein (Tir) into epithelial cells of the hosts small intestine. By traversing both sides of the host cell membrane, Tir is able to facilitate the generation of substantial cellular adherences by mediating interactions with both intimin molecules on the *E. coli* cell surfaces and a sturdy "pedestal" of host cell actin-associated proteins and actin filaments. Additionally, genera such as Listeria (AcTa protein(15), Figure 1.1.1), *Shigella* (LcsA protein (16)), *Rickettsia* (RickA protein (17)) and *Burkholderia* (BimA protein (18)) have also evolved ways to exploit the host cell actin related protein complex (ARP2/3) to conduct mass actin polymerisation at specific sites on their cell surface to generate propulsion through the host cytosol.

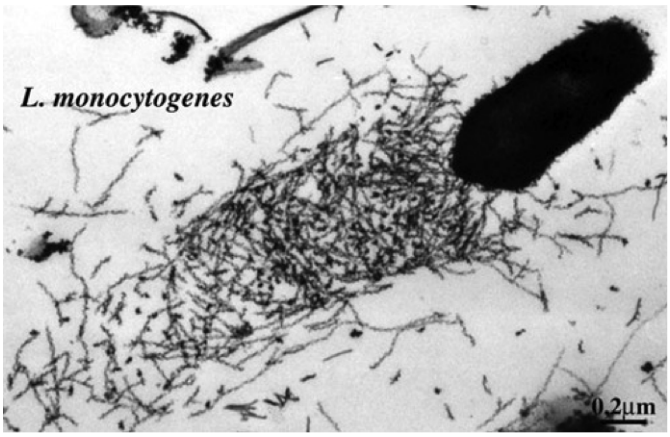

Figure 1.1.1: A transmission electron micrograph of *Listeria monocytogenes* utilising intracellular actin filaments to generate a motive force. Adapted from (17).

1.2 The *Phytophthora* genus and avirulence proteins

1.2.1 The *Phytophthora* genus

The pathogenic *Phytophthora* species belong to the class Oomycetes, a somewhat awkward taxonomic classification which describes a combination of both fungal (hyphal morphology, sporulation) and also protist (two distinct flagella, cellulose cell wall) characteristics (19). The *Phytophthora* genus is currently comprised of 116 known species(20), although theoretical predictions, and results obtained from sequencing the DNA of unculturable species, suggest that the final figure may actually fall somewhere between 200-600(21). Currently, all known species of *Phytophthora* are active pathogens and although host (and host tissue) promiscuity varies widely (22), many species have a particular host tropism towards the agriculturally important dicotyledonous plants. Examples of *Phytophthora* pathogen hosts include: carrot, chickpea, cotton, cucumber, pepper, potato, soybean, tobacco and tomato. Most recently, another species *Phytophthora ramorum* has been identified as responsible for the "sudden oak death" disease sweeping through the west coast of the United States and Europe (23).

Figure 1.2.1: An example of late blight tomato spoilage produced by *Phytophthora infestans*. Adapted from (24).

With such wide ranging and agriculturally important host plant species, many authorities (25, 26) and groups (27, 28) have labelled *Phytophthora* sp. as the main microbial threats to various agricultural industries and native species, with economic costs measured in billions of dollars (29, 30) from the resulting crop spoilage (Figure 1.2.1). Indeed, the devastating potential impact of *Phytophthora* can be recognised in history from the fact that the causative agent of late blight in potatoes, the principle cause of the infamous Irish potato famine of the 19^{th} century, was *Phytophthora Infestans*.

The *Phytophthora* lifecycle notably resembles that of typical fungi with distinct phases of sporulation, dissemination of spores and hyphal growth. Broadly speaking, the development cycle begins from either asexually or sexually fertilised diploid types of spore, termed either oospores or chlamydospores depending on the species and method of fertilisation. Oospores are hardy globular bodies that are able to weather nutrient deficient environments for long periods of time. Upon the onset of favourably hydrating conditions, these phytophthoral spores germinate into sporangia that are able to travel relatively great distances via wind or water. Sporangia contain a variable number of smaller, infectious particles termed zoospores that are released in favourable conditions. Zoospores are small, biflaggelate particles capable of chemotactically driven, short-range, movement towards host plants. Due to the differences in host specificity, multiple mechanisms are thought to exist across the *Phytophthora* genus for host recognition (22) but they generally not fully understood. One study has identified the metabolic intermediates daidzein and genistein as the chemoattractants facilitating, at least in part, the infection of soybean species by *P. sojae zoospores* (31).

Upon contact with the plant tissue, zoospores differentiate once more into adherent particles termed cysts that attach to the host plant surface (32). After encystation, a nascent germ tube emerges and presents a fixed platform from which hyphal extensions are able to grow and penetrate the plant exterior. Significant development of the hyphal network then allows extensive sequestration of the local plant cell nutrients through haustorial extensions from the hyphae.

The nutrients provided through the hyphal network then promote maturation and formation of haploid gametangia for either heterothallic (exterior to the plant) or homothallic (interior or exterior to the plant) oospore formation;

and the lifecycle continues.

1.2.2 *Phytophthora* pathogenesis and plant immunity

Perhaps the most well known and agriculturally important *Phytophthora* species is *P. infestans*, the causative agent of late blight in potatoes and tomatoes - as described in Section 1.2.1. In fact, *P. infestans* has been designated a model oomycete species, has had many experimental protocols optimised for laboratory manipulation (33) and is probably the most well documented *Phytophthora* species (34). Although pathogenesis can vary a great deal across the *Phytophthora* genus, *P. infestans* has been shown to share significant homology in mechanism, and levels and timings of gene expression, with fungal pathogens (35) and other *Phytophthora* species, including *P. capsici* (36).

The pathogenesis of *P. infestans* can be described as a two-stage process. After the initial interaction from the encysted particle, the biotrophic phase describes the asymptomatic growth of the hyphae and haustoria. Following as yet unknown determinants, *P. infestans* then switches to a necrotrophic phase whereby plant tissue necrosis is initiated, presumably to better facilitate nutrient absorption. The two pathological stages are clearly distinct and although advancements have been made, a thorough, descriptive model of pathogenesis does not currently exist (34). What is clear however for all of the known *Phytophthora* is that modulation of host cell processes, in particular the plant immune response, is vital for successful disease progression. The immune system of higher plants typically offers a multi-layered defence, ranging from basal, non-specific attributes (antimicrobial secretion, cell walls, external cuticle) to highly specific molecular interactions. However, without mobile, immunity-conferring agents, such as the vertebrate white blood cell, plants require all cells to be able to impart some level of immune response (37). Deemed part of the innate immunity, the pathogen-associated molecular pattern response (PAMP) is a type of plant immunity which functions through toll-like receptors to recognise microbial molecules or motifs such as bacterial flaggelin (38) and lipopolysaccharide (LPS) (39). Successful extracellular recognition of PAMPs then drives transmembrane signalling pathways and prompts immune responses, for example the generation of reactive oxygen species (ROS) (40) or programmed cell death (PCD) (41); a process termed PAMP-triggered immunity (PTI).

Similar to other plant pathogens, *Phytophthora* species evade this PTI response through the use of a suite of secreted effectors to maintain its asymptomatic biotrophic growth phase. For instance the cytosol-secreted *Phytophthora* avirulence effector AVR3b has been shown to decrease the production of ROS (42) (effector-triggered susceptibility, ETS). However, the plant:pathogen arms race has been a long fought one, and as such many plants have evolved a means to recognise these pathogenic effectors (effector triggered immunity, ETI) through the products of resistance R genes, to generate further immunity responses.

Perhaps due to the increased significance of intracellular, rather than extracellular threat recognition, the ETI response often results in the absolute action of the hypersensitivity response (HR), which gives rise to PCD (37). As a means of isolating and annexing the infected tissues, the HR-initiated PCD initially begins with a cellular influx of Ca^{2+} and H^{+} ions and an efflux of OH^{-} and K^{+} ions. This is then followed by increases in cellular ROS, leading to the breakdown of cellular membranes and the increased production of callose and lignin, which forms a defensive barrier alongside the adjacent cells.

The activity of many ubiquitin E3 ligases is heavily associated with successful activation of the HR response (43, 44), and although they are thought to facilitate downstream HR signalling, the pathways involved have not been detailed (45). A further complication of this escalating plant:pathogen conflict is that there are also many examples of secondary ETI effectors in pathogenic species able to reintroduce a secondary susceptibility by destabilising elements associated with the ETI recognition or signalling events. This evolutionary to and fro between plants and their pathogens can be described as a "zigzag" model (Figure 1.2.2) of cyclic coevolution; a relationship and mechanism well described by Flors gene for gene theory (46) between host and parasite.

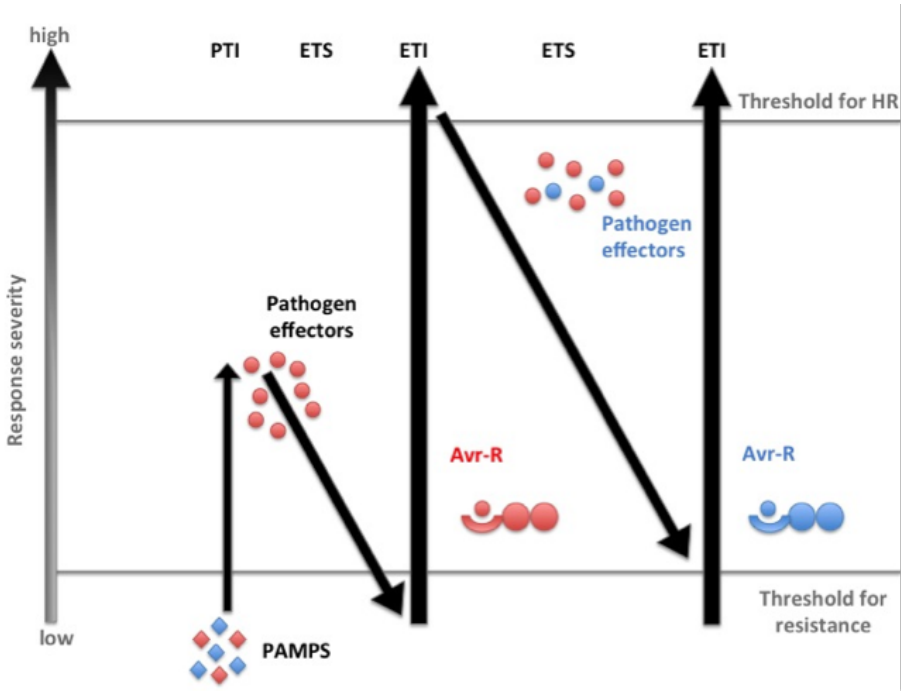

Figure 1.2.2: The zigzag model of plant:pathogen interaction. Initial PAMPs (red diamonds) are recognised by toll-like receptors giving rise to a low-amplitude PAMP-triggered immune (PTI) response. Pathogen effectors (red marks) subverting this process lead to effector-triggered susceptibility (ETS) unless recognised by resistance proteins (ovals) that can lead to effector-triggered immunity (ETI). Secondary pathogen effectors can inhibit this response until adaptations (blue marks) lead to additional effector recognition and secondary ETI. Adapted from (37).

1.2.3 RXLR effectors and the AVR3a11 protein

There is an increasingly large subgroup of oomycete effectors that are defined by a conserved Arg-X-Leu-Arg amino acid motif, where X can be any amino acid. Many RXLR motifs are also associated with a proximal downstream Asp-Glu-Glu-Arg (dEER; with a highly variable aspartic acid) motif. To date, hundreds of RXLR effectors have been directly identified in the main agriculturally important *Phytophthora* species such as *P. infestans*, *P. ramorum*, *P. sojae* and *P. capsici* (36, 47). These proteins, first recognised through genetic analysis of an oomycete pathogen of *Arabidopsis* spp. (48), can be described as modular, in that they possess an N-terminal localisation sequence (for extracellular secretion through the eukaryotic type-II secretion system) followed by the RXLR (+dEER) motif and then an effector domain of varying length. A broad range in function has been observed for RXLR effectors, however the

conserved RXLR motif has been shown to convey self- delivery into a wide variety of eukaryotic cells following secretion into the intercellular space; a hypothesis which was based on the function of homologous regions in effector proteins of the malaria-transmitting parasite *Plasmodium falciparum*. AVR3a (**Avir**ulence protein 3a) is one such RXLR effector protein identified in *P. infestans* (with homologues in *P. capsici* and *P. sojae* (49)) which has been identified as a secondary ETS protein (41, 42), that binds, stabilises and inhibits the action of the ubiquitin E3 ligase CMPG1, preventing signalling for PCD. Interestingly, two main natural alleles of AVR3a exist, termed $AVR3a^{KI}$ and $AVR3a^{EM}$ (describing Lys or Glu and Ile or Met at sequence positions 80 and 103, respectively). Both alleles confer virulence in plants missing the corresponding R3a resistance gene (i.e. resistant to the 3a AVR protein), however only $AVR3a^{KI}$ is recognised by R3a carrying plants. There are currently no studies that describe the direct interactions of either CMPG1 or R3a bound AVR3a and so the significance of these allelic forms is not explicitly clear. An X-ray crystal structure of an AVR3a homologue, AVR3a11 from *P. capsici*, has been published by our collaborators (Dr Mark Banfield, John Innes Centre, Norwich) (Figure 1.2.3), and defined a WY domain characterised by a conserved Trp and Tyr residues as part of the hydrophobic core. Through *in silico* comparisons it was shown that many (~44% (49)) of RXLR effectors share similar structural motifs to the AVR3a11 WY domain, which designates the core α-helical fold. The significance of this WY domain has been related to pronounced evolutionary plasticity, as significant mutations can be incurred without disrupting the overall fold. This would make RXLR effectors particularly suited to coevolution with host plants and agrees well with observations that RXLR genes also occupy particularly dynamic regions of the *Phytophthora* genome.

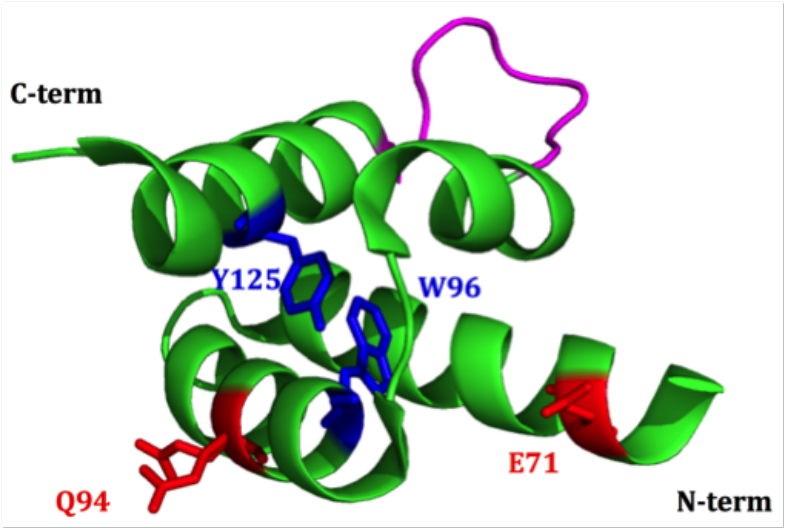

Figure 1.2.3: The four-helix bundle structure of the AVR3a11 effector domain from *P. capsici* as determined by X-ray crystallography (PDB code: 3ZR8 (49)). The sites of the allelic variations (E71/Q94 for AVR3a11, red), residues of the WY fold motif (blue) and the loop region that conveys variable resistance-protein recognition (magenta) are annotated. Figure created with the PyMOL molecular graphics program (50).

1.3 The *Chlamydia* genus and the TARP protein

1.3.1 *Chlamydia* and the chlamydial lifecycle

Chlamydiae are an example of actin-hijacking pathogens. The *Chlamydia* genus comprises three prokaryotic species, all of which are known intracellular parasitic pathogens and one of which, *Chlamydia trachomatis*, is known to preferentially infect *Homo sapiens*. The *Chlamydia* genus originally contained several more species but in 1999 a long-argued change was made to its taxonomic classification, based upon evidence supporting their genetic divergence (51). This placed six other pathogenic species into a new genus, the *Chlamydophila*, although partial conservation of their actual parasitic mechanism is evident (52). Both genera are still classified under the same bacterial phylum due to their ancestral genetics and unique biphasic lifecycle in which intracellularisation within a host cell is absolutely required for growth and replication. Interestingly, *Chlamydiae* have extremely small genomes of around 1.2 Mbp

and it is hypothesised that this occurred as an ancestral adaptation to intracellular habitation. Some degree of reliance is therefore expected upon the host cell proteinaceous machinery for a comprehensive metabolism. Further hypotheses have even suggested that *Chlamydiae* are distantly related to an intracellular parasite which ancestrally colonized early plant life (53) and evolved into the symbiotic chloroplast.

In an extracellular environment, *Chlamydiae* exist as elementary bodies (EB) which are metabolically inert yet infectious particles, 0.3 μm in diameter, with cell walls containing little peptidoglycan(54) (Figure 1.3.1). An abundance of disulphide bridging proteins have been described in their outer membrane and these could potentially be responsible for their extremely robust nature and particular resilience to ultraviolet radiation (55). Once chlamydial EBs have entered into a host organism, electrostatic interactions, mediated by negatively charged glycosaminoglycan moieties, facilitate nonspecific attachment to the membranes of a wide variety of potential host cell types (56). Although it has not been determined if these moieties originate from the host or parasitic cell, they are thought to be heparin and heparan sulphate due to the elimination of host cell attachment after heparinase treatment (57).

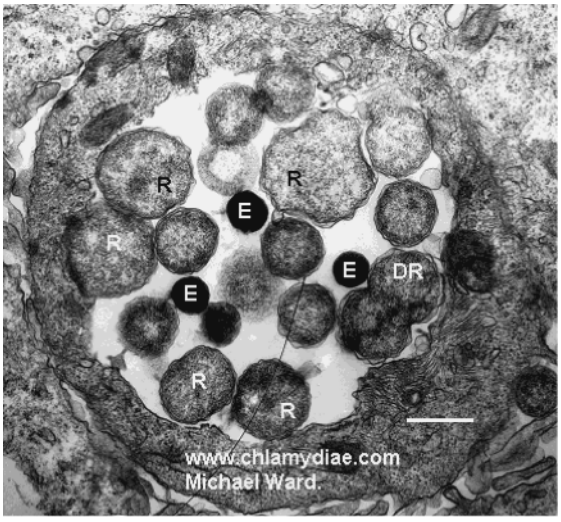

Figure 1.3.1: An electron micrograph of a *C. trachomatis* parasitic inclusion. Black elementary bodies (E), grey reticulate bodies (R) and a dividing reticulate body (DR) can be seen. The bar represents 1 μm (58).

When bound, *Chlamydiae* then translocate across the host cell membrane by a mechanism that has yet to be fully elucidated. The greatest experimental evidence suggests that clathrin-mediated endocytosis is partly required, however further evidence from the same study seems to imply that a mechanistic redundancy exists in certain species (59) which further complicates the issue.

Nevertheless, it is known that the ability to hijack the host cell actin cytoskeletal network is absolutely required for cell invasion. After initial cell contact, a 1005 residue protein termed translocated actin recruiting protein (TARP) is secreted into the host cell via a type III secretion system (T3SS) and accumulates on the host-cell side of the T3SS injection site. TARP is then able to bind actin monomers and cooperate with indigenous actin remodelling proteins such as Rac GTPase and the ARP2/3 complex (60) to generate significant actin filament polymerisation towards the site of infection.

The actin filament remodelling induced by these complexes then produces lamellipodia-like extensions (termed "pedestals") from the host cell that engulf the chlamydial EBs in a process called parasite-specified phagocytosis due the aesthetic similarity to the phagocytic immune response (Figure 1.3.2).

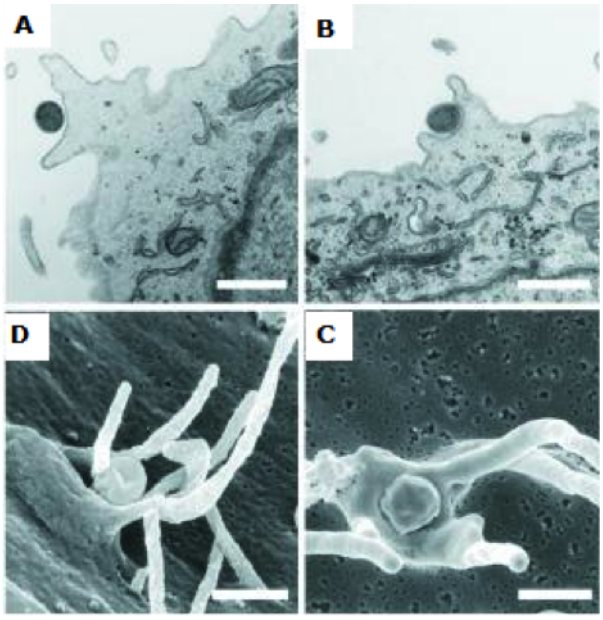

Figure 1.3.2: Lamellipodia-like extensions of the host cell membrane at different stages as viewed through a scanning electron microscope (SEM) (A,B,D) and a transmission electron microscope (C). Bars represent 0.5 μm (61).

Upon cell internalisation, the EB rapidly modifies its endocytotic vesicle with chlamydial proteins allowing it to both intercept Golgi-derived exocytotic vesicles and evade lysosomal fusion (62). It is also known that the parasitic vacuole can import host adenosine triphosphate (ATP) as a primary chemical energy source (63). During this time the EB cellular morphology also undergoes a thorough change resulting in a larger (~1 μm diameter), metabolically active cell that is able to undergo multiple rounds of binary fission (Figure 1.3.1). These are termed reticulate bodies (RB) and the chlamydial inclusion is the only known environment in which they will differentiate and replicate.

In response to an as yet unknown stimulus, the increased population of RB chlamydial cells will eventually begin to differentiate once again back to infectious EB particles. Exit from the host cell can then occur via two mechanisms, which for *C. trachomatis* occur equally under laboratory conditions (64). The "lysis" mechanism involves the rupture of the chlamydial inclusion and is swiftly followed by the lysing of the host nucleus and other cellular compartments. This is eventually followed by a break down of the host cell membrane resulting in host cell death and the release of the EBs into the extracellular environment.

The alternative release mechanism on the other hand does not result in the death of the host cell and the whole inclusion is effectively exocytosed from the host cell. Upon externalisation, the inclusion compartmentalises and individual EBs bud off from the main inclusion body.

1.3.2 Chlamydiaceae family pathogenesis

The greatest threat to humans from a Chlamydiaceae pathogen is the chlamydial species *Chlamydia trachomatis*, of which there are 15 known strains, or serovars. Serovars A, B, Ba and C probably signify the greatest direct threat to individuals in the developing regions of Asia and Africa where they are the causative agents of trachoma; a painful eye infection which when left untreated induces permanent blindness. It is readily spread by direct contact with the eye, helped by the ability of the chlamydial EBs to remain infectious on inanimate objects such as towels for long periods of time. Eye seeking flies also seem to be an important transmission vector for *C. trachomatis* and in some areas fly transmission has been shown to account for 75% of new cases (65). Early stages of *C. trachomatis* infection are easily treated with a single dose of

the antibiotic azithromycin, however due to the ease of transmission within impoverished areas, community based mass antibiotic treatment is recommended for thorough eradication (66) and this is rarely deemed monetarily feasible.

In comparison to the trachoma inducing serovars, serovars D through K of *C. trachomatis* preferentially inhabit the human reproductive tract and can cause a range of conditions in both men and women. Transmitted via unprotected sex, *C. trachomatis* is currently the most commonly diagnosed sexually transmitted disease (STD) in the western world (67) and, in the United States alone, 1.2 million people are estimated to be infected, with an annual increase of ~9% observed over the last 13 years (68). Males suffering with these serovars are usually asymptomatic, but can experience painful inflammations of the infected tissues with the conditions defined as urethritis, prostatitis and epididymitis. Females can also expect similar inflammations of the reproductive tissues such as cervicitis, cystitis, urethritis and pelvic inflammatory disease (PID). Both sexes can also experience proctitis though probably the most damaging presentation is none at all, as all *C. trachomatis* strains are capable of asymptomatic presentation. If a silent infection is untreated and spreads to the epididymis (males) or upper reproductive tract (females) the main symptom then becomes partial or irreversible infertility (69, 70). The causative pathology of infertility is associated with either a blocking of the fallopian tubes or the epididymis and is caused by tissue scarring and the accumulation of mucus and cell debris.

The final group of *C. trachomatis* serovars are those of the Lymphogranuloma venereum strains LGV1, LGV2 and LGV3. These are rarely found in patients from the developed world, although it is present, and even in the developing regions of Africa and Asia it is still relatively uncommon. Infection with any one of the LGV strains initially presents as genital ulcers or warts, something they are frequently mistaken due their low prevalence (71). LGV strains seem to particularly favour lymphoid cells and several weeks post infection LGV cells spread throughout the lymph system, causing swelling in potentially any or all of the lymph glands. If treatment is sought, the swollen lymph glands are usually drained to ease discomfort whilst antibiotics are also administered. If untreated, symptoms can eventually mimic those of Crohns disease with a wasting of systemic epithelium and lead to serious malnutrition and possibly even death.

The Chlamydiaceae family also contains the human pathogens *C. pneumonia* and *C. psittaci*. *C. pneumonia* is primarily a pathogen of the respiratory tract and although it has been documented as colonising other tissues, these have resulted in no deleterious symptoms. From within the respiratory tract however *C. pneumonia* can lead to various severities of bronchitis, pneumonia, chronic obstructive pulmonary disease and evidence even exists suggesting a possible role in the development of multiple sclerosis (72). *C. psittaci* is primarily an avian pathogen which infects all major systems and ultimately, without treatment, leads to death. There are a large number of cases however where serovars A and C-E have been known to zoonotically cross to humans, and present as a wide range of conditions; from flu-like symptoms to encephalitis and death.

1.3.3 The chlamydial TARP protein

Very little is currently understood about the host cell invasion of chlamydial species. It is known, however, that the secretion of a largely intrinsically disordered chlamydial protein, TARP, into the host cell is absolutely required. Orthologs of TARP exist in all *Chlamydia* and *Chlamydophila* species, albeit to a varying degrees of conservation. In *C. trachomatis* serovar LGV2, TARP is a 1005 residue protein able to recruit and nucleate filamentous actin that comprises 3 known functional domains: an N- terminal tyrosine rich repeat region, a proline rich region and a C-terminal actin binding domain (ABD) (Figure 1.3.3).

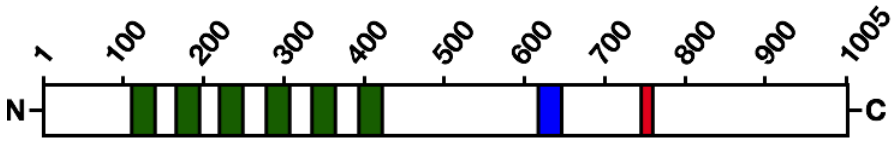

Figure 1.3.3: A scheme of the *Chlamydia trachomatis* LGV2 serovar TARP protein with tyrosine repeat regions (green boxes), proline rich region (blue box) and actin binding domain (ABD) red box). Adapted from (73).

The basic unit of the tyrosine rich repeat contains 50 amino acids and is present six times in the in *C. trachomatis* LGV2 serovar. The number of repeats varies in different species but interestingly this repeat region is missing completely in *Chlamydia muridarum*, *Chlamydophila caviae* and *Chlamydophila pneumoni* to no apparent detriment with respect to infection (74).

TARP orthologs from species containing the tyrosine repeat domain undergo phosphorylation by a host Src-family tyrosine kinase soon after secretion into the host cell (75). However, both the proline-rich domain and actin-binding domain are well conserved across all chlamydial species and have been shown to be absolutely required for the actin nucleation of TARP and subsequent cell invasion. The proline rich domain enables TARP to form homo-multimers to facilitate the process of actin filament nucleation at the site of chlamydial invasion. Because each *C. trachomatis* LGV2 TARP protein comprises only one actin binding domain it is hypothesised that the act of oligomerisation enables TARP to form an actin nucleating complex by colocalising multiple actin binding domains, each capable of binding an actin monomer.

This is further supported by evidence suggesting that TARP orthologs comprising more than one ABD have increased kinetics with respect to the actin nucleating reaction and in fact do not require oligomerisation to promote actin nucleation (76). The proposed actin binding helix is sited between A748 and K758 and shows striking sequence homology to a known actin binding helix in the Wiskott-Aldrich syndrome protein (WASP) homology domain-2 (WH2). Proteins containing the 35 residue WH2 globular actin binding domain are usually highly involved in the regulation of cytoskeletal growth and occur across a wide range of eukarya with the exception of plant species (77). The TARP ABD is therefore proposed to interact with actin in a similar fashion to other WH2 containing proteins.

Complete TARP has been shown to nucleate actin filaments *in vivo*, supporting its natural function; however it has also been shown that TARP proteins from *C. trachomatis* are able to interact with host actin regulating proteins and further contribute to actin polymerisation by subverting natural signalling pathways. Guanine exchange factors (GEFs) are well characterised proteinacious switches which are activated by the binding of guanosine triphosphate (GTP). Once active, GEFs are able to transfer the terminal phosphate group of GTP to their binding partners and lead to their activation. After passing on their terminal phosphate group, GEFs are rendered in an inactive state with bound guanosine diphosphate (GDP) but are free to reacquire GTP. TARP is able to interact with the well known GEFs Sos1 and Vav2 which are normally utilised during the signalling cascade responsible for the remodelling of the actin cytoskeleton with the eventual activation of cortactin and the Arp2/3 complex (78, 79). The ability of TARP to use the intracellular actin

remodelling system for its own means seems to be highly linked to chlamydial host cell invasion. It is interesting however that this functionality requires the phosphorylation of the tyrosine residues within the repeats of the tyrosine rich domains and so cannot therefore be utilised in the invasion of chlamydial strains lacking these domains. Because natural and artificial TARP proteins lacking this N-terminal tyrosine rich domain are still capable of invading host cells, it must therefore be hypothesised that mechanistic redundancy exists, for this case, potentially with its own actin nucleation function upon multimerisation.

It seems to be the case that chlamydial species present multiple examples of mechanistic redundancy in their methods of cell invasion (59), host kinase utilisation (60, 80), GEF activation (78), ATP acquisition (63) and host cell exit (64). This appears to suggest that *Chlamydia* sp., due to their absolute dependence upon an intracellular niche for replication, have largely evolved to support their main selection pressure: the invasion of host cells. It is therefore my opinion that TARP is a further example of this and an excellent example of a protein comprising evolutionary redundancy.

1.4 Aims of the thesis

It is the aim of this work to study both the structural and dynamic characteristics of the chlamydial protein TARP and phytophthoral protein AVR3a11, mainly through the application of solution state nuclear magnetic resonance (NMR) spectroscopy. A structure for the AVR3a11 effector could offer important insights into its function and work to this end will be detailed in Chapter 3. In Chapter 4 I will also document my efforts in describing the dynamic behaviour of AVR3a11. Currently very little is known regarding the TARP:actin interaction. In Chapter 5 I will discuss studies undertaken regarding the native state of TARP actin binding domain, both through *in silico* and experimental means and in Chapter 6 I will discuss work which attempts to describe its interaction with actin and present the construction of a theoretical model of the TARP:actin interaction. It can be supposed that at some point in the history of phytophthoral and chlamydial evolution, similar evolutionary advantages directed significant investment towards interspecies extra- and intra- cellular interactions.

The two genii however originate from different branches of life, which undermines a broad comparison between the biology of the two species. Nevertheless, the significance of studying these systems individually is obvious as a result of their continued damage to agricultural industries and human health throughout the world. It can also be said that a contrast between their respective mechanisms should prove an interesting observation with respect to the different biochemical approaches taken in subverting intracellular host processes.

Chapter 2

Materials and Methods

2.1 Experimental methods for AVR samples

2.1.1 AVR constructs

Constructs for both AVR3a11 and AVR3a11ΔN were provided with N-terminal hexa-His-tags, inserted into pOPIN-F plasmids (Novagen, USA), by Dr M. J. Banfield (John Innes Centre (JIC), Norwich, UK). The full-length AVR3a11 construct consisted of the 72-residue effector domain from the *Phytophthora capsici* protein AVR3a11. The "AVR3a11ΔN construct" bore a truncated version of the AVR3a11 effector domain, wherein the seven most N-terminal residues had been removed.

2.1.2 Overexpression of AVR3a11 and AVR3a11ΔN

2.1.2.1 Bacterial transformation

Cell transformations, based on a standard transformation protocol (81), were carried out by adding 1 μl pOPIN-F plasmid DNA (containing the construct for either AVR3a11 or AVR3a11ΔN) to 50 μl of competent Escherichia coli cells (strain BL21DE3) (see Section 2.3.1). Cell/plasmid mixtures were incubated on ice for 30 min before a 45 sec heat shock in a 42 °C water bath and were then returned to ice for a further 5 min. Reactions were then sup-

plemented with 0.5 ml Lysogeny Broth (LB) media (10 $g.l^{-1}$ tryptone, 5 $g.l^{-1}$ yeast extract, 10 $g.l^{-1}$ NaCl) preheated to 37 °C, and were incubated for 1 hr at 37 °C whilst shaking (200 rpm). Under sterile conditions, LB-agar-ampicillin plates (15 $g.l^{-1}$ agar, 100 $\mu g.ml^{-1}$ ampicillin) were then inoculated with 250 µl of the transformation reaction and left overnight in a 37 °C incubator.

2.1.2.2 Expression of isotopically enriched (^{13}C, ^{15}N) AVR3a11

All AVR3a11 samples were overexpressed and purified by Dr L. Boutemy (JIC, Norwich) as previously described (49). *E. coli* colonies from successful transformations were picked and used to inoculate 10 ml aliquots of LB-ampicillin media (100 $\mu g.ml^{-1}$ ampicillin), which were then incubated overnight at 37 °C with shaking (200 rpm). The following day, 8 ml of each saturated culture was used to inoculate a 100 ml aliquot of unlabelled minimal essential medium (MEM) (100 ml M9 salts, 0.4% glucose, 2 mM $MgSO_4$, 10 µM $CaCl_2$, 10 µM fresh $FeSO_4$, 1 x micronutrient solution, 1 x MEM vitamin mix (Sigma Aldrich, USA), 100 $\mu g.l^{-1}$ ampicillin. A 1000 x stock solution of micronutrient solution contained 3 µM $(NH_4)_2MoO_4$, 4 µM H_3BO_3, 30 µM $CaCl_2$, 10 µM $CuSO_4$, 80 µM $MnCl_2$ and 10 µM $ZnSO_4$. The stock "M9 salts" solution contained 42 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.5 mM NaCl, 18.7 mM NH_4Cl and was made to pH 7.8. Starter cultures were grown until saturated (~2.5 hr) and then harvested by centrifugation at 4,000 x g for 10 min, 20 °C. The resulting supernatant was discarded and each cell pellet was resuspended in 1 l of labelled MEM; wherein the NH_4Cl and glucose of the standard MEM recipe were replaced with equal concentration of $^{15}NH_4Cl$ and 0.2% ^{13}C glucose, respectively. Cultures were incubated at 37 °C with shaking (200 rpm) until an $O.D_{600}$ of 0.6-0.8 had been achieved. Overexpression was then induced with the addition of 1 mM IPTG to each 1 l culture and then cultures were incubated at 30 °C overnight with shaking (200 rpm). The following day, cells were harvested by centrifugation at 4,000 x g for 15 min, the resulting supernatant was removed and the cell pellets were stored at −20 °C.

2.1.2.3 Expression of isotopically enriched (^{15}N) AVR3a11ΔN

E. coli colonies from successful transformations were picked and used to inoculate 12 ml aliquots of LB-carbenicillin media (100 μg.ml^{-1} carbenicillin). Cultures were then grown at 37 °C with shaking (200 rpm) to an O.D$_{600}$ of ~1.5. Cells were harvested by centrifugation at 4,000 x g for 10 min, 20 °C. For each 12 ml starter culture, the resulting cell pellet was resuspended in 50 ml of N-5052 autoinduction media (82) (AIM) (50 mM Na_2HPO_4, 50 mM KH_2PO_4, 50 mM $^{15}NH_4Cl$, 2 mM $MgSO_4$, 10 μM fresh $FeSO_4$, 0.5% glycerol, 0.05% glucose, 0.2% lactose, 0.2 x trace metals). A 1000 x stock solution of trace metals contained 10 mM $MnCl_2$, 10 mM $ZnSO_4$, 2 mM H_3BO_3, 2 mM $CoCl_2$, 2 mM Na_2MoO_4, 2 mM $CuCl_2$, 2 mM $NiCl_2$, 2 mM Na_2SeO_4, 50 mM $FeCl_3$ and 20 mM $CaCl_2$.

Following inoculation, AIM cultures were incubated at 30 °C with shaking (200 rpm) and left overnight (~14 hr). Individual cultures were harvested by centrifugation at 4,000 x g for 15 min, 4 °C. The resulting supernatant was then discarded and cell pellets were stored at −20 °C. Overexpression was then checked using InstantBlue™ (Expedeon, UK) stained 15% SDS-PAGE gels (Section 2.3.2).

2.1.3 AVR3a11 and AVR3a11ΔN purification

Harvested cell pellets were thawed on ice and resuspended in 30 ml IMAC binding buffer (50 mM Tris-HCl pH 8.0, 0.5 M NaCl, 50 mM glycine, 5% glycerol) with the addition of an EDTA-free protease inhibitor (Roche, Germany). The resuspended cells were then lysed with three repeated sonication steps. Steps consisted of a 5 min incubation on ice followed by an 8 min 20 sec cycle with 30 W pulse at 200 ms.s^{-1} (Status 200, Phillip Harris Scientific, UK). Cell debris was then removed by centrifugation at 30,000 x g for 30 min (4 °C). The resulting cell lysate supernatant was then applied to a 10 cm disposable plastic chromatography column containing Ni^{2+} affinity resin (Prochem, USA) with a column volume (CV) of 5 ml that had been equilibrated with 4CV IMAC binding buffer. The flow-through was immediately collected on ice, followed by column washes of 2CV IMAC binding buffer, 3CV IMAC buffer (10 mM imidazole), 3CV IMAC buffer (50 mM imidazole) and 3CV IMAC elution buffer

(300 mM imidazole). The presence of AVR proteins was confirmed on silver stained (Section 2.3.3) 20% SDS-PAGE.

Elution fractions containing protein were concentrated to ~20 ml using a spin concentrator (Vivaspin 3 kDa molecular weight cut off (MWCO), Sartorius, Germany) and dialysed using 3.5 kDa MWCO tubing (Biodesign, USA) over 36 hr against two 5 l batches of PreScission^{TM} cleavage buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 1 mM DTT, pH 7.5). After 36 hr, the concentration of fusion protein was determined using UV-Vis spectroscopy (ε_{280nm}= 15930 $M^{-1}.cm^{-1}$ (His-AVR3a11 and His-AVR3a11ΔN) as calculated by the "Protein Calculator V3.3" web tool accessible at `www.scripps.edu/~cdputnam/protcalc.html`). To remove the hexa-His purification tag, PreScission^{TM} Protease (GE Healthcare, USA) was then added ($2U.mg^{-1}$ protein) and the digestion was incubated at 4 °C overnight with slight agitation.

Following confirmation of digestion using silver stained 20% SDS-PAGE gels, the sample was dialysed as before for 36 hr against two 5 l batches of IMAC binding buffer. The sample was then removed from dialysis and applied to a Ni^{2+} column, equilibrated as before. The flow-through, washes of 2CV IMAC binding buffer and 2CV IMAC buffer (300 mM imidazole) were then collected on ice. The purity of the final AVR sample was then confirmed on SDS-PAGE as before. AVR samples were then dialysed for 36 hr against two 5 l batches of AVR NMR sample buffer (0.12 mM NaH_2PO_4, 9.8 mM Na_2HPO_4, 50 mM Na_2SO_4) either at pH 8.8 (AVR3a11) or pH 6.8 (AVR3a11ΔN). Samples were then concentrated to 1 mM (Vivaspin 3 kDa MWCO) for NMR experimentation.

2.1.4 AVR3a11 NMR sample preparation

All NMR samples were made to a total volume of 500 µl and contained a final concentration of approximately 1 mM protein, 10% D_2O, 0.03% NaN_3 and 200 µM 4,4-dimethyl-4-silapentane-1-sulfonic-acid (DSS). Samples were spin filtered prior to use (0.22 µm, Corning Inc., USA) and were transferred to 5 mm NMR tubes (either: pp535, Wilmad, USA, or S-5-600-7, Norell, USA).

2.1.5 NMR spectroscopy of AVR3a11

All spectra relating to the full length AVR3a11 construct were acquired by Dr T. Blumenschein (UEA) and Dr L. Boutemy (JIC) on a Bruker Avance III 800 MHz spectrometer equipped with a triple resonance indirect detect TXI probe with Z-gradients at 298 K. Spectra acquired for backbone and side-chain assignment were: [^{1}H-^{13}C]-HSQC, [^{1}H-^{15}N]-HSQC, CBCA(CO)NH, CBCANH, CC(CO)NH, ^{1}H-^{1}H-^{15}N TOCSY, H(CCO)NH TOCSY, aromatic ^{13}C TROSY HSQC and an aromatic ^{13}C TOCSY (hnCBcgcdceHE). To obtain through-space distance measurements for the AVR3a11 structure calculation, ^{13}C NOESY- HSQC (120 ms mixing time) and ^{15}N NOESY-HSQC (100 ms mixing time) spectra were also acquired. All spectral parameters can be seen in Table 2.1.1.

Table 2.1.1: NMR acquisition parameters for experiments acquired at 800 MHz obtained for the assignment and structure calculation of AVR3a11. Experiments filtered with respect to aromatic moieties are denoted "(Ar.)". Dimensions different to those in the table header are labelled in superscript. All experiments had 1024 Complex points in the the F_1 ^{1}H dimension.

	Complex points			Spectral width (Hz)		
Experiment	Scans	^{15}N	^{13}C	^{1}H	^{15}N	13C
^{1}H-^{15}N HSQC	16	48	-	12019	2433	-
^{1}H-^{13}C HSQC	12	256	-	10416	-	33333
CBCA(CO)NH	32	32	60	12019	2433	15095
CBCANH	32	32	48	12019	2433	15095
CC(CO)NH	64	32	64	12019	2433	15091
^{15}N-TOCSY	32	32	$48^{^{1}H}$	12019	2433	$9603^{^{1}H}$
H(CCCO)NH	32	31	$64^{^{1}H}$	12019	2433	$12019^{^{1}H}$
^{13}C-TOCSY (Ar.)	1024	-	24	11161	-	3823
^{13}C-TROSY (Ar.)	32	-	128	11161	-	8081
^{15}N NOESY	16	32	$82^{^{1}H}$	12019	2433	$10416^{^{1}H}$
^{13}C NOESY	16	$96^{^{1}H}$	32	12019	$9603^{^{1}H}$	15092

2.1.6 Spectral processing

All spectra were processed with zero-filling and linear prediction (except directly detected dimension), Fourier transformed and phased using the NMR-Pipe software package (83). The DSS standard within each sample was used to reference the respective proton chemical shifts; all indirect dimensions were

referenced according to the ratio of their heteronuclear gyro-magnetic ratios and specific nuclear observation frequencies as described by Wishart et al. (84). IUPAC standard values (85) were taken for the gyro-magnetic ratios.

2.1.6.1 Manual backbone and side-chain assignment

All spectra were analysed in the CCPN Analysis software package (86). Backbone and side-chain assignments were initially characterised in terms of nondescript spin systems through correlations to amide proton resonances within the ^{15}N HSQC spectrum. A comprehensive comparison of secondary resonances to BMRB (87) standard values for residues in all structural configurations then enabled residue-type assignment.

The sequence specific assignment was achieved with the standard "backbone walk" technique (outlined in Chapter 3) by following the *i* to *i-1* connectivities for the Cα and Cβ atoms observable in the CBCA(CO)NH, CBCANH and CC(CO)NH spectra. Assignments were then verified by identifying *i* to *i-1* connectivities for the sidechain protons in the ^{15}N TOCSY and H(CCO)NH TOCSY spectra.

2.1.7 Structure determination

2.1.7.1 Restraint optimisation

The ATNOS/CANDID(88, 89) NOE peak picking and assignment algorithms within the UNIO (90, 91) software package were used to identify and assign NOE crosspeaks from both the ^{13}C and ^{15}N NOESY-HSQC spectra. Peak picking tolerances of 0.03 ppm (1H) and 0.4 ppm (^{13}C, ^{15}N) were used and based on the baseplane signal-to-noise ratio determined for the individual NOESY spectra, NOE crosspeak volumes were scaled and converted into upper limit distance restraints according to equation 2.1 (89):

$$b = \left(\sum_{k=1}^{n} \frac{I^p V_\kappa}{\sqrt{Q^S_{\alpha\kappa} Q^S_{\beta\kappa}}} \right)^{-(1/6)} \tag{2.1}$$

where b is the upper distance bound, I is the volume of the assigned NOE crosspeak p, Vk is a relative contribution score and Q is a calibration constant for the proton species (i.e. backbone ^{1}H or side-chain $^{1}H_2$ or $^{1}H_3$ methyl protons) of the α and β nuclei giving rise to the NOE crosspeak in spectrum S. If the assignment was found to be ambiguous with respect to n possibilities, the upper distance bound was taken as a superposition across the potential assignments until the possibilities were discarded according to incompatibilities with the initial covalent structure (UNIO, cycle 1) or the generated ensemble of the previous structural iteration (UNIO, cycle2 - cycle7). Input files consisted of an XEASY (92) formatted list of the AVR3a11 backbone and side-chain assignments (created with the CCPN format converter (86)) and CARA(93) formatted ^{15}N and ^{13}C NOESY-HSQC spectra.

In the same computational run as the ATNOS/CANDID automated NOESY crosspeak assignment, UNIO, in conjunction with the CYANA 2.1 (94) torsion angle dynamics algorithm, was used to calculate the initial restraint based models.

Input files consisted of a CYANA formatted amino acid sequence for AVR3a11, a list of φ and ψ backbone dihedral angles as predicted from the chemical shifts by the TALOS+ (95) web server, a list of all AVR3a11 NMR peak assignments and CARA formatted ^{13}C and ^{15}N NOESY-HSQC spectra. UNIO ran through a series of seven iterative cycles of simulated annealing governed by restraint optimisation and energetic minimisation of the resulting structures (Table 2.1.2). One hundred initial structures were calculated for each cycle and the top twenty, scored according to their CYANA target function (96, 97), were used to progress to the next iterative cycle along with the respective restraints list.

Table 2.1.2: The computational procedures undertaken with UNIO and CYANA 2.1 during the iterative restraint optimisation and initial structure calculation of AVR3a11.

Cycle(s)	Step #	Procedure	Parameters
1	1	Network anchoring	
2-7	1	Initial minimisation	
	2	High Temperature Dynamics	10 kK, 2 fs
	3	Slow cooling	
	4	Addition of protons	
	5	Final minimisation	8000 steps

2.1.7.2 Final calculation and refinement in explicit solvent

To generate a refined structural ensemble, CNS 1.3 (98) was used, with the EBI RECOORD scripts (99) and the fold-optimised restraints from cycle six of the UNIO/CYANA calculation (to maintain restraint ambiguity). The benefit to CYANA torsion angle sampling is that simpler potential energy terms and fewer degrees of freedom (and hence the extra benefit of being computationally faster) are better to optimise the starting restraints to best evade unfavourable local minima. However, once a set of self-consistent distances had been optimised and a core macromolecular fold defined, a more rigorous, atomic level, calculation could be carried out. The EBI RECCORD scripts were used to generate a final 100-model ensemble calculated in Cartesian space, in the presence of explicit H_2O solvent with the PARALLHDG 5.3 force field (Table 2.1.3).

Table 2.1.3: The computational procedures undertaken with CNS 1.3 as dictated by the RECOORD (99) scripts for the second simulated annealing step and water refinement of the AVR3a11 structural ensemble.

Phase	Step	Procedure
Simulated annealing	1	**High temperature torsion angle dynamics** 10 kK in 2000, 24 fs steps
	2	**Torsion angle dynamics cooling** 10 kK to 50 K in 2000, 24 fs steps
	3	**Cartesian dynamics cooling** i) 2 kK to 1 kK in 8000, 3 fs steps ii) 1 kK to 50 K in 8000, 3 fs steps
Solvated refinement	1	**Energy minimisation** 7 Å water shell immersion
	2	**Slow heating** 100 K to 500 K in 100 K, 3 fs, steps
	3	**Refinement** 500 K 4 fs , 2000 MD steps
	4	**Slow cooling** 500 K to 25 K, in 25 K incr. 4 fs
	5	**Final minimisation** 200 steps

2.1.7.3 Ensemble validation

The final ensemble of 100 structures were scored according to their lowest overall structural energies ($\sum(E_{bonds}+E_{angles}, +E_{improper}+E_{dihedral}+E_{VanDerWaals}+E_{elec}+E_{NOE}+E_{cdih})$) and the top 20 structures with zero NOE (>0.2 Å) or dihedral (>5°) restraint violations were compiled into a 20 model structural ensemble Protein Data Bank (PDB) file. The structural validation tools CING (100) and PSVS (Protein Structure Validation Suite (101), including analyses from the PROCHECK (102) and WHATIF (103) software packages) were used to assess the overall quality of the calculated models. If regions were found to be "disallowed" with respect to backbone φ and ψ angles, β atom deviation or side-chain rotamer geometry across a significant number of structural models, attempts were made to improve the overall ensemble by either relaxing the restraints to 5.5 Å or deleting them entirely if they were found to be incompatible with the overall fold. The calculation was then repeated with these alterations or omissions until no significant violations remained. The final structural ensemble has since been deposited with the RCSB under the accession code 3ZGK and will be released upon publication of the corresponding article.

2.1.8 AVR3a11ΔN NMR sample preparation

NMR samples for T_1, T_2 and hetNOE relaxation at 500 and 800 MHz were made to a total volume of 290 µl and contained a final concentrations of 1 mM protein, 10% D_2O, 0.03% NaN_3 and 200 µM DSS. Samples were spin filtered prior to use (0.22 µm pore size, Corning Inc., USA) and were transferred to 5 mm Shigemi tubes, susceptibility matched to D_2O, with an 8 mm spacer designed for use in Bruker probes (Shigemi Inc., USA). Tube inserts were positioned so as to fully remove any air above the sample and were fixed into position with Parafilm (Pechiney Plastic Packing Company, USA) to the top of the outer tube.

2.1.9 NMR relaxation of AVR3a11ΔN

For the AVR3a11ΔN construct, all ^{15}N relaxation NMR experiments were acquired at 293 K on Bruker Avance III (800 MHz) and Avance I (500 MHz)

spectrometers equipped with triple resonance indirect detect TXI probes with Z-gradients. All sets of T_1 (delays: 20 ms, 100 ms, 200 ms, 500 ms, 750 ms, 1 s (repeated) and 4 s) and T_2 experiments (delays: 17 ms (repeated), 51 ms, 85 ms (repeated), 136 ms, 170 ms (repeated), 204 ms and 254 ms) were acquired as interleaved pseudo-3D experiments with all relaxation delays acquired at each ^{15}N t1 increment. Heteronuclear (NH) steady state NOE (hetNOE) experiments were acquired, in triplicate, as interleaved pairs of HSQC experiments either with or without an initial period of proton saturation. Although only minor differences were observed between the AVR3a11 HN chemical shifts at 298 K, and AVR3a11ΔN at 293 K, a ^{1}H-^{15}N TOCSY-HSQC was also acquired at 293 K to facilitate the assignment of several peaks (Table 2.1.4). To ensure a consistent temperature between spectrometers and individual relaxation experiments, each spectrometer was temperature calibrated by a series of 1D ^{1}H spectra of a methanol standard according to the Bruker Variable Temperature Unit manual (Section 2.3.4.).

Table 2.1.4: NMR acquisition parameters for ^{15}N relaxation experiments acquired at 500 and 800 MHz for the AVR3a11ΔN construct. T_1, T_2 and hetNOE experiments were acquired with hsqct1etf3gpsi3d, hsqct2etf3gpsi3d and hsqcnoef3gpsi pulse programs, respectively and the HSQC-TOCSY was acquired with the dipsihsqcf3gpsi3d pulse program.Bracketed complex points and spectral widths apply to the F_3 ^{1}H dimension of the 800 MHz ^{1}H-^{15}N TOCSY-HSQC.

B_0 (MHz)	Experiment	scans	Complex points		Spectral width (Hz)	
			^{1}H	^{15}N	^{1}H	^{15}N
500	T_1	32	1024	64	7508	1519
500	T_2	32	1024	64	7508	1519
500	hetNOE	64	1024	32	7508	1519
800	T_1	32	1024	44	12019	2432
800	T_2	32	1024	64	12019	2270
800	hetNOE	64	1024	64	12019	2491
800	HSQC-TOCSY	64	1024 (48)	32	12019 (9603)	2432

2.1.10 NMR relaxation data processing

^{15}N relaxation spectra were processed as before (Section 2.1.1) and amide assignments not easily inferred from the AVR3a11 construct at 298 K were verified by comparing the chemical shifts of TOCSY-HSQC spectra at the two temperatures. The CCPN Analysis software package was used to extract the

peak intensities of peak maxima in all relaxation spectra and for T_1 and T_2 data; the relax (104) software package was used to fit, and minimise, the peak intensities to a two parameter exponential decay (equation 2.2):

$$I(t) = I_0 e^{-R_x t} \tag{2.2}$$

where *I(t)* is peak intensity at time point *t*, I_0 is the initial peak intensity and R_x is the relaxation rate. Baseplane noise, extracted from each spectrum with nmrDraw, was used to describe the initial peak intensity error and Monte Carlo simulations within the relax calculation generated final errors with respect to the deviation from the fit.

For hetNOE data, for each repeated pair of data sets, relax was used to calculate the hetNOE peak intensity ratio according to equation 2.3:

$$NOE = \frac{I_{sat}}{I_{ref}} \tag{2.3}$$

where I_{sat} refers to the peak intensities from spectra acquired with a preceding proton saturation pulse and I_{ref} refers to peak intensities from the standard reference ^{1}H-^{15}N HSQC. For analysis, and further calculations, the mean average of these triplicates at each field was taken, along with the standard deviation between the respective replicates, as an estimation of the error. The calculated R_1, R_2 and hetNOE relaxation values for AVR3a11ΔN at 500 and 800 MHz were then used as inputs, along with an estimated axially-symmetric diffusion tensor calculated from the 3ZR8 crystal structure using the program quadric (105, 106), to calculate the Lipari-Szabo model-free parameters with ModelFree 4.15 (107, 108) (see Section 4.3).

2.1.11 NMR H/D exchange of AVR3a11ΔN

A ^{15}N-labelled AVR3a11ΔN sample, prepared as before except with a 550 µl sample volume, was flash frozen in liquid nitrogen and lyophilised *in vacuo* overnight. The resulting protein was resuspended in 550 µl D_2O on ice and transferred to a 5 mm NMR tube (pp535, Wilmad).

A series of 28 SOFAST-HMQC (pulse program "sfhmqcf3gpph") experiments at 278 K were then acquired at 800 MHz (t_0 = 4 min) over 24 hr. Spectral widths of 12019 Hz and 3187 Hz and complex points of 1024 and 48 were used for the direct and indirect dimensions respectively. Initial experiments (<49 min) were acquired with 8 scans, and later experiments contained 16 scans to help account for the decrease in signal resulting from H/D exchange; each experiment had 96 indirect increments. A subsequent ^{15}N edited HSQC-TOCSY at 800 MHz was also acquired at 278 K to aid resonance assignment; with spectral widths of 12019 Hz, 2433 Hz and 9603 Hz with 1024, 32 and 48 complex points for the ^{1}H (F_1), ^{15}N (F_2) and ^{1}H F_3 dimensions respectively. The CCPN Analysis software package was used to extract peak intensity maxima which were then fitted to exponential decays using the Origin (109) software package.

2.2 Experimental methods for TARP samples

2.2.1 TARP constructs

2.2.1.1 TARP ABD (726-825)

A 100-residue construct of the actin-binding domain (ABD) of the TARP protein from *Chlamydia trachomatis* was kindly provided with an N-terminal glutathione-S-transferase (GST) tag, inserted into a pGEX-6P-1 plasmid, by Dr Ted Hackstadt (Rocky Mountain Laboratories, Montana, USA). The numerical descriptor "726-825" refers to the wild type, full-length TARP residue numbering.

2.2.1.2 TARP (726-808)

An 88 residue, C-terminally truncated, construct of the TARP actin-binding domain was also created by introducing a premature stop codon in the $TARP_{726\text{-}825}$ gene by site directed mutagenesis. This was achieved by using a Quikchange II^{TM} site directed mutagenesis kit (Agilent, USA), as per manufacturers instructions with custom primers (Sigma Aldrich, USA) (Table 2.2.1). The resulting construct was purified from successfully transformed *E.*

coli (strain XL1-Blue) using a NucleoSpin® plasmid midi prep kit (Macherey-Nagel, Germany) and sequenced (Genome Enterprise Limited, UK) using standard pGEX forward and reverse primers.

Table 2.2.1: The PCR primers used to incorporate a premature stop codon at residue position 809 in the TARP 726-825 gene by site directed mutagenesis.

Primer	Sequence (5'-3')
Forward	CCATGGAAAGGAAGCACGTCTTAAACGGGATCAGCAGGAG
Reverse	GGTACCTTTCCTTCGTGCAGAATTTGCCCTAGTCGTCCTC

2.2.2 Overexpression of TARP constructs

2.2.2.1 Bacterial transformation

Cell transformations of TARP constructs were carried out with the protocol outlined in Section 2.1.2.1.

2.2.2.2 Unlabelled TARP construct expression

E.coli (strain BL21 DE3*) colonies from successful transformations were picked and used to inoculate 10 ml aliquots of LB-ampicillin media (100 $\mu g.ml^{-1}$ ampicillin). After overnight growth at 37 °C with shaking (200 rpm), 5 ml of overnight culture was used to inoculate 0.5 l of LB-ampicillin media (2 l flasks), which were then incubated at 37 °C with shaking (200 rpm). Cultures were grown to an $O.D_{600}$ of 0.6-0.8 and then supplemented with 1 mM IPTG to induce overexpression. After 3-4 hours of continued growth, cultures were harvested by centrifugation at 4,000 x g for 15 mins, 4 °C. Following centrifugation, all supernatant was removed bar ~40 ml, which was used to resuspend the cell pellets. Cultures were then transferred to 50 ml centrifuge tubes and cells were harvested by centrifugation once more at 4,000 x g for 15 mins, 4 °C. All remaining supernatant was discarded and cell pellets were frozen at −20 °C.

2.2.2.3 Expression of isotopically enriched TARP constructs

All isotopically enriched TARP expressions were carried out according to Section 2.1.2.2, with either appropriate concentrations of $^{15}NH_4Cl$ and ^{12}C glucose or $^{15}NH_4Cl$ and ^{13}C glucose.

2.2.3 TARP purification

Harvested cell pellets were thawed on ice and resuspended in 35 ml ice cold phosphate buffered saline (PBS) (140 mM NaCl, 2.7 mM KCl, 10.1 mM Na_2HPO_4, 1.8 mM KH_2PO_4), pH 7.3. The resuspended cells were lysed with two sonication steps of 8 min 20 sec, by 30 W pulses for 200ms.s^{-1} (Status 200, Phillip Harris Scientific, UK) with a 5 min incubation on ice in between steps. Cell debris was then harvested by centrifugation at 30,000 x g for 30 min, 4 °C. Overexpression was checked using InstantBlue™ (Expedeon, UK) stained 15% SDS-PAGE gels. The resulting lysate was supplemented with 2 mM DTT and applied to a 20 cm glass chromatography column containing Glutathione Superflow GST-affinity resin, CV = 15 ml, (Generon, UK), that had been equilibrated with 4 CV PBS + 2 mM DTT. The column was then left to incubate, with agitation, at 4 °C for 2 hr. The flow-through was then collected on ice, followed by eight individual CV washes of PBS + 2 mM DTT. GST-tagged protein was then eluted with four individual CV washes of 10 mM glutathione, 50 mM Tris-HCl, 200 mM NaCl, pH 7.5. Elutions fractions were analysed with InstantBlue™ (Expedeon, UK) stained 15% SDS-PAGE gels and fractions containing GST-TARP constructs were pooled and dialysed exhaustively against two 5 l aliquots of PreScission™ cleavage buffer over 36 hr using 3.5 kDa MWCO tubing at 4 °C. Fusion protein concentration was determined by UV-Vis spectroscopy (GST:$TARP_{726\text{-}825}$/$TARP_{726\text{-}808}$ ε_{280nm} = 47650 M^{-1}.cm^{-1}) and PreScission™ protease was then added (2U.mg^{-1}) with the digestion incubated at 4 °C overnight with slight agitation. Silver stained 15% SDS-PAGE gels were then used to confirm digestion and samples were then concentrated to 1.5 ml with a 5 kDa MWCO concentrator.

The TARP samples were then finally purified using a 75 cm glass chromatography column packed with S75 Superdex size exclusion media (GE Healthcare, USA). Using an ÄKTA explorer purification system, samples were in-

jected onto the column (CV = 330 ml), equilibrated with 2 CV PreScission cleavage buffer, at 1 ml.min^{-1} and elution fractions were collected based upon the absorbance at 280 nm. The presence of pure TARP was established for experimental use by silver stained 15% SDS-PAGE gels. TARP samples were then dialyses exhaustively over 36 hrs against two 5 l aliquots of G-actin buffer (0.2 mM $CaCl_2$, 2 mM Tris-HCl, pH 7.5) either with 0.2 mM Na_2ATP for ITC and SRCD experimentation, or without for NMR experimentation.

2.2.4 Actin purification

G-actin was purified with a two-step protocol modified from that originally described by Pardee and Spudich (110) with the aid of Dr. Robin Maytum (University of Bedfordshire).

2.2.4.1 Preparation of actin-acetone powder

Two freshly killed chickens (A.D Harveys slaughterhouse, Norwich, UK) were collected within one hour of sacrifice and stored on ice during transit. All following work was carried out on the day of slaughter, in a 4 °C cold room, with all buffers and equipment prepared in advance and cooled to 4 °C. All obtainable breast muscle (~240 g) was dissected from both chickens and minced twice through a meat grinder until homogenous.

Mince was then stirred in 1 l of extraction buffer #1 (150 mM KH_2PO_4, 150 mM K_2HPO_4, pH 6.5) for 10 mins before being collected and drained through muslin cloth which had been sterilised by boiling in water for 15 min. The extraction process was repeated twice more, with extraction buffers #2 (50 mM $NaHCO_3$) and #3 (1 mM EDTA, pH 7.0), respectively. The extracted meat was then washed in two 2 l aliquots of H_2O for 5 min with stirring. The washed muscle was then moved to a room temperature fume cupboard and added to 1 l acetone, with stirring, for 10 min before being pressed dry with sterile muslin cloth. The acetone wash was repeated a total of five times and the resulting muscle was loosely broken up, covered with tissue and left to dry on a large evaporating dish overnight. The following day, all filamentous material was coarsely grated to remove large, insoluble matter and the resulting actin-acetone powder (21.13 g) was stored at −20 °C.

2.2.4.2 Actin purification

All buffers, and muslin cloth (sterilised by boiling for 15 minutes), were prepared 24 hrs prior so as to allow them to chill thoroughly to 4 °C. Packed in a mild salt:ice mix, 1.5 g of actin-acetone powder was stirred with 120 ml initial buffer (10 mM Tris-HCl, 0.5 mM Na_2ATP, 0.2 mM $CaCl_2$, 1 mM DTT, pH 8.0) for 30 min. The resulting solution was strained through two layers of sterile muslin cloth into a filter paper containing Buchner funnel on ice, under vacuum. The original beaker was rinsed with a further 30 ml initial buffer, which was then also filtered. Filtrate was then further clarified by ultracentrifugation for 1 hr at 92,650 x g, 4 °C. The G-actin-containing supernatant was decanted, supplemented with 2 mM $MgCl_2$ and 100 mM KCl to induce polymerisation and left to gently stir at room temperature for 1 hr. Polymerised actin was then harvested by ultracentrifugation for 3 hr at 92,650 x g, 4 °C. The supernatant was then discarded and F-actin pellets were resuspended in a total of 22 ml depolymerisation buffer (0.5 mM Tris-HCl, 0.2 mM $CaCl_2$, 1 mM NaN_3, pH 7.5). This was passed though a handheld homogeniser 10 times before being left to dialyse (3.5 kDa MWCO) against a further 4.5 l depolymerisation buffer for 12 hr at 4 °C.

The following day, the depolymerised actin was removed from dialysis and clarified once more from insoluble materials by ultracentrifugation for 1 hr, at 92,650 x g, 4 °C. The resulting solution was either diluted (depolymerisation buffer), or concentrated (Vivaspin, 10 kDa MWCO) to 1 $mg.ml^{-1}$ (Gallus gallus actin $\varepsilon 280nm = 1.028\ mg.ml^{-1}.cm^{-1}$), and the resulting solution supplemented with final concentrations of 5 μM Na_2ATP, 2 mM $MgCl_2$ and 100 mM KCl. Pure F-actin pellets were then harvested with ultracentrifugation at 92,650 x g for 3 hr, 4 °C. The pelleted F-actin was resuspended in a total of 10 ml G-actin buffer (0.2 mM Na_2ATP, 0.2 mM $CaCl_2$, 2 mM Tris-HCl, pH 7.5), passed through a handheld homogeniser 10 times and dialysed (3.5 kDa MWCO) for 36 hr against 5 l G-actin buffer. Actin samples were then diluted accordingly prior to use, and any excess was flash frozen, with the addition of 1 mM DTT, in liquid nitrogen and stored at −80 °C.

2.2.5 NMR sample preparation

All NMR samples were made to a total volume of 500 µl and contained final concentrations of approximately 1 mM protein, 10% D_2O, 0.03% NaN_3 and 200 µM DSS. Samples were spin filtered prior to use (0.22 µm, Corning Inc., USA) and were transferred to 5 mm NMR tubes (pp535, Wilmad).

2.2.6 NMR spectroscopy

NMR spectra for the $TARP_{726\text{-}825}$ and $TARP_{726\text{-}808}$ constructs were acquired at 298 K (unless otherwise stated), either at 800 MHz (using a Bruker Avance III spectrometer equipped with a triple resonance indirect detect TXI probe with z-gradients) or 500 MHz or 600 MHz (using Varian Inova spectrometers with 5 mm triple resonance z-gradient probes).

The following experiments were acquired by Dr. Tharin Blumenschein, UEA, using the $TARP_{726\text{-}825}$ construct, at 800 MHz: [^{1}H-^{15}N]-HSQC, 3D CBCA(CO)NH, 3D HNCACB, 3D HNN, ^{1}H-^{15}N TOCSY-HSQC and ^{1}H-^{15}N NOESY- HSQC (Table 2.2.2). Furthermore, series of variable temperature (278-307 K, in 5 K increments), pH (additions of 1 M NaOH or 1M HCl) and NaCl concentration (additions of 1 M or 5 M NaCl) [^{1}H-^{15}N] HSQC series were also acquired (Table 2.2.3).

Table 2.2.2: NMR acquisition parameters of experiments acquired at 800 MHz for the backbone assignment and analysis of the TARP ABD.Dimensions different to those in the table header are labelled in superscript.

Experiment	Scans	Complex points			Spectral width (Hz)		
		^{1}H	^{15}N	^{13}C	^{1}H	^{15}N	^{13}C
^{1}H-^{15}N HSQC	24	1024	128	-	11161	3250	-
CBCA(CO)NH	16	1024	32	90	9615	2595	15095
HNCACB	24	1024	32	70	9615	2595	15095
HNN	48	1024	41	32 ^{15N}	9615	2595	2595 ^{15N}
^{15}N NOESY	16	1024	32	96 ^{1H}	9615	2595	9615 ^{1H}
^{15}N-TOCSY	16	1024	32	92 ^{1H}	9615	2595	9615 ^{1H}

Table 2.2.3: NMR acquisition parameters for ^{15}N experiments acquired for the TARP ABD construct.All experiments except pH series (hsqcetf3gpsi) were acquired with the ghNhsqc pulse program.

B_0 (MHz)	Experiment	scans	Complex points		Spectral width (Hz)	
			^{1}H	^{15}N	^{1}H	^{15}N
500	T_1	16	1024	128	8000	1944
500	T_2	32	1024	128	8000	1944
600	NaCl series	32	1024	64	8358	2431
600	Temp. series	8	1024	256	6948	1630
800	pH series	8	3250	128	11161	2270

During the pH titrations, the method used to determine the exact pH was taken and modified from that described by Baryshnikova, O. K et al., (111), wherein the [HA]/[A-] ratio of the buffer species from are calculated from the 1D ^{1}H-NMR chemical shifts of the labile buffer protons (Equation 2.4).

$$pH = pK_a - log_{10}\left(\frac{\delta_{Obs} - \delta_A}{\delta_{HA} - \delta_{Obs}}\right) \tag{2.4}$$

where δ_{Obs} is the observed chemical shift (ppm) of the buffer species and δ_{HA} and δ_A are the predetermined chemical shifts with respect to the fully protonated and fully deprotonated states respectively.

The relaxation experiments carried out for the 105 residue construct of the TARP ABD were acquired at 500 MHz were using sets of individual HSQC experiments with randomly ordered sets of T_1 delay times (20 ms (repeated), 100 ms, 300 ms (repeated), 500 ms, 750 ms, 1 s, 2 s, 3 s, 5 s) or T_2 delays(10 ms (repeated), 50 ms, 70 ms, 110 ms, 150 ms (repeated), 210 ms, 250 ms, 310 ms) delays. All spectra regarding the TARP constructs were processed using NMRPipe and analysed with the CCPN Analysis and relax software packages as previously described (sections 2.1.1, 2.1.2 and 2.1.6 respectively).

2.2.7 Synchrotron circular dichroism

Synchrotron radiation circular dichroism (SRCD) experiments of the two TARP constructs, and their analysis, were acquired by our collaborators Dr Andrew Miles and Prof Bonnie Wallace (Birbeck college, UK) on beamline CD1 at the Institute for Storage Ring Facilities, University of Aarhus Denmark. Two

aliquots of both TARP constructs were shipped on ice at 1 mg.ml^{-1}, following dialysis against 5 l aliquots of either G-actin buffer (without Na_2 ATP) (20 °C experiment), or G-actin buffer (without Na_2 ATP), 10% glycerol (thermal denaturation experiments up to 85 °C). A 20 μM 3:1 TARP ABD:actin complex was also prepared, and shipped on ice, in the presence of G-actin buffer with 0.2 mM ATP. Measurements were made between 280 nm and 175 nm (1 nm step size) in 0.005 cm pathlength quartz Suprcil cells (Hellma, UK), with an averaging time of 2 s. Variable temperature SRCD was carried out from 20 °C to 85 °C in 5 °C increments, with each stage equilibrated with a 3 min hold-time. Three scans were acquired at each temperature and were averaged and subtracted from the average of three control scans of the relevant buffer. Spectra were calibrated against camphorsulphonic acid (112) and all processing was carried out with CDTool software(113). The Dichroweb analysis server was used for all secondary structure determination (114). Values from the CONTINLL (115, 116), SELCON and CDDSTR (117) (Reference dataset 6) algorithms were averaged.

2.2.8 Electrospray mass spectrometery

Mass spectrometry of TARP ABD was carried out by Dr Robin Maytum at the University of Bedfordshire using a Thermo Finnigan LCQ spectrometer, fitted with a metal needle electrospray ionisation option. A 1 mM sample of $TARP_{726\text{-}825}$ was buffer exchanged into 10 mM Ammonium Bicarbonate. Prior to experimentation, the sample was made to 20% methanol, 0.1% formic acid and the sample was applied by direct injection at a flow rate of 10 μl.min^{-1}. Data was recorded and averaged, over 5 min, and processed with the accompanying Thermo Xcalibur v1.3 software. The masses for each peak in the most intense multiply charged series were calculated using the equation $(m/z) * z$ and the average mass was calculated as the mean average with the standard deviation as the error.

2.2.9 Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) was carried out using a MicroCal 200 calorimeter (GE Healthcare) at the John Innes Centre, Norwich, UK.

Both the TARP ABD and actin samples were dialysed in separate dialysis bags against the same two 5 l batches of G-actin buffer for 36 hr. The concentrations of both samples were adjusted to the final values the morning of experimentation. Ligand binding experiments were carried out at 25 °C, with a series of 19, 2 µl injections of 175 µM $TARP_{726\text{-}825}$ (total syringe volume = 40 µl), into a 200 µl chamber of 10 µM G-actin with 1000 rpm stirring. Injections were made over 4 s, 120 s apart, with an initial injection point carried out (volume = 0.4 µl, duration = 0.8 s) to ensure the syringe was void of air. Three control experiments were also carried out in order to subtract the enthalpies associated with injection (buffer:buffer), actin dilution (buffer:actin) and TARP dilution (TARP:buffer). All analysis was carried out within the Origin software package using the integrated MicroCal ITC plugin.

2.2.10 TARP:actin docking simulations

All *in silico* docking simulations for the TARP:ABD and WAVE2 WH2 domains with monomeric G-actin were carried out on the FlexPepDock web server (118) (accessible: `http://flexpepdock.furmanlab.cs.huji.ac.il/`). Input files consisted of single PDB formatted coordinate files containing the G-actin monomer and relevant peptides as molecular chains A and S, respectively, with a single "TER" termination line in between. No small molecules or water molecules were included. The model for the G-actin monomer included in all simulations was derived from the WAVE2:actin crystal structure (PDB code: 2A40) (119).The reliability of the FlexPepDock methodology was tested with simulations of the 22 residue WAVE2 WH2 domain, for which the results could be compared to the 2A40 crystal structure. Three calculations were carried out for both the aligned 22 residue WAVE2 and TARP ABD WH2 domains. These were with both constructs either occupying a fully extended state (135° for all φ/ψ angles), conforming to ideal-geometry 10 residue α-helices (φ/ψ angles of $-57°$ and $-47°$ respectively) or containing the conformation of the actin-bound WAVE2 peptide. For the aligned TARP fragment, this was achieved by constraining the construct to the dihedral angles of the WAVE2 peptide from the 2A40 crystal structure. All WH2 chains were translated out of the actin hydrophobic cleft by 5 Å at 90° to the plane of the α-helical axis. The fully extended, ideal-geometry helices and WAVE2-adopted conformation (for the TARP ABD) were generated by standard scripts packaged with CNS version 1.3 (98), MODELLER (120) and UCSF-Chimera (121), respectively.

2.3 Common techniques

2.3.1 Preparation of competent cells

To prepare *E. coli* strains JM-109 and BL21-DE3 for efficient transformation, 5 ml of LB culture was inoculated from colonies picked from LB-agar plates and incubated overnight at 37 °C with 200 rpm shaking. The following day, 0.4 ml of overnight culture was used to inoculate 40 ml of fresh LB and cultures were incubated at 37 °C with 200 rpm shaking until the $O.D_{600}$ was approximately 0.4. Cells were transferred to a 50 ml centrifuge tube and harvested by centrifugation at 7,700 x g for 8 min, 4 °C. Following centrifugation, the supernatant was discarded and the cell pellet, incubated on ice, was resuspended in 8 ml ice cold Transformation Buffer 1 (100 mM KCl, 300 mM potassium acetate (pH 7.5), 13.5 mM $CaCl_2$, 15% (v/v) glycerol, 50 mM $MgCl_2$, pH 6.4, sterilised by autoclave). Cells were then incubated on ice for a further 15 min and then harvested as before. Supernatant was discarded and cells were resuspended in 4 ml ice cold Transformation Buffer 2 (10 mM KCl, 100 mM $CaCl_2$, 15% glycerol (v/v) and 20 mM 3-(N-morpholino)propanesulfonic acid (MOPS), pH 6.8, sterilised by autoclave). 50 µl aliquots of cells were transferred to 1 ml micro-centrifuge tubes, flash frozen in liquid nitrogen and stored at −80 °C for future use.

2.3.2 SDS-PAGE

All SDS-PAGE gels were run at 200 V until the leading edge of the Bromophenol blue sample buffer had flowed from between the gel plates. Cell lysate samples were prepared by harvesting cells from 1 ml of culture, removing the supernatant, and heating at 100 °C for 10 min in an appropriate (50 {µl pre-induction, 100 µl post-induction) volume of Bromophenol blue sample buffer (50 mM Tris-HCl pH 6.8, 10% (v/v) glycerol, 2% (w/v) SDS, 0.1% (w/v) Bromophenol blue, 0.7 M 2-mercaptpethanol). Protein samples were prepared by the addition of an equal volume of 2x concentrated sample buffer to sample. Following electrophoresis, protein bands were stained using InstantBlue™ coomassie stain (Expedeon, UK) until satisfactorily visible, and background stain was removed by rocking the gels over night in $_dH_20$.

2.3.3 SDS-PAGE silver staining

Silver staining was used to visualise SDS-PAGE protein bands that were ineffectively stained with InstantBlue due to a lack of aromatic amino acids. Following electrophoresis, gels were first removed from the casting plates and then washed for 10 min each in separate 50 ml volumes of: 50% methanol, 5% methanol, 40 μM DTT (30 s H_2O rinse) and 0.1% (w/v) $AgNO_3$ (30 s H_2O rinse). Protein bands were then visualised by washing in development solution (0.05% formaldehyde, 280 mM Na_2CO_3) until stained. Citric acid was then added in excess to stop the staining and background staining could be clarified, if necessary, by rocking in 50 ml H_2O for 12 hr.

2.3.4 NMR temperature calibration

The accuracy of the intended sample temperature within all NMR spectrometers was checked using the standard 100% methanol sample provided with Bruker spectrometers. 1D ^{1}H-NMR spectra were recorded at 293 K and temperature accuracy was assessed using equation 2.5 (122):

$$T = -23.832\Delta^2 - 29.46\Delta + 403 \tag{2.5}$$

where T is the temperature in Kelvin and $\delta\Delta$ is the difference in ppm between the CH_3 and OH ^{1}H-NMR peaks of the 100% methanol sample.

Chapter 3

Characterising the Structure of AVR3a11

3.1 Introduction

3.1.1 Principles of NMR spectroscopy

3.1.1.1 Matter and quantum spins

As a form of spectroscopy, nuclear magnetic resonance (NMR) deals with the interactions of matter and electromagnetic (EM) energy. Simply put, matter, as we know it, is composed of atoms, which are in turn comprised of nuclei and electrons. The nuclei can then be further divided into protons and neutrons and those in turn are comprised of quarks, without delving into the further sub divisions (Figure 3.1.1).

The quarks comprising the nucleons are either of the flavour up or down and it requires two down quarks and an up to form a neutron and two up and a down for a proton. This then results in a neutron with neutral electric charge and ½ spin and a proton with positive charge (1.602×10^{-19} C) and ½ spin. Spin is an odd concept to comprehend, as there is no macroscopic equivalent, but it is a fundamental property of all sub atomic particles such as the magnetic moment, mass, or charge.

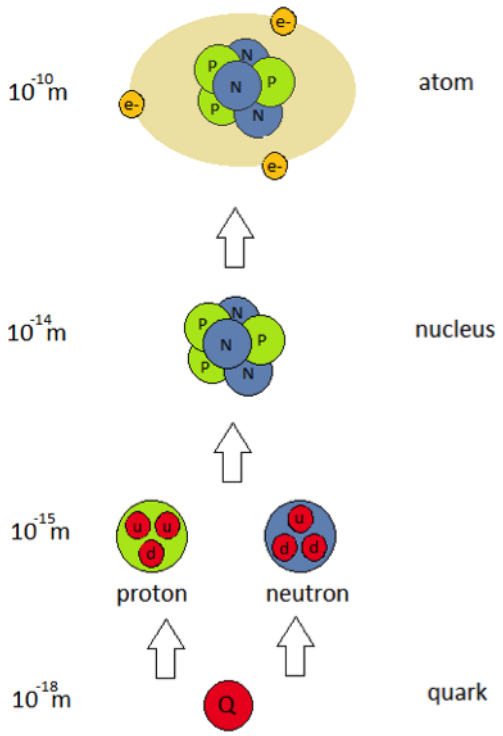

Figure 3.1.1: The composition of matter with an approximate diameter scale. Quarks (Q in red) are denoted with a U or a D for their up and down flavours, respectively.

It also does not mean an actual rotational movement but defines a tendency towards a particular angular momentum for a particle, regardless of its motion. Subatomic particles, atoms and molecules may have an actual rotational angular momentum obtained from physical collisions or electromagnetic manipulation but it is only subatomic particles that intrinsically possess an angular momentum, and this is determined by their spin. Spin, as a quantum property, is designated by S and is either expressed as integer terms (boson particles) or half-integer terms (fermion particles). As the physical rotational angular momentum at the atomic level is also quantised, the total angular momentum vector, J_3, for a particle can be deduced by arrangement of the two quantum components J_1 and J_2 (equation 3.1) (123).

$$J_3 = \begin{cases} |J_1 - J_2| \\ |J_1 - J_2| + 1 \\ \vdots \\ |J_1 + J_2| \end{cases} \tag{3.1}$$

Equation 3.1 is a good example of how quantum states can be combined through additive or subtractive means, and how further additions or subtractions of the relevant quantum integers are possible until the minimum state is achieved through sufficient reductions. In real environments a system will summate to the lowest energy state as dictated by its surroundings. This is described by the Aufbau and Pauli principles that explain how electron spins pair and reside in the lowest energy orbital locations (123). Atomic nuclei themselves also possess quantum spin, I. This is a culmination of their constituent proton and neutron spins and for hydrogen, (^{1}H), I = $^1/_2$ resulting from its single proton nucleus. For all other nuclei, which are comprised of multiple protons and neutrons (collectively termed nucleons), the potential values for the overall spin can be calculated in a similar fashion to Equation 3.1 for different combinations of the nucleon spin states. Although many potential configurations may exist, under standard conditions, the overall I will result from the lowest energy configuration, which is termed the ground state. There are no empirical methods of predicting specific ground state spins, and they are best thought of as properties inherent to each element or isotope.

Standard NMR techniques deal with nuclei of spin $^1/_2$. This is due to the fact that spin $^1/_2$ nuclei act as a single point charge and so subsequently exhibit a symmetrical electrical distribution. Electric effects do not then dictate the rotational movement of the nuclei and during an NMR experiment any apparent change can be thought of as induced by magnetic interaction. The utilisation of elements with spin $>^1/_2$ is possible with NMR but nuclei with largely homogenous electric fields are favoured due to their limited electronic interaction with the surrounding environment (124).

3.1.1.2 The Zeeman Effect and Nuclear Zeeman splitting

During an NMR experiment, the measured spectral energies arise from the emissions of perturbed nuclei relaxing to a lower energy state. The Zeeman Effect describes the initial splitting of these energy states due to the applied magnetic field on the sample (Figure 3.1.2, A & B). In the absence of a magnetic field any given particle will have $2n + 1$ alternate, constituent particle quantum state configurations which are degenerate; that is their energy levels are equal.

The Zeeman Effect has also been described for electrons (Figure 3.1.5, C)

(125) whereby the quantum arrangement of both their orbital arrangement or spin can occupy separate degenerate levels and for photons (126) which have degenerate quantum spins. NMR spectroscopy however utilises the Zeeman Effect of nuclei, the Nuclear Zeeman effect, wherein the constituent nucleon spins can be degenerately arranged. Although these configurations equate to the same degenerate energy, each arrangement will have its own specific angular momentum quantum state and hence its own magnetic moment as they are co-linear properties (Equation 3.2) (127):

$$\mu = \gamma\ I \tag{3.2}$$

where μ is the magnetic moment of the particle, γ is the gyromagnetic ratio of the particle and I is the nuclear spin.

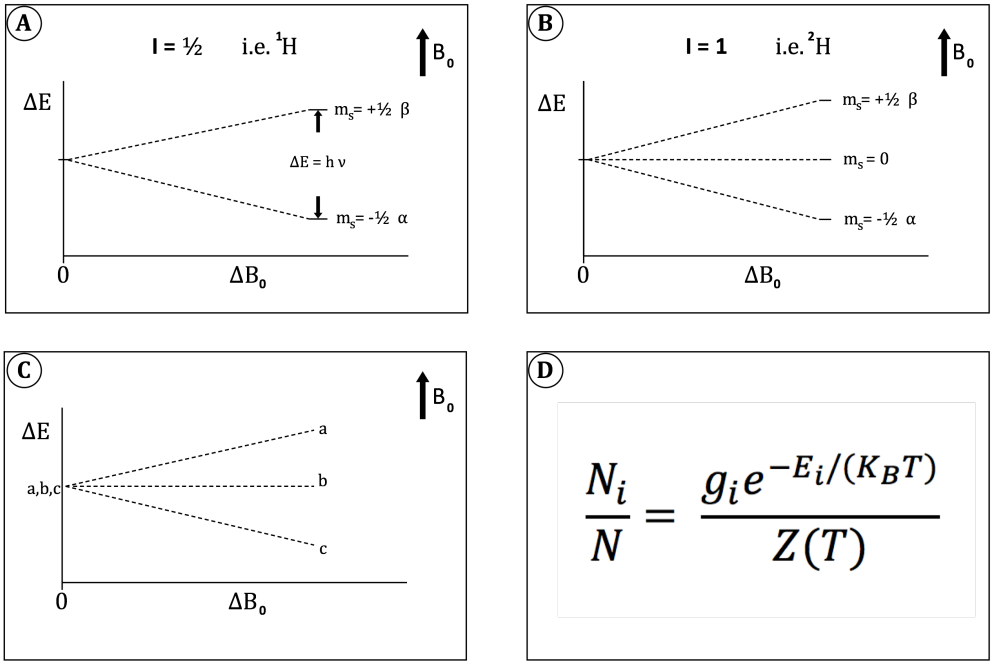

Figure 3.1.2: A). The Zeeman splitting of a spin $\frac{1}{2}$ nucleus, where $\hbar$ is Plancks constant (6.63x10^{-34} J.s), ν is the radiation frequency and E is emission energy. B_0 is the applied magnetic field and M_j is the azimuthal quantum number. B). The Zeeman splitting of a spin=1 nucleus. C). The Zeeman splitting of an electron orbital with three degenerate configurations. D). The Boltzmann distribution equation. Where N_i / N is the fraction of particles in each state, g_i is the number of degenerate states, E_i is the energy of the particles, Z (T) is the partition function, T is temperature in Kelvin and K_B is The Boltzmann constant (1.38065 10^{23} J.K^{-1}).

In the presence of a magnetic field the degeneracy of these energy states is removed as their respective magnetic moments align differently with the main magnetic field B_0. Once the magnetic field has been applied and the degeneracy of the energy states has been removed, whether or not the magnetic moment of a particular atomic nucleus will align with or against the field can be though of as a random process.

However, real-world samples are ensembles of atoms, and the fraction of the total population in each energy state is actually a function of Boltzmann distribution (Figure 3.1.2, D) (128). Related to the energy within the system, the Boltzmann distribution is a probability measure that can determine the ratio of high energy to low energy states in a particular system. Historically these states have been labelled β for high energy states (which align against the magnetic field) or α for the lower energy states (which align with the magnetic field).

3.1.1.3 Nuclear magnetism and precession

In a free moving state, atoms will tend to align their magnetic moment for a minimal magnetic energy along the axis of the external magnetic field B_0. Atomic magnetism arises from three components, two of which are inherent electron magnetism and inherent nuclear magnetism both intrinsic quantum properties of their respective particles. The other constituent is an induced magnetism dependent on electron orbital circulation and is wholly comparable to the phenomenon of electronic inductance.

The combination of these magnetic components then determines the overall magnetic moment for a given atom and therefore its behaviour within an external magnetic field. Magnetic moments with a dominant electron current component perturb the external magnetic field outwards (diamagnetic) whilst moments with dominant inherent magnetism attract it (paramagnetic). NMR, as the name suggests, deals with the inherent quantum magnetism of the nucleus, μ, and this is the summation of the inherent magnetism of the subnuclear particles. As shown in Equation 3.2, μ is also related to the nuclear spin I by a proportionality constant known as the gyromagnetic ratio, γ. This relationship not only defines how the magnetic moment and spin angular momentum magnitudes are correlated, but with respect to the plane in which they act, also explains the relationship between their directional components.

As co-linear properties, a nucleus with $\gamma > 0$ will have a parallel I and μ alignment whereas I and μ will be antiparallel for a $\gamma < 0$ (Figure 3.1.3, A). The gyromagnetic ratio is an inherent characteristic of sub atomic particles and, if known *a priori*, can be used to determine the behaviour of a particular nucleus within an applied magnetic field. Outside of a magnetic field, the atomic magnetisation of a normal material (i.e. not a fixed magnet) will be isotropic; it is equally distributed in each direction and has no net effect. In addition, the nuclear spin angular momentum will also be similarly isotropic. Upon the application of an external magnetic field (B_0), the individual nuclear magnetic moments will align, distributing themselves between energy configurations according to Boltzmann distribution (Figure 3.1.2 D).

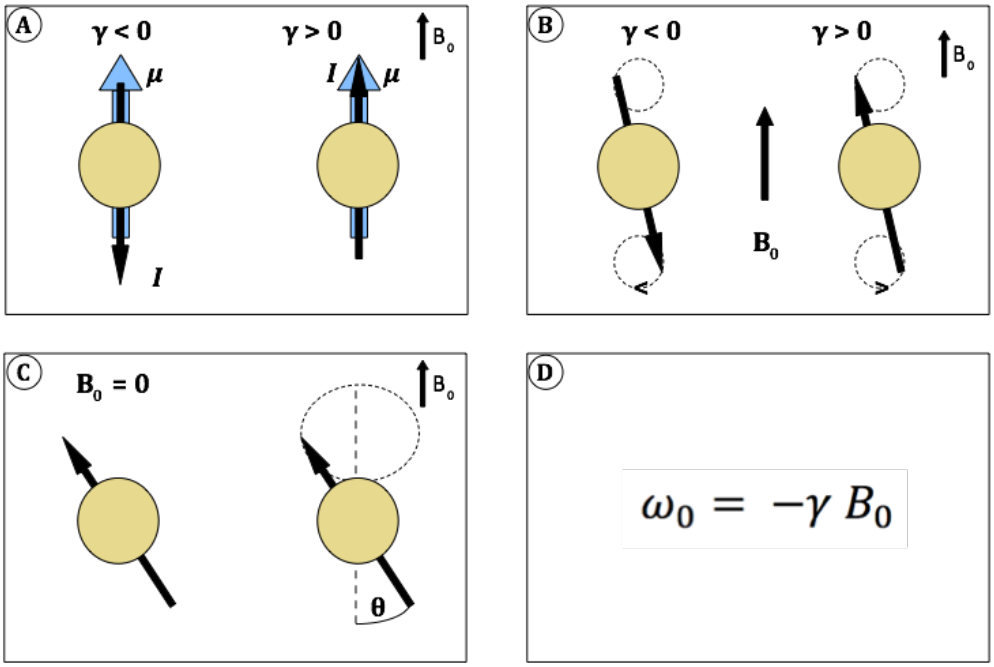

Figure 3.1.3: The relative directions of nuclear spin angular momentum (I) and magnetic momentum (μ) at the lowest energy configuration for nuclei with gyromagnetic ratios (γ) greater or less than zero. B) The torque generated nuclear precession in the presence of an external magnetic field C). The angle between precession and the B_0 axis immediately after application of the field relates to the nuclear directionality prior to B_0. D). The relationship between the frequency of precession (ω), the gyromagnetic ratio (γ) and the external magnetic field strength (B).

However the presence of the co-linear spin angular momentum complicates the actual behaviour in the magnetic field. The polarising effect of B_0 on the magnetic moment creates a torque with the spin angular momentum and the magnetic moments quickly begin to rotate, or precess, about the B_0 axis. Although the angle of precession will initially be in relation to the initial position of I with respect to the B_0 axis (Figure 3.1.3, C), the bulk magnetisation of the nuclei will quickly (ms) reorient to precess at a fixed angle (ϑ) (Figure 3.1.3, B) (Equation 3.3).

$$cos\theta = \frac{m_I}{\sqrt{I(I+1)}} \tag{3.3}$$

where m is the spin state of the nucleus, and I is the spin quantum number. The frequency of precession (The Larmor frequency, $\omega/2\pi$) around the B_0 axis (historically defined as the z axis) is characteristic of a particular elemental nucleus and is related to the strength of the applied field and the relevant gyromagnetic ratio (Figure 3.1.3, D). For nuclei with $\gamma > 0$ the Larmor precession will be clockwise about the z axis. For a $\gamma < 0$, Larmor precession will seemingly occur in an anticlockwise direction, although this is just a technicality owing to the observed frame of reference (+z axis), as the nuclei will have opposing signs of I and μ. This is akin to observing a clock from the opposite side of the clock face and watching the hands rotate anticlockwise. At this stage, when Equation 3.3 governs the angle of precession and the spin states have achieved Boltzmann distribution, the sample is said to be at thermal equilibrium with the system and the external magnetic field.

3.1.1.4 Generating the NMR signal

Before an experiment takes place, at equilibrium the aligned magnetic moments of all the spins in solution at equilibrium sum to create a definitive, net magnetisation vector along the B_0 z axis related to the temperature in the system. The time it takes to introduce this specific anisotropic magnetisation once the B_0 field is applied has been termed T_1, or the longitudinal relaxation time and was mathematically derived by Felix Bloch in 1946 (Equation 3.3).

$$M_{z_{(t,x)}} = \gamma(M_x B_y - M_y B_x) + \frac{M_0 - M_z}{T_1} \tag{3.4}$$

where M is the nuclear magnetisation for the subscript Cartesian axis (x, y and z), B is the magnetic field strength in the denoted plane (x and y), γ is the gyromagnetic ratio of the nucleus, M_0 is the steady state nuclear magnetisation and T_1 is the longitudinal relaxation time. An accurate measure of z axis bulk magnetisation is unfortunately difficult, if not impossible, to obtain owing to the huge dominance of the B_0 magnetic field. The majority of NMR experiments therefore work by measuring the net magnetisation and relaxation times within, and out of, the x-y plane after the application of a second rotating magnetic field, termed B_1. If this electromagnetic B_1 pulse is applied to the sample with the same frequency as nuclear precession, but along the x-y axis, the net magnetisation within a sample can be made to rotate away from z-axis equilibrium through an effective torque on the equilibrium bulk magnetisation vector. These EM pulses, due to the frequency range of nuclear precession, fall into the radio-frequency (RF) band of EM waves and in their simplest form are used to induce a 90° change in the net magnetisation. The subsequent change from the z-axis into the x-y plane enables the bulk magnetisation, now perpendicular to B_0, to be measured for the sample.

Once the desired axial changes in the bulk magnetisation vector have been induced, the pulse is turned off. Because the net magnetisation vector of the nuclei in the sample was aligned along the z axis prior to the pulse, the bulk magnetisation vector, having been rotated by the same degree, remains aligned but now in the x-y plane. At this time immediately after a pulse, the magnetic moments of the excited nuclei are said to be coherent, or in phase with one another, and so sum to create the xy term of the bulk magnetisation. Although the bulk magnetisation vector is now perpendicular to B_0, the thermal instability of the sample will eventually lead to the nuclear populations once again attaining thermal equilibrium with B_0. Concomitantly, because the frequency of precession is dependent on $\mu.B_0$, small spatial fluctuations in the magnetic field will cause some spins to precess at slightly different frequencies, therefore although the initial bulk magnetisation was transferred into the x-y plane, these different rates of precession will cause a loss of coherence of the net magnetisation over time. The time a sample takes to lose coherence and reach thermal equilibrium in the x-y plane is called the transverse relaxation time or T_2 and can yield important information regarding the dynamics of a particular species, as molecular motions can also cause variations in the locally experienced magnetic field. The same can also be said for the longitudinal relaxation time, which is the relaxation back to equilibrium along the

z axis (T_1). The signals acquired during an NMR experiment come from the bulk magnetisation and are measured by exploiting electromagnetic induction, which yields a flowing current in a coil, wound perpendicular to the B_0 magnetic field. The T_1 and T_2 relaxations to thermal equilibrium with B_0 yields a specific, yet typical, signal known as free induction decay (FID) which corresponds to the withdrawal of the magnetisation, from the observable x-y plane and consequently detection by the coil (Figure 3.1.4).

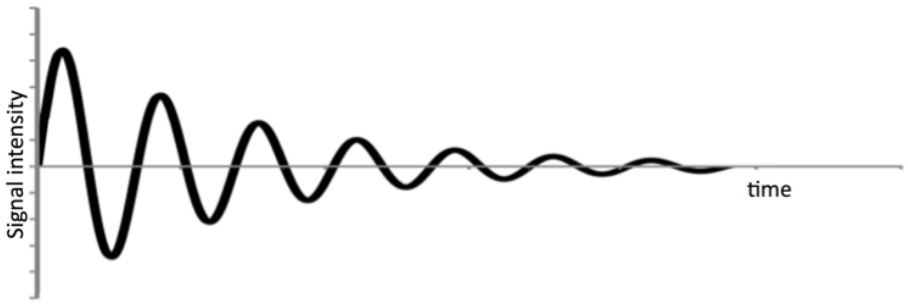

Figure 3.1.4: A typical free-induction decay (FID) signal acquired after a pulsed NMR experiment.

There are however multiple parts to an NMR dataset and this is dependent on the specific pulse sequence defined at the start of the experiment. Theoretically any spin $\frac{1}{2}$ nucleus will precess in a magnetic field, will be malleable with respect to B_1 and will have a detectable net magnetic moment. However the differences in the rate of Larmor precession, as dictated by the gyromagnetic ratios of the nuclei under observation, usually requires the use of dedicated circuitry (channels) for the excitation and detection of different isotopes.

3.1.1.5 The parts per million (ppm) chemical shift scale

Although NMR signals can then be used to ascertain a wide range of information about a sample, it is important to maintain a common measurement system to enable the comparative use of the acquired spectra. In this regard, when reporting chemical shift frequencies the B_0 field must be taken into account due to its proportionality to Larmor precession and the Zeeman splitting of energy levels. Consequently, the Fourier transformed absorption peaks measured at different field strengths will have a different resonant frequency for different B_0 magnetic field strengths.

For this reason, resonant frequencies are always described in relation to those of a reference compound so that they are independent of field strength. This chemical shift scale therefore enables results from different spectrometers to be readily compared, and are calculated according to the formula:

$$\delta = 10^6 \ \frac{\omega^0 - \omega^0_{Ref}}{\omega^0_{Ref}} \tag{3.5}$$

where δ is the chemical shift, ω^0 is the Larmor frequency of the nucleus under investigation and $\omega^0{}_{\mathrm{Ref}}$ is the Larmor frequency of a reference compound, usually tetramethylsilane (TMS) or 4, 4-dimethyl-4-silapentane-1-sulfonic acid (DSS). Reference compounds therefore designate a zero to the frequency domain and although different reference compounds can be used, as long the exact standard is known, it can be used to standardise the resulting spectra and allow comparative analysis. Due to the resulting numerical magnitude of most chemical shifts it has become standard to denote chemical shifts with units of parts per million ppm, hence why Equation 1.4 multiples the difference in frequency by 10^6.

3.1.1.6 The HSQC experiment

The 2D heteronuclear single quantum coherence (HSQC) NMR experiment, and many variants thereof, can be described as the workhorses of NMR investigations into protein systems. Its ubiquitous use stems from its ability to cleanly observe heteronuclear correlations with a relatively short acquisition time under standard operating parameters (potentially as quick as tens of seconds). This is most commonly acquired for H-N or H-C correlations, but does of course require the sample under investigation to contain spin = $\frac{1}{2}$ isotopes (^{13}C or ^{15}N) of the particular heteronucleus under observation. The ability to detect ^{1}H-^{15}N correlations for isotopically labelled samples is also particularly useful as the resulting spectra single peaks for the individual amide moieties of amino acids - with the exceptions of prolines (for which no amide is present) and NH_2 sidechain moieties (which appear as two peaks due to the individual protons each coupling to a shared nitrogen atom). If triple resonance experiments are also acquired, then amide peaks can be assigned to their corresponding amino acids within the primary sequence, allowing residue specific characterisation. The ^{1}H-^{15}N HSQC spectrum, is therefore described

as a "fingerprint" which reports on the conformational and biochemical profile of the proteins under investigation. The influence of variables, such as temperature, pH, or the addition of small molecules or macromolecular binding partners, can therefore be highlighted at a residue specific level. The ^{1}H-^{15}N HSQC experiment is also the principal building block of NH relaxation experiments that allow the elucidation of dynamic molecular phenomena (discussed later in Chapter 4). The HSQC pulse sequence (Figure 3.1.5) first begins with an insensitive nuclei enhanced by polarisation transfer (INEPT) magnetisation transfer that facilitates selective sensitivity enhancement by transferring magnetisation from the abundant, magnetically susceptible, I spins (^{1}H, large γ), to the directly bound, relatively insensitive S spins (^{15}N, small γ) (Figure 3.1.5, *1).

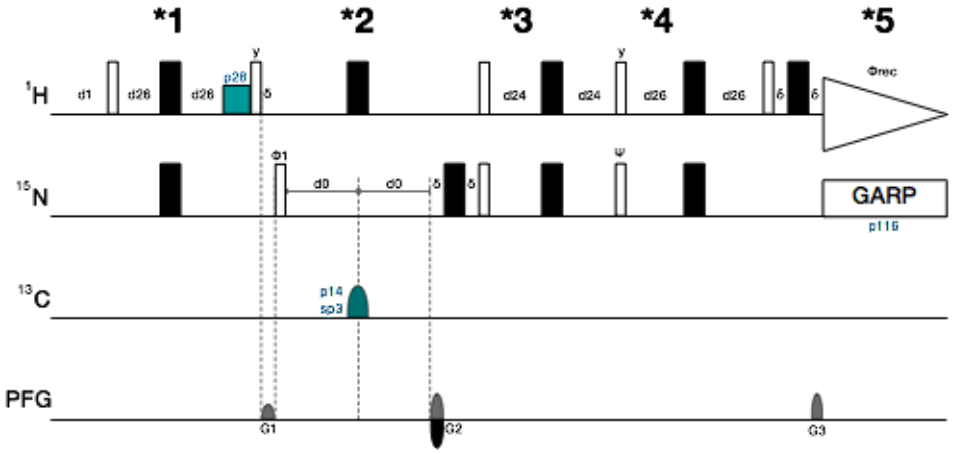

Figure 3.1.5: The pulse sequence for the gradient enhanced coherence selection, sensitivity improved hsqcetf3gpsi heteronuclear single quantum coherence (HSQC) experiment (130-132) routinely used for detecting ^{1}H-^{15}N correlations throughout these studies. Narrow hollow bars and wide black bars represent 90° and 180° pulses respectively and shaped pulses are denoted as rounded bars. All pulses were applied to the x phase unless otherwise stated. The GARP pulse sequence (133) was used for decoupling during signal acquisition (hollow triangle). Starred numbers denote the point of INEPT magnetisation transfer (*1), T_1 evolution period (*2), the reverse INEPT ^{15}N-^{1}H magnetisation transfer (3*), a PEP block to recover lost signal (*4) and GARP decoupling during acquisition (*5).

A T_1 period (Figure 3.1.5, *2), then follows where the magnetisation is left to evolve under the chemical shift of the S spins, and during which I magnetisation is refocused in the x-y plane by a 180° x phase pulse. If a double labelled sample is under analysis, an optional shaped pulse (Figure

3.1.5, p14) can be used to remove the ^{13}C-^{15}N coupling during the T_1 period. A reverse INEPT transfer block (*3) then transfers magnetisation back to the coupled I spins for enhanced detection (owing to the larger γ). This is followed by a "preservation of equivalent pathways" (PEP, (129)) sequence (*4) which salvages I magnetisation lost during the T_1 period due to the inability of the reverse INEPT transfer to propagate x-plane magnetisation. Incorporation of the PEP block can recover signal by as much as a factor of $\sqrt{2}$, therefore further increasing the sensitivity of the experiment. The resulting signal is then measured during decoupling (*5), so as to ensure singlet peaks for the detected HN correlations, simplifying the resulting spectrum.

3.1.1.7 Triple resonance experiments and assignment

Although 1D and 2D experiments, such as the aforementioned HSQC, can be employed in the characterisation of macromolecular biochemistry, the process of describing individual moieties specifically requires a more complex set of experiments which can observe particular nuclei based upon multiple types of different through bond correlation.

In particular, the repeating, and predicable nature of protein chemistry can prove effective "logic gates" for correlating through bond relationships between the inherent connectivities of specific hydrogen, carbon and nitrogen nuclei. Of course, for magnetisation transfer to occur, the exploited nuclei are required to be of I spin = $1/2$ and unfortunately this is not the case for the naturally abundant isotopes of carbon and nitrogen. However specific types of chemical synthesis and protein expression methods exist for introducing the spin = $1/2$ isotopes of ^{13}C and ^{15}N; although this can be an expensive and time-consuming process.

Termed "triple resonance", these three-dimensional (3D) experiments then provide the capability to select for, and resolve in 3D spectral space, the peaks of specific moieties such as protein backbone protons, nitrogens or carbons or sidechain atoms such as those of aliphatic or aromatic nuclei. The use of the information obtained from these spectra can be used to confidently add annotations (assignments) to other spectra, such as the HSQC, for which residue specific information is not inherently available. For instance, for any given amino acid (except proline) under ideal conditions, the HN correlation of its amide moiety will be observed as a single peak in the ^{1}H-^{15}N HSCQ spectrum.

Triple resonance experiments, which also utilise scalar magnetisation transfer pathways via these amide correlations, can then be acquired wherein the resultant signals also posses components modulated by the resonant frequencies (chemical shifts) of these HN nuclei. These frequencies should therefore match the ^{1}1H and ^{15}N chemical shift components observable in the ^{1}H-^{15}N HSQC spectrum.

Assignment itself usually requires multiple triple resonance experiments to be carried out which can correlate one, or more nuclei to the initial backbone amide. Because an atoms environment heavily influences the value of their chemical shift, peaks of a certain frequency can be indicative of particular residues, and residue-type assignment can then be made. This is due to the characteristic chemical shifts of some nuclei in particular amino acids; namely, alanine (high field Cβ), glycine (Cα ~45 ppm, lack of Cβ), serine and threonine (low field Cβ). A full atomic assignment however requires identifying connectivities between associated triple resonance peaks. Figure 3.1.6 describes this process for the commonly acquired HNCACB/CBCA(CO)NH pair of experiments.

As described in Figure 3.1.6, assignment is made possible with pairs of experiments which use similar magnetisation transfer pathways but enable the distinction between the nuclei of a particular residue (i) and those of the residue sequentially preceding it (*i-1*). Sequential connectivities are therefore identified with a "backbone walk" by matching Cα and Cβ chemical shifts between directly observed residues and those of the corresponding preceding residue within their own directly detected HN plane. Although a full resonance assignment is not always necessary (i.e. characterisation of all the observable nuclei) the assignment process is rarely trivial and to support this backbone walk methodology, other experiments can also be used which can characterise the side chain protons (H(CCO)NH), the CO chemical group (HNCO) or the amide ^{15}N (HNN) in a similar fashion between sequentially linked residues.

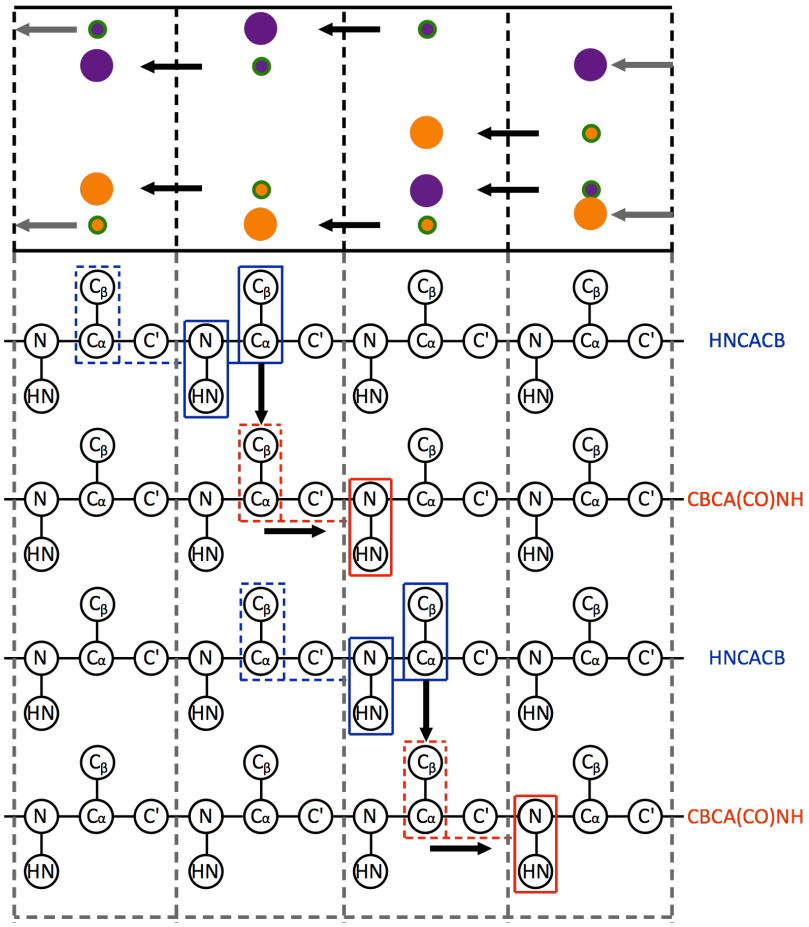

Figure 3.1.6: A schematic of the backbone walk methodology for NMR peak assignment (bottom) and typical carbon peak pattern connectivities in the relative HN plane of the directly observed residue (*i*) (top). HNCACB spectra visualise the Cα and Cβ nuclei from the *i* perspective (blue lines) and weaker peaks of the preceding residue (*i-1*, dashed blue lines) as opposing phases (orange and purple respectively). The CBCA(CO)NH experiment observes both the Cα and Cβ nuclei as a single positive phase (green peaks) for the *i-1* residue (dashed red) and confirm the *i-1* classification. Adapted from (134).

3.1.2 Experimental aims

By using the principles of NMR spectroscopy as explained above, the aim of the following chapter will be to document the work undertaken in the determination of a solution state structure for the AVR3a11 effector domain. The acquisition of NOESY-based experiments should allow pertinent distance measurements to be incorporated into a water-refined ensemble and could provide functional insight, either alone or in conjunction with the previously determined crystal structure for a shorter AVR3a11 construct (49).

3.2 Results

3.2.1 Purification

To determine the solution structure of the AVR3a11 effector domain, an 89 residue construct, comprising the entire secreted effector domain and an additional N-terminal hexa-His tag, was overexpressed. The purification tag was cleaved prior to experimentation, yielding 70 residues native to the 72 residues originally postulated as the effector domain, plus two N-terminal residues (GP) remaining from the PreScission^{TM} protease digestion. The cleaved construct had previously been shown unamenable to crystalisation (49). The ^{13}C, ^{15}N double labelled sample for the AVR3a11 effector domain construct was prepared by Dr Laurence Boutemy at the John Innes Centre, Norwich. Purification yielded >5 mg of pure protein, enabling the preparation of an ideal, 1 mM protein, NMR sample.

3.2.2 NMR resonance assignment

3.2.2.1 Manual peak assignment

Manual peak assignment was carried out with the backbone walk methodology originally described by Kurt Wüthrich in 1983 (135) (Section 3.1.1.7) that exploits the characteristic chemical shifts of certain moieties and the sequential connectivities of protein systems to yield information regarding the specific type and sequence position of the observed correlations.

In order to assign the chemical shifts of the AVR3a11 effector domain backbone nuclei, a 2D ^{1}H-^{15}N HSQC spectrum and 3D CBCA(CO)NH, CBCANH and CC(CO)NH TOCSY spectra were acquired. Aliphatic sidechain protons were assigned using ^{15}N-TOCSY and H(CCO)NH TOCSY spectra whilst aromatic sidechain protons required separate ^{13}C-TOCSY and TROSY HSQC experiments owing to their different chemical shift ranges (25 ppm - 40 ppm and 100 ppm - 140 ppm respectively, compared to the aliphatic ^{1}H shifts of -1 ppm - 8 ppm. Using this method, 77% of the amide residues (54 of 72, Figure 3.2.1), 69% of the remaining backbone nuclei and 71% of the non-labile protons could be assigned. A full resonance assignment for the full length

AVR3a11 construct is available at the BMRB (`http://www.bmrb.wisc.edu`), accession number 18910.

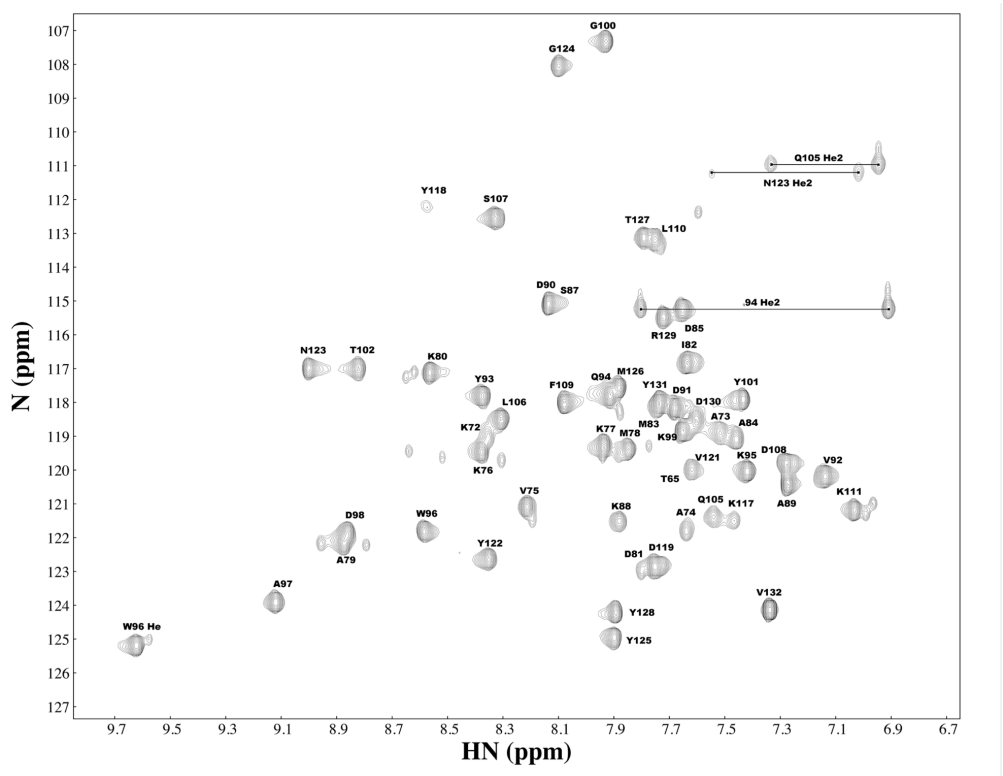

Figure 3.2.1: The assigned ^{1}H-^{15}N HSQC spectrum, acquired at 800 MHz for the full length construct of the AVR3a11 effector domain. Black lines indicate amide correlations with a shared nitrogen nucleus, as found in the NH_2 groups of amino acid sidechains.

Of the 72 residues within the AVR3a11 construct, only two (Pro2 and Pro26) were fully expected to be unobservable within the ^{1}H-^{15}N HSQC spectrum. All visible amide peaks were assigned except 15, however these were all were found to be minor alternate conformations of other assigned residues. This was confirmed by identical chemical shifts and peak patterns to those of other residues with respect to their *i* and *i-1* Cα and Cβ resonances.

3.2.2.2 Automated NOESY peak assignment

In order to describe the spatial restraints for NMR structure determination, the peaks present in the triple resonance nuclear Overhauser effect experiments (NOESY) must also be assigned. In principle, this can be accomplished through the sequential-walk methodology described above, as assigned root peaks (H-N or H-C scalar correlations) will show a correlation in the f_3 di-

mension to spatially close protons, which should then be identifiable from their specific chemical shifts. There are many examples of manual NOESY crosspeak assignments in the literature, however ^{13}C and ^{15}N edited HSQC-NOESY spectra can be notoriously crowded, as in theory, all protons within 5.5-6 Å should give rise to observable signals.

For this reason the UNIO automated assignment tool (90, 91) was used which, when given a list of direct resonance assignments and a set of referenced, processed HSQC-NOESY spectra, can objectively distinguish and assign the observable f_3 peak dimensions (and hence the proximal nuclei). A particular strength of the UNIO package is that it optimises the assignment list whilst performing a preliminary structure calculation for the protein in question. The assignments (and subsequently the structure) are then iteratively refined based upon which assignments are in fact possible from the current structural ensemble. It has been shown (88, 89) that the ATNOS automated peak picking algorithm is most reliable when >90% of the available non-labile proton chemical shifts have been assigned. However, depending on whether or not the assignments will give rise to meaningful NOE crosspeaks, this figure can be lowered to 70-75%. Although only 71% of the overall non-labile protons were assigned for AVR3a11 the impact upon the final structural ensemble will depend on the significance of the ultimately retained NOESY assignments. The original peak assignments were also used to generate a list of backbone dihedral angles from the TALOS plus web server (95). This generated 98 φ and ψ backbone angles found to be in good agreement with those in the TALOS+ angle database and were included as optional set of initial restraints with the UNIO calculation.

Using the manual assignments as input data, the ATNOS algorithm was able to identify a total of 2746 NOESY crosspeaks, of which an initial list of 345 and 1227 NOE crosspeaks were assigned from the ^{15}N and ^{13}C NOESY-HSQCs respectively. Upper limit distances were then calculated and combined if more than one distance existed for the same atom pair to generate a list of 879 initial distance restraints.

3.2.3 Structure determination

3.2.3.1 The initial AVR3a11 ensemble

After seven iterative cycles, the UNIO calculation resulted in a 20 model structural ensemble, restrained by 782 distance and 98 φ and ψ dihedral angular restraints (Figure 3.2.2). Common to the crystal structure of the shorter AVR3a11 construct (PDB code: 3ZR8 (49), Figure 1.2.3), the ensemble adopts a four helix bundle conformation with two pairs of anti-parallel alpha helices. The ensemble root mean squared deviation (RMSD) average distance violation was 0.0154 Å and 0.1764° for the final distance and angular restraints. All residues occupied "most favoured" (92.4%) or "allowed" (7.6%) Ramachandran backbone dihedral angles. The average deviation between the structural ensemble models was 2.2 Å across the protein backbone and although large deviations across an ensemble can be a reflection of poor restraint characterisation, it can also be seen as a measure of structural heterogeneity, as the data acquired fundamentally relates to an ensemble of molecules in solution.

Both facets to this structural deviation can arguably be seen for the AVR3a11 ensemble. On average there are 10.86 spatial restraints per residue and the four helices in particular are well represented (Figure 3.2.3). This is paralleled by a relatively small deviation in the atomic coordinates of just 1.1 Å RMSD for the backbone atoms of the helical regions. The N-terminal seven residues and the helix 3 to helix 4 loop however, with no spatial restraints, have larger ensemble deviations of 3.1 Å and 2.8 Å respectively. These regions can be said to have had their geometry determined solely by the tethering of their termini to the respective helices and the restrictions of ideal chemical bonding and geometry enforced during the calculation.

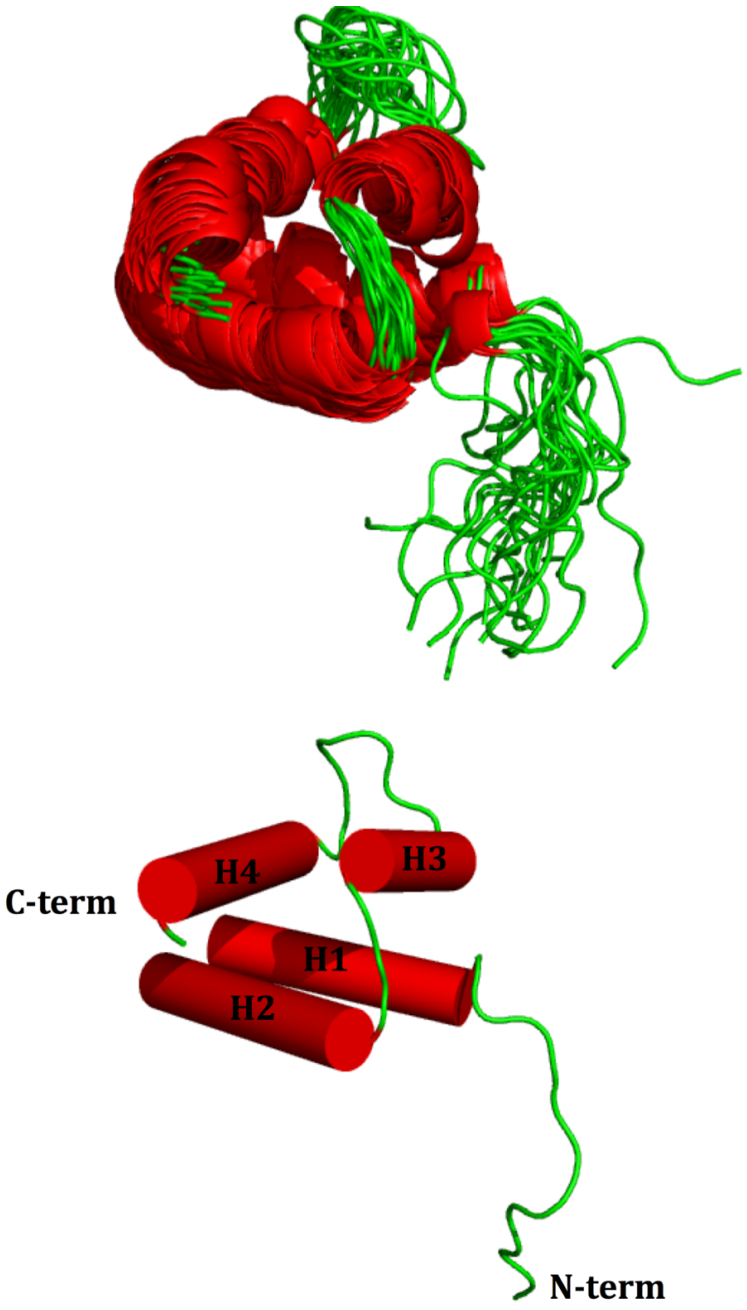

Figure 3.2.2: The Initial 20 model NMR ensemble for AVR3a11 with α-helices shown in red and loop regions in green (top). A cylindrical helix model is shown for clarity representing the overall average fold of the unrefined AVR3a11 NMR ensemble (bottom). Images were generated in the PyMOL molecular graphics program (50).

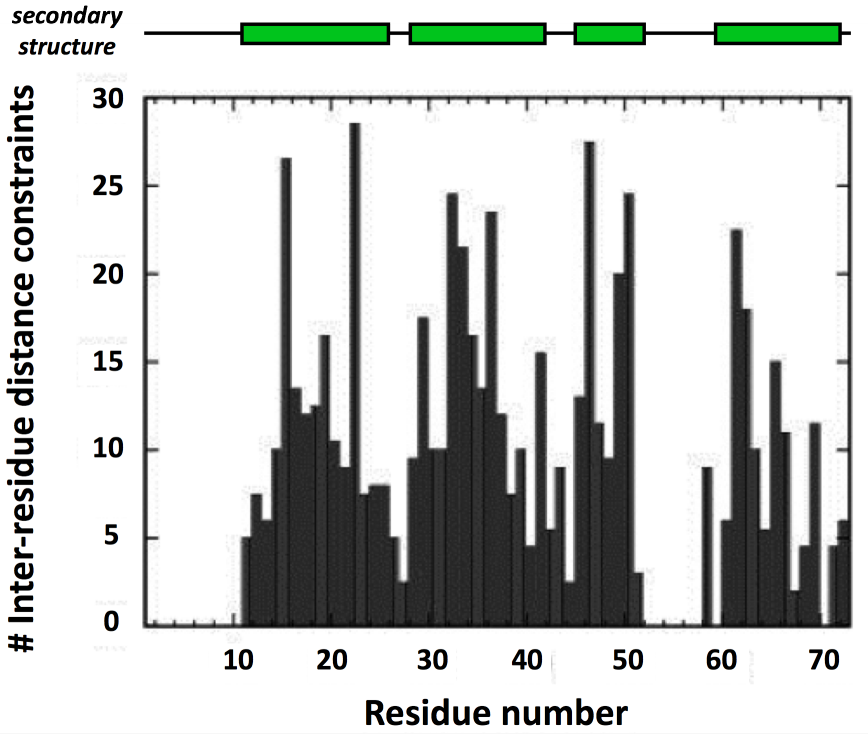

Figure 3.2.3: The number of NOE restraints per AVR3a11 residue. Residues K51 to K57 define the unassignable helix 3 to helix 4 loop. Plot generated from the output of the Protein structure validation suite (PSVS) (101).

The quality of a protein structure can be measured in many ways, and with respect to the overall quality of the UNIO calculation, there are conditions which should be satisfied in order to state whether a structure is of acceptable quality (88, 89). These are:

1. The initial cycle 1 average Target Function is less than 250 $Å^2$
2. The initial cycle 1 average RMSD is less than 3 Å
3. The RMSD between the first and last mean structures is less than 3 Å
4. The final cycle 7 Target function is less than 10 $Å^2$
5. The final ensemble should have less than 20% discarded long range NOEs
6. The final ensemble should have less than 20% unassigned NOEs

If the first three conditions are not met, due to the nature of the calculation itself whereby the result of one structural iteration is the input to the

next, it cannot be said that a poor initial model, and potentially inaccurate assignments, did not heavily influence the final structure. Condition 4 ensures the overall physicochemical quality of the final structure is not unrealistic. This is defined by the Target function score, an overall measure of quality (96, 97), which takes into account the number of distance and angular modelling restraints violated, the number of non-bound steric overlaps and the extent of any energy violations arising from NOE restraints, Van der Waals interactions or dihedral angles; a perfect score would be zero.

Conditions 1, 2 and 4 were satisfied during the AVR3a11 calculation (Table 3.2.1 & 3.2.2), and conditions 3 and 5 were met with a cycle 1:cycle 7 mean structure deviation of 1.824 Å and an overall value of 6.25% discarded long range NOE restraints. The 6th condition was not met, with 24.7% of all potential HSQC-NOESY peaks left unassigned.

Table 3.2.1: The target function and overall RMSD for each cycle's 20 highest scoring models. RMSD Drift is defined as the mean deviation from the cycle 7 ensemble.

Iteration	Target Function ($Å^2$)	RMSD (Å)	RMSD Drift (Å)	Constraints generated
1	191.169	2.351	6.070	879
2	72.478	1.067	6.303	1103
3	24.282	1.310	5.726	1022
4	6.913	1.229	5.565	969
5	4.236	1.716	1.836	958
6	0.817	1.988	1.238	870
7	0.544	2.151	0.000	782

The 6th condition however is intended to be an indirect measure of the signal to noise for a given NOESY spectrum (88), as it is assumed that the ATNOS algorithm within the UNIO package will be able to assign peaks for which root assignments were also available.

Albeit correct, this assumption has the potential to be influenced by unassignable initial residues that then correspond to visible HSQC-NOESY peaks. Since 29% of all non-labile protons could not be assigned, this is a realistic possibility. The observation that minor conformer peaks exist also suggests that this measure should not impact the quality of the UNIO ensemble alone, as these do correlate to NOESY peaks which would have been picked by ATNOS, but which would have been unassignable as a result of the

differences in the root chemical shifts.

Table 3.2.2: The calculated and refined NOE distances as determined by UNIO for the AVR3a11 structure calculation. Sequential NOEs are defined as i-j=1.

	Number of determined NOE distances				
Iterative cycle	Intra residue	Sequential	Medium range (≤4 Å)	Long range (≥5 Å)	Total
1	221	205	295	158	879
2	197	313	391	202	1103
3	191	266	357	208	1022
4	169	250	343	207	969
5	157	254	344	203	958
6	158	227	296	195	870
7	143	193	265	181	782

3.2.3.2 The refined NMR ensemble of AVR3a11

The methodology described above for simultaneous NOE peak assignment and structure determination is a powerful tool for identifying the tertiary fold of proteins and their molecular restraints. However, the advantageous speed of the calculation (~3 hours for a 10 kDa protein using a ~3 GHz, quad core based CPU) is achieved by simplifying the modelled degrees of freedom within a protein about their dihedral angles (torsion angle dynamics). Although it is more computationally time consuming, a more rigorous method of structural modelling however can be achieved by modelling the individual atoms and interatomic forces (Cartesian dynamics). Similarly, standard structural modelling programs do not take account of protein solvation or (as was the case for the initial UNIO modelling) apply a simplified global hydrophobic term. Therefore, a realistic depiction of a proteins structure, in terms of its biologically relevant solvated state ideally requires additional modelling to best reflect models from solution state data.

To this end, the penultimate (cycle 6), structurally guided, UNIO restraints list for the AVR3a11 construct was taken to generate a refined structural ensemble in explicit solvent with a simulated annealing procedure utilising Cartesian dynamics. The penultimate restraints list was chosen as it still contained restraint ambiguities and although the calculated upper bound limits remained the same as cycle 7, the inclusion of additional restraint possi-

bilities allowed additional freedom to the refinement calculation. As stated in previous work (136), structure minimised modelling restraints, such as those from UNIO, should in principle be relaxed by 0.2 Å to account for a structural bias in the restraints themselves incurred during the iterative calculations. Similarly, the lower bound limits should also be reduced to 0 Å (instead of the minimum proton:proton distance of 1.8 Å) to allow diastereotopic substituents to be modelled more precisely without the need for modelled pseudo atoms.

The refinement calculation of the AVR3a11 effector domain utilised the CNS 1.3 software suite (98) with the RECOORD (99) parameter set and CNS scripts. RECOORD parameters are a published set of CNS force constants (bonding, angular restrictions) and a tailored simulated annealing calculation (as described in Section 2.1.3.2) that have been devised for NMR based structure calculation by optimising the recalculation of over 500 PDB published NMR structures. 100 individual refined models were calculated for the AVR3a11 effector domain and the top 20 scoring models, as determined by the CNS total energy scores, with no restraint violations were compiled into a structural ensemble (Figure 3.2.4). Although it is important to describe NMR derived structures as ensembles to reflect the heterogeneity of the molecule, the exact number of models varies widely in the literature. A value of 20 was chosen as this is the most commonly submitted number in the Protein Data Bank (137).

The refined structural ensemble has a visible improvement in the deviation between the four -helices, with much improved RMSD values of 0.7 Å and1.4 Å for the backbone and heavy atoms respectively. Concomitantly, the effect of the solvated molecular dynamics run has also served to amplify the heterogeneity across the unstructured N-terminus and helix 3 to helix 4 loop. This should not be thought of as decreasing the significance of the model, but better reflecting the potential for conformational heterogeneity in those regions, given the lack of rigid secondary structure. These heterogeneous regions were also those unobservable within the NMR spectra acquired, and as dynamic phenomena occurring in the intermediate exchange regime can act to broaden peaks by averaging the signals from their many conformations, the resulting models can be seen as an accurate depiction of the solution state(s).

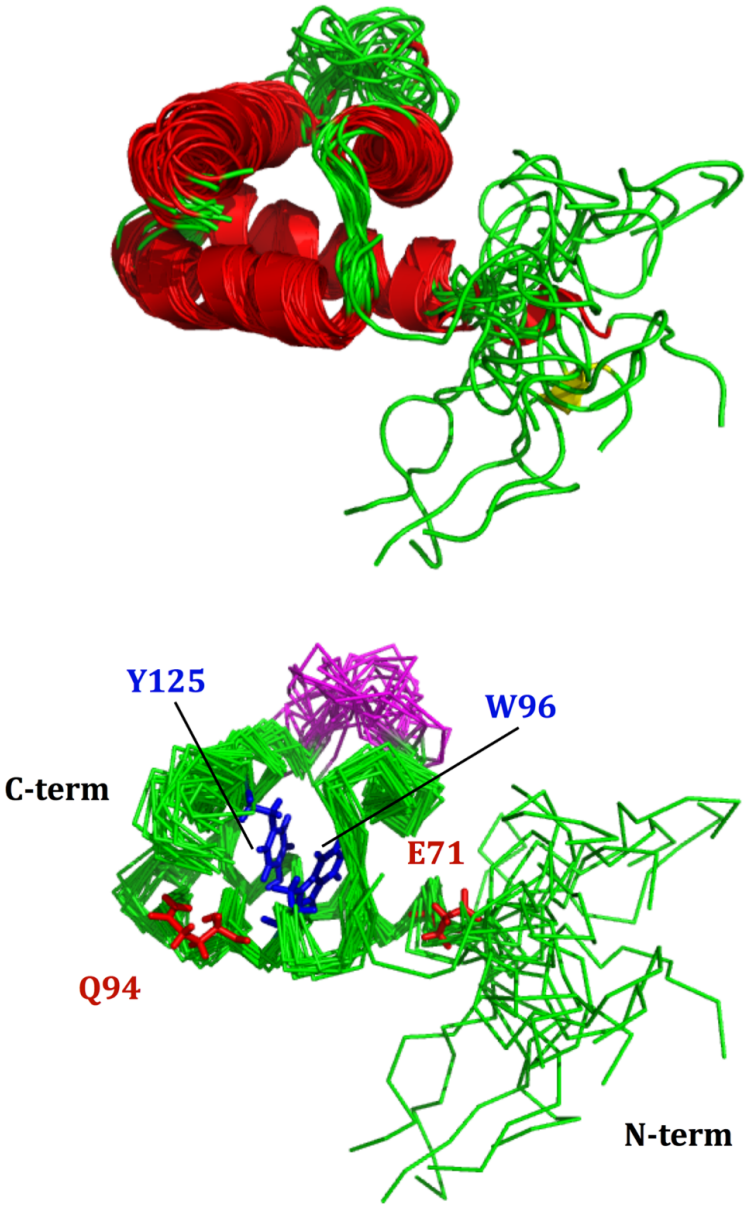

Figure 3.2.4: The final, 20 model, water refined ensemble for AVR3a11, with water molecules omitted for clarity (top). A ribbon representation of the refined 20 model NMR ensemble highlighting the functional residues (Q94 and E71, red) and residues comprising the WY motif (W96, Y125, blue). Images were generated in the PyMOL molecular graphics program (50).

The refined solution model can also be compared against that of the previously published crystal structure for the shorter AVR3a11 construct (PDB code: 3ZR8) (Figure 3.2.5) which is missing the N-terminal 7 residues. The mean refined AVR3a11 ensemble structure is in good agreement to that of 3ZR8 with the two aligning to within 1 Å RMSD across the protein backbone.

Although this is not alone sufficient to validate the accuracy of the generated model, it certainly strengthens the significance of the solution ensemble which, even with 71% assignment, can be seen to map extremely well to the crystal structure which was solved to 0.9 Å.

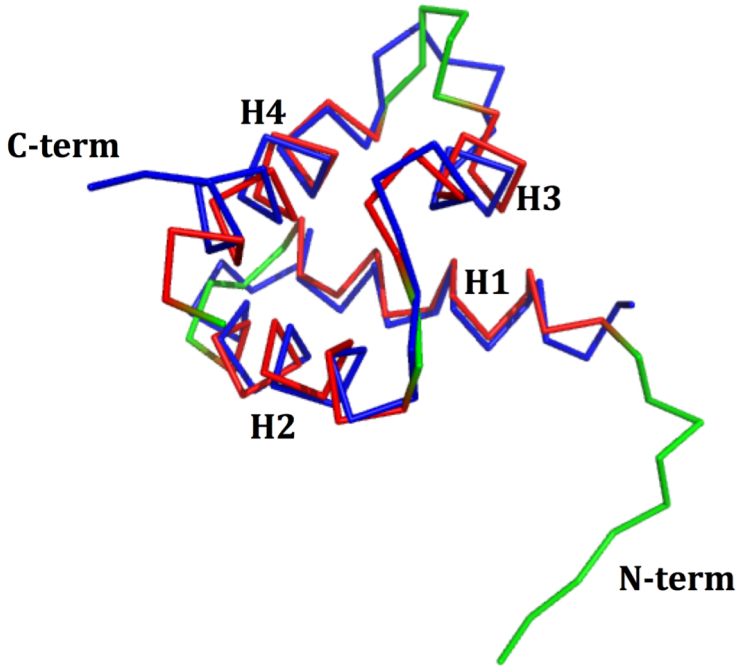

Figure 3.2.5: The average refined AVR3a11 Cα backbone trace (red helices/green loops) aligned against the shorter AVR3a11 construct crystal backbone trace (PDB code: 3ZR8 (49)) (blue). Image and alignment generated with the PyMOL molecular graphics program (50).

The level of conformational agreement can further be seen from the trace of the C atoms for all NMR models in Figure 3.2.6 as compared to the 3ZR8 crystal structure, and serves to highlight both the similarity between the helical conformations, the overall fold and particular the structural divergence across the unrestrained residues occupying the N-terminus and the helix 3-4 loop. Equally, the overall depiction of the heavy atom RMS deviations between the most representative NMR model (as determined by MolProbity) and the 3ZR8 structure (Figure 3.2.7, blue) displays the striking agreement between helices 2 and 3 (2-5 Å), the looser but still agreeable alignments of helices 1 and 4 (5-10 Å), and the relative divergence between the termini and the helix 1-2 and 3-4 loop regions (10-20 Å). Furthermore, all of these trends can be seen for the averaged, all-atom, RMS deviations per residue between the 20 models of the NMR ensemble (Figure 3.2.7, red). Although this seemingly depicts a poorer extent of agreement (5-15 Å across the helices), as the overall trend is

consistent (as compared to the most representative model:3ZR8 comparison), it can still be said that the conformational heterogeneity of the ensemble is depicting that of the 3ZR8 conformation.

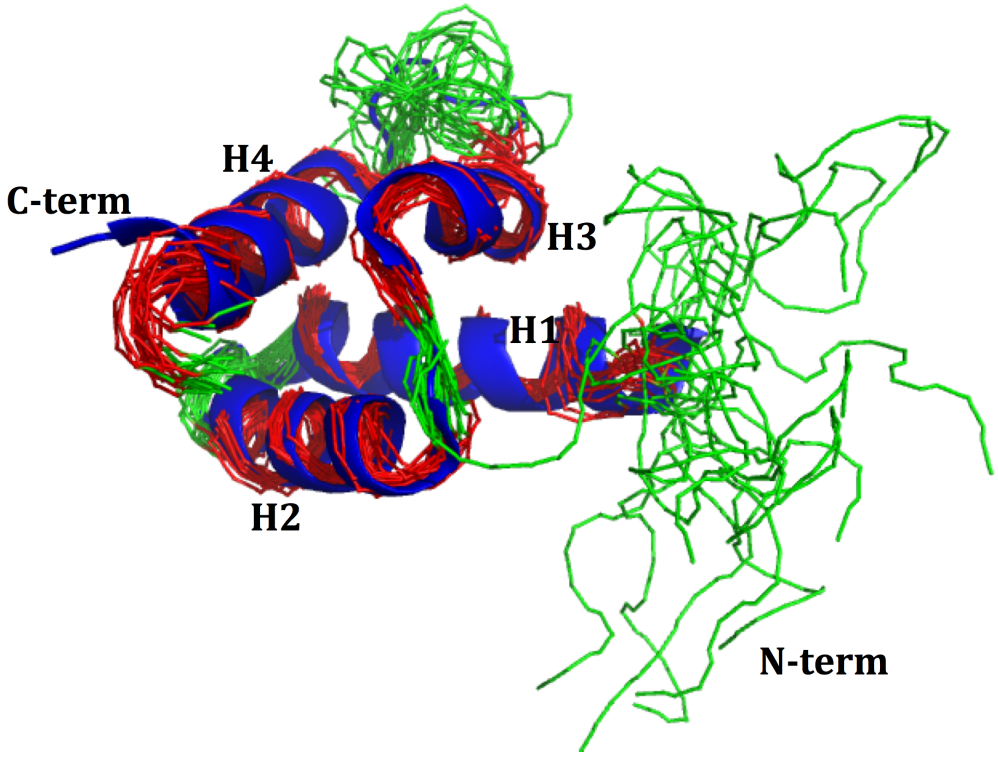

Figure 3.2.6: The Cα backbone traces of all NMR ensemble models (red/green) aligned with the 3ZR8 (49) crystal structure of AVR3a11ΔN (blue). Image and structural alignment generated with the PyMOL molecular graphics program (50).

Common characteristics of four helix bundle conformations (138) include a significant hydrophobic core amongst the helices and the presence of a number of salt bridges which act to stabilise the bundle by restraining the anti-parallel pairs of helices. For the refined AVR3a11 ensemble, a considerable area of the bundle interior is occupied by hydrophobic residues (Figure 3.2.8). However, no residues can be seen which would be capable of forming salt bridges, even at the maximum theoretical limit for such interactions of 4 Å (139). Other AVR homologues such as AVR3a4 (*Phytophthora capsici* [PDB code: 2lC2, currently unpublished] and AVR3a (*Phytophthora infestans* (49)) also share this four helix bundle fold containing a significant hydrophobic core whilst lacking the potential to form salt bridges.

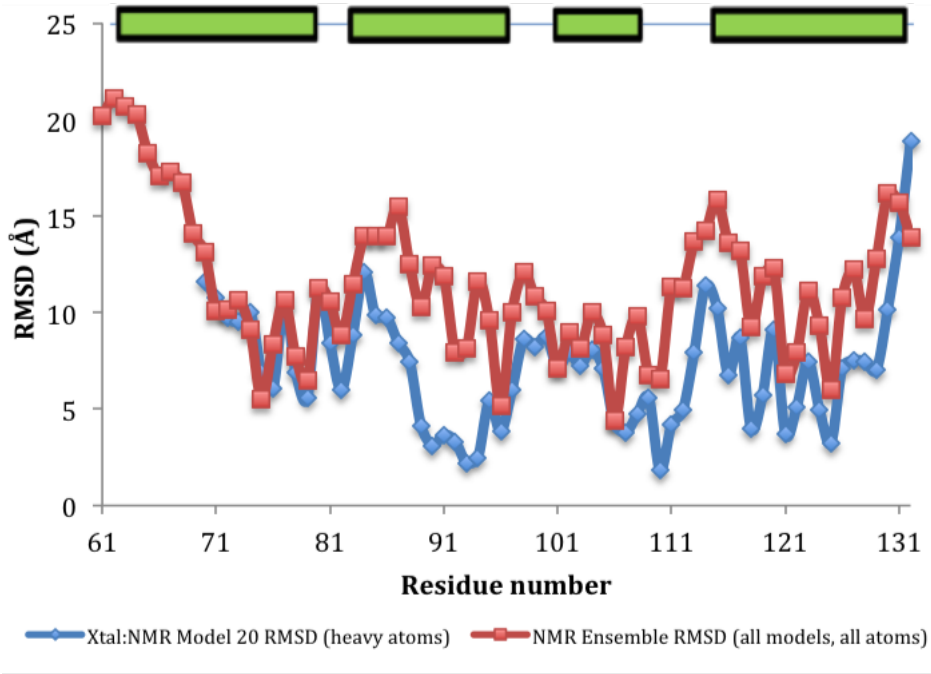

Figure 3.2.7: The average RMSDs per residue between the heavy atoms of the most representative NMR model (model 20) and the 3ZR8 (49) crystal structure (blue) compared to the average RMSD for all atoms across all models of the AVR3a11 NMR ensemble (red).

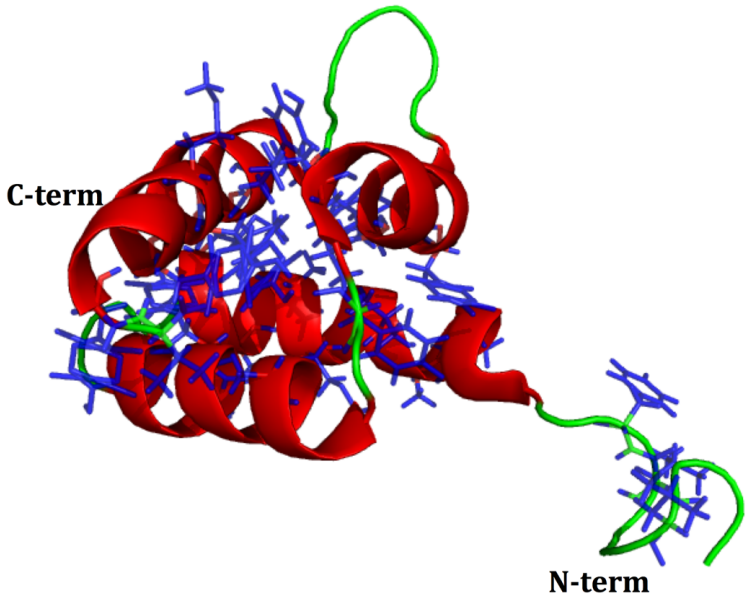

Figure 3.2.8: The average AVR3a11 ensemble structure highlighting the hydrophobic core within the four helix bundle. Blue regions belong to the hydrophobic sidechains of alanine, isoleucine, leucine, methionine, phenylalanine, tyrosine and valine residues.

3.2.3.3 Ensemble Validation

To fully describe the global quality of the refined AVR3a11 ensemble, the use of specific validation tools is required in order to compare the modelled bond lengths, bond angles and sidechain geometries against ideal values whilst also ensuring the final ensemble is still an accurate reflection of the initial data used to restrain it. The PSVS (101) and CING (100) validation suites, with the latter also including analyses from the PROCHECK (102) and WHATIF (103) analysis tools, are two such sets of tools which score NMR, X-ray or homology based models according to theoretically ideal bonding geometries, potential energy terms and comparisons to precise high resolution X-ray data.

The 20-model AVR3a11 ensembles from the unrefined final UNIO calculation and the water refined CNS calculation were analysed with the CING and PSVS suites and the outputs are summarised in Table 3.2.3. Generally, the refined ensemble scores can be seen to score better than that of the UNIO structure and both structures are good representations of the starting data, with only 5 restraint violations present within all models of initial ensemble, and zero restraint violations for the refined model. Of course, an absence of restraint violations was a determinant in choosing the top 20 models for the CNS ensemble, so this is not necessarily a self-consistent check.

The Ramachandran plot analysis (140) is a measure of assessing the physical validity of the φ and ψ backbone angles. By plotting the respective angles against one another for individual residues, the resulting plot describes the distinct zones of α helices and β sheet secondary structure elements. Both structures are can be seen to adopt highly idealistic backbone angles from the PSVS analysis. However, this approach has recently been expanded to include residue-specific plots to aid visual analysis, as the typical φ vs. ψ 'structured zones' for these residues are distinct from those of other residues (141). The Molprobity webserver (142) was used for this analysis (accessible: `http://molprobity.biochem.duke.edu/index.php`). For the AVR3a11 refined ensemble, 97.6% of all residues can be seen to adopt favoured or allowed backbone angles (Figure 3.2.9). Although this is less favourable that the PSVS score, this is still a respectable result and suggests that, in conjunction with the other validation scores, the significant majority of the ensemble backbone accurately depicts a valid solution state ensemble.

The PROCHECK and WHATIF tools score protein structures based upon a number of restraint-independent Z-scores, which depict the RMS deviation between various characteristics of the queried models to those obtained form high resolution X-ray structures. For WHATIF analysis, favourable scores are shown as positive values. Similarly, the PROCHECK scores weight the characteristic variance between the WHATIF scores and those of the ideal structures so that an ideal PROCHECK Z-score is equal to 1.0. Under or over- restrained characteristics are then defined as values greater than or less than 1.0, respectively.

Although there is clear improvement after the refined Cartesian modelling, both AVR3a11 ensembles have less than ideal rotamer geometries. Although scores between -3 and +3 are commonly deemed acceptable (143), a score of -3.903 $\pm$ 0.574 could signify a potential problem. However, significant time was spent manually investigating specific sidechain assignments, restraints and individual violations and no consistent errors were apparent, suggesting that this is simply a reflection of an average deviation from ideal crystalline geometry. The refined ensemble of the AVR3a11 effector domain has been deposited in the Protein Data Bank (137) (PDB code: 3ZGK) and the AVR3a11 chemical shifts have been deposited in the Biological Magnetic Resonance Bank (BMRB) (87) (accession number: 18910). Both datasets will be made freely available on the 18th of December 2013.

Table 3.2.3: The average validation scores for the AVR3a11 structural pre and post refinement in explicit solvent. The software/webtools used for specific validations are denoted with superscript terms. The WHATIF Z-scores are deemed favourable when positive and the PROCHECK Z-scores should ideally be close to 1.00.

	Unrefined	**Refined**
Average violations per model		
>0.2 Å NOE violations	2^{Cyana}	0^{CSN}
>5° degrees dihedral angles	3^{Cyana}	0^{CNS}
Backbone deviation (RMSD) **PSVS**		
All residues	2.2 Å	2.9 Å
Ordered	0.4 Å	0.7 Å
Ramachandran angle statistics **PSVS**		
Favoured regions	95.20%	94.40%
Allowed regions	4.80%	5.60%
WHATIF Z-Scores **iCING**		
1st generation packing	-0.216 ± 0.786	0.028 ± 0.628
2nd generation packing	0.602 ± 1.173	2.006 ± 1.514
Ramachandran plot appearance	-6.205 ± 0.432	-4.726 ± 0.613
χ_1/χ_2 rotamer normality	-7.643 ± 0.354	-3.903 ± 0.574
Backbone conformation	-0.648 ± 0.355	-1.961 ± 0.589
PROCHECK Z-Scores **iCING**		
Bond lengths	1.002 ± 0.0	1.111 ± 0.009
Bond angles	0.209 ± 0.0	0.540 ± 0.025
Omega angle restraints	0.014 ± 0.01	0.659 ± 0.084
Sidechain planarity	0.037 ± 0.002	1.085 ± 0.217
Improper dihedral distribution	0.253 ± 0.001	0.965 ± 0.057
Inside/Outside distribution	1.076 ± 0.029	1.126 ± 0.045

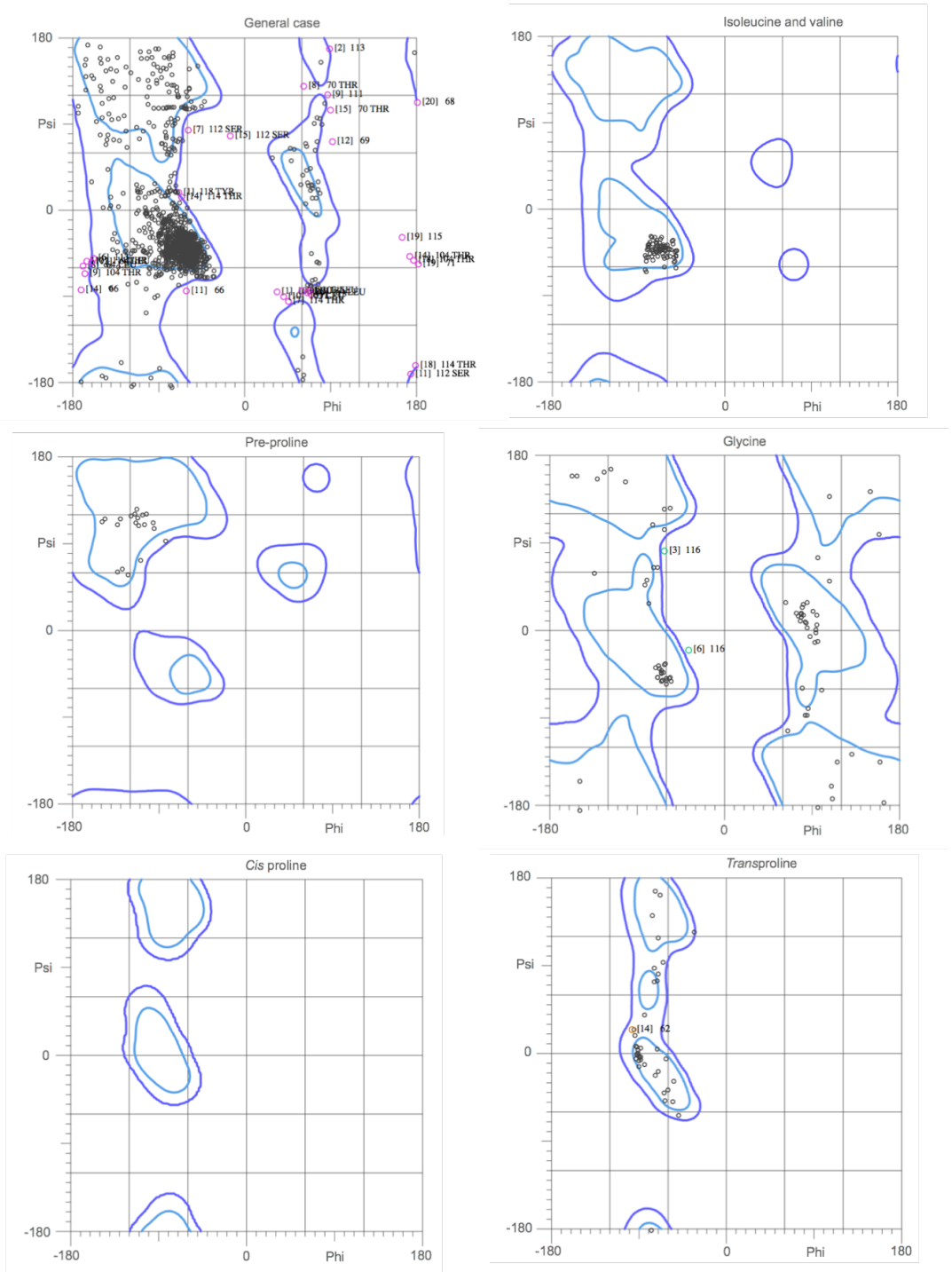

Figure 3.2.9: The Ramachandran plots representing the refined AVR3a11 20 model ensemble. The light blue contours indicate regions highly favoured whilst the dark blue corresponds to allowed regions. Residues adopting disallowed backbone angles are represented by magenta data points, with the relevant ensemble model noted in brackets. Generated with the MolProbity web tool (142).

3.3 Discussion

Using NMR data alone, a water refined, solution ensemble has been calculated for the AVR3a11 effector domain. Although an X-ray crystallography derived structure was accomplished and published before that described here (PDB entry: 3ZR8), detracting somewhat form the novelty of the NMR based structure, it does not reduce the overall significance of the NMR ensemble and, at the very least, both structures should be regarded as complementary evidence in future investigations of the AVR3a11 protein. Although the completeness of the NMR peak assignments was less than ideal, the automated restraint optimisation and initial calculation were shown to be of sufficient quality as defined by the guidelines set out by of the UNIO developers. The number of restraints generated per residue, the iterative improvement in the structures, and the reproduction of the overall fold, as compared the 3ZR8 crystal structure, also support the worth of the resulting models and the concomitantly guided structural restraints.

The lack of any observed peaks for the seven N-terminal residues, which in effect limits the restraint based model to the regions which comprised the truncated crystal structure construct also does not reduce the significance of the ensemble. Although these seven residues were unrestricted and therefore allowed to evolve freely during the simulated annealing calculations, the final model confirms that the main tertiary fold of the structure remains unchanged in the presence of the N-terminal residues and as such adds extra credibility to the X-ray crystallography based structure.

The water refinement procedure undertaken also generated a generally well scoring structural ensemble. Although ideal rotamer geometry was not obtained, the validation scores are in line with those seen in other NMR derived structures within the PDB (144) and our laboratory (143) (data not shown). One caveat to this type of comparative validation is the inherent differences between crystalline characteristics and those determined from either solution state data or molecular dynamics. Although the accuracy of X-ray crystallography based structures is not being called into question, the potential for a broader range of dynamic, greater than ground state, configurations to exist in solution implies that non-crystalline molecular characteristics should not necessarily be absolute measures of poor structural quality.

In terms of biological significance, identification and characterisation of the wild type alleles of RXLR AVR effectors have shown that the residues that map to positions 71 and 94 within AVR3a11 can influence effector recognition by host cell resistance proteins, (Section 1.2.3) (49). The construct used in this study represents the AVR3a11 allele shown to evade the host immune response and the NMR ensemble clearly describes a four helix bundle conformation with a similar positioning of the residues E71 and Q94 as seen in the 3ZR8 crystal structure (Figure 3.3.1). It has previously been hypothesised that the localisation of these residues to a single molecular face suggests they are involved in the binding interface to host resistance proteins. The NMR ensemble therefore also supports this hypothesis as the accessible conformational scope, as highlighted by the structural heterogeneity at residue positions 71 and 94, could facilitate such a mechanism.

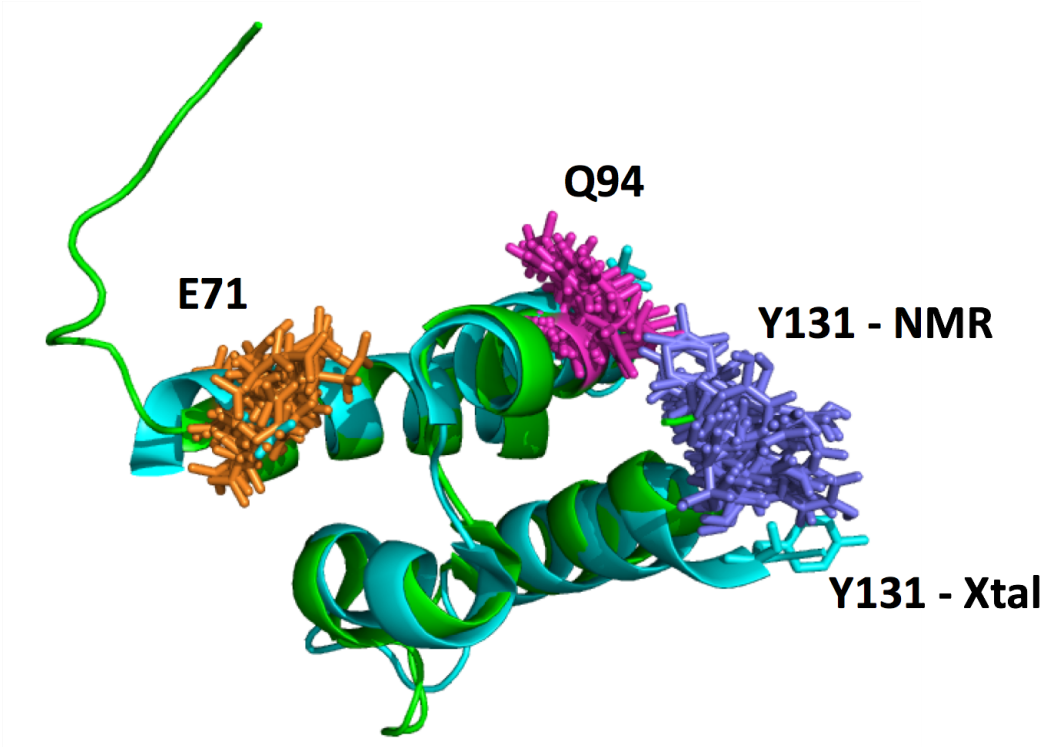

Figure 3.3.1: The aligned structure of the AVR3a11 effector domain as solved by X-ray crystallography (cyan, pdb code: 3ZR8) and the average structure as determined by NMR spectroscopy (green cartoon representation). The sidechains of the biologically relevant residue E71 (orange), Q94 (magenta) and Y131 (purple) are overlaid from all models.

A further residue position, Y131 - the last helical residue of helix 4 has also been shown to be important in determining inhibition of pathogen initiated

programmed cell death (PCD) through interactions between host cell proteins and the co-secreted oomycete protein INF1 (145). It is currently not known whether residue positions 71, 94 and 131 represent a single binding interface for both resistance protein recognition and INF1 inhibition, or if multiple binding sites exists with differing dependencies on these residue positions. However, an interesting distinction between the NMR ensemble and the X-ray structures can be seen at residue Y131, whereby the crystal structure seemingly deviates from the linear helical arrangement and adopts a more open conformation (Figure 3.3.1). The source of this discrepancy was identified by visualising the molecular symmetry of the 3ZR8 structures crystal packing (Figure 3.3.2).

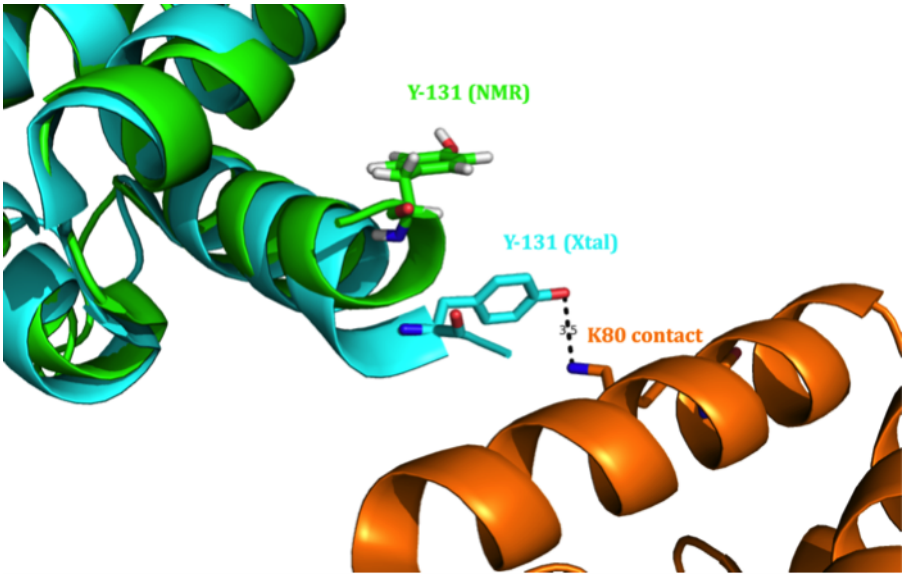

Figure 3.3.2: The aligned structure of the AVR3a11 effector domain as solved by X-ray crystallography (cyan, pdb code: 3ZR8) and the average structure as determined by NMR (green cartoon). The relative positions of the Y131 sidechains are shown and an identified crystal contact salt bridge between K80 and Y131 between the crystal structure and a symmetric partner (orange) has been highlighted at 3.5 Å.

It was determined that the Nζ terminal nitrogen of the K80 sidechain from an adjacent unit cell lies within 3.5 Å of the sidechain hydroxyl oxygen of Y131, implying that the crystal lattice is most likely distending the protein conformation. Although this is only a small difference; the change in position places Y131 on the same molecular face as that occupied by positions 71 and 94, and being sufficiently close to residue 94 so as to be involved in the same ligand interactions.

Chapter 4

Dynamic Characterisation of AVR3a11

4.1 Introduction

4.1.1 NMR relaxation

As touched upon in Section 3.1.1.4, the relaxation of excited spin states to thermal equilibrium is of fundamental importance to NMR, for without it no NMR signal would be observed due to the equalisation of spin state populations. In contrast to other spectroscopic methods, excited state nuclei do not spontaneously emit excess energy across a practical timeframe but dissipate it through discrete pathways via stimulations resulting from fluctuations in the locally experienced magnetic fields, creating an effective magnetic field B_{eff}.

The two concomitant pathways that facilitate the relaxation of NMR excited states back to equilibrium by energy transference to the surrounding physicochemical system are longitudinal relaxation and transverse relaxation. Using a Cartesian description of the bulk magnetisation vector, longitudinal relaxation describes the relaxation processes responsible for returning magnetisation to equilibrium along the z-axis (axis of the B_0 magnetic field). Transverse relaxation describes the loss of coherence in the x-y plane following excitation (Figure 4.1.1).

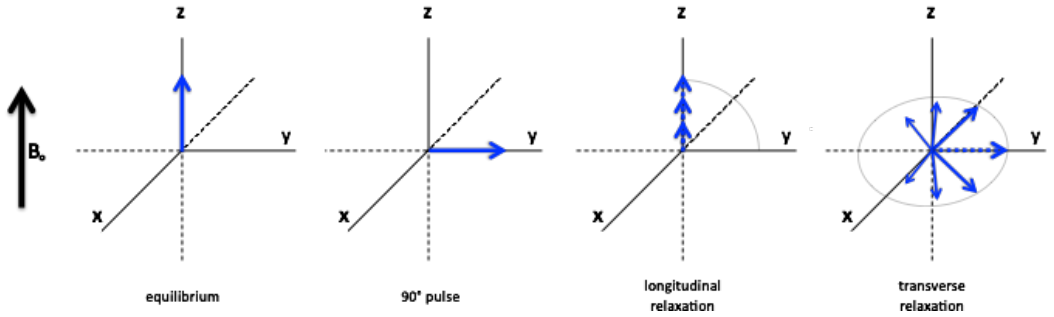

Figure 4.1.1: The general phenomena of longitudinal and transverse relaxation with respect to the bulk magnetisation vector (blue) and the z-axis of the applied magnetic field (B_0) after an example 90° pulse to y.

Longitudinal relaxation is also called spin-lattice relaxation due the nature of the underlying mechanisms of energy transfer. Although the term "lattice" was originally described for the relaxation of solid materials fixed within a molecular lattice, in its solution-state interpretation "lattice" refers to both bound and unbound neighbouring molecules within the sample (123). The complex vibrational and rotational motions of the surrounding lattice influence the B_{eff} experienced for any given spin. If these field oscillations occur perpendicular to the z-axis and at a frequency proportional to the energy difference between the ground (α) and excited (β) states, then the allowable quantised energy transfer leads to the enthalpic restoration of the equilibrium Boltzmann distribution of spin populations.

The associated mean lifetime for net longitudinal relaxation to occur is termed the T_1 time constant and can be described by the exponential equation 4.1:

$$M_z = M_0(1 - 2e^{-\tau/T_1}) \tag{4.1}$$

where M_z denotes the bulk z-axis magnetisation after a time τ for a given T_1 relaxation time and a initial ($\tau = 0$) starting z-axis magnetisation (M_0).

Transverse relaxation is also termed spin-spin relaxation because the dissipation of coherence occurs though an inter-spin flip-flop mechanism wherein the loss of energy of an excited state (β) to ground state (α) transition induces a reciprocal α-to-β transition for an additional spin.

The same fundamental mechanisms that cause longitudinal relaxation may also induce transverse relaxation, with the addition of longitudinal field effects from both inhomogeneities in the B_0 field and those arising from molecular motions within the sample. However the resulting pathway relates to processes that cause a loss of magnetisation coherence following a pulse. In this sense, transverse relaxation is not an enthalpic process, as it does not disturb the energetic populations, but an entropic decay wherein the magnetisation of excited stated diffuses to full isotropy within the x-y plane.

The apparent transverse relaxation can also be described by an exponential time constant (T_2) (equation 4.2):

$$M_{xy} = M_{xy_0} e^{-\tau/T_2^*}) \tag{4.2}$$

where M_{xy} is the resulting net magnetisation in the x-y plane after a time τ for a given T_2^* relaxation time constant and an initial x-y magnetisation (M_{xy0}). However, T_2^* represents the mean lifetime associated to effects from both external field inhomogeneities and molecular motions. To report solely on the extent of transverse relaxation accountable to the inherent molecular motions (T_2) the external field inhomogeneities can be compensated for with a spin echo pulse (146) (Figure 4.1.3).

Assuming that the present B_0 field inhomogeneities remain static, the acquisition of experiments with multiple delay times (τ) can be used to observe the loss of coherence as a function of the refocusing pulse, thereby negating the field effects. The same effect is exploited in the investigation of heteronuclear relaxation phenomena with a CPMG sequence.

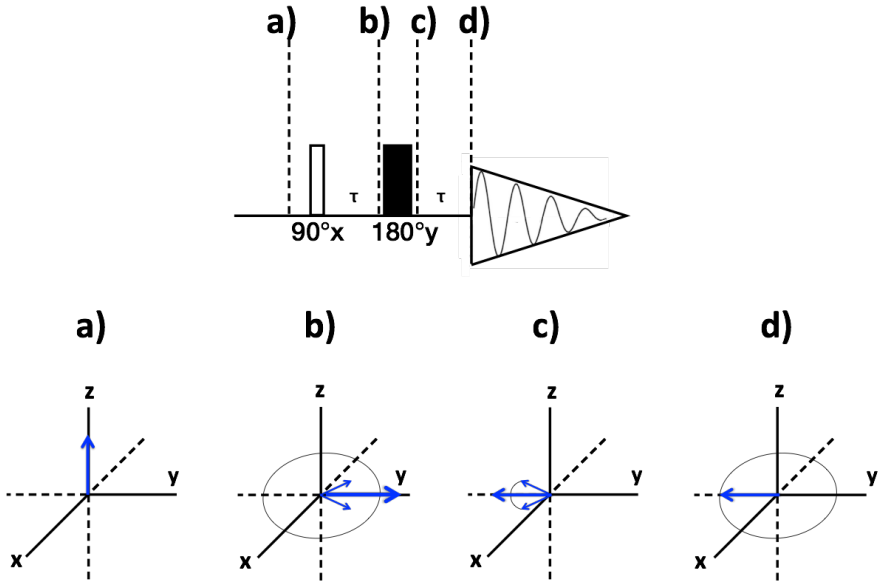

Figure 4.1.2: The spin echo refocusing pulse (top) and its effect on the net magnetisation (bottom). Following the net magnetisation at equilibrium (a), a 90° pulse (hollow bar) of x phase reorients the net magnetisation along the y-axis. After a time, τ, field inhomogeneities act to dephase the coherent magnetisation, with some spins beginning to precess faster or slower than net (b). After the application of a 180° pulse (solid bar) of y phase, the resulting 180° flip of all magnetisation vectors to the y-axis (c) results in the eventual refocusing of the net magnetisation after a second period of τ (d).

4.1.2 Relaxation mechanisms

Alluded to in the above descriptions of the relaxation-stimulating B_{eff} effects, the molecular motions with frequencies proportional to the energy difference between the α and β energy states (Larmor frequency), or combinations of states, can facilitate relaxation through a number of distinct mechanisms.

4.1.2.1 Dipole-dipole interactions

Dipole-dipole interactions are the primary mechanism responsible for both longitudinal and transverse relaxation pathways. Every magnetic nucleus within a sample will possess a magnetic moment, and for the predominantly observed spin ½ nuclei, this will be dipolar in nature. The ability of a magnetic dipole to interact (couple) to that of another nucleus depends predominantly on the distance between them and the strength of the individual magnetic moments (γ) as described by equation 4.3:

$$D = \frac{\mu_0}{4\pi} \frac{\hbar\gamma_I\gamma_S}{r_{IS}^3} (3cos\theta - 1) \tag{4.3}$$

where D is the strength of the interaction (in units of rad.s^{-1}), μ_0 is the permeability of free space, $\hbar$ is the reduced Planck constant, γ_I and γ_S are the gyromagnetic ratios of the coupling nuclei, r is the distance between the two nuclei and ϑ is the angle between the respective dipoles. For this reason, directly bound protons (with a large γ) offer the greatest potential for dipole-dipole interactions and this is especially true for large biological molecules with relatively slow correlation times.

4.1.2.2 Chemical shift anisotropy (CSA)

The particular chemical shift measured during an NMR experiment for individual types of nuclei will depend on the magnetic shielding available from their immediate local environment and the overall orientation within B_{eft}. However, we observe singular chemical shifts in solution state NMR for distinct species as a result of isotropic molecular tumbling averaging out these variations over the course of the experiment. Nonetheless, at the level of the individual spin, the anisotropic fluctuations in individual chemical shifts also present an accessible relaxation mechanism. Equation 4.4 (127) describes the summation of the average isotropic chemical shift tensor, σ:

$$\sigma = \frac{\sigma_{xx} + \sigma_{yy} + \sigma_{zz}}{3} \tag{4.4}$$

where xx, yy and zz refer to the respective spatial components of the chemical shift tensor, σ. The chemical shift (CSA) anisotropy term itself ($\Delta\sigma$) is defined by equation 4.5:

$$\Delta\sigma = \sigma_{xx} - \frac{\sigma_{yy} + \sigma_{zz}}{2} \tag{4.5}$$

4.1.2.3 Chemical exchange

The interchange of chemical species between specific magnetic environments during the acquisition of an NMR experiment is a complex phenomenon. Simply put, chemical exchange refers to the inter- and intra- molecular exchange of chemical species or conformations and the ability to distinguish the individual spin environments depends on how the rate of exchange (k) compares to the difference in precession at the two sites ($\Delta\omega$).

However, in a continually exchanging system, the difference in Larmor frequency between the two sites of exchange ($\Delta\omega = \omega_2$-ω_1) will cause an effective Larmor precession to be experienced (ω_{eff}) which will be dependent on the rate of exchange, k. Over time, the effective Larmor frequency will be seen to fluctuate about ω_{eff} by the half the difference between the two sites ($\Delta\omega/2$) giving rise to a true effective frequency of precession of $\omega_{eff} \pm \Delta\omega/2$. The longitudinal magnetic component of this oscillation of the Larmor frequency can then act as an effective transverse relaxation mechanism for surrounding nuclei. However, the frequency of oscillation is directly related to k and minimal effects will be seen when k << $\Delta\omega$ (slow exchange) or when k >> $\Delta\omega$ (fast exchange) and the greatest effect will be when k ω (intermediate exchange).

4.1.2.4 The heteronuclear Overhauser enhancement

The nuclear Overhauser effect, as first described in 1962 (147) essentially describes the rate of a dipole-dipole internuclear relaxation process (σ_{IS}) whereby a β state spin is able to relax by inducing the simultaneous excitation of an α state spin (Figure 4.1.3).

When discussing the interactions between the spin states of two nuclei, the possible transitions fall into three classes. Single quantum transitions (W_1) represent the change of one spin state at any one time (i.e. a single α to β transition without perturbing the neighbouring nuclei) whereas double quantum transitions (W_2) and zero quantum transitions (W_0) occur via through space cross-relaxations with the spin states of neighbouring nuclei.

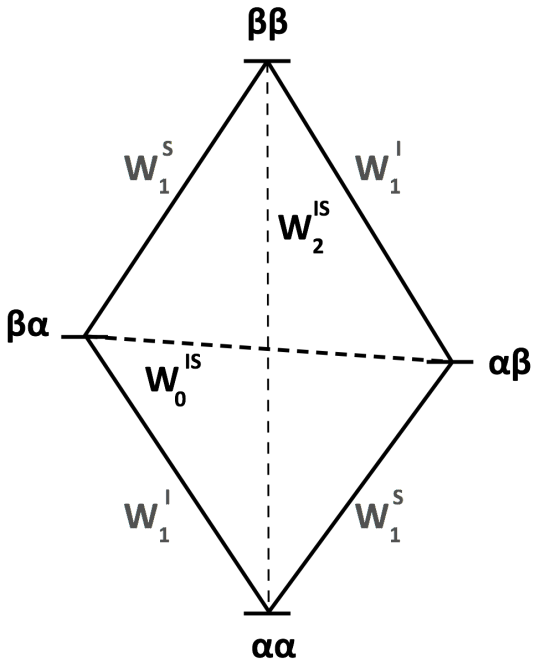

Figure 4.1.3: An energy level diagram, for two heteronuclei (I and S) and the spin state transitions (w_0, w_2) accessible to the nuclear Overhauser effect.

The probability of either a W_0 or W_2 transition predominating depends on the whether the frequencies of the experienced molecular motions correspond to either of the transition state energies. In small molecules, W_2 transitions prevail due to rapid molecular tumbling; whereas macromolecular tumbling typically promotes W_0 transitions. If a particular set of nuclear spins (S) are saturated, (i.e. equal spin state populations), W_0 and W_2 zero and double quantum transitions result in a measurable change (with respect to the net effect) in I spin magnetisation. The nuclear Overhauser enhancement (η_{IS}) is then said to be either positive (augmented M_I, W_2 dominates) or negative (attenuated M_I, W_0 dominates). Although the nuclear Overhauser effect is related to the distance between two spins, the enhancement alone, η_{IS}, is not directly proportional to the inter-spin distance, and only qualitatively reports on whether the spins were close enough to magnetically couple. Nonetheless, an observation of η_{IS} can enable the characterisation of the NOE as a relaxation mechanism and therefore yield information on the magnitude of the molecular motions proportional to the zero quantum of double quantum transitions.

4.1.2.5 Paramagnetic relaxation

Paramagnetic relaxation is the magnetic dipolar coupling between a nuclear magnetic moment and that of an unpaired electron. This effect is not seen for electron pairs as the antiparallel spin arrangements of the constituent electrons reduce the effective magnetic moment to zero. The magnetic moment of an unpaired electron is approximately 2000x greater than that of common nuclei and so if free radicals are present in solution, they can readily provide a form of dipole-dipole relaxation that dominates the other aforementioned mechanisms. As an inherently dipolar interaction, the extent of paramagnetic relaxation is dependent on the distance between the coupling species but owing to the significantly larger magnetic moment, paramagnetic relation can act over larger distances than inter-nuclear dipolar coupling. This effect is exploited in the structural investigations of proteins by measuring the paramagnetic relaxation enhancement (PRE) that arises from cognate metal ions or artificially bound paramagnetic labels. As a relaxation enhancement, the PRE effect acts to reduced the signal intensity of proximal spins and when explicitly measured can provide inter-atomic distance information through equation 4.6:

$$r = \left[\frac{{}^{1}/_{15} S(S+1)\gamma^2 g^2 \beta^2}{R_2^{SP}} \left(4\tau_c \frac{3\tau_c}{1+\omega_H^2 \tau_c^2} \right) \right]^{1/6} \tag{4.6}$$

where r is the nuclear-electron distance, S is the electron spin quantum number, γ is the nuclear gyromagnetic ratio, g is the electron g-factor, β is the Bohr magneton, $R_2{}^{SP}$ is the enhanced paramagnetic transverse relaxation rate, τ_c, is the correlation time for the electron-nuclear interaction and ωH is the nuclear (proton) Larmor frequency.

4.1.2.6 Other relaxation mechanisms

Three other mechanisms can facilitate the relaxation of excited, coherent spin states, although in biological solution state NMR, they are rarely deemed significant.

Spin rotation relaxation applies to small molecules and, in rare cases, the freely moving portions of larger molecules without other accessible relaxation mechanisms. Due to the rapid tumbling of small molecules in solution, the motion of the associated electrons acts to create a local magnetic field. Molecular collisions then cause oscillations in these local fields, and if the frequency is proportional to the Larmor frequency, relaxation can occur.

The mechanisms discussed so far directly apply to commonly observed spin half ($I = 1/2$) nuclei that have a dipolar magnetic moment. However if $I \leq 1/2$, and a significantly asymmetric nuclear charge distribution exists then coupling can occur between the surrounding electrons and the quadrupolar nuclear electric field. This can then lead to accessible relaxation if molecular motions perturb the quadrupolar coupling at a rate comparable to the Larmor frequency.

Scalar relaxation is a consequence of the scalar couplings (J-couplings) between covalently attached nuclei, mediated by the individual nuclear hyperfine couplings to the bond electrons. Scalar couplings are of course exploited in the process of magnetisation transfer between attached spins and are thought of as orientation independent, with respect to B_0. However, although the effect averages en mass, at the level of the individual spin, they are not independent of internal molecular motions of bond vibrations and rotations. Again, therefore, if these internal motions act to modulate the effective inter-particle couplings on at timescale proportional to Larmor frequency of proximal spins, an energetic pathway is said to be accessible for the relaxation of excited nuclear spin states.

4.1.3 The study of protein dynamics by NMR relaxation

The above descriptions of NMR relaxation mechanisms should highlight the many ways in which molecular motions enable the restoration of equilibrium spin populations and isotropic magnetisation for an ensemble of excited spin states. Fortunately for NMR spectroscopists, the reciprocal statement also holds true and from the measurement and analysis of relaxation times, information on the scope, magnitude and type of molecular motions accessible to a sample can be revealed.

The broad timescales across which different NMR techniques report upon can provide atomic level insights into the conformational behaviour and biological function of proteins in particular (Figure 4.1.4).

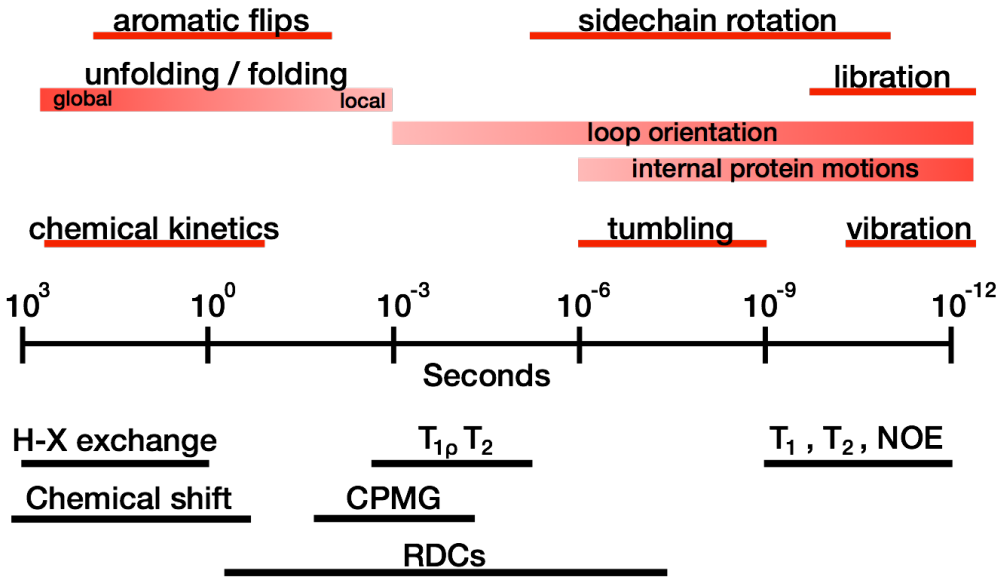

Figure 4.1.4: The timescales associated with protein dynamic phenomena (red bars) and the NMR techniques used for their detection (black bars). CPMG and RDC refer to Carr, Purcell, Meiboom, Gill relaxation dispersion and residual dipolar couplings respectively. T_1, $T_{1\rho}$, T_2 and NOE refer to longitudinal, longitudinal (rotating frame), transverse relaxation and nuclear Overhauser enhancement, respectively.

A fundamental caveat to the investigation of proteins however is one of molecular size. The correlation time (τ_c) is defined as the average time taken for a molecule to rotate through one radian. This is of course affected by the viscosity and temperature of solution, but is principally a factor of molecular size and shape. From Figure 4.1.5, it can be seen that as the correlation time increases, the transverse relaxation time, T_2, becomes faster. Equation 4.7 describes how the linewidth ($\nu 1/2$) of Fourier transformed peaks is proportional to the T_2 time.

$$\Delta\nu_{1/2} = \frac{1}{\pi T_2} \tag{4.7}$$

NMR experiments of large molecules (>30 kDa) therefore result in broad spectral lines, and combined with the fact that larger molecules generally yield more complex spectra, this quickly limits the effective size of the molecules that

can be studied with NMR although experimental techniques such as transverse relaxation optimised spectroscopy (TROSY) and spin $^1/_2$ labelling schemes do exist which can which simplify the spectra.

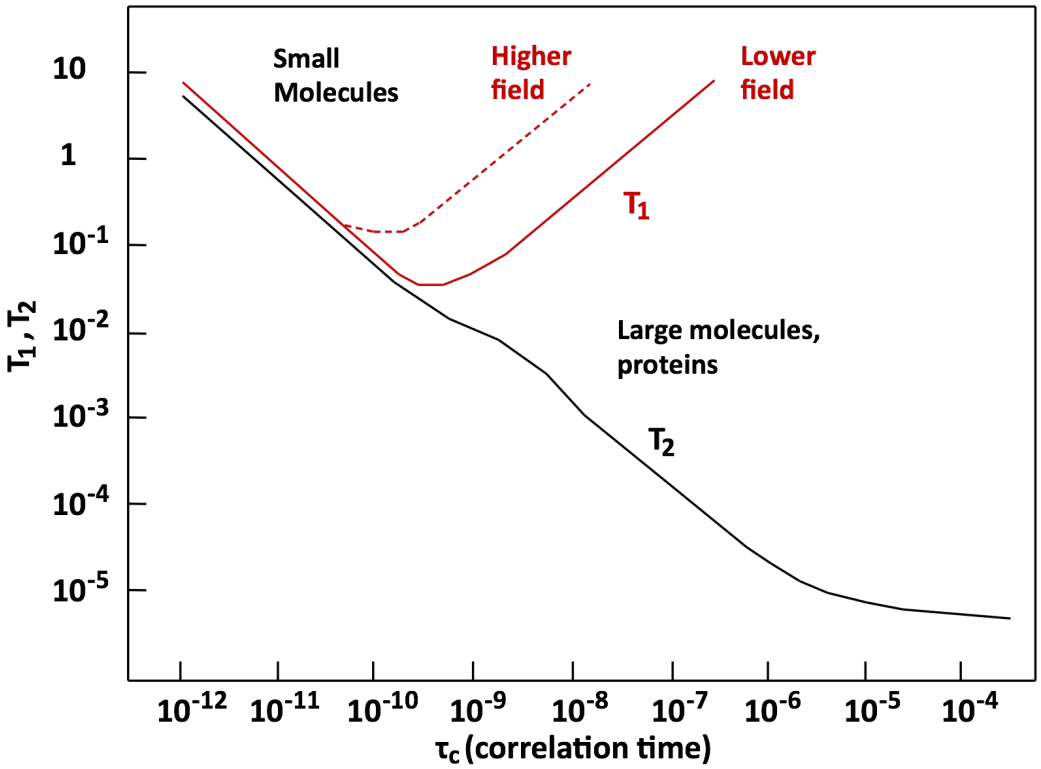

Figure 4.1.5: The general relationships between T_1 and T_2 relaxation times as a function of the applied B_0 magnetic field and the molecular correlation time (τ_c).

Although one-dimensional ^{1}H relaxation is occasionally used in the study of proteins, even without the complications from T_2 broadening, significant spectral overlap exists. For this reason, the relaxation of heteronuclear spins is frequently used to study protein dynamics due to the inherent improvement in resolution of 2-dimensional acquisition. Modifications of the ^{1}H-^{15}N HSQC pulse sequence (section 3.1.1.6) that enable the measurement of longitudinal, transverse and cross-relaxation rates are particularly advantageous for proteins as they are able to effectively probe the scope of relaxation across the ps-ns and μs-ms timescales for each residue in the protein backbone (excluding proline).

The potential for ^{15}N relaxation to report on dynamic phenomena is encapsulated in equations 4.8 - 4.12 (148):

$$\frac{1}{T_1} = \frac{d_2}{4}\left[J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)\right] + c^2 J(\omega_N) \tag{4.8}$$

$$\begin{aligned}\frac{1}{T_2} = \frac{d^2}{8}\big[4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)\big] \\ + \frac{c^2}{6}\big[4J(0) + 3J(\omega_N)\big] + R_{ex}\end{aligned} \tag{4.9}$$

$$NOE = 1 + \frac{d^2}{4}\,\frac{\gamma_H}{\gamma_N}\big[6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)\big]\; T_1 \tag{4.10}$$

where J is the spectral density function evaluated at the frequencies ω_H-ω_N, ω_N, ω_H+ω_N, ω_H and 0, with ω indicating the Larmor frequency of the subscript nucleus,

$$d^2 = \left(\frac{\mu_0}{4\pi}\right)^2 \left(\frac{\gamma_H \gamma_N \hbar}{2r_{NH}^3}\right)^2 \tag{4.11}$$

$$c^2 = \frac{1}{3}\left(\omega_N \Delta\sigma\right)^2 \tag{4.12}$$

where μ_0 is the permeability of free space, $\hbar$ is the reduced Planck constant, r_{NH} is the heteronuclear bond distance, γ_H and γ_N are the gyromagnetic ratios of the superscript nuclei and $\Delta\sigma$ is the nitrogen chemical shift anisotropy. The terms c and d represent the contributions to relaxation from chemical shift anisotropy and dipole-dipole mechanisms respectively. R_{ex} is defined as the rate of any exchange process (section 4.1.2.3) that contributes to transverse relaxation. In the case of standard macromolecular ^{15}N relaxation, these are the only relaxation mechanisms presumed significant.

Interestingly, a noteworthy question exists surrounding the exact values chosen for $\Delta\sigma$ and r_{NH}. Standard practice assumes values of -172 ppm and 1.02 Å for the ^{15}N CSA and N-H bond length respectively. However experimentally determined deviations from standard values have been reported in the literature for both the N-H bond length (~1.04 Å (149)), and the ^{15}N CSA (-172 ppm ±13 (150)). Although it is not routine practice, due to the in-

creased significance of the 15N $\Delta\sigma$ term at high fields, changing this value can sometimes be used as a technique to improve the compatibility of calculated molecular motions.

4.1.4 Spectral density mapping

Spectral density mapping is thought of as the most direct extraction of motion characterisation from a set of relaxation data. By acquiring T_1, T_2, hNOE and two-spin relaxation experiments, the evaluation of the spectral density at five frequencies - proportional to those observed (i.e. J(0), J(ω_N), J(ω_H+ω_N), J(ω_H) and J(ω_H-ω_N)) can detail the magnitude of motions acting across those frequencies which contributed to relaxation. This method also benefits from being independent of the overall rotational motion (see equation 4.1.5) and requires no *a priori* assumptions regarding the diffusive models (151). For instance, for acquisition of a relaxation dataset at 500 MHz the evaluated frequencies of 0 MHz, 50 MHz, 550 MHz, 500 MHz, and 450 MHz would sample the extent of molecular motions acting across those reciprocal timescales.

As spectrometer time or access is rarely abundant, standard relaxation investigations are only typically concerned with the measurement of T_1, T_2 and hetNOE relaxation. This therefore leaves the methodology mathematically underdetermined (i.e. 3 measurements, with 5 equations) and without incurring assumptions regarding J(ω), the methodology falls down.

However, a simplified form of the above, termed reduced spectral density mapping (152), only requires three measurements to be made and makes just one assumption, that J(ω_H) J(ω_H+ω_N) J(ω_H-ω_N) = J(0.87ω_H). By rearranging equations 4.8-10, the interrogation of the spectral density function at the three frequencies J0.87(ωH), J(ωN) and J(0) can be directly related to the relaxation measurements by Equations 4.13-4.15:

$$J(0.87\omega_H) = \frac{4\sigma}{5d^2} \tag{4.13}$$

$$J(\omega_N) = \frac{4R_1 - 5\sigma}{3d^2 + 4c^2} \tag{4.14}$$

$$J(0) = \frac{(6R_2 - 3R_1 - 2.72\sigma)}{3d^2 + 4c^2} \tag{4.15}$$

where:

$$\sigma = R_1(NOE - 1)\ \frac{\gamma_N}{\gamma_H} \tag{4.16}$$

and c and d are defined by Equations 4.11-12.

4.1.5 Model free analysis

As alluded to already, the mechanisms which underpin nuclear relaxation are dependent on the fluctuations of the local magnetic fields surrounding a given spin, which are in turn dependent on molecular motions. For ^{15}N relaxation, the N-H bond vector is a strong indicator of these motions and by assuming a fixed N-H bond length and a fixed motional scheme (or model) such as free conical diffusion, then the extent of specific motions can be estimated.

However, in assuming any one particular diffusive model for the N-H vector, the extraction of motional characteristics from experimental data is restricted to that model alone. The model-free approach, first described by Lipari and Szabo in 1982 (153, 154), does not make assumptions regarding the specific diffusive model and can therefore describe general characteristics applicable to different motional schemes. This approach relies on the ability to summarise individual contributions to relaxation through the deconvolution of the correlation function that describes the overall Brownian molecular diffusion.

In the case of isotropic diffusion, the spectral density of the correlation function J(ω) decays as a function of the isotropic correlation time (τ_m) and is calculable from equation 4.17:

$$J(\omega) = \frac{2}{5}\left(\frac{\tau_m}{1 + (\omega\tau_m)^2}\right) \tag{4.17}$$

which can then be employed in the characterisation of the generalised model free order parameter, S , with the following formula:

$$J(\omega) = \left(\frac{S^2 \tau_m}{1 + (\omega\tau_m)^2} + \frac{(1 - S^2)\tau}{1 + (\omega\tau)^2} \right) \tag{4.18}$$

whereby τ, the effective correlation time:

$$\frac{1}{\tau} = \frac{1}{\tau_m} + \frac{1}{\tau_e} \tag{4.19}$$

where τ_e is correlation time describing intramolecular motions which lead to N-H vector reorientation. The generalised order parameter S (commonly referred to as S^2) is a measure of the spatial restrictions of the N-H bond vector and has values between 1 (full restriction) and 0 (no restriction).The isotropic case of macromolecular diffusion however is usually an idealistic description, and in the case of anisotropic diffusion the overall diffusion cannot be encompassed within a single correlation time. Internal motions are independent of overall molecular diffusion (and so the term remains unchanged) but the model free formalism needs to include additional terms to describe the separate axial components (Equation 4.20, (155, 156)):

$$J(\omega) = S^2 \sum_{k=1,\ldots,5} A_k \left[\frac{\tau_k}{1 + (\omega\tau_k)^2} \right] + \frac{(1 - S^2)\tau}{1 + (\omega\tau)^2} \tag{4.20}$$

where,

$$\tau^{-1} = 6D + \tau_e^{-1} \tag{4.21}$$

where A and τ are defined as:

$$A_1 = 6m^2n^2 \qquad \tau_1 = (4D_x + D_y + D_z)^{-1}$$

$$A_2 = 6l^2n^2 \qquad \tau_2 = (4D_y + D_x + D_z)^{-1}$$

$$A_3 = 6l^2m^2 \qquad \tau_3 = (4D_z + D_x + D_y)^{-1}$$

$$A_4 = d - e \qquad \tau_4 = \left[6(D + (D^2 - L^2)^{1/2}\right]^{-1}$$

$$A_5 = d + e \qquad \tau_5 = \left[6(D - (D^2 - L^2)^{1/2}\right]^{-1}$$

with l, m and n representing the direction cosines of the NH vector with respect to the Cartesian axes x, y and z respectively.

D and L are defined as:

$$D = \frac{1}{3}\ (D_x + D_y + D_z) \tag{4.22}$$

$$L^2 = \frac{1}{3}\ (D_x D_y + D_x D_z + D_y D_z) \tag{4.23}$$

d and e are defined as:

$$d = \frac{\left[3(l^4 + m^4 + n^4) - 1\right]}{2} \tag{4.24}$$

$$e = \frac{\delta_x(3l^4 + 6m^2n^2 - 1) + \delta_y(3m^4 + 6l^2n^2 - 1) + \delta_z(3n^4 + 6l^2m^2 - 1)}{6} \tag{4.25}$$

and

$$\delta_i = \frac{D_i - D}{(D^2 - L^2)^{\frac{1}{2}}} \tag{4.26}$$

The ratio of R_2/R_1 relaxation times for residues that are not heavily perturbed by internal motions or exchange processes (155) can be used to good effect, in combination with Equations 4.8, 4.9 and 4.17 (which determine the spectral density $J(\omega)$ as a function of R_2/R_1), to report on the overall rotational correlation time τ_c. In the case of isotropic diffusion, Equations 4.27-28 describe the relationship between the correlation time and the isotropic diffusion tensor:

$$\tau_c = \frac{1}{6D_{iso}} \tag{4.27}$$

with D_{iso} equal to:

$$D_{iso} = \frac{kT}{8\pi\eta r^3} \tag{4.28}$$

where k is the Boltzmann constant, T is the temperature in Kelvin, η is the viscosity of solution and r is the effective molecular radius experiencing isotropic tumbling.

This approximation of D_{iso} does not of course apply in the anisotropic case. However, given a high-resolution three-dimensional structure, the D_{xx} D_{yy} and D_{zz} components of either axially symmetric or fully anisotropic diffusion tensors can be calculated in relation to the observed R_2/R_1 ratios with programs such as Modelfree (107). Variations in specific model free formalism do exist such as that presented by Clore et al. (157) which enables the extraction of two independent internal motions (S_f^2, S_s^2); however the ultimate goal of model free analysis is to utilise the numerically quantified parameters calculated above (R_{ex}, τ_e and S^2 (or S_f^2 and S_s^2)) to determine the most suitable dynamic model for the observed relaxation data.

Five standard models exist to which the obtained parameters can be fit, and statistically optimised for isotropically diffusing molecules:

$$\begin{aligned} model1 =& S_2 \\ model2 =& S_2 + \tau_e \\ model3 =& S_2 + R_{ex} \\ model4 =& S_2 + \tau_e + R_{ex} \\ model5 =& {S_f}^2 + {S_s}^2 + \tau_e \end{aligned}$$

Model 1 is indicative of fast internal motions ($\tau_e < 10$ ps) and describes the amide vector solely as a factor of its S^2 order parameter whereas model 2 includes an additional internal correlation time. Models 3 and 4 are extensions of models 1 and 2 respectively, with the inclusion of a chemical exchange term R_{ex} to account for slow (ms) contributions to T_2. The fifth model is based upon the work of Clore et al., (157) and accounts for two distinct internal motions acting across timescales of at least an order of magnitudes difference.

4.1.6 Experimental aims

Building upon the principles of NMR spectroscopy that were outlined in Chapter 3, the aim of Chapter 4 will be to document the investigation into the dynamic behaviour of the AVR3a11 effector domain. The potential for NMR to report on a wide range of dynamic molecular phenomena at the level of the individual amino acid could help to elucidate any pertinent molecular motions that may underpin the functional interactions of the AVR effector.

4.2 Results

4.2.1 Purification and resonance assignment

For dynamic analyses, a shorter construct of the AVR3a11 effector domain (C-terminal 65 residues) was chosen to enable direct comparison in between obtained NMR relaxation parameters and the 3ZR8 crystal structure (49). The purification of ^{15}N isotopically enriched samples of the truncated AVR3a11ΔN construct was both straightforward and generated comparable yields of pure protein to that of the full length construct, enabling the creation of 1 mM protein NMR samples.

Figure 4.2.1 shows example silver-stained SDS-PAGE gels of the initial Ni^{2+} affinity column purification for the extraction of the hexa-His tagged fusion protein from cell lysate and the subsequent PreScission protease digestion to remove the purification tag. The purity of the sample was such that a further gel filtration step was not required for the removal of any contaminant proteins and a repeated run through the Ni^{2+} affinity column yielded sufficiently pure, protease-free, protein.

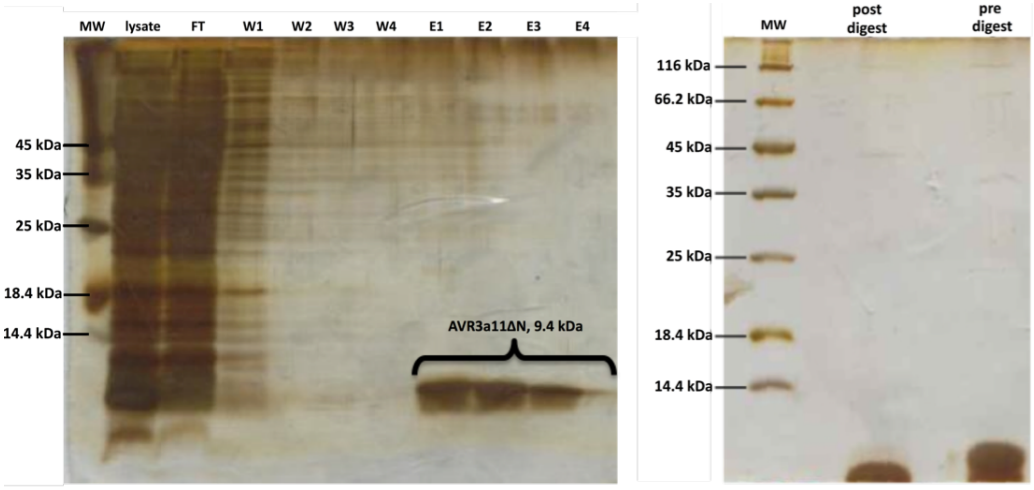

Figure 4.2.1: An example silver-stained SDS-PAGE gel of the resulting Ni^{2+} affinity purification step (Left). Lanes contained samples of cell lysate (lysate), column flow through fractions (FT), column washes 1-4 (W1-4) and the elution fractions (E1-4) containing 300 mM imidazole. An example silver-stained gel following hexa-His tag cleavage is also shown (Right).

4.2.2 Initial NMR assignment

The initial ^{1}H-^{15}N HSQC of the AVR3a11ΔN construct shared significant similarity to that of the full-length construct (Figure 4.2.2), allowing a direct mapping of most of the HN assignments. A small number of residues however were not directly assignable and required analysis of a ^{15}N edited HQSC- TOCSY experiment, in comparison to that acquired for the full-length construct, to complete resonance assignment. Perhaps unsurprisingly, the residues which experienced the greatest chemical shift differences to those of the full length construct (K67, V75, K77, A79, K80) were contained within the N-terminal α-helix 1 (Figure 4.2.3, red residues). It is unlikely that these residues have structurally relevant implications with respect to the four helix bundle conformation.

Although this was not tested with the acquisition of triple resonance data for structure determination or chemical shift indexing, it is more likely that the increase in relative conformational freedom of the N-terminal resides affected the average chemical shifts.

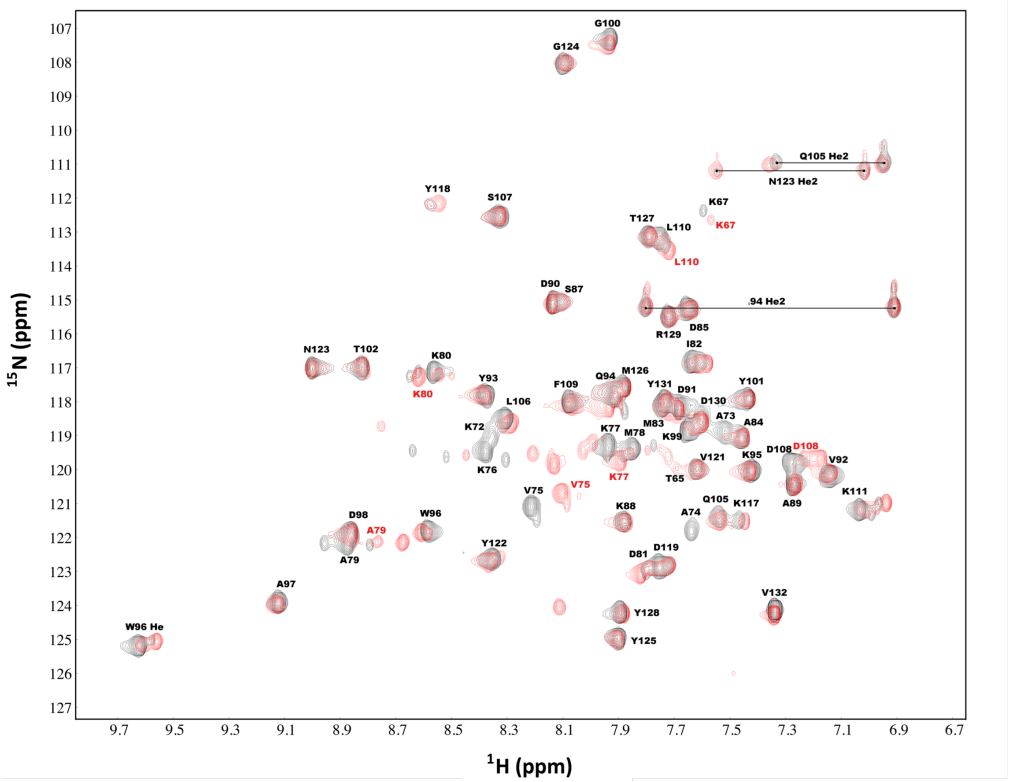

Figure 4.2.2: The ^{1}H-^{15}N HSQC spectra of the AVR3a11 full length effector domain construct (black peaks, black assignments) and the truncated AVR3a11ΔN construct acquired at 800 MHz. AVR3a11ΔN peaks with significantly different chemical shifts to those of the full length construct (red peaks, red assignments) were proximal to the truncated region. Black lines indicate the NH_2 groups of amino acid sidechains.

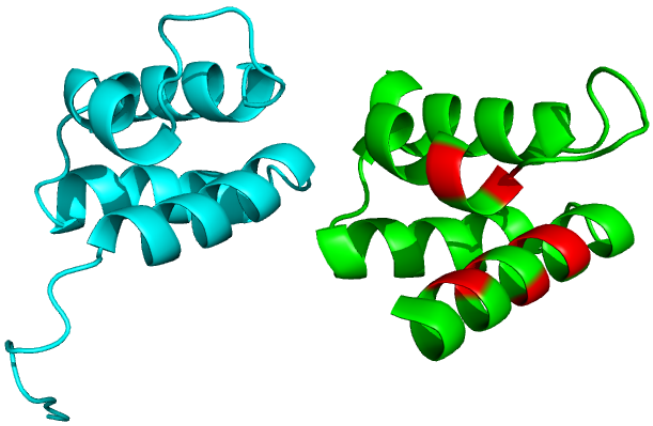

Figure 4.2.3: The top scoring ensemble model from the full length AVR3a11 NMR solution structure (cyan, left) an the offset 3ZR8 crystal structure (green, right), with red residues displaying notable HN chemical shift differences in initial ^{1}H-^{15}N HSQC spectra.

The observation that residues D108 and L110 also experienced chemical shift perturbations from the N-terminal truncation also agree with this hypothesis, as their positioning at the C-terminus of α-helix 3 would also place them within range of the conformational potential of the extended N- terminus.

4.2.3 AVR3a11 Hydrogen-Deuterium exchange

To probe the slow timescale dynamics of the AVR3a11ΔN construct, a series of Hydrogen-Deuterium (H/D) exchange NMR experiments were carried out. To this end, following the lyophilisation of a pre-existing ^{15}N labelled AVR3a11ΔN sample and resuspension in 500 μl D_2O, a series of ^{1}H-^{15}N SOFAST HSQC spectra were recorded at increasing time intervals. Although the bulk, protonated, solvent will be removed through lyophilisation, any protons attached to the protein should remain intact. Because ^{1}H-^{15}N HSQC based experiments only give rise to NMR signals following successful ^{1}H-^{15}N magnetisation transfer (and back again), as the resuspended protein begins to exchange with the deuterons of the solvent, the rate of exchange can be identified from the decrease in NMR peak intensity over time. As well as being a measure of solvent accessibility and hydrogen bonding, determination of amide exchange rates can also help to describe slow timescale dynamic phenomena such as protein breathing or the large conformational rearrangements seen in unfolding/folding pathways (158).

An important experimental caveat to H/D exchange experiment is the lag period between resuspension of the sample and subsequent analysis; as the bound protons will immediately begin to exchange with the solvent following resuspension. It can be relatively time consuming to ensure that all dehydrated protein has been resuspended, transferred to an NMR tube and placed within the magnet before experiment acquisition can begin. The lag period of the acquired H/D exchange experiments was reduced to 4 minutes by resuspending the sample close by to the NMR laboratory, pre-tuning the spectrometer to the sample prior to lyophilisation and only performing a brief, automated, shimming procedure for the z-axis shims. Nonetheless, preliminary H/D exchange experiments at 298 K, pH 8.8 yielded no signals, indicating that all protons had thoroughly exchanged (data not shown).

However, as a kinetic process with a temperature dependency, by reducing the acquisition temperature to 278 K, preparing the sample on ice with pre-chilled D_2O and lowering the pH of the sample buffer to 6.8, the loss of peak intensity became measurable over a longer timescale, with two signals in particular persisting up to 23 hours 50 minutes (Figure 4.2.4).

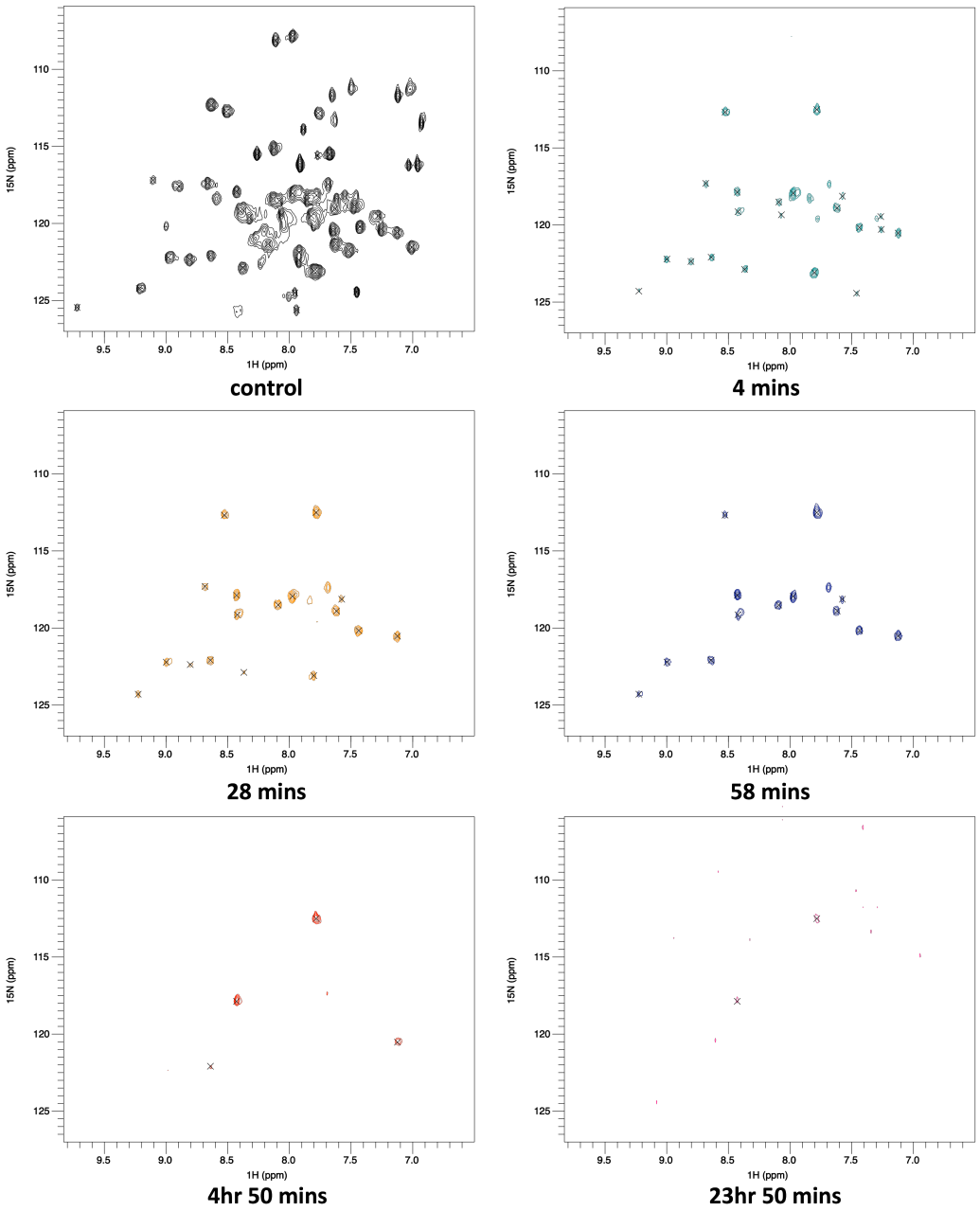

Figure 4.2.4: Example ^{1}H-^{15}N SOFAST HSQC spectra acquired at 800 MHz for the AVR3a11ΔN construct before resuspension in D_2O (control), and 4 mins, 28 mins, 58 mins, 4hr 50 mins and 23hr 50 mins post-resuspension.

The residue assignments of the ^{1}H-^{15}N HSQC spectra at 278 K required an additional ^{15}N-edited HSQC-TOCSY experiment to be acquired and only a small number of peaks became unobservable due to the effective increase in viscosity as a result of the decreased temperature.

Upon analysis of the H/D exchange spectra it became apparent that twenty residues had exchanged beyond observation before the first time point (4 mins). The lifetimes of the measurable peaks were calculated by fitting the loss of peak intensity over time as describe in (Section 2.1.7) to the exponential equation $y = y_0 + A^{(-x/t)}$. Figure 4.2.5 presents the log of these lifetimes (for simplification of the plot) and although this appears to be a seemingly sparse dataset, the behaviour of the persistent amides is explainable when mapped onto the structure of the truncated AVR3a11 domain (Figure 4.2.6).

The results suggest that the AVR3a11ΔN four-helix bundle can be described as a particularly stable structural element with respect to slow timescale motions. Although the H/D exchange rates of many residues were not apparent, the positions of these residues can be seen to map to the surface of the folded structure and therefore represent the residues with the greatest solvent accessibility and, subsequently, those which would be expected to rapidly exchange with the solvent (Figure 4.2.6, cyan).

In contrast, the amide protons of residues which can be seen to line the internal core of the fold were observed to posses the longest observed lifetimes. This is in agreement with the description of the AVR3a11 bundle possessing a significantly hydrophobic core (Figure 3.2.8, (49)) and suggests that significant, large scale, conformational changes do not regularly occur. Interestingly, this effect is not globally true across all four helices and a number of residues constituting the interior face of the N-terminal helix can be seen to have fully exchanged before the first time point, which indicates significant solvent accessibility. It may well therefore be the case then, that the N-terminal helix is more prone to slow timescale reorientations or transient dissociations from the hydrophobic core.

Because the process of protein H/D exchange is inherently linked to protein conformational stability, accurate H/D exchange measurements can also be used to draw direct conclusions regarding the free energy of unfolding. As was the case for the AVR3a11ΔN experimentation, standard experimental protocols serve to measure an overall rate constant of exchange (k_{ex}). However,

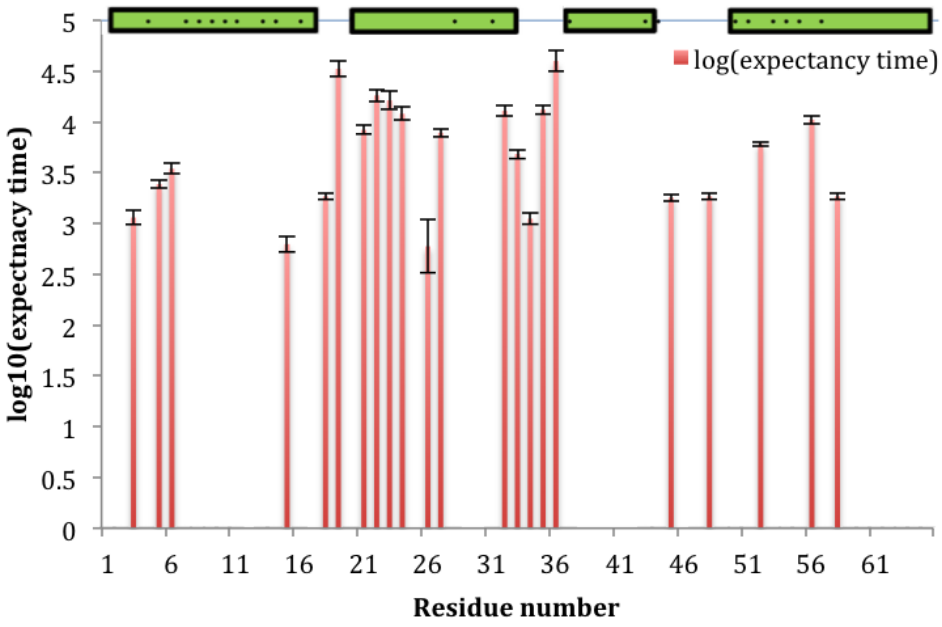

Figure 4.2.5: The calculated amide proton exchange lifetimes across the AVR3a11ΔN backbone. The secondary structure, as described by the 3ZR8 crystal structure (49) is presented above. Resonances that were assignable in the standard AVR3a11ΔN ^{1}H-^{15}N HSQC, but which had no signal at the first time point are denoted with * in the secondary structure.

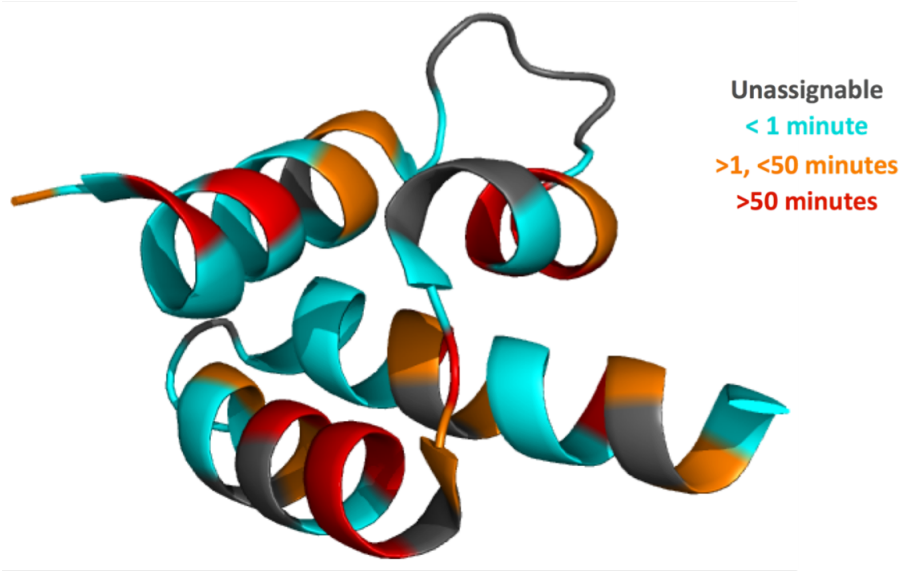

Figure 4.2.6: The structure of the 3ZR8 AVR3a11ΔN crystal structure (49) coloured by the calculated amide proton exchange lifetimes. Unassignable residues = grey, residues for which no signal was seen, and hence the expectancy time was 1 minute, (cyan), residues with amide expectancy times >1 and <50 minutes = orange and >50 minutes = red.

H/D exchange is best described as a two step process (159), wherein kex is comprised of transitions between closed:open conformational states, as well as the rate of chemical H/D exchange (Equation 4.29).

$$(N-H\cdots O=C)_{cl} \underset{k_d}{\overset{kop}{\Longleftrightarrow}} (N-H)_{op} \overset{k_{rc}}{\Longrightarrow} N-D \tag{4.29}$$

where (N-H$\cdots$O=C)$_{\mathrm{cl}}$, (N-H)$_{\mathrm{op}}$ and N-D refer to the structurally closed, open and solvent-exchangeable states, respectively, k_{op} (open) and k_{cl} (closed) are the associated rate constants for exchange between the structural forms and k_{rc} (random coil) is the intrinsic H/D exchange rate constant.

As described by Bai et al., (160), the intrinsic rate of exchange from the open state is generally predictable when information regarding the sample temperature, pH and salt concentration are known. Interestingly however, k_{rc} is also dependent on the local chemical environment, as defined by the protein primary sequence (steric blocking, inductive effects), and this also has to be taken into account by modulating the intrinsic exchange rate for each amide by incorporating the potential of the flanking sidechains to inhibit or facilitate exchange.

Theoretical k_{rc} values for the AVR3a11ΔN construct at pH 6.8 and 278 K were calculated with this methodology as described by Bai et al., (160) wherein k_{rc} is the sum of the residue-specific acid, base and water catalysis rate constants ($k_{rc} = k_{acid} + k_{base} + k_{water}$) (Figure 4.2.7).

Compared to the relatively tentative description of H/D exchange from Figure 4.2.5, Figure 4.2.7 highlights the relative stability of the central 2 helices in direct comparison to the exchange rates that would be expected for a random coil conformation (k_{rc}). Furthermore, although k_{ex} was unable to be measured for large sections of the terminal helices, it cannot be said that this therefore indicates a significantly open conformation at these positions. This is because exchange, slower than that of the expected k_{rc}, could still have occurred during the lag time of the experiment (note the separate axes of Figure 4.2.7). This therefore indicates that AVR3a11ΔN is more often occupying a closed conformational state than one able to exchange with the solvent; however to extract meaningful conclusions regarding protein stability the precise balance of k_{cl}, k_{op} and k_{rc} exchange processes must be ascertained.

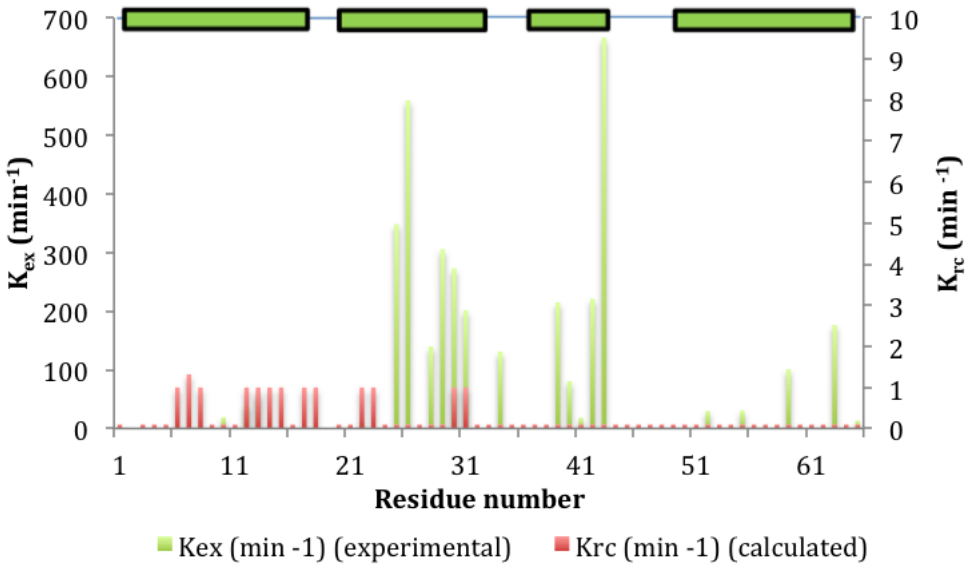

Figure 4.2.7: The experimental H/D exchange rates (k_{ex}) for AVR3a11ΔN (green) and the calculated intrinsic rate constants (k_{rc}) (red) using the methodology of Bai et al.,(160).

In the case of AVR3a11ΔN, where $k_{op} << k_{cl}$, the underlying exchange can either be defined by k_{op} alone (i.e. $k_{cl} << k_{rc}$, termed EX1) or by both k_{op} and k_{cl} (i.e. $k_{cl} >> k_{rc}$, termed EX2).

Explicit methods exist for determining whether a particular system is experiencing EX1 or EX2 exchange, however the most accurate requires that multiple data sets be acquired at different pH to identify any pH rate- dependency (i.e. if $k_{cl} << k_{rc}$, then $k_{ex} = k_{op}$; which should not be directly pH dependent = EX1). However, there was unfortunately not time to carry out this experimentation during this work. A rough approximation does exist (159) which states that residues with similar free energy of exchange values (ignoring the exchange model) will have similar k_{rc} values (EX1), whilst the reverse is applicable to EX2 conditions. This approximation was attempted by using the experimental k_{ex} and calculated k_{rc} values to derive $\Delta G_{HD} (\Delta G = -RT\ ln \frac{(k_{ex}}{k_{rc}})$ however, it was unfortunately the case that a mixture of the aforementioned scenarios were present and so no further analysis could be carried out.

4.2.4 ^{15}N backbone relaxation

Longitudinal, transverse and hetNOE ^{15}N relaxation measurements were acquired for the AVR3a11ΔN construct at both 500 and 800 MHz. Relaxation rates R_1 and R_2 and the hetNOE enhancement were determined according to Section 2.1.6 and the calculated relaxation rates are presented in Figures 4.2.8 and 4.2.9.

The residues of the AVR3a11ΔN construct that were assignable within the ^{1}H-^{15}N HSQC spectra were those pertaining to the well folded regions of the molecule, as observed from comparisons to both the AVR3a11ΔN crystal structure and the NMR ensemble of the full length effector. The regions that would be expected to experience the greatest dynamic phenomena (such as the helix3-helix4 loop) could therefore not be measured. The relaxation rates of three N-terminal residues (L5, A6, A7) also could not be ascertained due to significant spectral overlap. The immediate observation from all of the ^{15}N relaxation measurements ascribes a relatively uniform dynamic behaviour over the molecular system as seen from wholly consistent relaxation rates. The only consistent deviation from this regularity is the very C-terminal residue V65 (V132, full length protein numbering), which is a common observation based upon the lack of motional restriction at either terminus. The mean averages, errors and standard deviations of all the relaxation values (excluding the C-terminal V65) further highlight this consistency (Table 4.2.1).

Table 4.2.1: The mean average values, errors and standard deviations (σ) for the calculated relaxation rates at 500 and 800 MHz of all residues, excluding the C-terminal residue V65.

Relaxation	Avg. (s^{-1}), hNOE)	Err. (s^{-1}, hNOE)	σ (s^{-1}, hNOE)
R_1 500 MHz	2.20	0.03	0.09
R_1 800 MHz	1.38	0.02	0.06
R_2 500 MHz	9.57	0.13	0.72
R_2 800 MHz	12.17	0.27	1.20
hNOE 500 MHz	0.73	0.02	0.05
hNOE 800 MHz	0.82	0.01	0.05

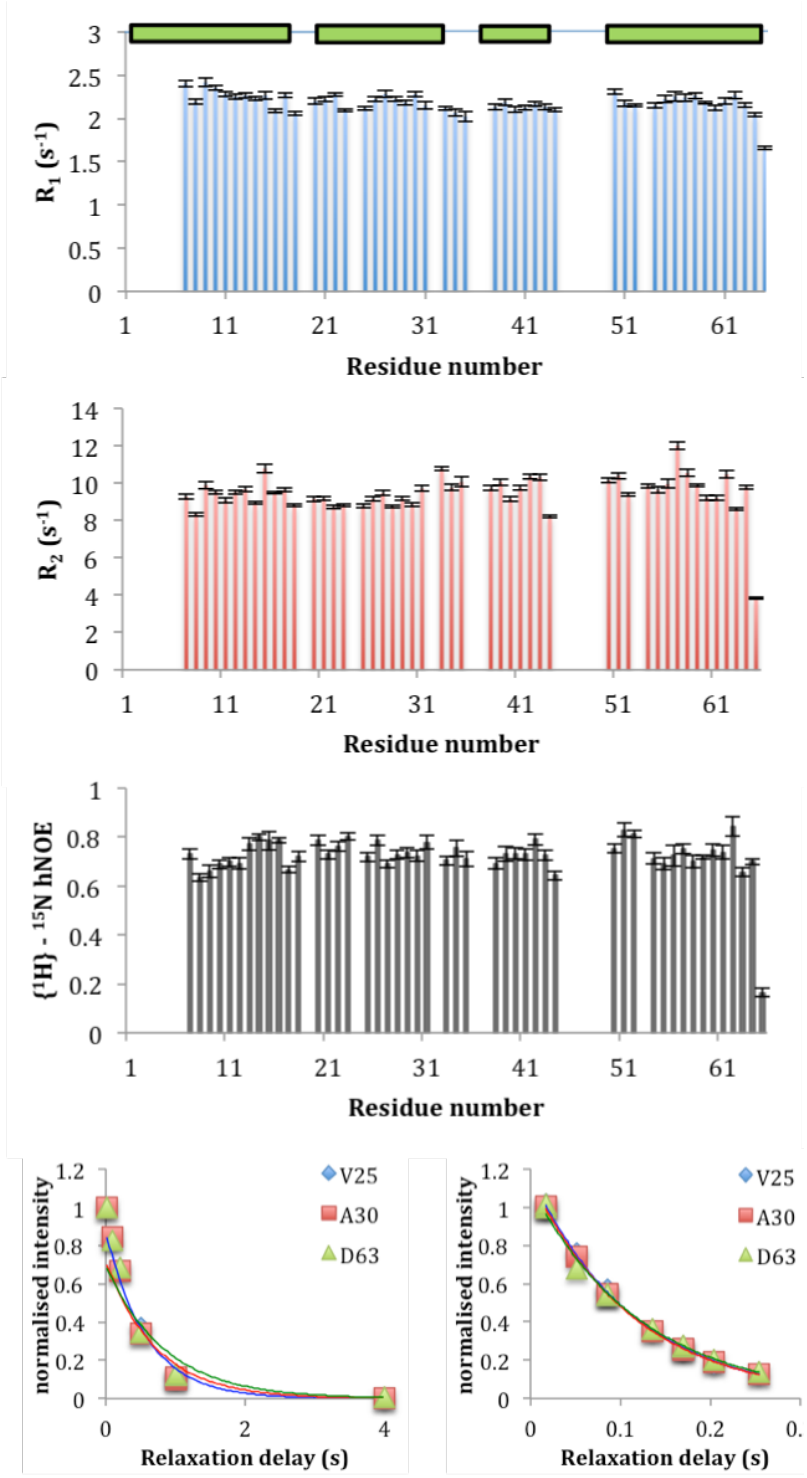

Figure 4.2.8: The R_1 and R_2 ^{15}N relaxation rates and the ^{1}H-^{15}N heteronuclear NOE enhancements at 500 MHz across the AVR3a11ΔN backbone. Errors were calculated from the error of the fit with respect to signal to noise. Examples of the mono-exponential decays of normalised peak intensities for R_1 (bottom left) and R_2 (bottom right) rates for three residues are shown, with numbers relating to the truncated construct.

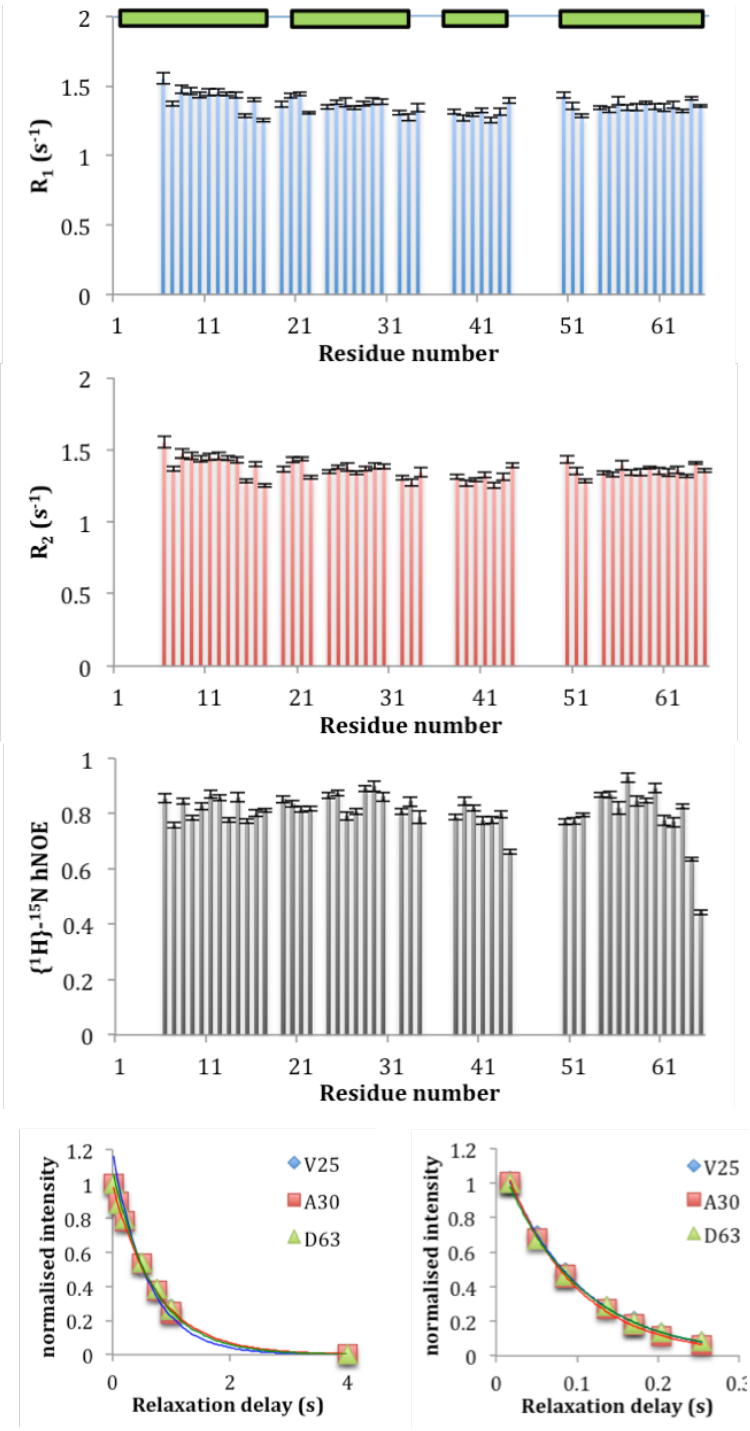

Figure 4.2.9: The R_1 and R_2 ^{15}N relaxation rates and the ^{1}H-^{15}N heteronuclear NOE enhancements at 800 MHz across the AVR3a11ΔN backbone. Errors were calculated from the error of the fit with respect to signal to noise. Examples of the mono-exponential decays of normalised peak intensities for R_1 (bottom left) and R_2 (bottom right) rates for three residues are shown, with numbers relating to the truncated construct.

Of the 48 residues analysed in all datasets, approximately 30 residues were within ±1 (terminus-excluding) standard deviation of the calculated averages across all datasets (Table 4.2.1); however, although relaxation values were relatively uniform, only eight, scattered, residues were consistently shown to be within these parameters across all datasets (K21, Q27, D31, Q38, D41, N56, M59 and Y61). The calculated errors presented in Table 4.2.1 represent percentage errors of 1.4%, 1.3% and 3.0% for the R_1, R_2 and hetNOE measurements at 500 MHz and 1.4%, 2.2% and 1.5% for the respective values at 800 MHz; which are comparable to values reported in similar ^{15}N relaxation studies found in the literature (161164). Figures 4.2.8-9 also highlight the dependence of R_1 and R_2 on the magnetic field strength, with observations of shorter R_1 rates and longer R_2 rates at 800 MHz compared to 500 MHz. This is due to due to the increase in Larmor frequency of the respective nuclei (and subsequently $J(\omega)$). The increase in hNOE at 800 MHz is also as expected due to its dependence on R_1.

The R_2/R_1 ratio is frequently used as a qualitative probe of molecular motions. This arises from the fact that slow exchange processes serve to increase R_2, but not R_1 (R_{ex}, Equation 4.9), consequently increasing the ratio. Conversely, residues experiencing significant internal motions can be identified, as they will serve to decrease the R_2/R_1 ratio.

The R_2/R_1 ratios presented in Figure 4.2.10 indicate that, as sampled at 500 MHz, the observable residues of the AVR3a11ΔN construct are experiencing relatively uniform contributions to relaxation from slow and fast timescale motions. However, the relative decrease in the R_2/R_1 value for the C-terminal residue V65 is suggestive of enhanced relaxation due to rapid internal dynamics. Although this description of V65 is mirrored in the R_2/R_1 value at 800 MHz, the remainder of the protein is distinctively less uniform. In particular, a number of residues in the C-terminal helix 4 are suggestive of influence by slow timescale exchange, which was not directly evident from either the relaxation rates alone, or the ratios at 500 MHz.

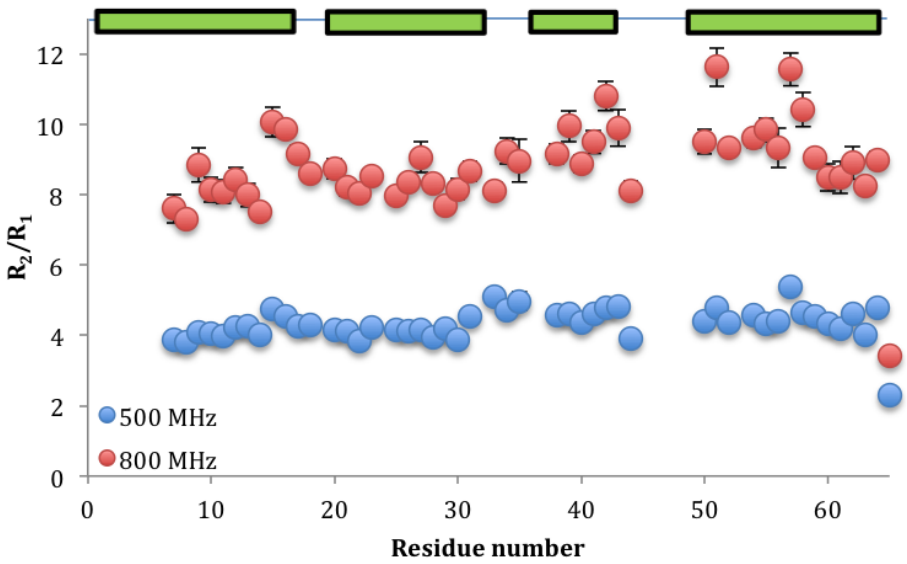

Figure 4.2.10: The R_2/R_1 ^{15}N relaxation ratios at 500 MHz (blue) and 800 MHz (red) for the AVR3a11ΔN construct.

Whilst the above descriptions are rather vague with respect to specific motional characteristics, R_2/R_1 ratios can also be used as an approximate measure of the overall rotational correlation time, τ_m, in the case of isotropic diffusion (Equation 4.17). For this approximation to hold, Tjandra et al., have defined two criteria (161) so that the relaxation times of any residues deemed to be experiencing significant fast internal motions (as evident from hNOE $<$ 0.65) or substantial exchange processes (equation 4.30) can be discarded due to their inability to report on an overall correlation time.

$$\frac{(\langle T_2\rangle - T_{2,n})}{\langle T_2\rangle} \; - \; \frac{(\langle T_1\rangle - T_{1,n})}{\langle T_1\rangle} > 1.5 \; SD(\sigma) \tag{4.30}$$

where the T_1 and T_2 relaxation times of individual residues, n, should be evaluated against the average relaxation times $\langle T_2\rangle$ and $\langle T_1\rangle$ (as calculated from the subset of residues for which hNOE > 0.65), and the standard deviation (σ) is with respect to $((\langle T_2\rangle - T_{2,n})/\langle T_2\rangle \; - \; (\langle T_1\rangle - T_{1,n})/\langle T_1\rangle)$.

These Tjandra-Bax criteria can also be employed in the characterisation of anisotropic diffusion although a singular correlation time does not apply in this instance. Because the R_2/R_1 ratio will depend on the orientation of the NH bond vector due to the dipolar interaction (Equation 4.20), with an *a priori* description of the diffusive mode, estimations can be made regarding the principal axes of the diffusion tensor.

In the case of the AVR3a11ΔN, the relative principal axes of the diffusion tensor (D_x, D_y and D_z) were first estimated with pdbinertia (accessible: `http://www.palmer.hs.columbia.edu/software/diffusion.html`) using the 3ZR8 crystal structure of the AVR3a11ΔN construct as input. The result of this calculation generated an inertia tensor with lengths of the relative principal axes equal to: 1.00:0.93:0.75 which suggests that either an axially symmetric of fully anisotropic model of diffusion should, in principle, be observed. Although this has previously been shown a valid approximation (161), the calculated inertia tensor does not take account of any asymmetric drag due to protein hydration, as the calculated inertia tensor only describes the 3D distribution of angular momentum as a result of molecular size and shape.

HYDRONMR (165) offers an alternative approach which, provided a 3D structure is available, is able to both extract the NH bond vectors and estimate of the T_1, T_2 and NOE relaxation rate (and also the reduced spectral density terms J(0), J(ω_N) and J($0.87\omega_H$)) at a particular magnetic field strength by modelling the effect of hydration on diffusion (accessible: `http://leonardo.inf.um.es/macromol/programs/hydronmr/hydronmr.htm`). By redefining every non-hydrogen atom within the structure as a series of beads with increased atomic radii (defined as the atomic element radius, AER) the program attempts to account for the additional contribution of the hydration shell to diffusion. The AER value can then be optimised to account for the hydration shells of individual proteins.Figures 4.2.11-12 present the R_2/R_1 ratios for the AVR3a11ΔN construct (excluding those significantly affected by fast and slow motions, as defined by Equation 4.30) and the theoretical ratios as calculated by HYDRONMR at 500 MHz and 800 MHz respectively for five values of AER bead size. Calculations were carried out between 2.6 and 3.5 Å in 0.10 Å increments, with a ^{15}N chemical shift anisotropy of -172 ppm, with an NH bond length set to 1.02 Å and at a temperature of 293 K with a standard value of viscosity taken to be 1.0 x 10^{-3} N s m^{-2}.

Unfortunately, no agreement was observed between the experimentally determined relaxation values and those theoretically calculated with the various bead sizes. All calculated values were significantly lower than those measured. As would be expected, the increase in bead size corresponded to increasing values of R_2/R_1 as a result of slower effective diffusion.

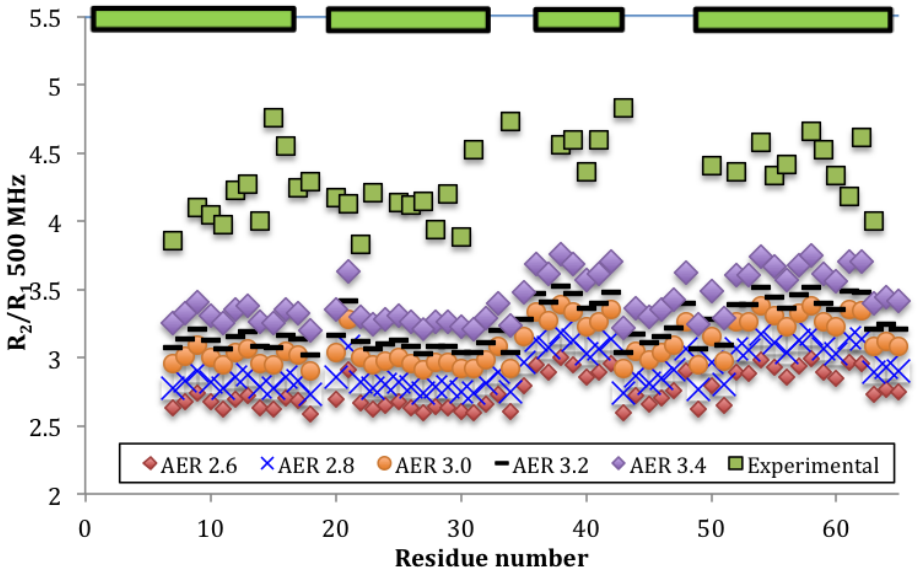

Figure 4.2.11: The AVR3a11ΔN R_2/R_1 ratios at 500 MHz which satisfy the Tjandra-Bax criteria (161) (green squares) and the HYDRONMR theoretical ratios for multiple AER values of 2.6 (red), 2.8, (blue), 3.0 (orange), 3.2 (black) and 3.4 Å (purple).

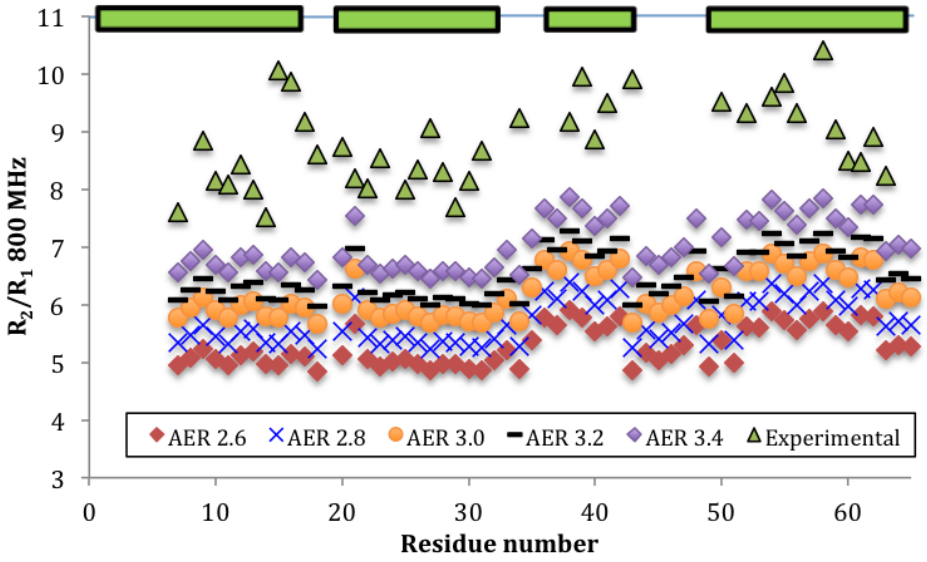

Figure 4.2.12: The AVR3a11ΔN R_2/R_1 ratios at 800 MHz which satisfy the Tjandra-Bax criteria (161) (green squares) and the HYDRONMR theoretical ratios for multiple AER values of 2.6 (red), 2.8, (blue), 3.0 (orange), 3.2 (black) and 3.4 Å (purple).

However this trend could not be explored further to better match the experimental results as the overlap between the beads resulted in premature termination of the calculation (data not shown).

The inability to recreate the high R_2/R_1 ratios from the experimental results suggest that the residues chosen for comparison are experiencing significant exchange processes (166). Although this subset of real data should have been devoid of slow dynamic phenomena through application of the Tjandra-Bax criteria, a potential criticism of the method should be highlighted. The selection process assumes that any residues experiencing significant fast or slow exchange will be distinct from those not, by lower or higher deviations from the mean R_2/R_1 values respectively. However, in the instance where both fast internal motions and slow exchange processes affect the majority of the protein backbone, a situation can be envisaged wherein different proportions of these exchange processes can culminate in a seemingly consistent R_2/R_1 values. Arguably, the requirement for residues with an hNOE > 0.65 should account for this effect due to its dependence on significant fast motions, and so suggests that slower timescale motions are most likely perturbing the analysis.

4.2.5 Reduced Spectral Density Mapping

Although unfortunate, the above description of slow exchange processes disrupting characterisation of the molecular diffusion from R_2/R_1 ratios is catered for by work from Chang et al., (167). This states that a better method of identifying rotational correlation times is to directly compare *in silico* calculated m values to those that have been calculated from spectral density calculations and modified to mathematically cancel the internal motions component ($(1 - S^2)\ \tau_e$) from the initial equations (Equation 4.31):

$$\tau_m = \left(\frac{J(0) - J(\omega_N))}{(J(\omega_N) - J(\omega_H))} \ \frac{(0.87\omega_H - \omega_N)^2}{\omega_N^2} \ -1 \right)^{-1/2} / \omega_H \qquad (4.31)$$

Therefore, using the spectral density equations, presented here as Equations 4.13-4.1.5, the spectral densities were evaluated at $J(0)$, $J(\omega_N)$ and $J(0.87\omega_H)$ for both 500 and 800 MHz magnetic fields for all of the AVR3a11ΔN residues. Equation 4.30 was then used to calculate local τ_m values for each residue and these were then compared to the *in silico* modelled τ_m values, as determined by HYDRONMR for the calculated values between 2.6 and 3.5 AER.

Using this methodology, the AER value which best represented the experimental results was determined to be 2.9 Å at both magnetic fields, as the local τ_m values for this bead size possessed the lowest mean deviation from the experimental results. A comparison between the manually calculated m values at both magnetic fields and those calculated by HYDRONMR can be seen in Figure 4.2.13.

The local τ_m values that were calculated to be independent of internal motions of the acquired ^{15}N relaxation data are in good agreement with that determined by HYDRONMR with a bead size of 2.9 Å, with even small trends in τ_m visible such as the marginally higher rotational correlation time for helix 3. Due to the success of this calculation in mirroring internal motion independent data, it can be said with some confidence that the other results from the HYDRONMR calculation should also describe the dynamic behaviour of how AVR3a11ΔN would behave if it were indeed a rigid body.

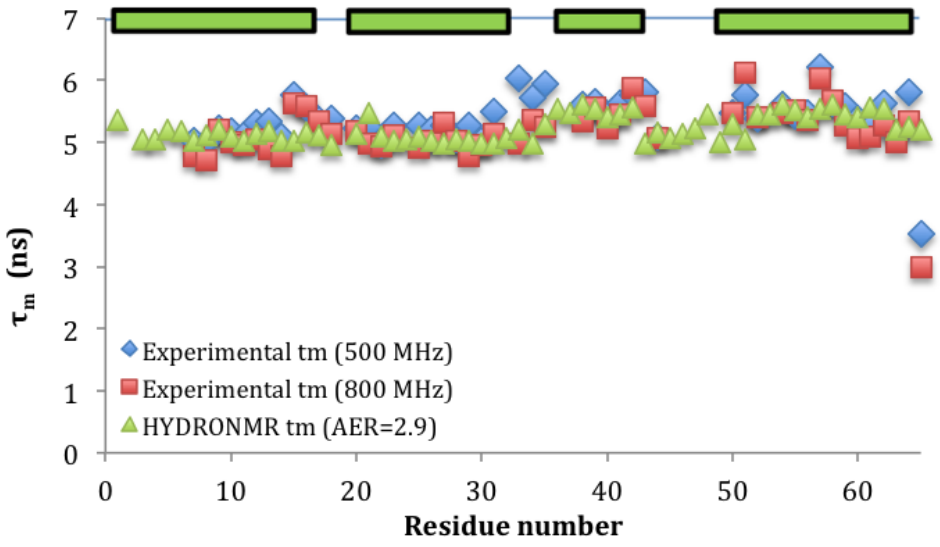

Figure 4.2.13: The local τ_m values as calculated at 500 MHz (blue) and 800 MHz (red) from ^{15}N relaxation data of all AVR3a11ΔN residues and the local τ_m values as determined by HYDRONMR for an AER value of 2.9 Å.

The HYDRONMR calculation also determined an average mean rotational correlation time for the whole protein of 5.22 ns, which is reasonable given the potential for exchange and anisotropic diffusion to increase the rule-of-thumb (the rule being 0.6 × MW (kDa) = τ_c (ns)) correlation time, which for AVR3a11ΔN = 4.49 ns. Furthermore, relative lengths of the anisotropic diffusion tensor were determined to be 1.00:0.77:0.85 from the PDB file, and although these are with respect to a rigid structure, they suggest a greater deal of axial symmetry than the relative lengths of the inertia tensor calculated by pdbinertia.

Figure 4.2.14 then compares the experimentally calculated reduced spectral density parameters to those from the HYDRONMR outputs at 500 and 800 MHz, corrected by a factor of 2/5 given that the equations used by HYDRONMR differ from equations 4.13 to 4.17. The differences between the evaluated frequencies provide an interesting comparison as it can be assumed that the HYDRONMR data represents that of a rigid AVR3a11ΔN protein. The observation of higher J(0) values suggest that there are significantly greater slower frequency internal motions across the entire protein, and specifically for helix 4, which confirms that a significant Rex exchange term would have skewed conclusions from the R_2/R_1 ratio alone.

The faster internal motions probed by $J(\omega_N)$ and $J(0.87\omega_H)$ however are in contrast to one another and suggest that on the faster nanosecond timescale, faster internal motions of 50 MHz are particularly more prominent. Conversely, the faster motions at the ps timescale, probed by $J(0.87\omega_H)$, shows little divergence from a relatively rigid body.

Finally, although model free analysis was attempted with the raw ^{15}N relaxation data, all calculations either failed to converge or yielded thoroughly uninterpretable results (data not shown). Although unfortunate, it would seem that the inherent level of conformational exchange, as determined from comparisons to the HYDRONMR calculations, were most likely complicating the dynamic characterisation where levels of exchange were mistaken for anisotropy, therefore failed to fit to the structure and dynamics of the system as a whole.

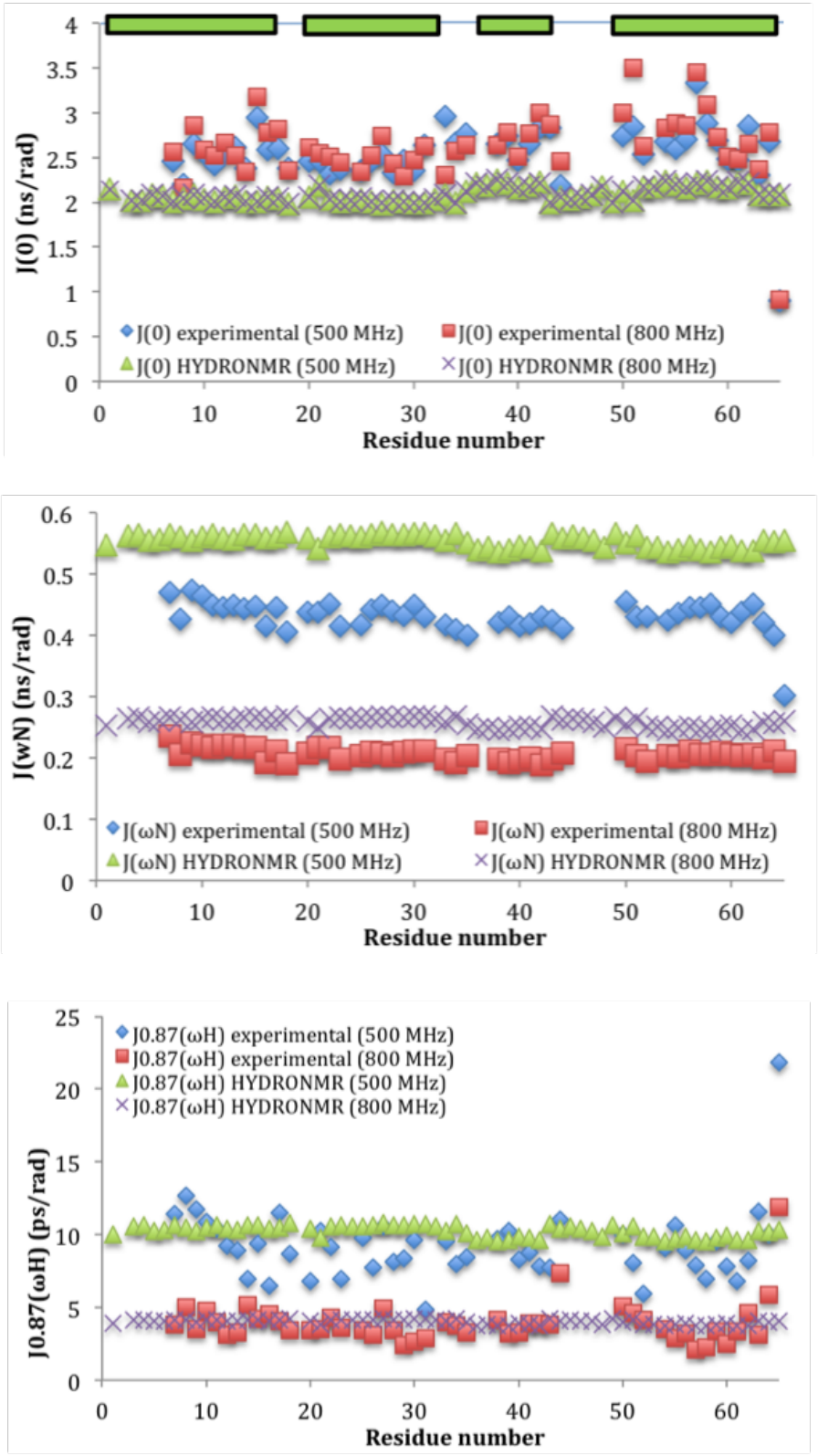

Figure 4.2.14: A comparison between the spectral densitys evaluated at J(0) (top), J(ω_N) (middle) and J($0.87\omega_H$) (bottom) of the values calculated from ^{15}N data and those determined by HYDRONMR for an AER value of 2.9 Å

4.3 Discussion

Particular regions of the AVR3a11 effector have been shown to be without fixed secondary structure, as observed from both the calculated X-ray structure and the NMR ensemble, and so could therefore be susceptible to dynamic phenomena. However, due to the inability to resolve the amide peaks of these residues, no conclusions could be made from experimental data about the level of dynamics at these sites. This could have been particularly insightful, as mutations in the helix 3-to-4 loop have been documented in the literature as capable of reintroducing plant recognition, via the R3a resistance protein, for the AVR3a^{EM} allele of *Phytophthora infestans* for which there is no native interaction (145).

The results of the hydrogen/deuterium exchange experiments suggest that as a whole, the 4-helix bundle fold of the effector domain is largely stable with regard to slow dynamic phenomena such as folding-unfolding reactions or changes in helical rearrangement. However, the observation that three residues of the N-terminal helix-1 interior face have rapid exchange times, akin to those directly on the surface of the protein proves interesting. Although other RXLR effectors, including the AVR3a4 protein from *P. capsici*, are also shown to form a four-helix bundle conformation, RXLR effectors in general have a distinct overall lack of sequence homology. In trying to describe similarities between RXLR effectors, the actual conserved structural motif has been defined as that of a three helix bundle conformation termed a WY domain (168), which constitute helices 2, 3 and 4 of the AVR3a11 fold. The N-terminal helix therefore represents an addition to the central WY domain fold, although it clearly has functional significance as it contains residue position 71. Therefore it may well be that the hydrophobic interactions which drive the compactness of the fold are not as favourable for the N-terminal helix than for helices 2-4, causing a relative lack of affinity for the hydrophobic core which would allow its residues specifically greater solvent accessibility.

Reduced spectral density analysis was used to analyse the extent of fast and slow internal molecular motions across the protein backbone. Although initial interpretations of the fast timescale motions with ^{15}N relaxation data alone proved to be complicated by the presence of slow timescale exchange process(es), after extraction of these terms by use of Equation 4.31 it became

obvious that in relation to a relatively rigid model, AVR3a11ΔN as a whole displays a greater level of slow motions across the entire backbone in comparison to that of a rigid macromolecule. It could be argued that the level of slow exchange observed was in fact an artefact resulting from either weakly associating dimers or inconsistencies between the acquired data sets at the two magnetic fields. Common problems include either the setup of the particular pule sequences used or, in particular, differences in the concentrations of the samples used which can impact the viscosity of solution and hence impact the rate of diffusion. However although minor differences in the sample concentrations cannot be ruled out entirely, previous work has explicitly highlighted that AVR3a11ΔN does not form dimer species in solution (49). Furthermore, a comparison of the field independent J(0) values calculated across the protein (Figure 4.3.1) at each field exhibit significant correlation to one another, indicating that the data acquired was in fact comparable.

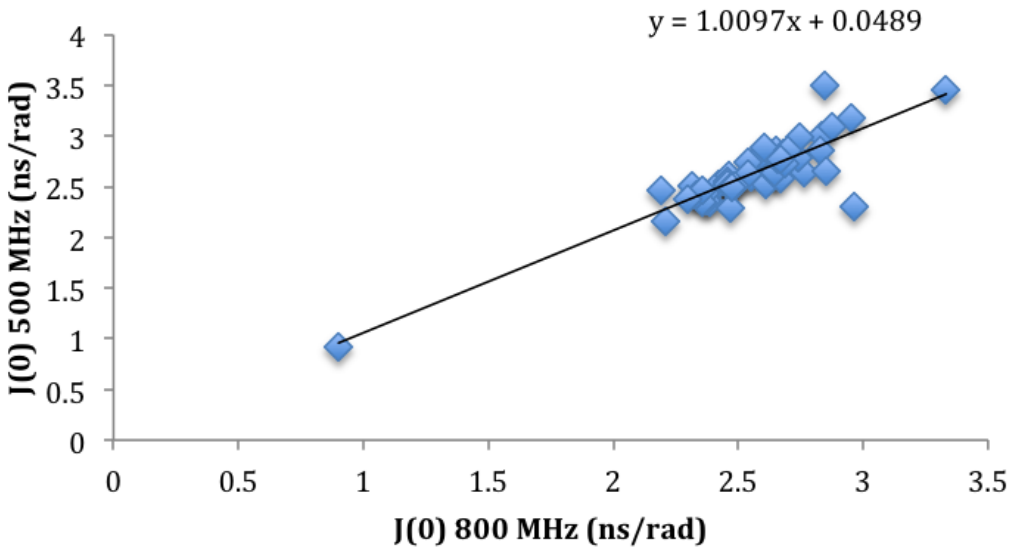

Figure 4.3.1: The field independent motions for the AVR3a11ΔN construct at both 500 and 800 MHz plotted together. The linearity of the plot can be used as a measure of the consistency between the datasets.

Although not seen as an explicit characterisation of dynamic phenomena, the B-factor, or temperature factor, obtained during the processing of X-ray crystallography data has classically been used as a tentative reporter of protein dynamics. Evaluations of B-factors are frequently carried out in the literature regarding protein structure-function investigations and, in particular, are rou-

tinely compared to NMR-derived dynamic data (169, 170). However, although crystallographic B-factors are related to the additional x-ray scattering that arises from molecular motion, they can equally reflect the level of static disorder within the crystal lattice; and at present the two phenomena are unable to be meaningfully deconvoluted.

With this is mind, recent works (171, 172) suggest that in actuality no general conclusions can be made from comparing crystallographic B-factors to explicit dynamic observations and that molecular dynamics calculations are actually a far better complimentary method to aid the characterisation of molecular motions. Although the relevant Gaussian Network Modelling (GNM) calculations (171) were outside the scope of this work, for completeness, Figure 4.3.2 shows the B-factors of the backbone Cα and N atoms extracted from the 3ZR8 crystal structure as compared to the R_2/R_1 ratio for the AVR3a11ΔN construct at 800 MHz.

Indeed, the NMR and crystallographic data are in poor agreement, with no visible trend in B-factor correlating to the helix 3-4 loop for which NMR data could not be obtained, presumably due to unfavourable dynamics. Conversely, the largest Cα and N B-factors correlate precisely to helix three for which no noteworthy, experimentally determined, dynamics had been shown, and for which no greater-than-average backbone deviations across the AVR3a11 solution-state NMR ensemble were apparent. Although the NMR derived S^2 order parameter is perhaps better compared to crystallographic B-factors (173), because this was unable to be determined in this work, the R_2/R_1 ratio perhaps stands as the next best measure to highlight any significant trends between both slow and fast timescale motions. Therefore, it is entirely more probable that the observed B-factors are in fact highlighting significant static disorder for helix 3 rather than an important dynamic process, and subsequently, further investigation of the AVR3a11ΔN B-factors, should not be carried out with a view to functional interpretation.

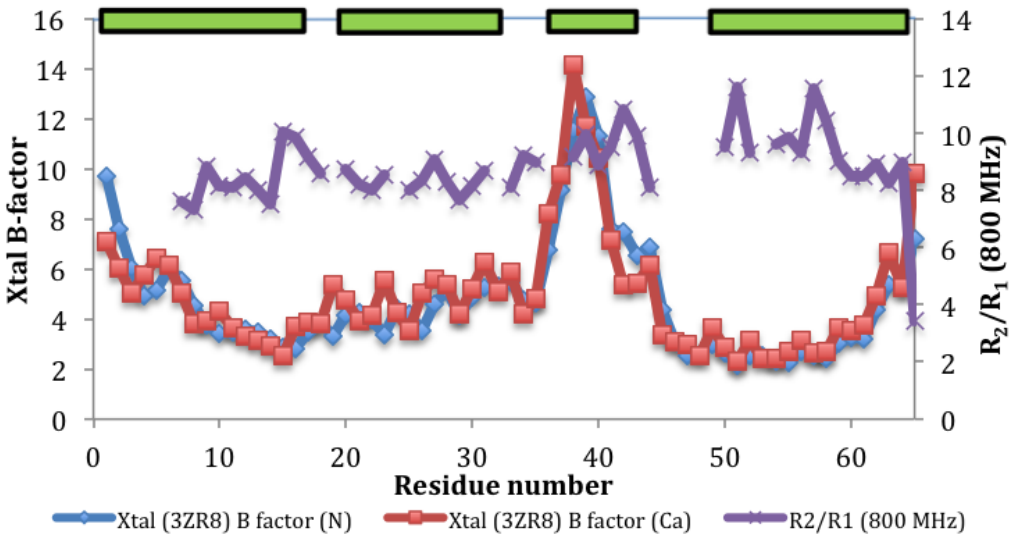

Figure 4.3.2: The B-factors of the backbone N (blue) and Cα (red) atoms of the AVR3a11ΔN crystal structure (3ZR8) compared to the experimental R_2/R_1 ratio at 800 MHz (purple).

Chapter 5

Characterising the TARP ABD

5.1 Introduction

5.1.1 Protein intrinsic disorder a brief history

Protein structure is classically defined by four, increasingly complex, hierarchical levels: primary structure (order of amino acids), secondary structure (α-helical, βstrand, coil and turn elements), tertiary structure (association and folding of the secondary structure elements) and quaternary structure (the association of multiple units of tertiary structure). For many decades, the classical understanding of the structure-function paradigm was applied to molecular biology with an understanding that the breadth of specific and selective function, as required by the interactions of complex life, was only achievable through the three dimensional complexity accessible to the higher tiers of protein structure. The last decade or so however has seen an increasing amount of evidence supporting the idea that extended, unstructured, regions or indeed whole disordered proteins, not only exist in nature but that they also have distinct functional roles.

A great deal of scientific study has of course focussed on induced-disorder states, for instance in elucidating the fundamental principles of protein folding. Unfortunately, it can be said that this early research into enzyme chemistry, central to the conceptualisation of the molecular structure-function paradigm, also lead to the temporary disregard of disordered protein states (174).

Undeniably, it is difficult to contemplate reaching an alternate conclusion, given the scientific zeitgeist of the mid-to-late 20^{th} century, with the accumulating evidence linking precise structural arrangement to selective reactivity and, concomitantly, chemical or thermal denaturation to functional ablation. This emerging interest in protein structure also helped seed the advent of routine X-ray crystallography and indeed for many years this generated a somewhat inevitable bias in protein science towards structured, crystallisable proteins. Distinct disordered regions were regularly dismissed as functionless linkers, phasing errors or simply products of an exogenous environment; that is if they were not removed entirely to facilitate crystalisation.

Over time, evidence of functional, natively disordered proteins has been supported through a combination of functional assays and the application of structure-probing solution state techniques such as NMR and small angle x-ray scattering (SAXS).

These observations, for a time, unsettled the classical structure-function paradigm; indeed, functionality with a lack of distinct 3D structure is in direct opposition to its understanding. But recently a general description of protein function is starting to be put forward which suggests an understanding of disordered functional domains should supplement the classical paradigm, rather than redefine it (174). It is widely accepted that any degree of protein structure, whether it be none at all, is a consequence of the order and composition of the primary sequence amino acids (175, 176). Natively disordered proteins have therefore been termed "intrinsically disordered proteins" (IDPs) so as to accurately describe their inherently predetermined lack of structure. Also, functional observations of ordered proteins continue to be satisfactorily explained (generally) by induced-fit or conformational selection mechanisms which are facilitated by active site dynamics and supported by precise, coded, structural arrangement (177).

The distinction begins to arise from the fact that whilst structure determines function, an infinite variation in 3D structure can not necessarily allow for infinite possible function; and broadly speaking, the majority of ordered proteins have evolved to be highly selective, specific, tight binding macromolecules. Conversely, disordered proteins regions are more commonly associated with the highaffinity lowspecificity or lowaffinity high specificity protein-binding events involved in signal transduction and transcriptional regulation

(178) (Note: To date, only one example of a disordered catalytic enzyme has been documented (179)). Consequently, protein disorder can be thought of as expanding the functional repertoire of proteins, into territory that the very nature of 3D structure precludes.

An appreciation of IDPs may also help to explain the recently described paradox in molecular biology which shows an unexpected lack of correlation between an organisms complexity and its genome size (180). Many attempts have been made to rationalise this observation, such as with microRNAs or alternative splicing, but a significant positive correlation also exists between an organisms complexity and the number of disordered proteins encoded by its genome (181). It may well be that cellular complexity is greater than the sum of all these biological parts, but with IDPs proving central to functional regulation and with being increasingly linked to involvement in human disease states (182), an understanding of the long dismissed disordered state may well prove vital as we begin to decipher increasingly complex biochemical problems.

5.1.2 Characterising protein disorder

Much work has focussed on describing protein structure by both direct and indirect biochemical and biophysical observations and this is also true for the characterisation of disordered proteins. One of the most telling biophysical measurements in this regard, observable through a number of techniques, is that of a proteins effective hydrodynamic diffusion profile (Stokes radius (R_S)).

By assuming an average folded density for proteins, (1.37 g.cm^{-3} (183), although specific densities can be calculated as a function of molecular weight (184)) for a given molecular weight (and hence the effective radius), the diffusion of a spherical body can be calculated through use of the Stokes-Einstein equation (eq 5.1.1):

$$D = \frac{\kappa T}{6\pi\eta R_S} \tag{5.1}$$

where D is the diffusion constant, ϰis the Boltzmann constant (1.3806 x 10^{23} J.K-1), T is the temperature of the system (in Kelvin), η is the viscosity of solution and R_S is the radius of the diffusing sphere. When applied to biological molecules, the effective R_S can be increased through solvent interactions, but

the effect is essentially uniform and therefore allows for the determination of either theoretical molecular weight or diffusion rate when given the converse information.

Techniques such as dynamic light scattering (DLS), pulsed field gradient (PFG) NMR and gel filtration chromatography are commonly used to measure the diffusion characteristics of proteins in solution, and do so by comparison to the diffusion of either small molecules, or standard calibration curves.

The inherent extended structure of IDPs (and hence relatively low densities and variable Stokes radii) limits the success of molecular weight determination with these methods. The level of divergence from spherical- molecule calibration curves can then be used as a measure for the amount of disorder within a protein, and net approximations of R_S can be made by comparing diffusive characteristics to those of ideal spherical bodies with a given R_S (185). Interestingly though, disorder regions do not commonly share the diffusion characteristics of denatured proteins of similar sizes. This is thought to be as a result of their unusually low hydrophobicity and high net charge as compared to globular proteins (see Figure 5.2.2), and can cause IDP regions to be relatively more compact compared to fully denatured states (186). This overall high net charge of IDPs is also thought responsible for the observation that disordered regions tend to migrate as unusually large molecules by SDS-PAGE. For ordered proteins, the relative proportion of hydrophobic residues is mainly proportional to molecular weight, allowing proportional binding of SDS and comparative PAGE migration. The unusually high net charge of IDPs however, lowers their affinity for SDS, and they subsequently experience lower net negative charge, slower migration through the acrylamide matrix and appear to migrate as larger molecules (187).

The extended structure of IDPs also has direct biochemical consequences. Enhanced proteolysis is commonly observed due to the large accessibility of intrinsically disordered domains to proteases. Techniques that measure protein folding-unfolding events such as differential scanning calorimetry (DSC) or other methods of thermal or chemical denaturation also typically yield atypical results as the there are non-existent, or limited, structural elements to monitor.

Small-angle X-ray scattering (SAXS) is perhaps the most direct measure of a proteins overall size and shape, with the potential to also characterise the conformational distribution sampled by a particular molecule (188). A wide

variety of parameters can be extracted from SAXS data, such as molecular weight, diameter and radius of gyration (R_G) that can help to characterise the level of disorder within a protein. Furthermore, although the generation of accurate structural models from SAXS data is currently quite limited due to the relatively low resolution of the data, computational advances in the processing of SAXS data are starting to significantly enhance it (189).

The many information-rich facets of NMR make it perhaps the most insightful tool for studying proteins, ordered or disordered, in solution albeit with a size limitation of ~40-100 kDa depending on the system and the available hardware (190). NMR spectroscopy provides the unique ability to monitor the structure, dynamics and interactions of proteins in solution at the level of the individual amino acid, and indeed the (magnetically active) nuclei within them. As previously discussed (Sections 3.1 and 4.1 respectively), the chemical shifts of specific moieties, and their relaxation to equilibrium in a magnetic field, are heavily influenced by their covalent and spatial surroundings and their local and overall mobility. The higher levels of protein structure therefore have a major impact on the observed chemical shift dispersion, linewidth and relaxation times measured during NMR experiments. In contrast, the lack of structure within IDPs consequently yields largely distinct data to those of globular proteins enabling the characterisation of their dynamic properties and latent structural propensities.

Analysis of NMR chemical shifts alone is usually sufficient evidence to state if disorder is prevalent within a sample. For instance, minimal chemical shift dispersion within the 1H dimension of ^{1}H-^{15}N HSQC spectra is indicative of an absence of the heterogeneous spatial surroundings involved in secondary structure. Comparisons to the chemical shifts of amino acids occupying known secondary structure elements (chemical shift indexing, CSI) can then provide a quantitative measure of the extent and type of order within a sample (84). For IDPs, this is usually more informative than traditional NOESY-based inter-nuclei distance measurements as the significant backbone dynamics can preclude distinct magnetisation transfer through the nuclear Overhauser effect.

The scope allowed in the analysis of NMR relaxation data for IDPs is somewhat restricted compared to the analysis of globular proteins because a single isotropic correlation time is not sufficient to describe their molecular tumbling in solution. Nonetheless, by recording backbone amide ^{15}N longitu-

dinal (R_1), longitudinal rotating frame ($R_{1\rho}$) and transverse (R_2) relaxation rates and steady state heteronuclear NOE enhancements, dynamic phenomena across fast (ps-ns) and slow (μs-ms) timescales can be probed (see Figure 4.1.4).

Furthermore, use of these relaxation rates can be applied to reduced spectral density mapping (RSDM) to determine the spectral density function $J(\omega)$ at the frequencies of $J(0)$, $J\omega_N$ and $J(0.87\omega_H)$ as previously mentioned. As measures of the extent of relaxation at theses frequencies, they are used as common measures to compare the internal molecular motions of different systems (152), and can help determine latent or partial secondary structure elements in IDPs (191). Kinetic phenomena can also be probed with NMR and, similar to Fourier- transform infrared spectroscopy (FTIR) and mass spectrometry (MS) techniques, experiments designed to monitor the chemical exchange between protons and deuterons can provide residue specific insights into the level of solvent accessibility and the slow time scale motions (seconds-hours) which are indicative of large conformational changes or folding:unfolding events.

5.1.3 Experimental aims

As a result of the relatively sparse information currently available about the TAPR protein, and in particular the biochemical detail of its actin interaction, this chapter will server to document an initial *in silico* investigation into the actin binding region of the TARP protein and the biophysical characterisation regarding its native state.

5.2 Results

5.2.1 *In silico* characterisation

Because it can be said that a proteins structure and function are downstream results of its initial primary amino acid composition and sequence, the long pursued Holy Grail of structural biology is to be able to identify these characteristics from sequence information alone. However, the complexities of folding landscapes, exchanging and minor-populated conformations, post-translational

modifications and interactions with solvent, small molecules or other proteins make this an extremely ambitious goal. Nevertheless there are a growing number of sequence-related trends that can currently be identified, either through sequence comparison or *ab initio* calculations, that are supported with biochemical data. Owing to the relative simplicity of using these web-based tools (although not necessarily the interpretation of the results) *in silico* methods are increasingly being used in the initial characterisation of proteins.

The 100 residue sequence of the TARP actin binding domain (ABD), which had previously been shown to be the minimum actin binding construct (73), was used for all calculations with the omission of the five N-terminal residues remaining from GST cleavage which were present in the physical construct. Initially, the PSIPRED webserver was used (Version 3.3, accessible at: `http://bioinf.cs.ucl.ac.uk/psipred/`) to identify elements of secondary structure in the TARP ABD construct. Methodologically, PSIPRED works by building a sequence homology matrix from multiple protein alignments to the query sequence. This matrix profile is then ranked as a function of regions of solvent accessibility (or conversely, hydrophobicity) and the original sequence is then retrospectively scored according to the structural propensity of the aligned homology profile (192). Results are then returned on a residue- by-residue basis with the most likely type of secondary structure (helix, strand or coil) and an associated confidence value. This approach has scored consistently well in the CASP assessment of *in silico* tools (193).

The PISPRED prediction for the TARP ABD (Figure 5.2.1) shows two regions of α-helix and a short stretch of β-strand, with the remaining protein predicted to be in a random coil conformation. The first region, a 12 residue helix (full length numbering: A749-H761), has previously been identified as homologous to that of other WH2 actin binding helices (73). The second predicted helical region (L783-T792) has not been previously described, with respect to structure or function, but shows very little homology to other characterised proteins, as evidenced from searches carried out with the Basic Local Alignment Search Tool (BLAST, accessible at `http://blast.ncbi.nlm.nih.gov/Blast.cgi`) (Data not shown). The predicted region of β-strand (T799-V781) is particularly short at three residues and with low confidence values, its existence should be treated tentatively.

Disordered protein regions were historically described by an absence of sequence or biochemical characteristics that matched those of known structural elements. An approach derived in 2000 has been able to specifically identify disordered regions by correlating their overall hydrophobicity with their net charge at physiological pH (194). Within this charge-hydrophobicity phase space, a visible threshold distinguishes known ordered proteins from predicted and experimentally characterised disordered (Figure 5.2.2).

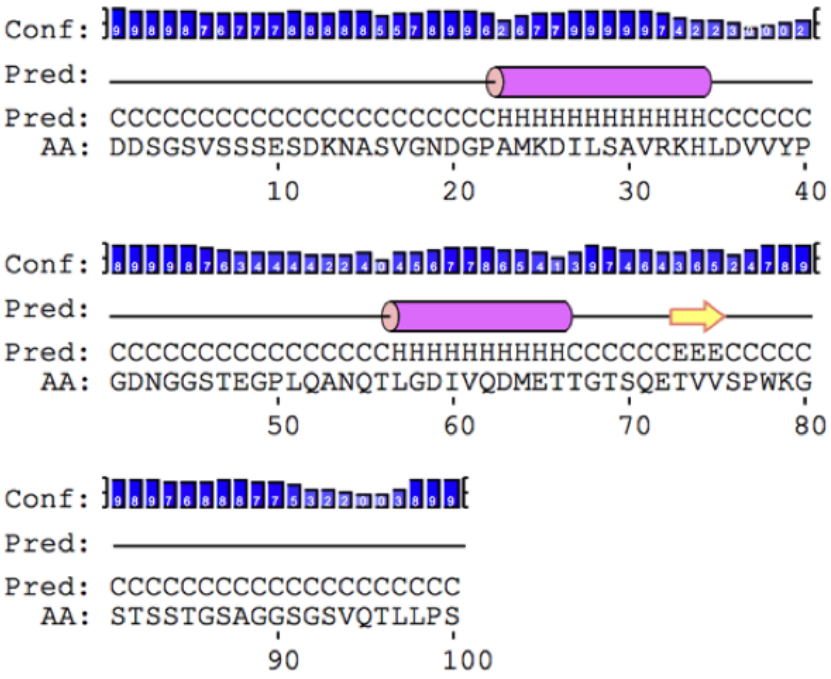

Figure 5.2.1: The adapted output from a PSIPRED v3.3 secondary structure prediction concerning the 100 residue TARP ABD. Annotated helices (H, pink bars), strand (E, yellow arrow) and coil (C, black line) are shown below their respective confidence value (blue bars).

The threshold between order and disorder, as defined by charge-hydrophobicity phase space if define by equation 5.2.1:

$$\langle R \rangle = 2.785 \langle H \rangle - 1.151 \tag{5.2}$$

where <R> and <H> refer to mean net charge and mean Kyte-Doolittle (195) hydrophobicity respectively. The position of the TARP ABD in charge-hydrophobicity phase space (<R> = 0.074, <H> = 0.444, calculated at pH 7.0) lies almost exactly at the interface between ordered and disordered proteins (Figure 5.2.4, yellow star). Neither fully ordered or disordered by this

definition, this could suggest that the TARP ABD contains either an equal proportion of order and disorder promoting regions (a theory not supported by Figure 5.2.1) or that regions with propensity for either order or disorder are detracting from fully indicative characteristics within charge-hydrophobicity phase space.

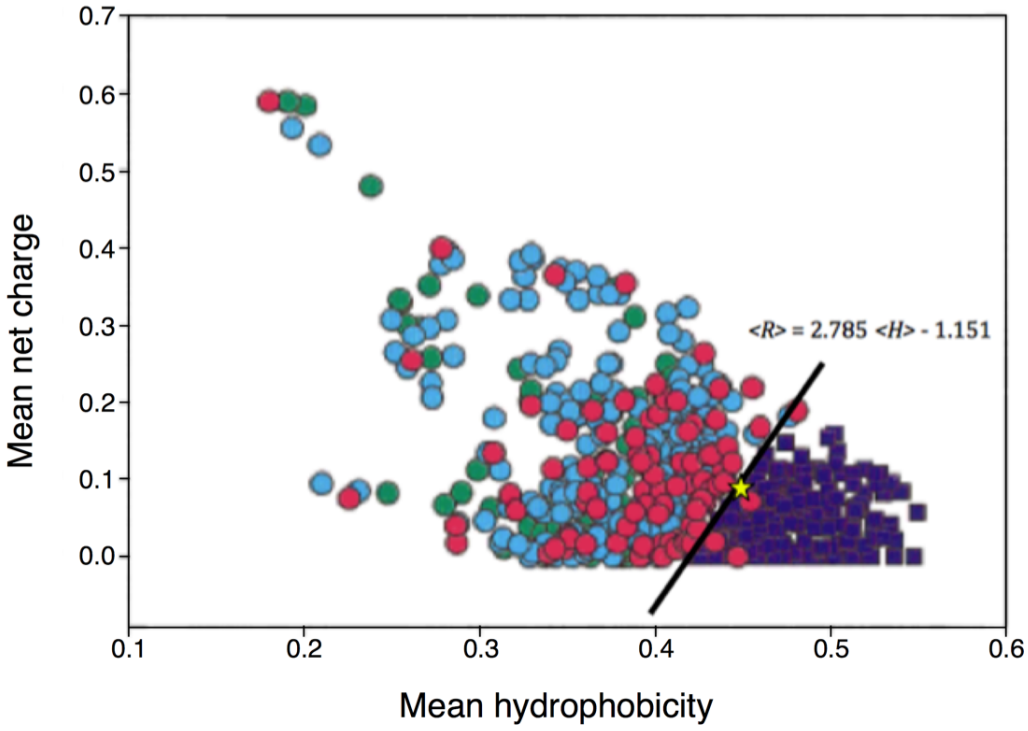

Figure 5.2.2: A mean net charge vs. mean hydrophobicity plot (adapted from (194), figure 3d) with the TARP ABD superimposed (yellow star). Known folded proteins are depicted as squares (dark blue), coloured circles represent known disordered proteins (pink), their sequence based homologues (cyan) and proteins predicted to be disordered (green). The order:disorder boundary is defined, with <R> and <H> denoting mean net charge and mean net hydrophobicity respectively.

A number of methods also exist for locating disordered regions specifically within a sequence, rather than treating a whole domain as an individual entity. The methodologies of these disorder prediction tools can differ significantly through the inclusion of different, or mixed and weighted, components such as physical principles (i.e. charge-hydrophobicity phase space), amino acid composition, homologue based position-specific matrices or conversely predictions of secondary structure or solvent accessibility (196). There are also tools, termed meta-predictors, which seek to combine the results of multiple approaches to best describe the disorder within a system.

The PONDR-FIT web tool (196) is one such meta-predictor (accessible: `http://www.disprot.org/pondr-fit.php`) shown to significantly improve initial predictions (as compared to experimental structural determination) by incorporating multiple sequence-based features. The PONDR-FIT prediction of disorder within the TARP ABD (Figure 5.2.3) agrees well with the PSIPRED structural prediction (Figure 5.2.1) and also a prediction of the average area buried (accessible: `http://web.expasy.org/protscale`); a traditional measure of folded structure elements (197).

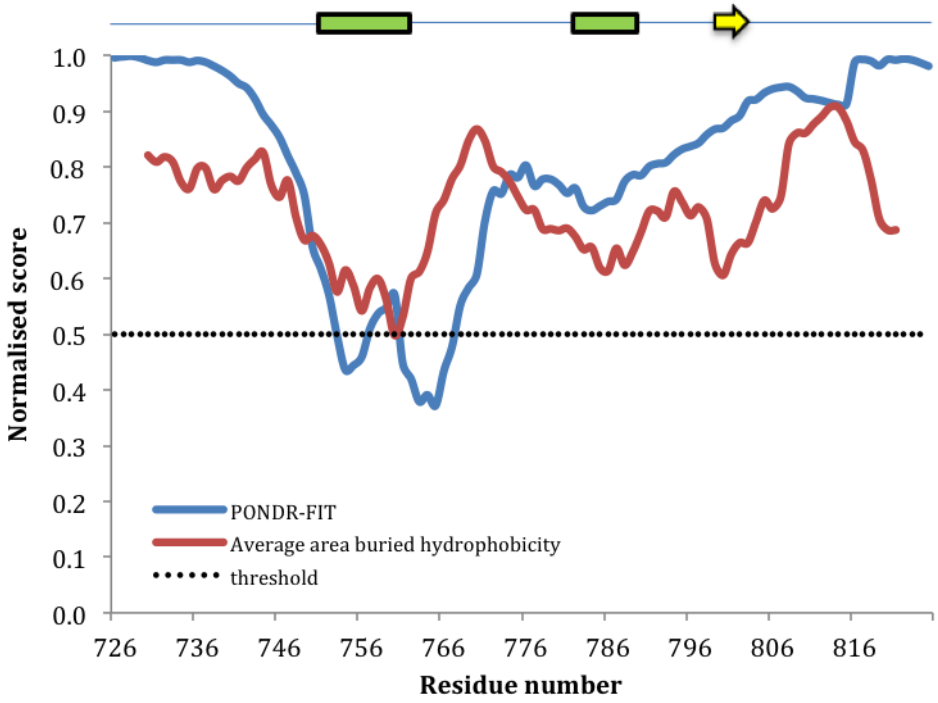

Figure 5.2.3: A comparison of the *in silico* structure predictions for the TARP ABD. The PSIPRED prediction (Figure 5.2.3) is represented as green bars (α-helices) and yellow arrows (β-strand) (top). The PONDR-FIT disorder prediction (blue) and an average area buried hydrophobicity plot (red) (197) of the TARP ABD overlaid. Scores have been normalised, between 1 (full disorder) and 0 (full order). The threshold between order and disorder (PONDR-FIT) is shown (dashed black).

Although the PSIPRED-predicted structural elements match trends in both hydrophobicity (and hence propensity to fold) and a lack of disorder, the relative divergence from a fully disordered score should be taken into account. In this sense, only the presence of the WH2-homologous α-helix 1 is supported by all calculations carried out, as seen by a PONDR-FIT score below the disorder:order threshold of 0.5 (196) (Figure 5.2.3). Nonetheless, it is only a slight, and partial, deviation from disorder.

Although experimental observations should ideally complement *in silico* analysis, this result may also explain the inconclusive position of the TARP ABD within the charge-hydrophobicity phase space. It is clear that the majority of the construct resembles that of largely disordered protein. It may well prove that the regions of predicted secondary structure (in particular helix 1) actually only correspond to propensities for secondary structure. This would agree with observations of other disordered protein-binding domains (174) that present partially formed or non-existent secondary structure elements, which only truly fold upon ligand binding.

5.2.2 Purification and size characterisation

There were no significant practical obstacles to overcome during the purification steps of the TARP ABD (Section 2.2.3). Initial visualisation of the GST-TARP fusion protein by SDS-PAGE (Figure 5.2.4) indicated that the unlabelled fusion protein migrated as a 43.77 kDa protein, although theoretical prediction (`http://www.scripps.edu/~cdputnam/protcalc.html`) estimated the 331 residue protein to be 36.82 kDa.

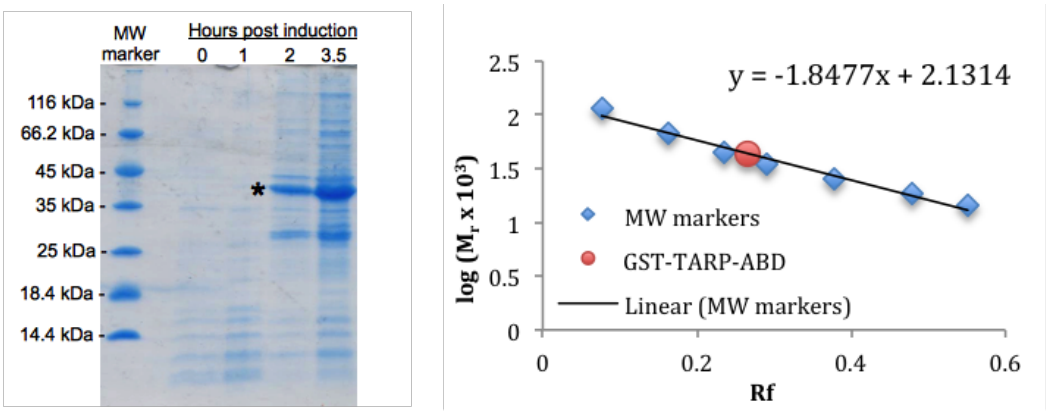

Figure 5.2.4: An example 20% SDS-PAGE gel of the overexpression of the GST-TARP ABD (*) after induced overexpression with 0.8 mM IPTG (left). The apparent molecular weight of the fusion protein was determined by graphical analysis of the fusion protein SDS-PAGE migration (red) relative to the mobility of the molecular weight standards (blue). (Rf = band migration / dye front migration) (right).

Following GST affinity-purification and cleavage of the fusion protein, the molecular weights of the eluted species were assessed. Although free GST is known to dimerise in solution (198), suggesting a molecular weight migration of 52.86 kDa could be expected, analysis of the size exclusion purifica-

tion chromatogram indicated free GST migrated as a 66 kDa protein (Figure 5.2.5, top and bottom right). Analysis of the size exclusion fractions by SDS-PAGE, with its denaturing sample conditions, identified the GST monomer as 29.1 kDa (Figure 5.2.5, bottom left), a figure in better agreement with its expected monomeric weight of 26.43 kDa. Migration of the TARP ABD implied a diffusive profile greater than that of its expected 10.4 kDa by both size exclusion chromatography and SDS-PAGE, observed as a 39.75 kDa and a 15.1 kDa protein respectively (Figure 5.2.5). Both results confirm that the TARP ABD has a larger radius than that expected for a globular protein of the same molecular mass.

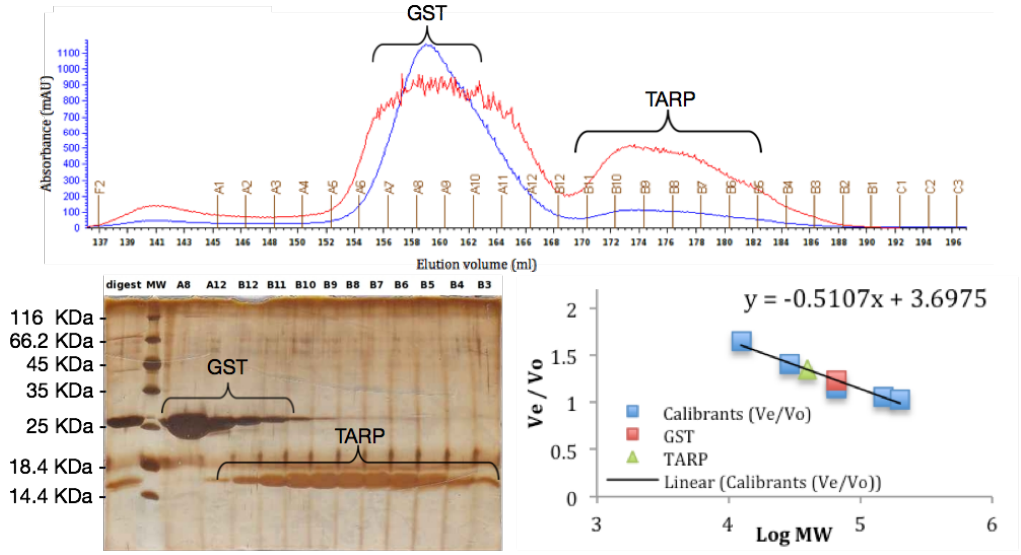

Figure 5.2.5: An example size exclusion chromatogram at 280 nm (blue) and 214 nm (red) of the cleaved GST and TARP ABD sample (Top). Orange alphanumerics denote individually collected 2 ml fractions. Standard protein calibrants (cytochrome c, carbonic anhydrase, bovine serum albumin (BSA), alcohol dehydrogenase and β-amylase) were used to calibrate the S75 column across a molecular weight range of 12.4 kDa to 200 kDa. The linear relationship between elution volume and molecular weight was used to evaluate sample size (Bottom right). An example silver stained 20% SDS-PAGE gel of the cleaved GST, TARP ABD size exclusion fractions. Alphanumeric descriptions correspond to those from the size exclusion purification (Bottom left).

The full length TARP protein has previously been shown to migrate aberrantly on SDS-PAGE (74) and a fragment containing the TARP ABD has also shown a larger than expected molecular weight by size exclusion (73).

The gel filtration elution volumes of the protein calibrants were also used to determine the effective Stokes radius of the TARP ABD. This was carried out by exploiting the correlation between the Stokes radius of a diffusing body and the elution volume of a given sample, combined with the properties of the specific gel filtration column (and packed media) through calculation of the partition coefficient (K_{AV}) (199, 200).

The TARP ABD was shown to diffuse with an $R_s = 27.2$ Å, another measurement far larger than that expected for a spherical protein of 10.4 kDa. In contrast, the most similar diffusing protein, carbonic anhydrase - a 29 kDa spherical protein - has an $R_s = 24$ Å indicating that the TARP ABD is indeed sampling a significantly extended conformation in solution.

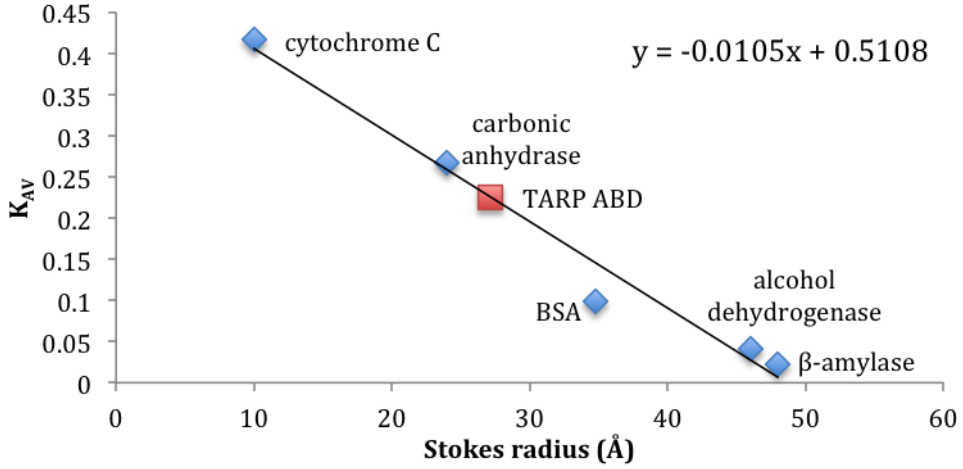

Figure 5.2.6: Determination of effective Stokes radius for the TARP ABD. The known Stokes radii of the protein standards (201), blue diamonds (β-amylase 48 Å, alcohol dehydrogenase 46 Å, BSA 34.8 Å, carbonic anhydrase 24 Å and cytochrome C10 Å) were correlated to the partition coefficient K_{AV}. $K_{AV} = (V_e \ V_0)/(V_t\text{-}V_0)$ where V_e is the elution volume of the sample, V_0 is the void volume of the column, as determined with Blue dextran, and Vt is the total volume of the column, calculated from the height of the resin.

Unfortunately, the migration characteristics of dimerised GST and the TARP ABD limited the efficiency of the size exclusion purification step, with numerous TARP fractions shown to contain GST impurities (Figure 5.2.5. bottom left). However, pooled pure fractions from 1 l of culture consistently yielded $>$ 5 mg of the TARP ABD, which was sufficient for a 1 mM, 0.5 ml NMR sample. Furthermore, the TARP ABD was later shown to be particularly stable, so impure fractions containing notable proportions of the TARP ABD were frozen and added to subsequent purifications prior to size exclusion.

5.2.3 Biophysical characterisation

5.2.3.1 Mass spectrometry

Because a larger molecular weight than expected was seen during purification and initial observations of ^{15}N-HSQC NMR spectra of the TARP ABD yielded significantly fewer amide-correlation peaks than expected (Section 5.2.3.3), priority was given to ascertaining the molecular weight of the purified sample. Electrospray ionisation mass spectrometry (ESI-MS) was used to check the molecular mass of the purified TARP ABD construct (Section 2.2.5) (Figure 5.2.7).

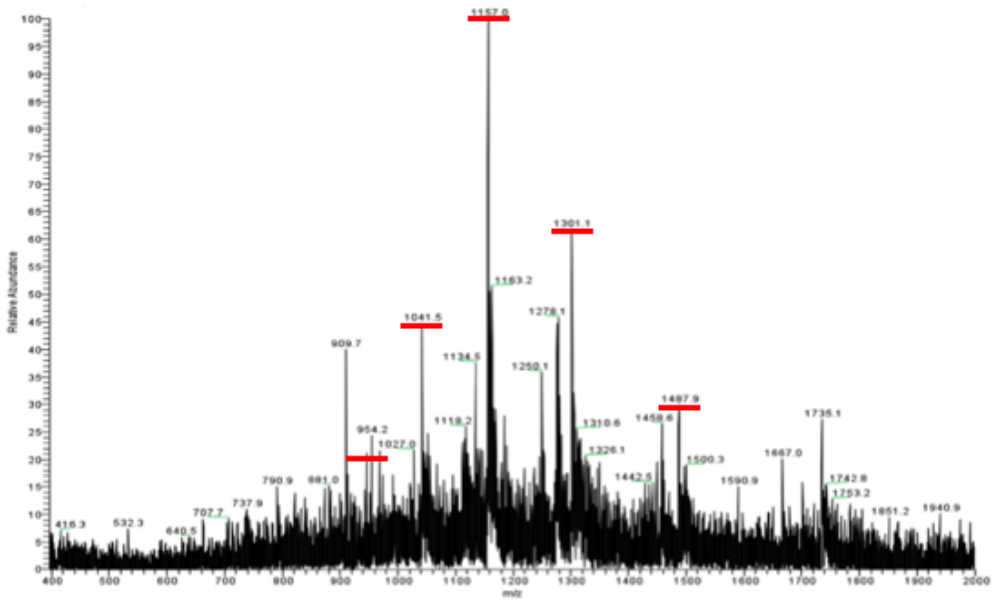

Figure 5.2.7: The electrospray ionisation mass spectrum of the TARP ABD. Highlighted peaks correspond to integer m/z ratios for a mass of 10403 ± 3 Da, in agreement with that of the theoretically predicted molecular weight of 10411 Da.

The sum of the predominant mass:charge peaks (Figure 5.2.7, red bars) indicate a total mass of 10403 Da for the 105 residue TARP ABD construct; a figure in agreement with the *in silico* calculated molecular weight of 10411 Da (calculated as in Section 5.2.2).

5.2.3.2 Synchrotron circular dichroism

The secondary structure of the TARP ABD construct was also probed with synchrotron radiation circular dichroism (SRCD). Standard bench-top circular dichroism measures the differences in the absorption by a sample between left and right plane polarised light, most commonly across the far-ultraviolet (UV) region of the EM spectrum (190-250 nm). For protein samples this can yield quick and non-destructive structural information resulting from the different absorptive properties of the secondary structure elements. There are also rich structural data available in the vacuum ultraviolet (VUV) EM region (150-200 nm), but this is not practically accessible with standard laboratory light sources due to inabilities in maintaining low wavelength flux (202). Modern synchrotron light sources can generate plane-polarised light with significantly increased flux (~103 photons.sec^{-1}) across the VUV region, allowing these extra data to be acquired and simultaneously providing a higher overall signal to noise (for a given sample concentration). SRCD spectra were acquired in collaboration with Professor Bonnie Wallace and Dr Andrew Miles, Birkbeck College, University of London. The SRCD spectra of the TARP ABD suggest that the construct is a largely disordered protein (Figure 5.2.8, Table 5.2.1). Positive changes in ellipticity are stereotypically indicative of helices (195 nm, 208 nm, 222 nm) and strands (185 nm, 192 nm, 212 nm). All SRCD spectra for the TARP ABD are largely devoid of positive signal. The SRCD temperature series presents a relative increase in ellipticity at ~222 nm between 20 °C and 85 °C, indicating an increase in helical structure. Although counter-intuitive with respect to globular proteins (with regards to thermal denaturation), the observation of increasing helical content with increased temperature matches reports in the literature for other IDPs (203205). The phenomenon has not been thoroughly explained (203), and competing theories have been put forward. One hypothesis states that an increase in temperature results in the strengthening of general hydrophobic interactions which facilitates folding. Other research however suggests that increasing temperature is favourable to polyproline helix II formation (which do not necessarily require proline, as would be the case for the TARP ABD) and that this is often mistaken for helical content with CD methods (206, 207).

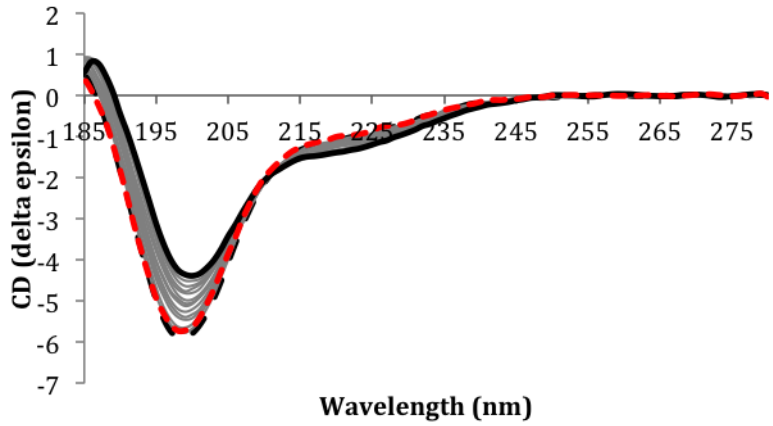

Figure 5.2.8: Synchrotron radiation circular dichroism (SRCD) spectra of the TARP ABD measured at 20 °C (dashed black line). Experiments were repeated in 5 °C increments (solid grey lines) up to 85 °C (solid black line). To check sample stability, the red line indicated the final spectrum, after returning to 20 °C.

Table 5.2.1: The average secondary structure content of the TARP ABD calculated from SRCD spectra at 20 °C and 85 °C (Figure 5.2.3). Values are averaged from the CONTINLL, SELCON and CDDSTR algorithms.

Secondary structure	20 °C (%)	85 °C(%)	Δ%
α-helix	5 ± 1	8 ± 2	3
β-strand	10 ± 1	14 ± 2	4
Disordered	76 ± 1	66 ± 6	-10

5.2.3.3 NMR resonance assignment

The NMR resonances of the TARP ABD domain were assigned using the same backbone walk methodology outlined in Section 3.2.1, by utilising the HNCACB/CBCA(CO)NH and HNCO/HN(CA)CO experimental pairs. The observable TARP ABD backbone amide peaks exhibited extremely limited spectral dispersion (Figure 5.2.10), which hampered the assignment process. To resolve these difficulties, an HNN triple resonance experiment was also acquired. The HNN experiment can aid in the sequential assignment of peaks in crowded spectral regions by visualising the *i-1* and *i+1* amide correlations in the same ^{15}N plane as the *i* amide resonance. The increased spectral dispersion of ^{15}N species then usually allows for a clearer identification of connected spin systems (Figure 5.2.9).

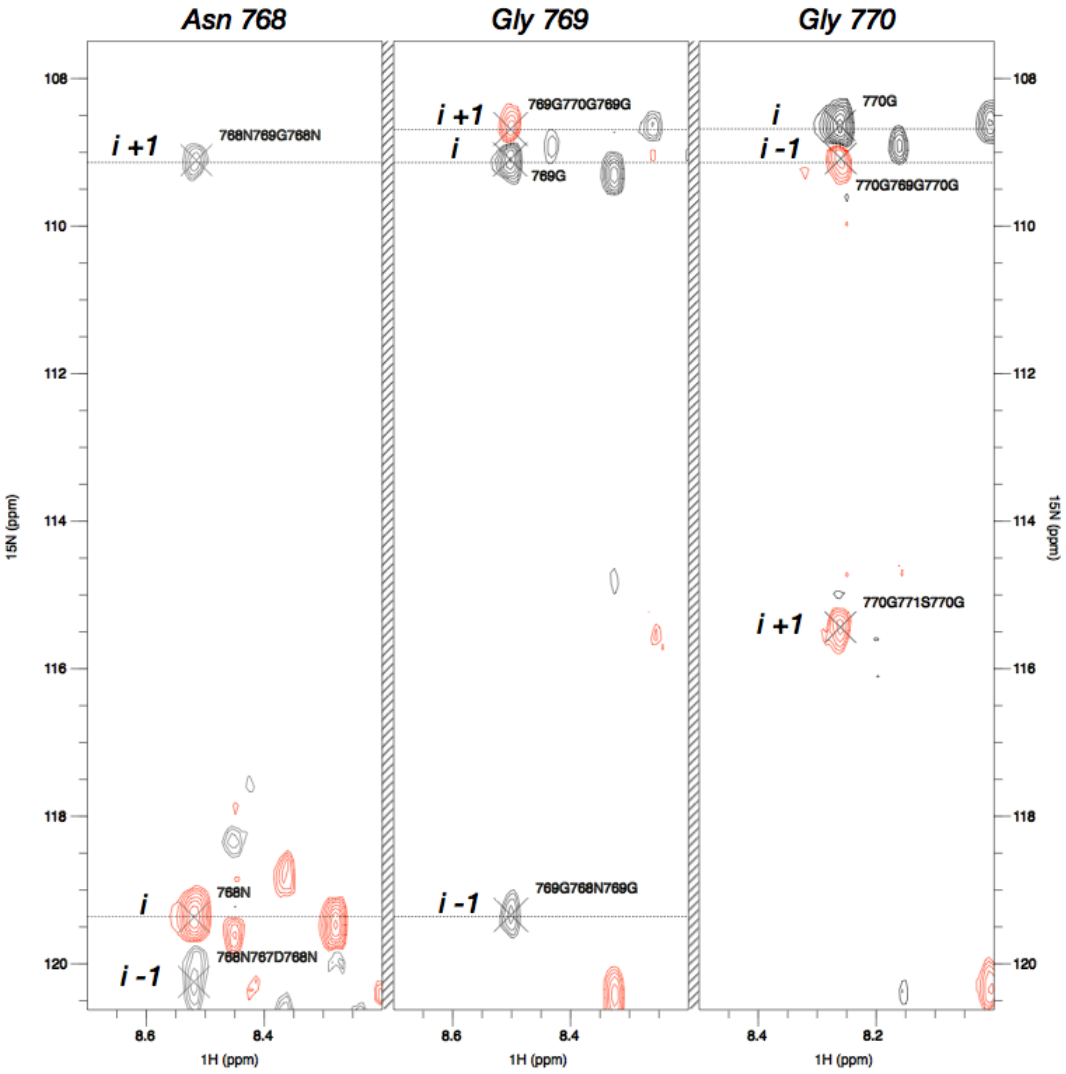

Figure 5.2.9: An example of the backbone-walk assignment process carried out for the TARP ABD construct, using ^{1}H-^{15}N strips from the HNN triple resonance NMR experiment. Crosspeaks correspond to the backbone amide group of either the directly observed residue (*i*), or sequentially preceeding (*i-1*) or proceeding (*i+1*) residues. The ambiguous assignment of *i*, *i-1* and *i+1* peaks was carried out according to the methodology described by Panchal et al. (208) regarding the phase of the peak (red = positive, black = negative) and the residue type, if known.

In total, 80% of the TARP ABD backbone (Asp726 - Ser808) was assigned sequentially (Figure 5.2.8), excluding direct assignment of prolines 747, 765, 775 and 802, although the chemical shifts of their Cα Cβ and C resonances were observed from an *i+1* perspective. A full list of the assigned residues and resonance is included in Appendix I. The were no other peaks in any spectrum acquired for the TARP ABD that could even partially account for the remaining C-terminal 17 residues (Ser809 - Ser825) and hence were unable

to be assigned. Eleven peaks were identified which corresponded to minor conformations of assigned residues, as evidenced by identical chemical shifts and *i* and *i-1* peak patterns in the HNCACB/CBCA(CO)NH and HNCO HN(CA)CO spectra.

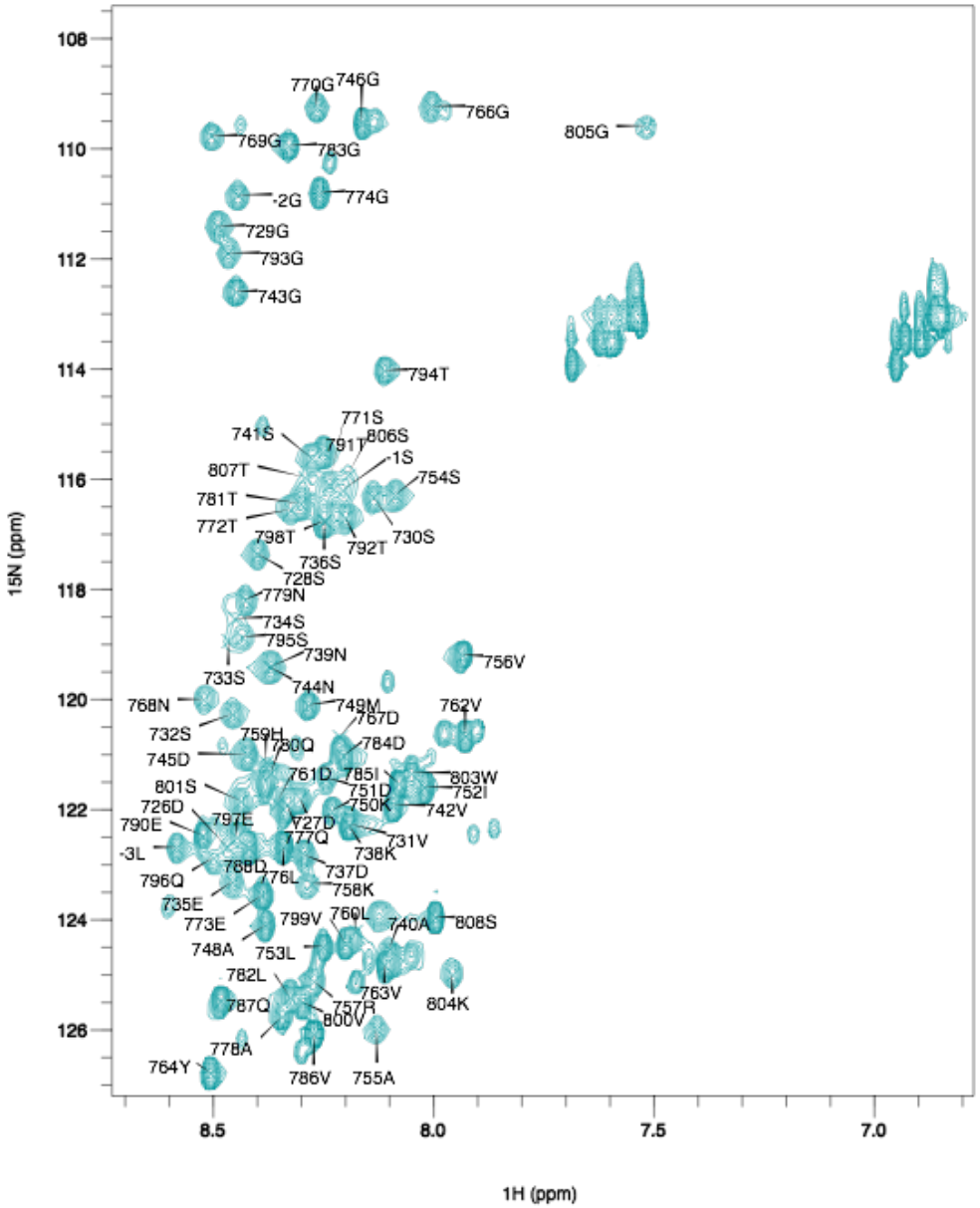

Figure 5.2.10: The assigned ^{1}H-^{15}N HSQC spectrum at 800 MHz of the TARP ABD. The spectrum has been cropped and the Trp 803 side chain has been omitted for clarity. Side chain resonances (1H 6.8-7.7 ppm) were not assigned.

The failure to observe the 17 C-terminal residues of the TARP ABD with standard double and triple resonance NMR is most likely explained by peak broadening resulting from an intermediate chemical (inter- or intra-molecular

interactions) or conformational exchange process. That is, that the exchange process between two or more species is occurring on a timescale highly similar to the reciprocal of the difference in chemical shift frequency between the two states, as observed by NMR at a particular magnetic field strength.

Both chemical and conformational exchange phenomena are of course kinetic in nature, however the populations of the interconverting states will be governed by thermodynamics, and as such are open to manipulation. With this in mind, a series of ^{1}H-^{15}N HSQC experiments were carried out, using the TARP ABD, between 5 and 35 °C (283-298 K), in 5 °C increments (Figure 5.2.11).

The aim of the series was to try and force the exchange processes into relative slow (low temperature) or fast exchange (high temperature) and therefore observe the individual (slow exchange limit) or averaged (fast exchange limit) signals. Unfortunately, no extra NMR signals peaks were observed, and at, or approaching 308 K, a number of peaks were lost, presumably from the increased solvent exchange. Although the C-terminus of the TARP ABD does not contain any significantly hydrophobic or charged patches, because the temperature series was not successful in identifying any extra peaks, pH and NaCl titrations were also carried out to try and inhibit any potential inter- or intramolecular interactions (Figure 5.2.12). Both titrations did not significantly alter the conformational behaviour of the TARP ABD as evidenced by consistent NH chemical shifts. The [NaCl] titration did not yield any new peaks and all peaks remained observable throughout the titration. Unfortunately, the pH titration also failed to introduce any new peaks with respect to the C-terminal region of the TARP ABD, although after and including pH 8.6, a number of minor degradation peaks were observed.

Although the amide resonances remained fixed and observable between pH 4.5 to 7.7, as of the pH 8.6 titration point only V762, V763, V786 and W803 were visible, and by pH 9.9 no assigned peaks were present. Similar to the temperature titration, this loss of signal is due to enhanced proton exchange with the solvent, as the pH approaches the pKas of the amide protons. It is interesting to note the longevity of a number of residues. The amide pK_as for valine and tryptophan (9.62 and 9.39 respectively) are not particularly high compared to other amino acids. The disappearance of other valine cross peaks is also interesting.

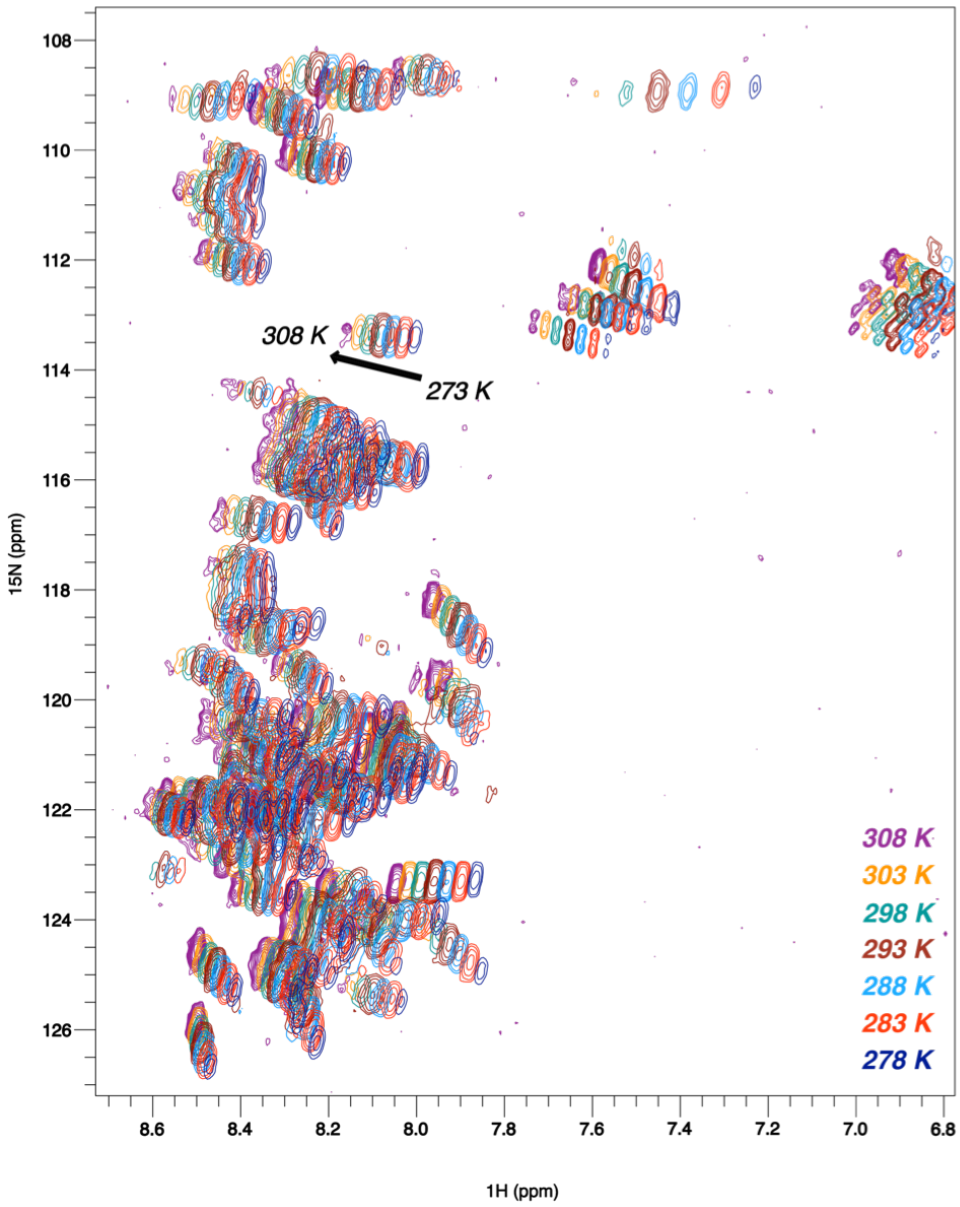

Figure 5.2.11: The ^{1}H-^{15}N HSQC temperature series at 600 MHz of the TARP ABD between 278 and 308 K. The direction of the upfield shift associated with increasing temperature is marked. Assignments have been omitted for clarity.

One hypothesis is that these amide protons are participating in hydrogen bonding, potentially altering their effective pK_as and leading to increased endurance in a high pH environment. It is worth noting that V762 and V763 are just C-terminal to the WH2 homologous helix 1. It could therefore be the case that hydrogen bonding from some partial helical content is causing these observations, but the possibility of long-range H bond interaction between or amongst any or all of these residues cannot be ruled out.

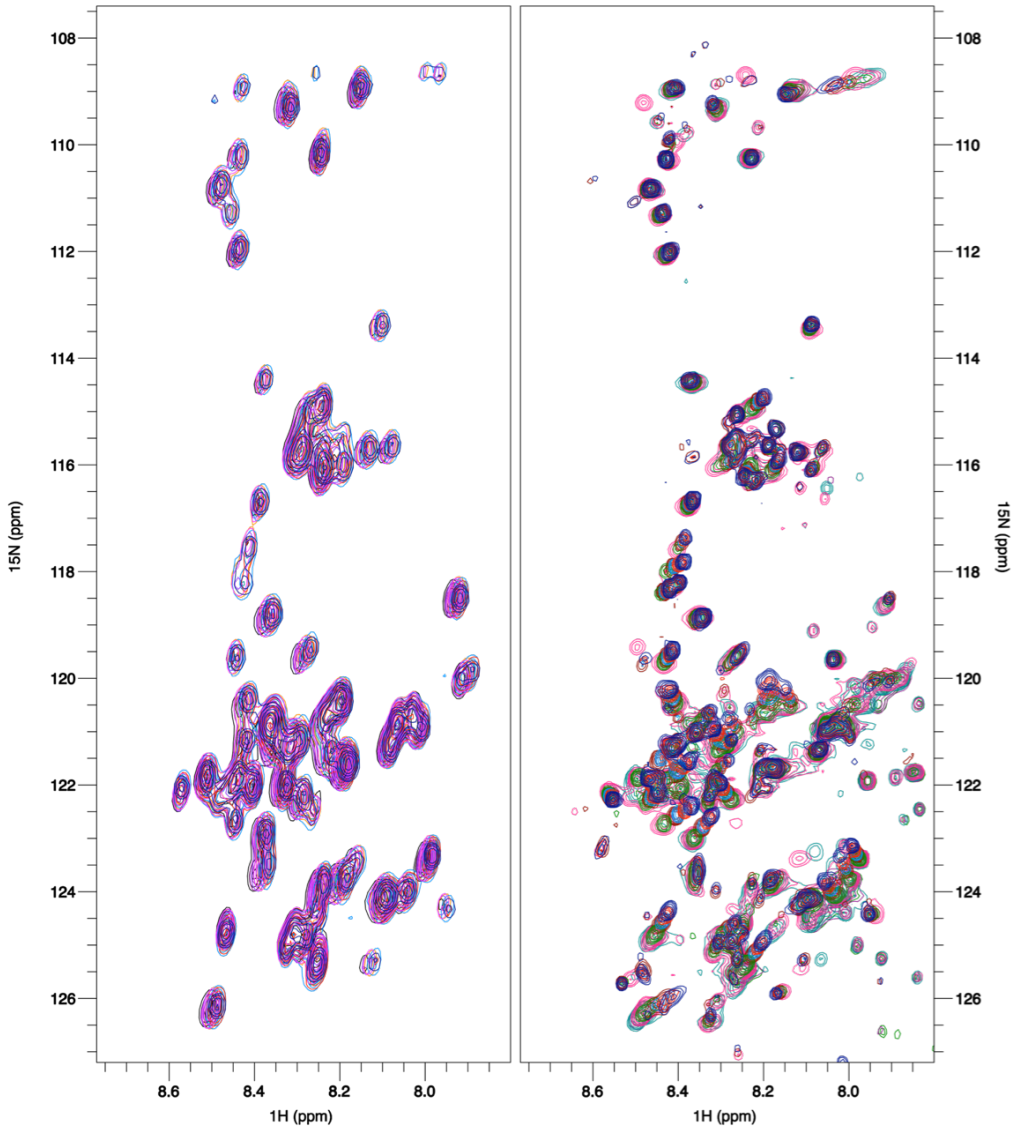

Figure 5.2.12: Cropped backbone amide regions of overlaid ^{1}H-^{15}N HSQC [NaCl] (left) and pH (right) titration series of the TARP ABD at 600 and 800 MHz. Concentrations of NaCl used were: 0-mM (Navy), 10 mM (red) , 20 mM (sky blue), 30 mM (maroon), 50 mM (teal), 75 mM (orange), 0.1 M (magenta), 0.15 M (mauve), 0.2 M (purple) and 0.25 M (black). The pH for each experiment was calculated after the addition of either 1 M NaOH or HCl (see Section 2.2.3): pH 4.5 (navy), 4.7 (red), 5 (sky blue), 6.4 (green), 6.5 (maroon), 7.3 (pink), 7.7 (teal), 8.6 (purple), 8.8 (green) and 9.9 (black).

Upon searching the literature, studies were found which conducted ^{1}H-^{15}N HSQC temperature series to investigate hydrogen bonding in proteins (209, 210). By correlating the overall, isotope weighted, changes in chemical shift across the temperature range, assumptions can be made about the level of hydrogen bond participation by particular amide protons. Amino acid

temperature coefficients for the TARP ABD (Figure 5.2.13) were therefore calculated from Figure 5.2.9. The result suggests that a number of residues may be engaged in hydrogen bonding. For comparison to the literature, chemical shifts were reported as (-δ) and a threshold value of -4.7 ppb . K^{-1} is stated as an 85% probability threshold for hydrogen bond participation. Eleven resides had temperature coeffients more positive than the 85% probability threshold (Figure 5.2.13, Table 5.2.2).

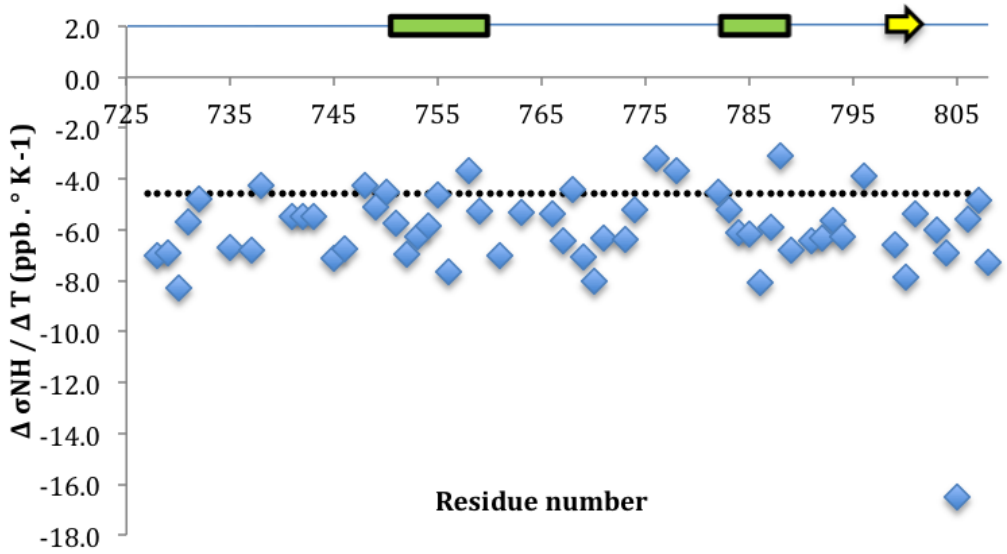

Figure 5.2.13: The amide proton temperature coefficients for the TARP ABD as calculated from the ^{1}H-^{15}N HSQC temperature series (Figure 5.2.9). 20 residues have been omitted because they could not be assigned at all temperatures. Chemical shift changes are reported as σ (-δ). Temperature coefficient more positive than -4.6 ppb K^{-1} (black dashed line) indicate the likelihood of hydrogen bonding exceeds 85% (209). The PSIPRED secondary structure prediction is overlaid.

Interestingly, the residues tolerant to high pH did not all correspond to those for which hydrogen bond formation was likely. However, five residues matched regions of predicted helical secondary structure.

Table 5.2.2: The TARP ABD residues with temperature coefficients greater than the 85% probability threshold for hydrogen bonding.

Residue	$\Delta\ \sigma_{NH}$ / ΔT (ppb.K^{-1})	Predicted region
L738	-4.27	-
A748	-4.26	-
K750	-4.54	Helix 1
A755	-4.64	Helix 1
K758	-3.70	Helix 1
N768	-4.42	-
L776	-3.19	-
A778	-3.68	-
L782	-4.55	Helix 2
D788	-3.08	Helix 2
E796	-3.88	-

5.2.3.4 NMR comparison of TARP constructs

Because of the inability to resolve the C-terminal region of the TARP ABD with NMR, a shorter, GST-tagged construct was generated (Section 2.2.1.2) that omitted the C-terminal 17 residues. The purpose of the construct, $TARP_{726\text{-}808}$, was to observe whether the C-terminus was important to the behaviour of the larger TARP ABD. Figure 5.2.14 shows the overlaid ^{1}H-^{15}N HSQC spectra of the two constructs.

The ^{1}H-^{15}N HSQC spectra of the two constructs are largely identical. The only differences were two peaks, unassignable in the TARP ABD construct, which were not observed in the spectrum of the shorter construct. Triple resonance peaks did exist for the two unassigned HSQC peaks, but they were indicative of threonine residues (Cα δ = 63.17, 63.30 ppm and Cβ δ = 70.77, 70.63 ppm) and as no threonines remained unassigned, they were thought to have arisen from a significantly different minor populated state. If the inability to observe the C-terminus of the full length protein was indeed a result of dynamic phenomena, the removal of this potential for conformational destabilisation could result in a restriction of the conformational sampling or a minimisation of the conformational populations. Nonetheless, the assignable region of the TARP ABD was not perturbed in its major-form behaviour by the truncation.

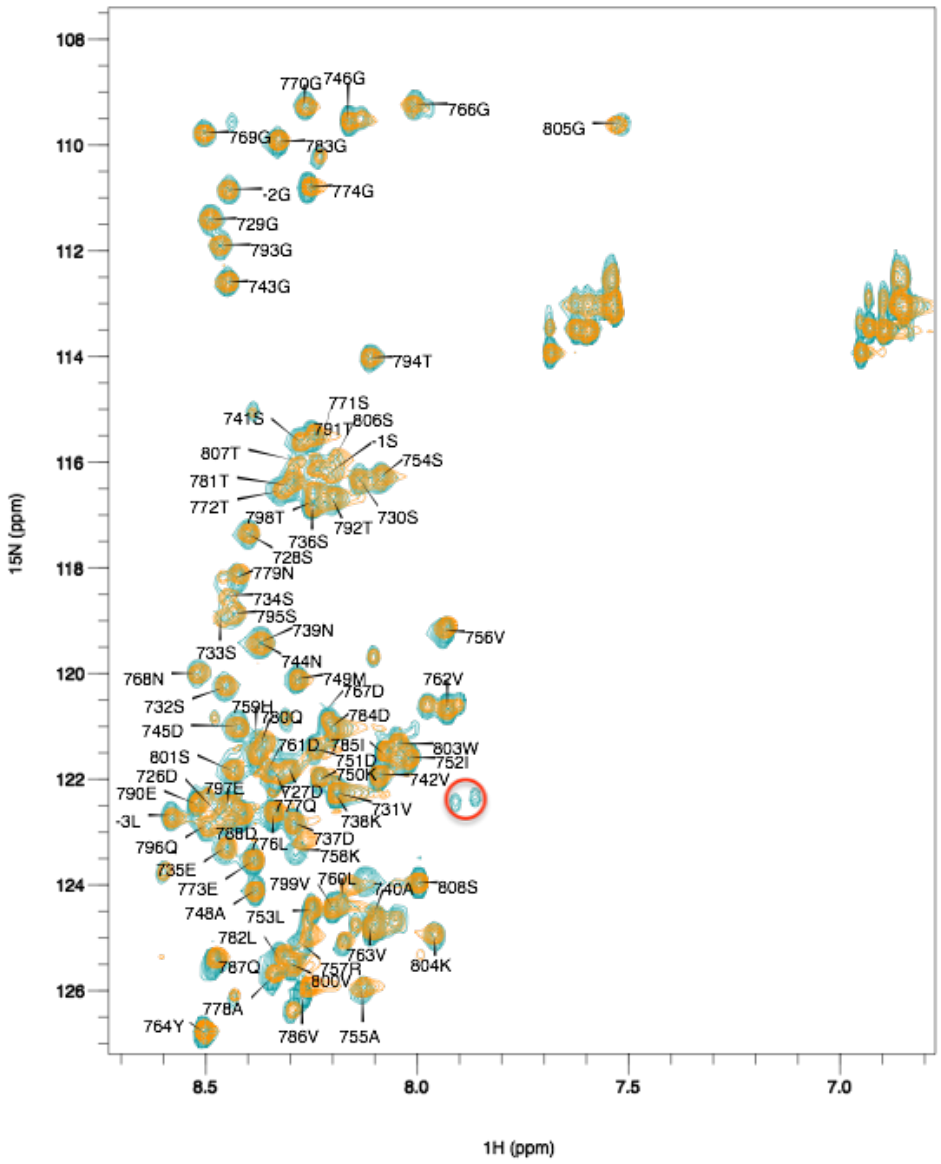

Figure 5.2.14: The assigned ^{1}H-^{15}N HSQC spectrum (800 MHz) of the TARP ABD (teal) and the overlaid ^{1}H-^{15}N HSQC spectrum (800 MHz) of the $TARP_{726\text{-}808}$ construct (orange). The spectrum has been cropped and the Trp 803 side chain has been omitted for clarity. The two peaks unobservable in the $TARP_{726\text{-}808}$ have been circled in red.

The secondary structure content of the shorter construct was also probed with SRCD (Table 5.2.3, Figure 5.2.15).

The results of the SRCD analysis suggests that there was not a significant difference between the structural content of the two constructs. Although the relative size of the shorter construct is 7.6% smaller than that of the TARP ABD, the expected increase in relative helical content (assuming the C-

terminus was indeed disordered, as predicted) falls within the standard error of the measurements. It should also be noted that for largely disordered proteins, content reported as β-strand can also be indicative of disorder (Dr Andrew Miles, personal communication 04/05/2012).

Table 5.2.3: The average secondary structure content of the TARP ABD and $TARP_{726\text{-}808}$ construct as calculated from SRCD spectra at 20 °C (Figure 5.2.15). Values are averaged from the CONTINLL, SELCON and CDDSTR algorithms.

Secondary structure	TARP ABD (%)	$TARP_{726\text{-}808}$ (%)
α-helix	5 ± 1	5 ± 2
β-strand	10 ± 1	14 ± 1
Disordered	76 ± 1	69 ± 3

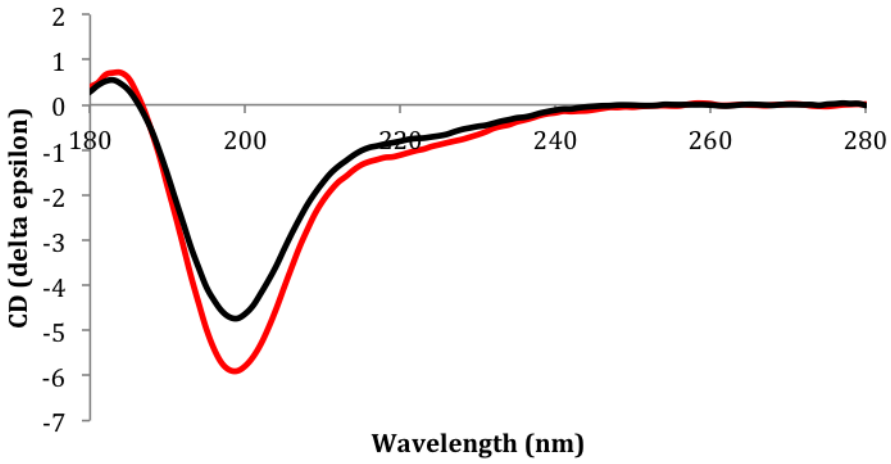

Figure 5.2.15: Synchrotron radiation circular dichroism (SRCD) spectra of the TARP ABD (red) and the shorter $TARP_{726\text{-}808}$ construct (black) measured at 20 °C.

The inability to resolve the TARP ABD C-terminus with NMR was thus deemed inconsequential and further experimentation planned to investigate the actin:TARP ABD interaction by using the larger TARP ABD construct. If the C-terminus were to be involved in the actin-binding function of TARP, then firstly this would be observable to other biophysical techniques (ITC, SRCD), and secondly it may even be that the actin bound state stabilises the C-terminus into an observable population with NMR.

5.2.3.5 Chemical shift indexing

As previously mentioned (Section 5.1.2), a detailed analysis of a proteins NMR chemical shifts can highlight regions of secondary structure by comparison to the chemical shifts of residues in fully folded, or random coil (and hence disordered), regions. To this end, the standard chemical shift indexing (CSI) technique compares whether a given resonance is upfield or downfield shifted (the structural implications differ between moieties) compared to a reference random coil value (84). Addressing structural content, and especially structural propensity, is therefore heavily reliant on accurate random coil chemical shifts. However, for proteins with significant disorder, a recent paper has shown that determination of structural content can be greatly improved by redefining the reference random coil values through incorporating a shielding term to account for the chemical shift perturbations, away from true random coil, resulting from their flanking residues (211).

This neighbour-corrected IDP (ncIDP) analysis was carried out for the TARP ABD by uploading the backbone ^{15}N and ^{13}Cα chemical shifts to the ncIDP webserver (accessible at: `http://nmr.chem.rug.nl/press/?p=1`) (Figure 5.2.16, green and red respectively). ^{15}N secondary chemical shift were included as negative values to simplify Figure 5.2.16 so that the 1 and -1 thresholds correspond to fully formed α-helix and β-strand respectively for both nuclei. Only the backbone ^{15}N and ^{13}Cα resonances were included in the analysis due to their apparent sensitivity to IDP structural propensities (211), although Hα, ^{13}Cα, ^{13}Cβ and ^{13}C$'$ resonances are usually deemed more sensitive, and hence are more commonly used, in the CSI analysis of predominantly folded proteins.

For comparison, a further method of chemical shift based structure assessment was carried out for the TARP ABD. The SSP method (212) creates a residue, and moiety-weighted structural propensity score, based upon how sensitive the individual chemical shifts are to secondary structure. All assigned TARP ABD chemical shifts were used, with the SSP program (212) (freely available at: `http://pound.med.utoronto.ca/software.html`) (Figure 5.2.17). Similar to the ncIDP analysis, SSP scores above 1, and below -1, signify fully formed α-helices or β-strands respectively.

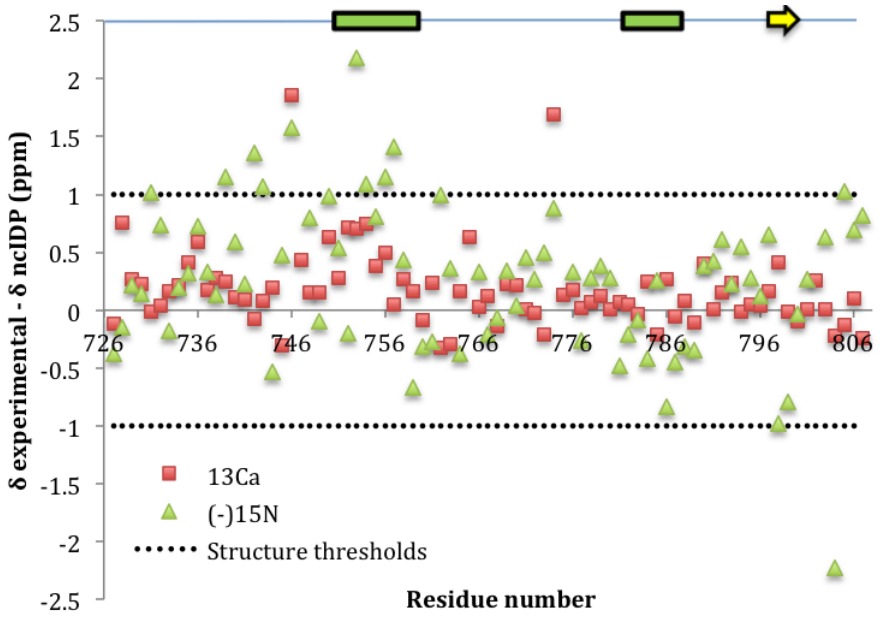

Figure 5.2.16: The ncIDP chemical shift analysis carried out for the TARP ABD using ^{13}Cα and ^{15}N backbone resonances. The ^{15}N secondary chemical shifts are reported as negative so that the structural thresholds are comparable. SSP scores $\geq$ 1 are indicative of fully formed α-helix and scores $\leq$ -1 indicate fully formed β-strand. The PSIPRED structure prediction of the TARP ABD is overlaid.

Both chemical shift analyses suggest that a level of α-helical content exists across the M749-D761 region, corresponding to the WH2-homologous, *in silico* predicted helix. With respect to the ncIDP analysis, the chemical shifts of four residues (L753, S754, V756 and R757) are clearly beyond the -helical threshold and K750 is only marginally below, with an ncIDP Δ chemical shift of 0.985. The ncIDP secondary chemical shifts for the other helix (G783-T792) remain largely random coil, although two residues (T798 and V799) have shifts that would seem to correlate to the PSIPRED predicted β-strand region.

The SSP analysis is not as explicitly suggestive of any fully formed secondary structure elements within the TARP ABD. SSP scores are weighted so as to create an effective probability describing the distribution of conformational states, as they are based on the assumption that secondary chemical shifts represent the average conformational state (212, 213). The magnitude of the SSP scores for the TARP ABD suggest that most residues are largely occupying random coil conformations, with an SSP score of 0 $/pm$ 0.1. The most prominent patch of apparent helical content matches that of the WH2 homologous, PSIPRED predicted α-helix; albeit with only a partial, 20%, conformational population. However, when summed and averaged with respect

to the 105 residue construct, the overall helical propensity shown by SSP of 3.2%, although slight, is in good agreement with the value of 5% ±1 of the overall helical content as determined by SRCD.

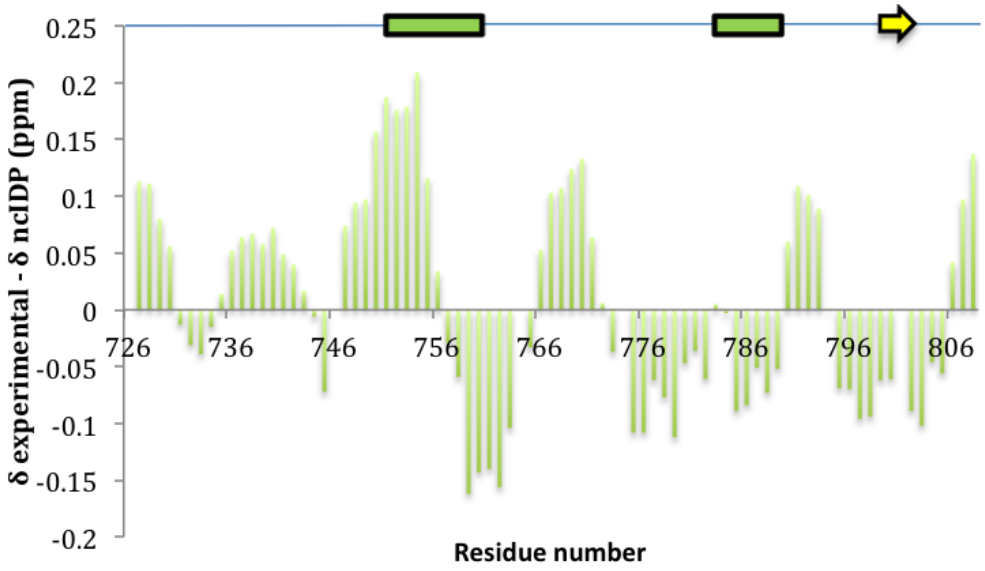

Figure 5.2.17: The SSP score chemical shift analysis carried out for the TARP ABD, using all assigned resonances. The SSP authors (212) describe thresholds of 1 and -1 for fully helical or strand conformations respectively. The PSIPRED structure prediction of the TARP ABD is overlaid.

5.2.3.6 Relaxation analysis

Chapter 4 describes how ^{15}N NMR relaxation measurements can report on the level and frequency of dynamic phenomena within a molecular system. The aforementioned experiments using the TARP ABD construct have shown a significant preference for conformational disorder across the protein, with small regions of variable secondary structure propensities. However, owing to sensitivity across faster timescales, ^{15}N relaxation can also provide a different measure of structural propensities by distinguishing between the reduced mobility of folded regions and the comparative freedom of disordered ones.

To this end, T_1 and T_2 ^{15}N relaxation experiments were to be carried out, with modified ^{1}H-^{15}N HSQC experiments (Section 4.3), for the TARP ABD at 500 MHz. Due to the restricted spectral dispersion, some residues were unable to be analysed as the degree of crosspeak overlap was so severe that it would be impossible to separate the decay in intensity for each peak for relaxation

analysis. The analysed R_1 and R_2 values throughout the protein are shown in Figures 5.2.18-19 respectively.

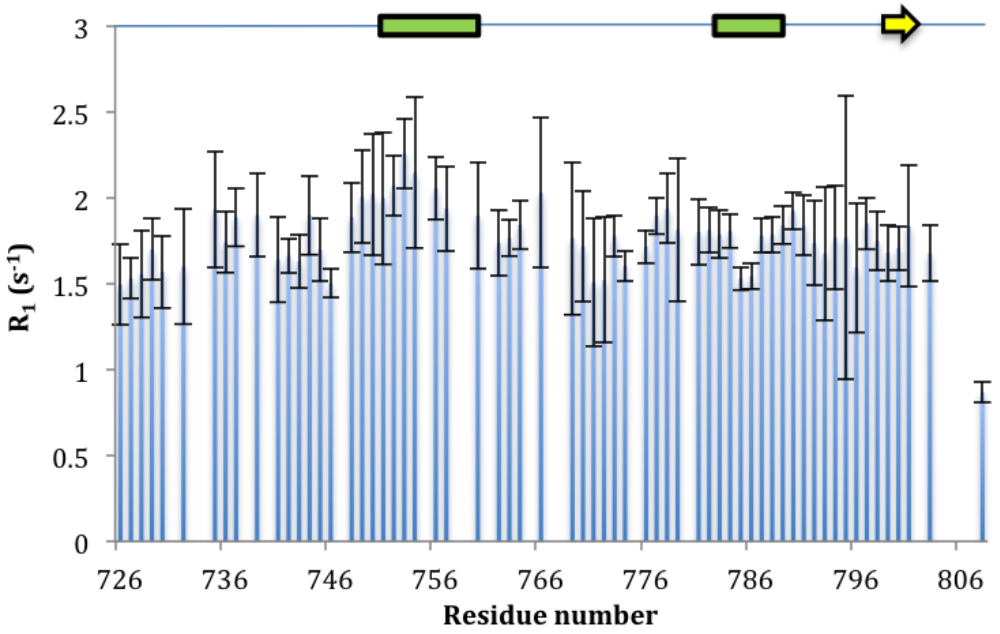

Figure 5.2.18: The R_1 rate at 500 MHz for the TARP ABD construct. Errors in rate were calculated from the standard deviation in peak height from repeated delay times using the relax software package (104).

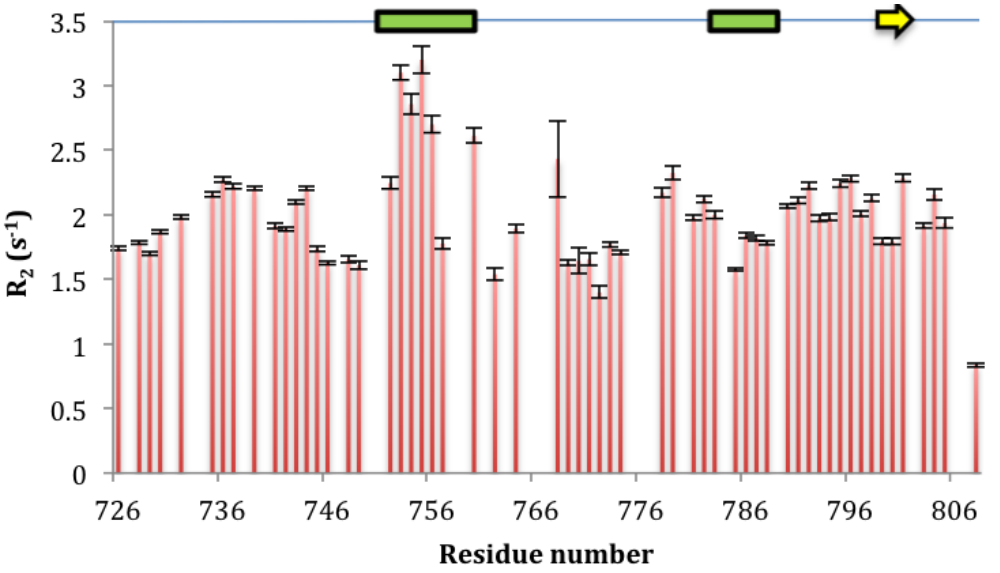

Figure 5.2.19: The R_2 rate at 500 MHz for the TARP ABD construct. Errors in rate were calculated from the standard deviation in peak height from repeated delay times using the relax software package (104).

The relaxation data acquired for the TARP ABD shows an inherent problem in that the R_2 relaxation rate (average $R_2 = 2.00 \pm 0.04$ s^{-1}) was compa-

rable to, if not slower than, that of the calculated R_1 rate (average R_1 = 1.77 ± 0.22 s^{-1}) for the majority of the analysed residues. Owing to the nature of longitudinal and transverse relaxation, wherein x-y plane magnetisation must return to equilibrium (T_2) before the z-component (T_1), this observation is theoretically impossible (214). Nonetheless, the observed T_2 exponential decays (Figure 5.2.20, right) suggest that abnormal peak height decays perturbed the accurate exponential fitting of the data, and that this lead to an improper determination of T_2 relaxation times. Further analysis of the relaxation data was therefore not possible, and additional experiments were not acquired.

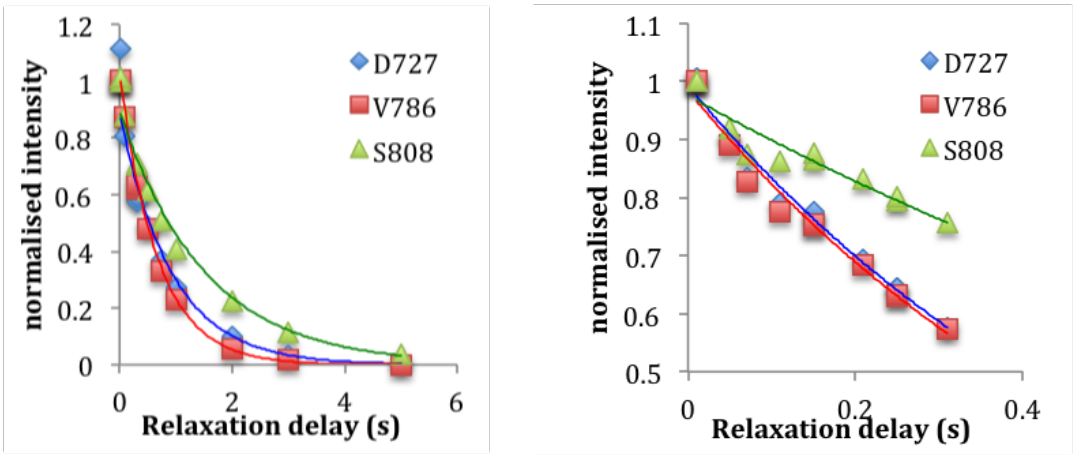

Figure 5.2.20: Examples of T_1 (left) and T_2 (right) peak intensity decay curves (normalised to the first time point) fit to mono-exponentials for the residues D727, V786 and S808 of the TARP ABD.

5.3 Discussion

The analyses undertaken for the TARP ABD were undertaken with a view to characterising its structural content and behaviour to better understand its interaction with actin monomers. Described in detail in Chapter 6, many WH2 domain-containing proteins comprise a short α-helix which binds to actin across a hydrophobic cleft that lies between two actin subdomains (215).

The multiple *in silico* analyses carried out suggest that the TARP ABD is largely occupying a disordered conformation. Perhaps unsurprisingly, the most significant prediction of secondary structure (and lack of disorder) agrees well with the region of the sequence shown homologous to other WH2 heli-

cal domains. However, the calculated confidence levels, and proximities to theoretical thresholds, seem to indicate that, at best, the A749-H761 region contains only partial helical content, and that fully formed helix should not be expected from experimental data. The position of the TARP ABD within the charge-hydrophobicity phase space also suggests that neither fully disordered or ordered amino acid sequence determinants predominate.

Analyses of the gel filtration purification for the isolated TARP ABD construct suggested that it diffused in solution with a significantly greater effective RS than that expected for a protein of 10.4 kDa. According to Csizmok et al. (216), the extent of conformational extension for an IDP can be predicted based upon the ratio of its theoretical molecular weight verses that calculated from gel filtration elution volumes; wherein a MW_{app} / MW of 2-3 indicates a pre-molten globule state and a MW_{app} / MW of 4-6 is indicative of a protein largely sampling a fully extended, random coil conformation. With a MW_{app} / MW of 3.82, I propose that the vast majority of the TARP ABD construct is principally sampling a fully extended random coil conformation. Assuming that any level of secondary structure across the A749-H761 region would result in a certain level of conformation restriction, partial helix formation may explain why the MW_{app} / MW of the average conformation ratio is slightly below the random coil range.

One potential apparent criticism to this hypothesis is the possibility of TARP ABD multimerisation. This possibility was not directly analysed, but both electrospray ionisation mass spectrometry (ESI-MS) and NMR data allude to a monomeric TARP ABD species. Primarily, because ESI-MS applies a soft-ionisation technique, m/z peaks from covalently or electrostatically attached multimeric species are commonly observed (217, 218); and this was not the case in the TARP ABD ESI-MS spectrum. Similarly, multimeric species of any multiplicity were not observed via SDS-PAGE, although non-denaturing native PAGE was not explicitly carried out. Secondly, the NMR spectrum of the TARP ABD construct did not contain any extra peaks or many peaks, which were determined to be from minor or alternate conformations that could possibly suggest a multimeric species. If the TARP ABD were indeed amassed as a ~39.75 kDa species as suggested by gel filtration, the associated increase in T_1 and T_2 relaxation rates would have severely hampered NMR analysis due to poor signal to noise and broadened signals respectively; and these traits were also not observed. This argument could be countered with the hypothe-

sis that the TARP ABD C-terminus, unamenable to NMR, may have provided a site of dynamic dimerisation - hence occluding itself from observation due to exchange broadening and allowing relatively symmetric behaviour across the remaining N-terminus. Yet, the near-identical ^{1}H-^{15}N HSQC spectra of both the TARP ABD and the shorter 726-808 construct for the C-terminal peaks 806-808 (which would be perturbed by such an interaction) also rules this out.

The dominance of conformational disorder is also apparent from both SRCD and NMR spectroscopic observations. SRCD analysis indicates that the TARP ABD is ~95% disordered, assuming that the β-strand contributions to absorbance at 185 nm, 192 nm, 212 nm are indeed indicative of disorder as has been stated. It could be that the calculated 10% β-strand exists, yet it was neither predicted by *in silico* means or, more importantly, discernible in the NMR chemical shift analysis. Interestingly, only 5% of the TARP ABD was suggestive of an α-helical conformation. The WH2-homologous region was predicted to exist as a 12 residue α-helix by PSIPRED and, within a 105 residue construct, this represents a theoretical prediction of 11.4%. The acquired SRCD data therefore strengthens the argument that helical content is present but only partially formed at any one time.

Studies of the TARP ABD with NMR spectroscopy offered information at a residue-specific resolution on the conformational behaviour and secondary structure content of the construct. The initial ^{1}H-^{15}N HSQC spectrum displayed an extremely reduced spectral dispersion throughout the ^{1}H dimension which is entirely characteristic of a predominantly disordered protein (219) (Note: Figure 3.3.1 presents an example of the spectral dispersion present in a ^{1}H-^{15}N HSQC spectrum of an ordered protein).

At a more direct, albeit semi-quantitative, level, the employed chemical shift indexing methods, specifically designed to account for the properties of IDPs, also imply that the TARP ABD construct is largely disordered in solution. Both techniques predicted some level of helical content aligning to the region of WH2 homologues, although there was some disparity regarding the degree of propensity for helix formation between the two methods. The ncIDP analysis suggested that seven ^{15}N chemical shifts were highly indicative of helical structure, although only one residue (Gly746) possessed both ^{13}Cα and ^{15}N ncIDP secondary chemical shifts above the structural threshold. This may be a measure of partial helix formation although a more rigorous approach would

be to discuss all the available data with weighted scores such as that provided by the SSP program. The region associated with helical propensity from SSP analysis matches that predicted by both *in silico* calculations and ncIDP analysis yet the result is far less evocative of true helix formation. Although the ncIPD analysis takes into account the chemical shift shielding from preceding and proceeding residues, the unification of all available chemical shifts into a single weighted scores is perhaps the more accurate method as in this instance, the depiction of SSP partial helical content is also in agreement with both SRCD data and *in silico* results.

The investigation into the predisposition for hydrogen bond formation yielded no conclusive results. There were no definitive areas of significant propensity for hydrogen bond formation within the TARP ABD, however the observation that most residues experienced temperature associated shifts close to the threshold for significant hydrogen bond formation also correlates with the idea of partial WH2 helix formation.

The erroneous T_2 relaxation decays for the TARP ABD, which inhibited dynamic analysis with NMR, were unfortunate and puzzling. Figure 4.1.5 describes the general relationships between T_1 and T_2 as a function of the correlation time of the molecule and the strength of the applied B_0 magnetic field. Theoretically, T_1 and T_2 are only ever seen to be of a similar magnitude when the apparent local molecular motions, which significantly contribute to T_2 relaxation, are akin to that of small molecules at relatively low magnetic field strengths. At 500 MHz field strength, the observed T_1 and T_2 relaxation times should have been wholly distinct for a macromolecule of the TARP ABDs size. Errors in experiment acquisition, spectral processing or relaxation rate analysis could possibly account for this problem, but although all computational methodologies were checked and repeated multiple times, no obvious source of error could be determined.

Ultimately, the quality of the observed T_2 decay curves suggest that the calculated T_2 values do not reflect a real, single, relaxation process, as the data does not conform to a single exponential decay. Fitting to double exponentials was attempted, but no statistical improvement in the fittings was observed (data not shown).

Chapter 6

Characterising the TARP:actin Complex

6.1 Introduction

6.1.1 The controversy surrounding intrinsic disorder

The notion of intrinsic protein disorder is, in theory, well supported by recent experimental data as described in Section 5.1. However, there are still scientists that argue about the functional significance of intrinsic disorder *in vivo*, primarily because most IDP binding regions can be shown to stabilise into fully formed secondary structure elements through ligand interactions (220). In this regard, it is suggested that the observation of persistent intrinsic disorder is in fact an artefact of non-host *in vitro* purification and that IDPs would be better termed "proteins waiting for a partner" (PWPs) (221). The inference being that PWPs, in the unbound, anti-folded state, never truly persist in the complex and crowded *in vivo* environment and therefore have no functional significance until a binding partner is present. The 3D structure acquired upon binding, and how it facilitates the interaction, is then thought of as the principally significant functional component, superseding that of the initial conformation. Disorder, as measured *in vitro*, is therefore only regarded as a reflection of an atypical flexibility *in vivo*, and that this is simply an extension of the well documented examples of protein flexibility, in fully folded proteins, which facilitate induced fit binding.

In my opinion, and those of other IDP proponents (222), this is somewhat of a semantic argument designed to maintain the classical structure-function paradigm. By all definitions it is agreed that on a fundamental level, IDPs represent a distinct class of protein species and that the majority of IDPs do indeed transition to folded states upon ligand binding. But owing to the prevalence of IDPs throughout Eukarya, it is reasonable to assume that intrinsic disorder, as a characteristic, has evolved due to the type of functionality it can bestow. Undeniably, this logic does not necessarily identify "disorder" specifically as the sought after trait, but although the importance of fold-mediated interactions should not be disregarded, the unique biophysical traits of disordered regions, as well as their potential to augment the interactions of their folded states, suggests that protein disorder does indeed have distinct functional significance.

For instance, the extended nature of disordered proteins can itself be a direct functional asset. One example can be witnessed from microtubule associated proteins (MAPs), a group of proteins which surround the cytoskeletal microtubule filaments and project a disordered domain out into solution. Using atomic force microscopy, the coat of MAP projection domains has been shown to exert to a long-range resistance to compression whilst maintaining a phosphorylation-dependent fixed length. These "entropic brushes" therefore act to maintain microtubule spacing as a direct consequence of disorder (223).

Specific ligand binding mechanisms have also been shown to exist which exploit the lack, or transient nature, of structural elements in IDPs. IDP:protein "fuzziness" (224) describes the spectrum of bound state polymorphism which can range from static fuzziness, wherein a folded IDP element can bind in multiple different conformations, to dynamic fuzziness, where the bound state of the IDP also samples a significant conformational ensemble. Similarly, experimental observations (225) in support of the "fly- casting" binding mechanism proposed by Shoemaker et al. (226) imply that initial low affinity / high specificity interactions can facilitate the folding of high affinity interaction sites within IDPs. By reducing the potential for conformational search, this mechanism acts to smooth the energy landscapes of interaction and folding, leading to enhanced binding kinetics.

But of course, the folding of any protein is in fact a molecular binding reaction (227), albeit an intra-molecular one, with a precise balance between

the associated entropic and enthalpic changes. However, perhaps as a result of the enormous body of work undertaken in understanding ordered proteins, folding and inter-molecular interactions have almost become conceptually independent, when in fact they are analogous. This statement emphasises the potential for cohesion between a theory of disordered proteins and the traditional structure function paradigm. Although IDPs reside in a unique region of the structure-function continuum, they share the fundamental principles of primary sequence encoded function, chemical constituency and the scope for chemical interactions. Their distinct biophysical characteristics simply offer an alternate solution to molecular challenges and although not every disordered region will confer function, a figurative dismissal of intrinsic disorder as "awaiting significance" may well prove short sighted.

6.1.2 Actin, actin binding proteins and the WH2 domain

Actin is a highly conserved, 42 kDa protein which can self associate into extensive polymerised chains to form the microfilament components of the eukaryotic cytoskeleton (228). In a standard cytoplasmic environment, actin microfilaments provide a comprehensive morphological scaffold and facilitate vesicular trafficking by providing transit routes. The molecular forces generated by rapid filament polymerisation *in vivo* also enable dynamic cellular morphologies, such as those involved in migration or phagocytosis (229). But perhaps the most well known function of actin filaments is as a component of the contractile myofibril units within muscle tissues, which generate practical forces from the ATP dependent interactions of actin (thin) and myosin based (thick) filaments.

Monomeric, globular, actin (G-actin) is comprised of four subdomains in a planar arrangement, which impose a distinct structural and functional polarity between opposing molecular faces (Figure 6.1.1).

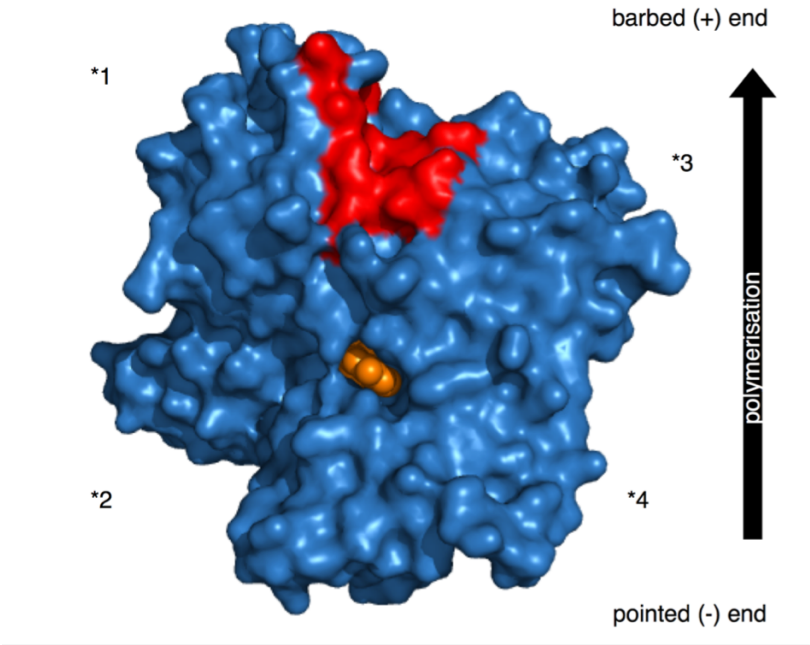

Figure 6.1.1: A surface representation of the 42 kDa G-actin monomer as derived by X-ray diffraction to 1.54 Å (PDB code: 1J6Z) (230), with its four labelled subdomains (*). The barbed-end direction of actin filament extension is shown and the ADP binding site (ADP = orange spheres) and the residues that line the hydrophobic cleft (red, residues 143,144,146,148, 168, 341, 345, 346, 349, 315 and 355) are displayed. Created with the PyMOL molecular graphics program (50).

The pointed (or minus) end lies at the apical interface between subdomains 2 and 4 and forms an ATP binding site, whilst the barbed (or plus) end lies at the opposing face between subdomains 1 and 3, creating an "actin minus end" binding site (Figure 6.1.1). The structural polarity therefore also applies to actin filaments, which extend in a barbed endtopointed end fashion, although a slight offset of 166°, which results in the actin filament aesthetically resembling a helix of two twisted linear filaments (231).

G-actin is a catalyst of ATP hydrolysis in the presence of Mg^{2+} ions, albeit a slow one, with a standard ATP turnover of 0.3 s^{-1} (232), however, hydrolysis itself is a key step in the actin polymerisation:depolymerisation mechanism. Following G-actin:ATP association, the resulting conformational change facilitates the accommodation of an additional ATP-bound actin monomer at its barbed end. Once actin has self-associated, the hydrolysis of ATP to ADP + P_i then occurs, making the process of filament elongation incredibly energy expensive, especially as the products of hydrolysis remain confined to

the nucleotide binding cleft as a result of the arrangement between associated actin monomers. For the formation of new actin filaments (F-actin) to occur, nucleation complexes of trimeric F-actin are required to overcome the initial destabilisation of simultaneous depolymerisation (233) and once formed, the concentration of free G-actin:ATP (the critical concentration, Cc) must also remain high to continually favour K_{on} over K_{off}. Once polymerisation has reached a steady state equilibrium ($K_{on} \approx K_{off}$), actin filaments are said to "treadmill"; an analogy which describes a relatively fixed filament length whilst the constituent actin monomers physically pass along the filament with a barbed-to-pointed end directionality due to filament association and dissociation occurring at different ends.

Actin polymerisation does not readily occur *in vitro* at low concentrations, although in the presence of Mg^{2+} and ATP, it can self-polymerise at concentrations above 100 μM, wherein the rate of successful binding collisions supersedes the depolymerisation rate. However, actin polymerisation never truly occurs unchecked *in vivo* due to a plethora of regulatory proteins which function to modify the inherent rates of polymerisation or depolymerisation. Broadly termed "actin binding proteins" (ABPs), both G-actin- and F-actin-specific ABPs act to modulate actin polymerisation (or depolymerisation) kinetics to maintain the constitutive functional roles actin provides (234).

One such group of ABPs are the actin related proteins (Arp) 2 and 3, which share 45% sequence identity to actin and are structurally very similar. Arp 2 and 3 are able to form a stable complex together with five other proteins (actin related protein complex (ARPC) 1-5) that serve to mimic the actin nucleating homotrimer. The Arp2/3 complex then effectively caps the minus end of the filaments and stops their decay, promoting rapid elongation. The Arp2/3 complex is also capable of nucleating actin filaments from the sides of pre-existing filaments at an included angle of ~70° aiding in the creation of vast branching networks (235, 236). Conversely, the cofilin family of proteins are actin depolymerisation factors and are present in all eukaryotic cells. Cofilins can bind along the length of the actin filament and induce a tighter helical conformation to subsequently destabilise the filament. The kinetic off-rate at the minus end is then greatly increased and swift degradation of the filament ensues (237).

A particular subgroup of ABPs was first identified as a result of studying the underlying mutations responsible for Wiskott-Aldrich syndrome (238), a condition which primarily impairs the human immune system and results in eczema, thrombocytopenia and recurrent infections. Interestingly, the isolated protein in question (Wiskott-Aldrich syndrome protein (WASP)), although unknown at the time, was shown to share striking sequence homology to other, recognised, G-actin binding proteins and significant sequence conservation throughout the eukaryotic kingdom (77). The actin binding domain was termed Wiskott-Aldrich homology domain 2 (WH2). The most investigated WH2-ABP is thymosin-β 4, a G-actin sequestering protein, which is the only known WH2 domain containing protein to promote filament depolymerisation. However, in all WH2 variants experimentally analysed, a short N-terminal α-helix is observed that is comprised of a highly conserved amino acid motif: MxxIxxfxkxxLK that is shown to span the actin hydrophobic cleft (I136-G146) (Figure 6.1.2).

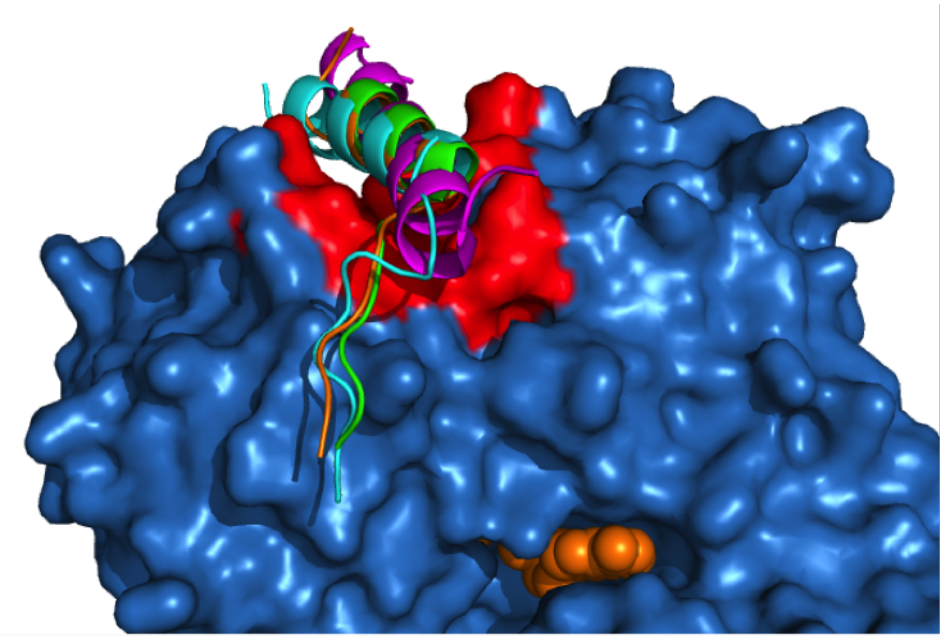

Figure 6.1.2: A surface representation of a G-actin monomer (blue, pdb: 1T44 (240)) and superimposed actin-aligned structures of the WH2 domains from the actin binding proteins spire (orange, pdb: 3MN7 (239)), WASP (green, pdb: 2A3Z (119)), thymosin-β 4 (magenta, pdb: 1T44) and ciboulot (cyan, pdb: 1SQK (241)). The actin hydrophobic cleft is highlighted in red and the bound ATP nucleotide is shown as orange spheres. Created with the PyMOL molecular graphics program (50).

Of the conserved residues, M1 and I4 locate to the actin facing side of the α-helix, whilst L12 and K13 localise to the C-terminal extended region of the WH2 domain that interacts with actin towards the proximal regions of subdomains 2 and 4 (239). Lowercase f7 and k9 refer to residues only conserved in thymosin-β 4 homologues and are thought to impart the increased actin affinity necessary for G-actin sequestration (77). The TARP ABD contains a homologous region to the WH2 domain, and is therefore expected to bind actin in a similar manner.

6.1.3 Experimental aims

The TARP protein, and the 100 residue construct chosen for analysis in Chapter 5, have previously been shown to bind actin and serve as a platform for actin polymerisation; however no biochemical or structural information exists for the interaction itself. The aim of Chapter 6 will be to describe the actin interaction of the TARP actin binding domain with NMR and isothermal titration calorimetry (ITC) and to calculate and discuss potential *in silico* models for the actin bound state of the TARP WH2-homologous region.

6.2 Results

6.2.1 NMR spectroscopy

Owing to its potential for residue specific insights, NMR spectroscopy was used in an attempt to characterise the actin interaction of the TARP ABD. Although the TARP ABD:actin K_d was not known prior to experiment design and acquisition, in order to skew the interaction in favour of a majority bound state, an excess of G-actin was required. However, because G-actin self- polymerisation occurs in the presence of ATP at ~100 μM and Mg^{2+}, a severe limitation became the useable concentration of TARP ABD. Unfortunately, triple resonance data could therefore not be acquired for the bound state owing to unrealistic experiment time as a consequence of the low TARP concentration. Nonetheless, a 2D ^{1}H-^{15}N HSQC spectrum was acquired for a sample containing 31 μM TARP ABD and 85 μM G-actin, using 256 scans to best match the signal to other TARP HSQC spectra (Figure 6.2.1).

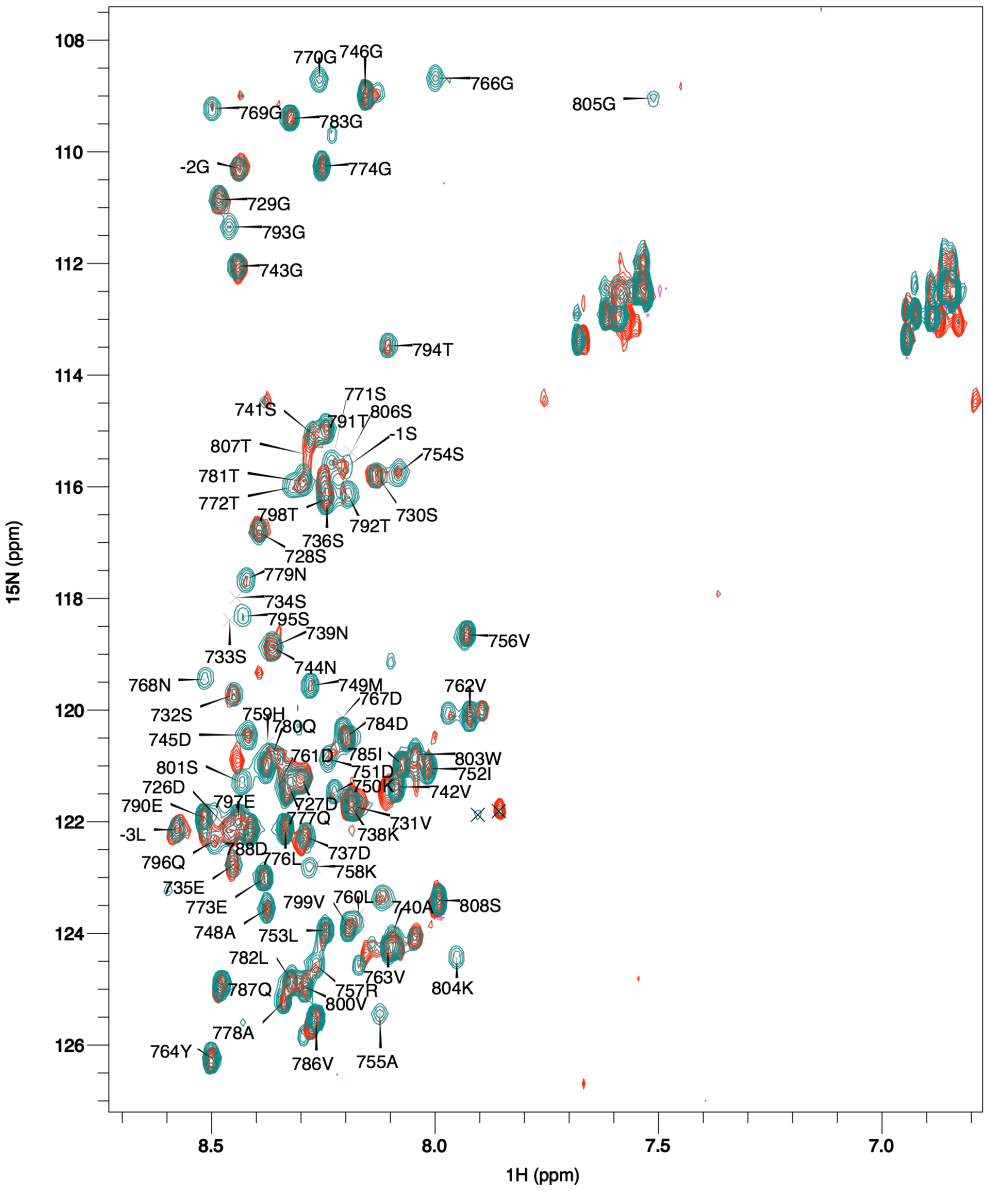

Figure 6.2.1: The assigned ^{1}H-^{15}N HSQC spectrum at 800 MHz of the TARP ABD (teal) and the overlaid ^{1}H-^{15}N HSQC of a 56 μM 1:1 ratio of the TARP ABD and G-actin (red).

The amide crosspeaks that were observable in the $_{1}$H-$_{15}$N HSQC spectrum of the TARP ABD in the presence of actin match previously assigned chemical shifts; however a number of peaks had lost relative intensity or become completely unobservable and no additional peaks were observed. Jewett et al. have show in previous work that the TARP ABD fragment is able to bind actin without inducing filament polymerisation (76). The observations of the $_{1}$H-$_{15}$N HSQC spectrum of the TARP ABD in the presence of actin could be explained by either incomplete binding in intermediate exchange in the NMR

timescale or an increased T_1 time for the bound residues due to the influence on molecular tumbling from the 42 kDa actin monomer.

In an attempt to extract some information regarding the residues of the TARP ABD involved in the actin interaction, the relative peak intensities of the two spectra were compared. Peak intensities from each spectrum were first internally scaled as a ratio of their intensity compared to a glycine residue that had been assigned to the N-terminal GST-cleavage site ("-2G" Figure 6.2.1). This was deemed the best possible reference peak because of a similar intensity in both spectra and a consistent chemical shift, suggesting it was unperturbed by the presence of actin. The respective peak height ratios were then compared to analyse the residue specific differences. As a result of peak overlap, 12 peaks were not included in the analysis due to an inability to confidently distinguish individual intensities (Figure 6.2.2).

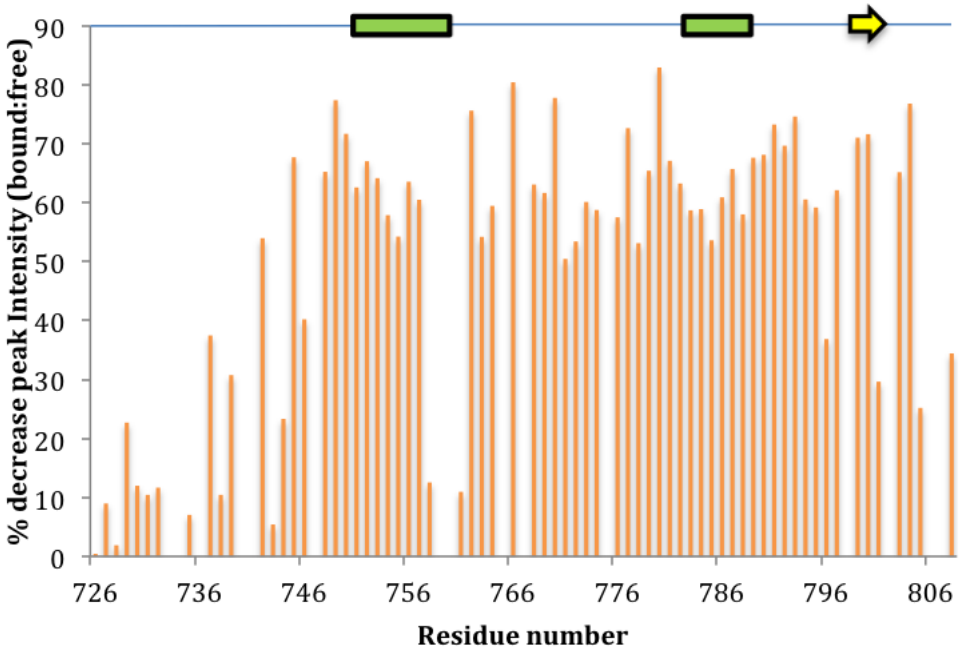

Figure 6.2.2: The observed changes in ^{1}H-^{15}N HSQC peak intensities, of the assigned TARP residues, in the presence of excess actin. The PSIPRED secondary structure prediction of the TARP ABD is overlaid.

The peak intensity analysis suggests that the region N-terminal to the predicted WH2 α-helix (D726-G743) does not interact with G-actin to any significant extent. A clear increase in consistent peak perturbations correlates to the N-terminus of the predicted WH2 α-helix, suggesting an interaction is does indeed taking place. Interestingly, a similar level of peak perturbation

persists across the remaining residues of the construct, indicating that the C-terminal region may also interact with the G-actin monomer to a significant extent.

6.2.2 Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) provides a means of measuring the enthalpic response (ΔH) associated with molecular interactions. Although the magnitude of these heat changes is frequently on the order of 10×10^{-10} °C, by accurately following ΔH as a function of the ratio of ligand and receptor concentrations, thermodynamic parameters can be determined for an observed reaction by use of the following equations and data.

As an example of an ITC titration under ideal conditions (i.e. when ΔH is measurable and saturation of binding occurs), Figure 6.2.3 shows how n (reaction stoichiometry), ΔH and K_a (association constant) can be determined from the integration of the isotherm:

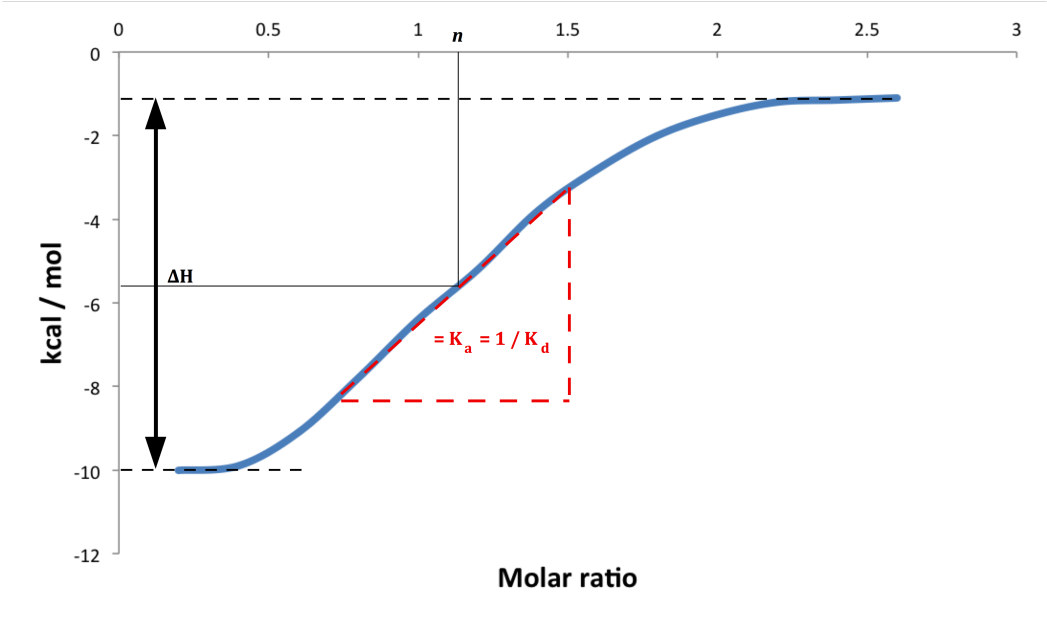

Figure 6.2.3: An example of an integrated ITC binding isotherm. The methods of ascertaining preliminary values for ΔH (overall energy change), n reaction stoichiometry (molar ratio at $^1/_2\Delta$H) and the equilibrium constant (either K_a or $1/K_d$ as a function of the gradient) are shown.

Equations 6.2.1 and 6.2.2 then show how, with the measured enthalpic change (ΔH) and the experimentally set temperature (T), the entropic change (ΔS) and Gibbs free energy (ΔG) of binding can be determined:

$$\Delta G = -RT\ ln\ K_a \tag{6.1}$$

$$\Delta G = \Delta H - T\Delta S \tag{6.2}$$

where R is the defined as the ideal gas constant (8.314 J K^{-1} mol^{-1}) and K_a is the association constant of binding. For an assumed binding mechanism (e.g. single binding site), the total heat of association can then be theoretically determined for each titration point as a function of free and bound protein and ligand to theoretically describe the overall binding reaction. The theoretical isotherm is then optimised to best fit the experimental data as a mixed function of K_a, stoichiometry (n) and ΔH.

In terms of experimental design, assuming the molecules involved release or absorb measurable heat upon binding, the main consideration is taking the binding constant into account, so that concentrations of protein and ligand can be chosen to observe saturation over the course of the experiment. The TARP ABD: actin interaction had not previously been measured in any detail. Although many ITC investigations carry out test titrations for this reason, owing to machine time and material constraints, a more informed method of initial estimate was sought. Table 6.2.1 shows the sequence alignments between the TARP ABD WH2 predicted helix and other WH2 domains for which an experimentally derived actin binding constant had been determined. Although WH2 domain containing proteins have not been exclusively shown to interact with actin solely by their WH2 domain helix, as the prominent, conserved, interaction site, it was assumed that a significant degree of the protein-specific actin affinity arose from the specific WH2 domain helices. The TARP ABD WH2 region best aligned with the tightest known actin binding WH2 domain (K_d = 52 nM), found in WAVE2, a protein which nucleates the branching of actin filaments in humans as part of the ARP2/3 complex (119). However, as the next best aligning WH2 domain had a K_d = 230 nM, a conservative estimate of 100 nM was chosen as the theoretical association constant for the TARP ABD:actin interaction. Figure 6.2.4 shows an example of the resulting

ITC dataset collected, fit to a single-site model.

Table 6.2.1: Pairwise sequence alignments between the TARP ABD WH2-homologous region and other WH2 domains for which an actin-only associated K_d had been published (119). Alignments were carried out using the T-COFFEE webserver (242, 243) with FASTA formatted sequences (accessible: `http://tcoffee.crg.cat/apps/tcoffee/do:regular`).

% conserved	WH2 domain	T-COFFEE alignment	actin K_d
con : 91	$TARP_{726-825}$ WAVE2 433-464	---AMKDILSAVRKH----------LDV VSDARSDLLSAIRQQGFQLRRVEEQREQEKRDV	52 nM
con : 89	$TARP_{726-825}$ MIM 725-755	---AMKDILSAVRKHLD----------V DTPQGEDMLNAIRRGVKLKKTTTNDRSAPRFS	230 nM
con : 87	$TARP_{726-825}$ WASP 430-458	AMKDILSAVRKHL----------DV GRGALLDQIRQGIQLNKTPGAPESSALQP	250 nM
con : 87	$TARP_{726-825}$ WIP 29-46	---AMKDILSAVRKHLDV EQAGRNALLSDISKGKKL	2.15 µM
con : 79	$TARP_{726-825}$ WAVE2 450-464	AMKDILSAVRKHLDV LRRVEEQREQEKRDV	26.53 µM
con : 56	$TARP_{726-825}$ WIP 46-63	AMKDIL---SAVRKHLDV LKKTVTNDRSAPILDKPK	> 100 µM
con : 49	$TARP_{726-825}$ WIP 29-60	---AMKDILSAVRK----------HLDV EQAGRNALLSDISKGKKLKKTVTNDRSAPILD	160 µM

In total, four TARP ABD:actin titrations were carried out, along with control titrations to determine the heat of actin dilution, TARP ABD dilution and potential buffer mismatches between the two samples. Only dilution of the TARP ABD (Figure 6.2.5) proved significant and the average value was therefore subtracted from each TARP ABD:actin dataset prior to analysis (Table 6.2.2).

The calculated Kd of 102 nM ± 33 nM for the TARP ABD:actin interaction is of a similar magnitude to those of other WH2:actin interactions and all repeated titrations lead to negative values for ΔG, indicating that the binding interaction is favourable. The reaction stoichiometry of 1:1 also agrees other WH2 domain:actin studies. Furthermore, the consistently positive value for $-T\Delta S$ indicates that a significant decrease in entropy (disorder to order) is concomitant with binding. This is commonly observed when proteins, in particular IDPs (174), undergo a folding transition upon binding and is further evidence supporting the fact that the TARP ABD WH2 helix is not actually fully formed in the unbound state.

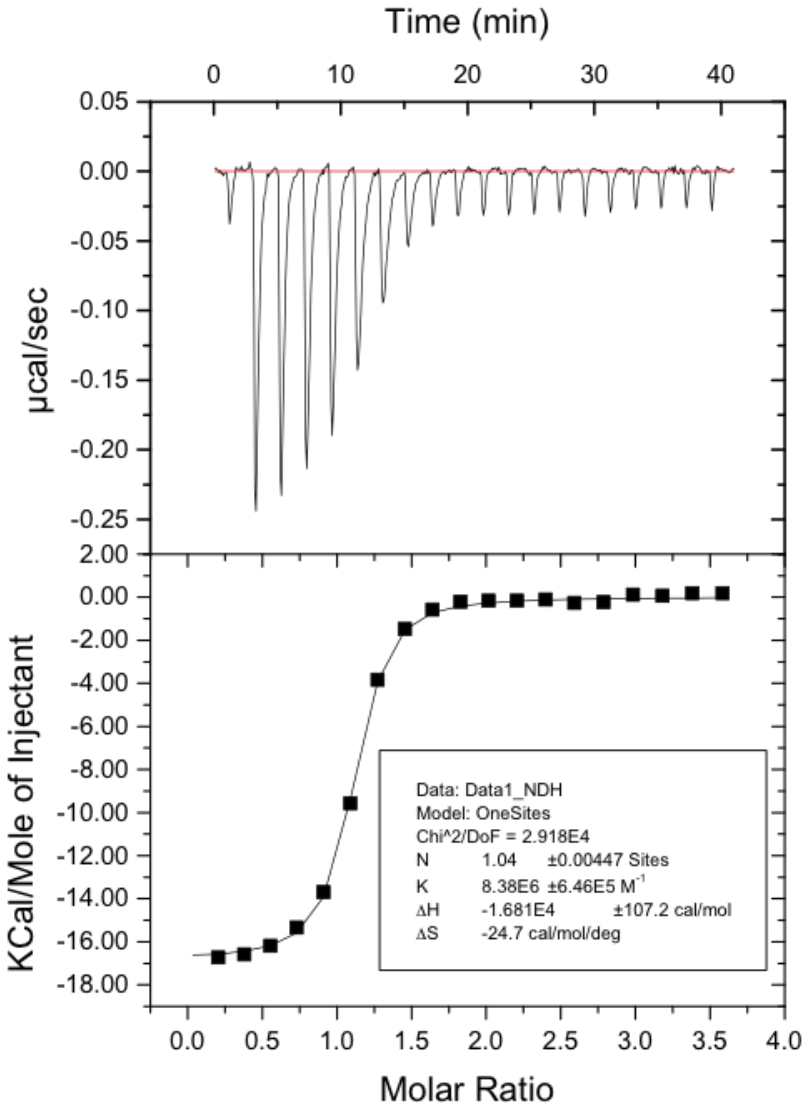

Figure 6.2.4: The first measured ITC binding isotherm between the TARP ABD and G actin (top) and the integral of the heat change fitted to a single binding site model (bottom). The enthalpies associated with the dilution of TARP were subtracted prior to integration. Fitting was carried out with the MicroCal ITC plugin for Origin (109). All fitting parameters were left free to optimise until the chi-squared value no longer improved. Experiments were carried out with a 200 µl chamber of 10 µM G actin thermostatically held at 25 °C. A preliminary injection of 0.4 µl 175 µM TARP was followed by 19 injections of 2 µl, giving a final TARP:actin molar ratio of 3.36:1.

When looking at the aligned sequence characteristics of documented WH2 domains and that of the TARP ABD (Figure 6.2.6), the sequences and binding affinities qualitatively suggest a correlation between actin binding affinity and the number of hydrophobic residues present.

Chereau et al. (119) also describe the significance of the two arginine residues in the WAVE2 LRRV motif of extended structure, just C-terminal to the WH2 helix, which form salt bridges to actin residues D24 and D25. Although these residues are not conserved in the TARP ABD, the corresponding amino acids (V763-Y764) could contribute to the interaction through additional hydrophobic interactions and may explain the similar binding affinity.

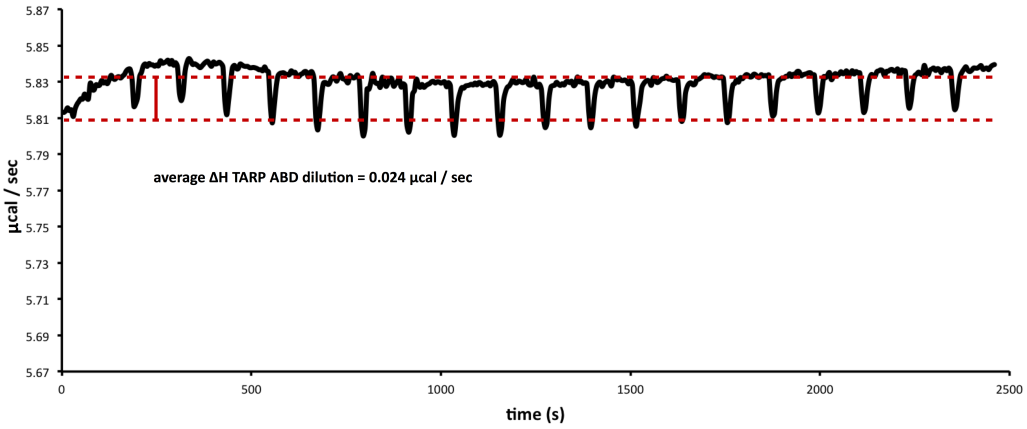

Figure 6.2.5: The ITC isotherm used to determine the heat of dilution for 175 µM TARP ABD injections. The observed average heat change across 19 2 µl injections of 175 µM TARP, of 0.024 µcal.sec^{-1}, was subtracted from all ITC datasets prior to analysis.

Table 6.2.2: The optimised thermodynamic parameters obtained by fitting the TARP ABD:actin ITC isotherms to a single biding site model. Gibbs free energies were calculated using equation 6.2.2 for a set temperature of 298 K.

ITC run	n	K_d (nM)	ΔH (kcal.mol^{-1})	$-T\Delta S$ (kcal.mol^{-1})	ΔG (kcal/mol)
1	1.04	119	-16.8	7.36	-9.44
2	0.935	134	-16.6	7.27	-9.33
3	1.06	58.8	-15.3	5.48	-9.82
4	0.982	94.3	-15.7	6.14	9.56
Avg.	1.00 ± 0.01	102 ± 32.9	-16.1 ± 0.71	6.56 ± 0.91	-9.54 ± 0.21

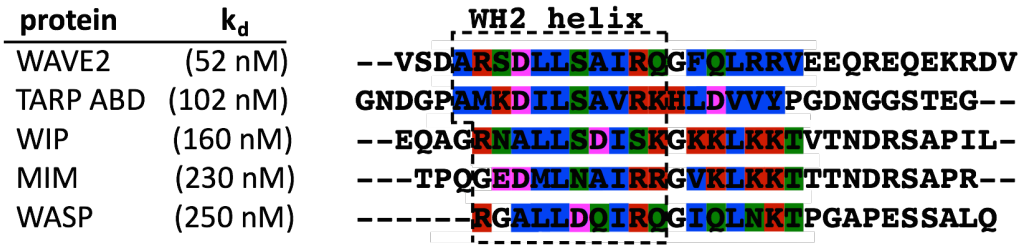

Figure 6.2.6: An EBI-MUSCLE (accessible: `http://www.ebi.ac.uk/Tools/msa/muscle`) (244) multiple sequence alignment (gaps removed as of WH2 helix, dashed black box) for WH2 domains sorted by their actin binding K_d (119). Residues have been coloured according to their chemical characteristics: blue = most hydrophobic (195) (A,C,F,I,M,L,V,W,Y), green = polar / uncharged (N,Q,S,T), magenta = acidic (D,E) and , red = basic (K,R,H).

6.2.3 Synchrotron radiation CD spectroscopy

The analysis of NMR data for the TARP ABD in the presence of actin did not yield any extra chemical shift data to help characterise the interaction. Nonetheless, to further support the presence of the WH2 α-helix in the TARP ABD bound state, synchrotron radiation circular dichroism (SRCD) spectroscopy was used to try and determine any changes in secondary structure content in the presence of actin. Although actin is a relatively large, heavily folded protein and the expected increase in TARP secondary structure would be minor, due to the summative nature of the CD signal, and the favourable signal to noise from a synchrotron light source, even small changes in the structural content of a sample should be apparent. SRCD spectra were acquired for the TARP ABD, G-actin and a ~3:1 molar complex of the TARP ABP and G-actin (Figure 6.2.7) in collaboration with Professor Bonnie Wallace and Dr Andrew Miles, Birkbeck College, London. A theoretical SRCD spectrum was also calculated to hypothetically describe the expected SRCD spectrum if no changes in secondary structure were observed in either component of the TARP ABD:actin complex. This was achieved by summing the observed ellipticity, ε, from the individual SRCD spectra of the TARP ABD and actin by equation 6.2.3 (245) to take account of the dominant CD absorption of the peptide bond:

$$\Delta\epsilon_{T:A} = \frac{(\Delta\epsilon_A \times N_A) \; + \; (\Delta\epsilon_T \times N_T)}{N_A + N_T} \tag{6.3}$$

where $\Delta\varepsilon$ is the CD ellipticity, N is the number of peptide bonds (number of residues 1) and subscript A, T and T:A refer to actin, the TARP ABD and the TARP ABD:actin complex respectively. To accurately take account of the significant disordered and structural content in the individual spectra of the TARP ABD and G-actin respectively, different reference data sets (db6 and sp175t respectively) were used on the recommendation of Dr Andrew Miles, Birkbeck College (personal communication, 03/08/2013).

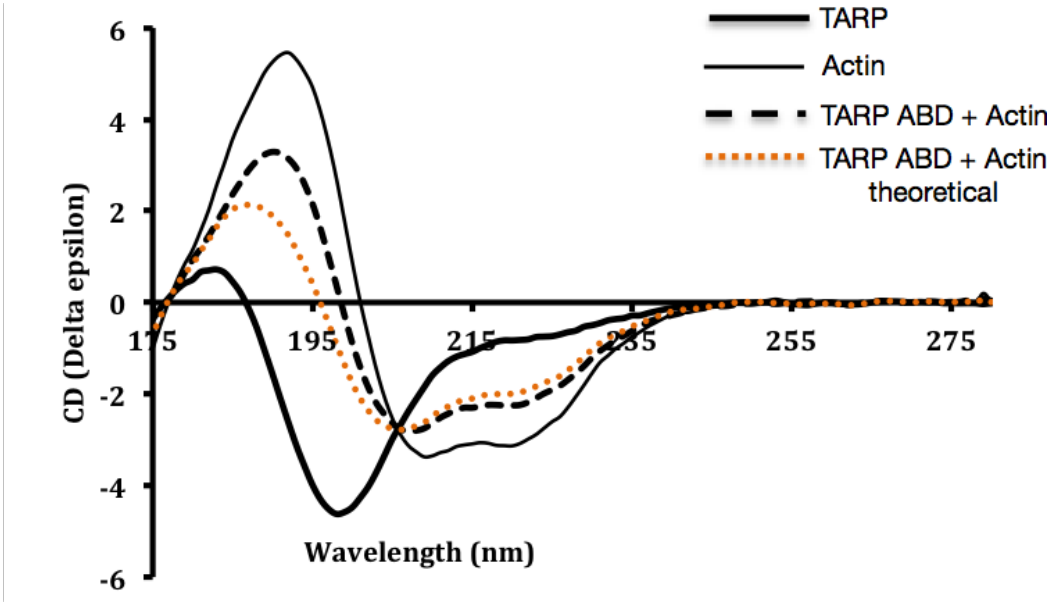

Figure 6.2.7: The SRCD spectra of the TARP ABD (black, bold), G-actin (black) and a complex of the TARP ABD:actin in a 3:1 ratio. The theoretical TARP:actin spectrum (assuming no change in secondary structure at a 1:1 molar ratio) is also shown (orange).

The structural content of the free TARP ABD suggests 5% of α-helix is present in solution (Table 6.2.3) in agreement with that of the previous analysis undertaken (Figure 5.2.15). In the bound state the overall helical content of the complex is 4% greater than that expected from the mixture of actin and the TARP ABD, assuming no structural change takes place. Ignoring potential changes in the secondary structure of actin, for which observation of free and WH2-bound actin structures are not significantly different, this suggests that the increase in helical content arises from the conformation of the bound state TARP ABD.

The actin bound state of the TARP ABD therefore suggests that in total 9% ± 3.7 of α-helix is present. If the TARP ABD does have native propensity for a 12 residue α-helix to form, which would match that seen for the actin

bound state of the homologous WAVE WH2 region, then 11.4% of α-helix would be expected. The calculated helical increase, with its associated error could then account for a helix of this size.

Table 6.2.3: The average secondary structure content of the TARP ABD, G-actin, a 3:1 molar ratio complex of the TARP ABD:actin and that of the theoretical 1:1 combination of the TARP ABD and G-actin assuming no secondary structure changes. SRCD spectra were acquired at 20 °C and values are averaged from the CONTINLL, SELCON and CDDSTR algorithms using the db6 database (D) for TARP analysis and the SP175t database (S) for actin containing samples.

Secondary structure	% Secondary structure content				
	TARP^{D}	G-actin	Theoretical^{S}	TARP:actin^{S}	$\Delta_{\text{Theor:meas}}$
α-helix	5 ± 3	28 ± 0	17 ± 2	21 ± 1	4 ± 2.24
β-strand	16 ± 0	24 ± 3	29 ± 2	29 ± 1	0 ± 2.24
Disordered	67 ± 4	34 ± 1	39 ± 2	37 ± 2	-2 ± 2.83

6.2.4 Modelling the TARP ABD: actin interaction

Although this investigation was unable to provide adequate experimental data to describe the TARP ABD:actin interaction at a molecular resolution, by combining the qualitative and quantitative low resolution data with respect to the TARP ABD, the TARP ABD:actin complex and previously published information regarding WH2 domain:actin interactions, an *in silico* approach was undertaken in an attempt to model the actin bound TARP ABD.

As previously shown in Figure 6.1.3, existing structures of actin bound WH2 domains typically present the WH2 helical domain binding to the actin hydrophobic cleft in between subdomains 1 and 3 and the remaining extended residues interacting through hydrophobic interactions and salt bridges across actin subdomains 2 and 4. The NMR data presented in Section 6.2.1 (Figure 6.2.2), suggests the majority of the TARP ABD is perturbed by actin and so could support this type of actin binding mechanism. However, the likely structural rearrangement of the TARP ABD upon binding, as evidenced from SRCD and ITC data, ruled out the use of traditional rigid-body *in silico* docking methods, which do not account for significant protein dynamics.

The Rosetta based FlexPepDock web server (118) provides a freely available docking methodology whereby only one molecule is portrayed as rigid and

an approximate peptide structure of up to 30 residues can be docked under high resolution low energy refinement by incorporating full peptide backbone flexibility into the calculation. The published articles corresponding to the FlexPepDock methodology (118, 246) state that during their own testing, near native binding conformations were able to be modelled when the starting conformation of the docked peptide was up to 5.5 Å RMSD away from that of the native fold and fully extended. The description of "near native" conformation is defined as an impressive < 2 Å backbone peptide RMSD and < 1 Å sidechain RMSD from that of the experimentally determined bound conformation.

Prior to FlexPepDocking the TARP ABD with actin, I felt it necessary to test the methodology, and the method of input file preparation, with the WAVE2:actin structure (PDB code: 2A40). WAVE2 was specifically chosen as it had the most similar sequence and actin binding affinity to that of the TARP ABD compared to other WH2 domains (Figure 6.2.1). Three control simulations were run, with each being scored against the experimentally determined 2A40 crystal structure (with small molecules and the cocrystalised DNase I removed). The chosen starting conformations, all based around the 22 residue WH2 fragment from the 2A40 structure (Figure 6.2.8, orange), were an *in silico* extended peptide ($\pm 135°$ for all φ/ψ angles) in the same orientation as the bound state (Figure 6.2.8, green), an *in silico* generated 10 residue ideal α-helix with a linear conformation (Figure 6.2.8, cyan) and the conformation of the bound state both with a 5 Å translation out of the actin cleft at 90° to the axis of the WH2 helix (Figure 6.2.8, magenta).

The FlexPepDock calculation itself begins with a steric energy minimisation for both the receptor and the ligand sidechains. 200 receptor:ligand structures are then generated in the presence of increased attractive, and decreased repulsive, van der Waals terms to prevent separation of the two chains. This also facilitates the generation of diverse starting conformations under random optimisation. The initial ensemble of structures is then refined over 10 cycles of iterative optimisation throughout which the van der Waals terms are gradually returned to those outlined by the standard Rosetta framework parameters (247). Traditional rigid-body docking occurs during each cycle and is followed by Monte-Carlo energy minimisation of the peptide backbone and sidechain geometries to generate a final set of 200 high resolution output models. When using the FlexPepDock server, the top 10 low energy structures are returned with a plot correlating the RMSDs from a reference structure of

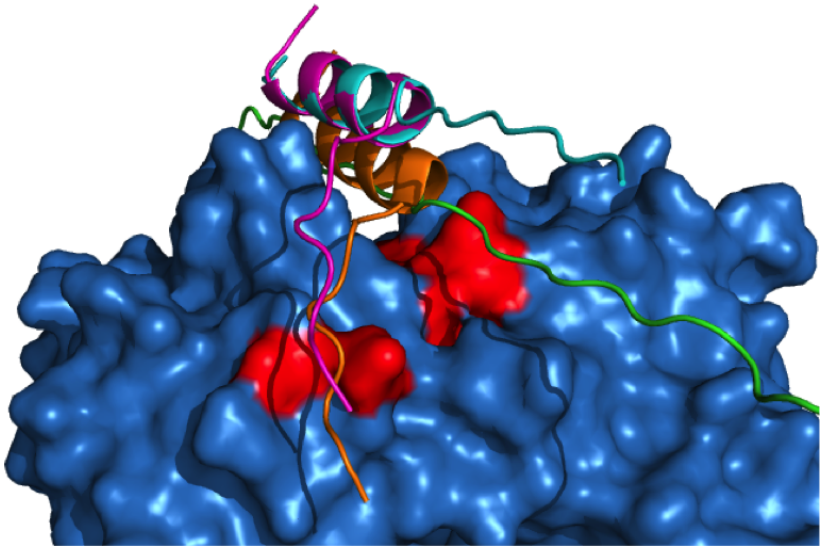

Figure 6.2.8: The WAVE2:actin structure (PDB code: 2A40) (orange cartoon = experimentally determined bound state of WAVE2, blue surface = G-actin) overlaid with the starting positions of the WAVE2 FlexPepDock tests for the extended structure (green) and an extended structure with an ideal-geometry 10 residue α-helix (cyan) and native state conformation (magenta), both translated 5 Å out of the actin binding cleft.

the individual models, if provided (otherwise the server determines the RMSD from the input structure), against their overall Rosetta energy score (lower is better).

6.2.4.1 WAVE2:actin docking test

The combined results from the three WAVE2:actin FlexPepDock calculations can be seen in figures 6.2.9-10. Contrary to the results presented in the article originally describing the FlexPepDock methodology (246), the modelling of the WAVE2:actin complex, as compared to that of the experimentally determined 2A40 crystal structure (119) was significantly influenced with respect to the starting conformation of the docking peptide.

When initiated from a fully extended conformation (Figure 6.2.9, cyan) neither formation of the WH2 helix, or localisation of the C-terminal extended region were comparable to the crystal structure, with average backbone RMSDs of between 13.7 - 17.9 Å (Figure 6.2.10, green) between the top scoring models and that of the 2A40 conformation.

When initiated with an ideal-geometry α-helix for the natively helical residues, the docking procedure fared better energetically and in relation to the 2A40 conformation (Figure 6.2.9, yellow and Figure 6.2.10, red). However the extended C-terminus also failed to localise to the native position. The best approach was clearly that where the docking procedure was started with the native conformation already present, albeit with the peptide removed by 5 Å from the actin surface (Figure 6.2.9, magenta and Figure 6.2.10, black), which generated the most favourable total Rosetta scores and average backbone RMSDs of 0.6 3.7 Å across the top 10 scoring models. The C-terminal extended region can also be seen to generally maintain its conformation and the sidechains of the residues shown to form salt bridges with actin residues D24 and D25 locate to within 3 Å of the native conformation for 9 of the 10 top scoring models.

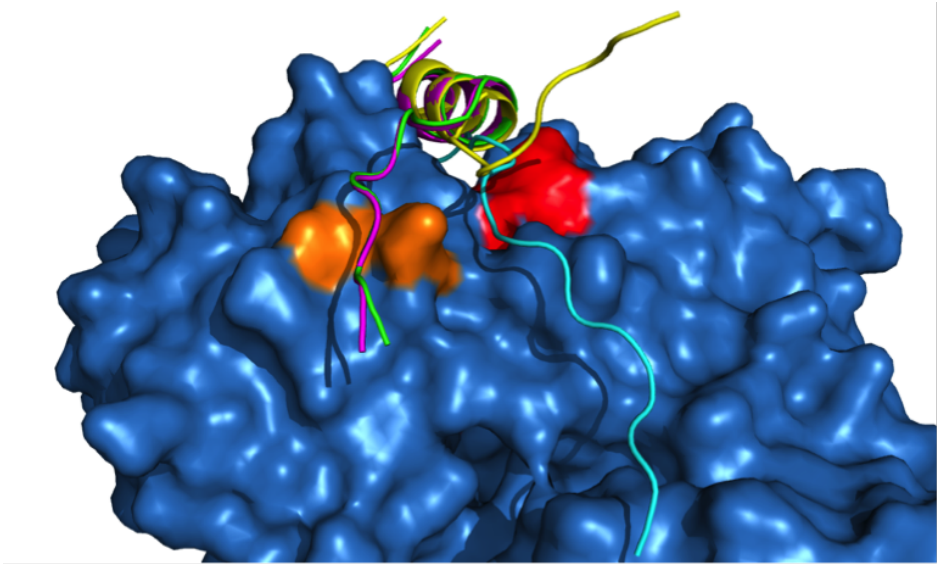

Figure 6.2.9: The average structures of the top 10 scoring models from the fully extended (cyan), ideal-geometry helix (yellow) and actin-bound WAVE2 conformations (magenta), with the latter two removed 5 Å from the actin hydrophobic cleft, WAVE:actin FlexPepDock docking calculations. The simplified 2A40 crystal structure of the bound WAVE WH2 domain (green) and a surface representation of the actin monomer (blue) are shown, with highlighted residues depicting the hydrophobic cleft (red) and residues involved in salt bridge formation (D24 & D25, orange).

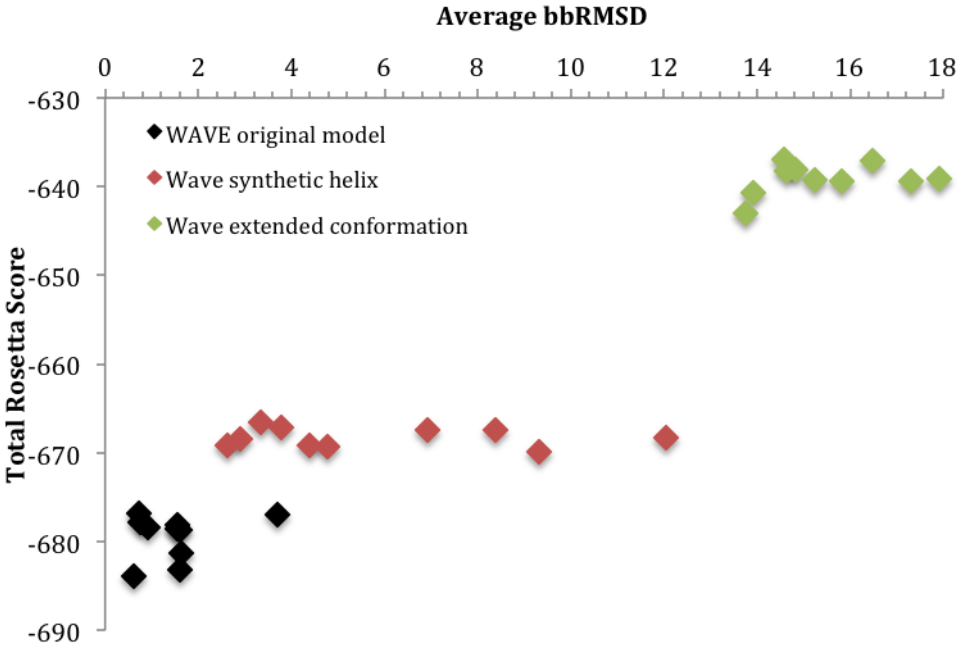

Figure 6.2.10: The total Rosetta energy scores and backbone RMSDs to the experimentally determined WAVE:actin complex (modified 2A40 pdb file (119)), for the top ten scoring models from the WAVE:actin FlexPepDock actin docking calculations for extended (green), ideal helix (red) and original model (black) starting conformations.

6.2.4.2 TARP ABD:actin docking calculations

Although the WAVE2:actin docking test suggested that a near-native starting conformation generated the most favourably docked structures, for comparison, three TARP ABD:actin docking calculations were carried out for starting conformations similar to those used in the WAVE test. The results from these FlexPepDock calculations can be seen in Figures 6.2.11-12.

As seen during the WAVE docking test, the calculation starting from a fully extended conformation generated the poorest scoring results (Figure 6.2.12, green), again, with neither formation of the WH2 helix or an actin-proximity for the C-terminal extended residues (Figure 6.2.11, cyan) which could account for the perturbations in NMR peak heights as seen in Figure 6.2.2. The second starting model however, managed to both maintain the conformation of the WH2 helix throughout the calculation and favourably position the C-terminal residues proximal to the actin surface (Figure 6.2.11, yellow), albeit differently to that of the native WAVE conformation.

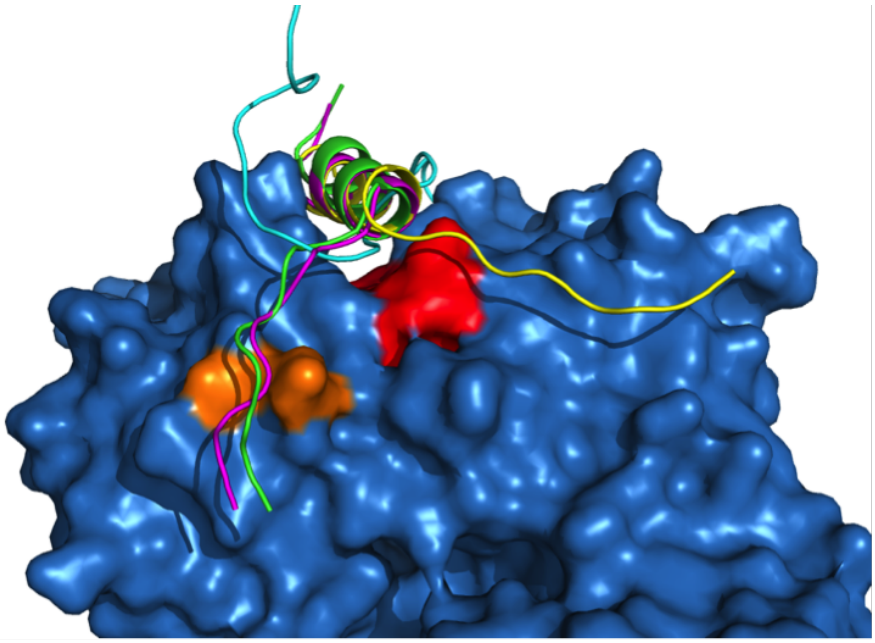

Figure 6.2.11: The average structures of the top 10 scoring models from the fully extended (cyan), ideal-geometry helix (yellow) and WAVE2-adopted conformations (magenta), with the latter two removed 5 Å from the actin hydrophobic cleft, TARP ABD:actin FlexPepDock docking calculations. The simplified 2A40 crystal structure of the bound WAVE WH2 domain (green) and a surface representation of the actin monomer (blue) are shown, with highlighted residues depicting the hydrophobic cleft (red) and residues involved in salt bridge formation (D24 & D25, orange).

The final starting conformation, with an initial input structure modelled on the backbone dihedral angles of the 2A40 crystal structure actin-bound WAVE2 domain, perhaps unsurprisingly contained the most favourable backbone RMSD to that of the native actin bound WAVE2 conformation (Figure 6.2.11, magenta and Figure 6.2.12, black). The total Rosetta scores were also favourable and comparable to those obtained for the native conformation WAVE2 model.

However, interestingly, it was not the highest scoring calculation. Even small improvements in Rosetta scores have previously been shown to be a good reflection of enhanced docking calculations with respect to generating near-native models (248). Owing to the lack of conformational bias in the input structure, the result of this docking calculation should not therefore be ignored in favour of that more similar to the actin-bound WAVE conformation.

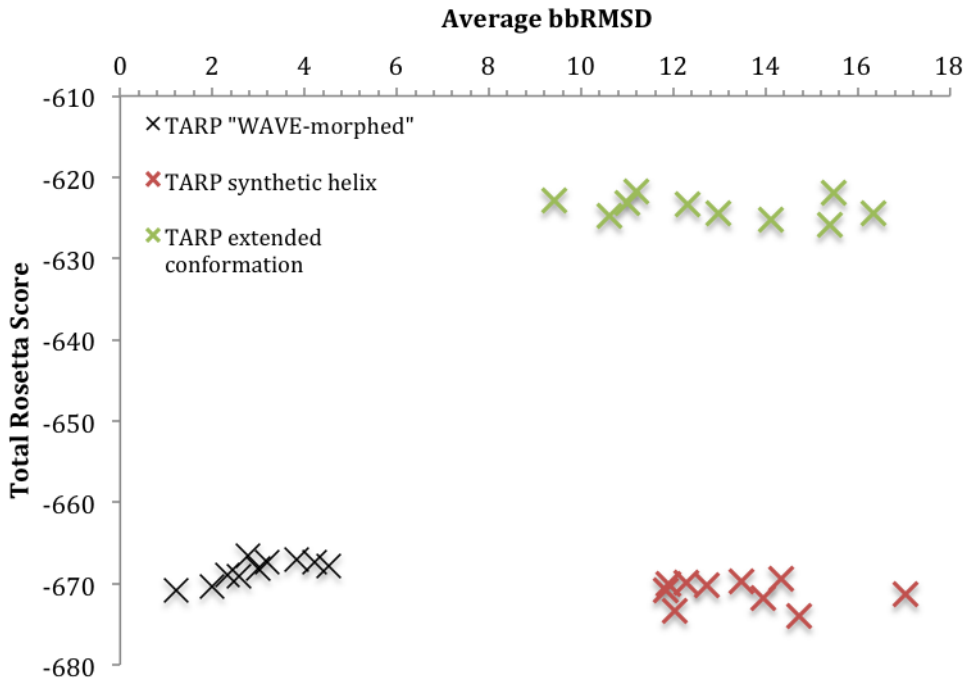

Figure 6.2.12: The total Rosetta energy scores and backbone RMSDs to the experimentally determined WAVE:actin complex (modified 2A40 pdb file (119)), for the top ten scoring models from the TARP ABD:actin FlexPepDock actin docking calculations for extended (green), ideal helix (red) and WAVE-adopted conformation (black) starting models.

Although experiential validation is outside the scope of this work, it may prove true that, similar to other IDP proteins, the TARP ABD may exhibit a degree of static "fuzziness" in the bound state and that the results of these docking calculations highlight the potential for a broad binding interaction.

The intermolecular interactions describing the actin bound WAVE2 conformation have been previously described (119) in terms of the WH2 helix:hydrophobic cleft and the salt bridge interactions between the actin residues D24 and D25 and those of the LKKV motif within the WAVE2 C-terminus. Figure 6.2.13 highlights the structural observations from the high scoring TARP ABD docking models of the unbiased input conformation. All models (Figure 6.2.13, red) can be see to adopt a similar α-helical position to that of the native WAVE2 conformation (Figure 6.2.13, green), with the backbones of the 12 helical WAVE2 residues aligning to within 0.26 Å RMSD of the average TARP ABD helix.

Excluding model 10, the C-termini of the top nine scoring TARP ABD models all show a similar localisation (backbone RMSD of 1.3 Å), which is sig-

nificantly different to that obtained from the unbiased helical WAVE2 domain that was shown to protrude away from the actin surface (Figure 6.2.9, yellow). In looking for potential interaction sites, the actin residue R147 sidechain (Figure 6.2.13, magenta) was identified as lying within 3 Å of the sidechain of the TARP ABD residue D762 in the majority of calculated models.

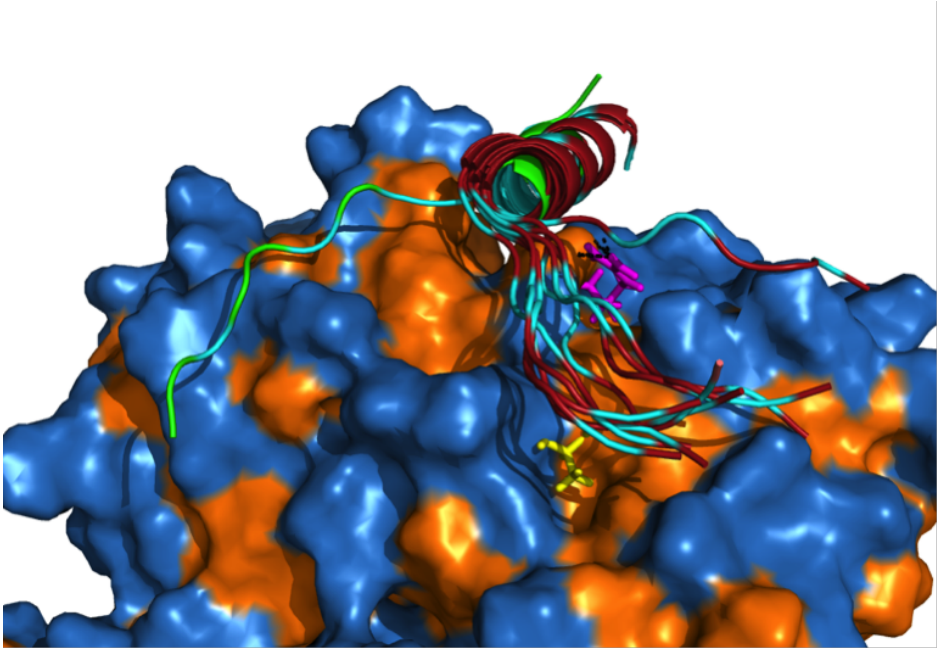

Figure 6.2.13: The structure of the simplified 2A40 crystal structure of the WAVE WH2 domain (green) and an actin monomer (blue surface) overlaid with the top ten models of the TARP ABD docking calculation started from an overall linear conformation with an ideal α-helix (red). Hydrophobic residues on the actin surface and the WH2 proteins are coloured orange and cyan respectively. Actin residues R147 and A331 are coloured magenta and yellow respectively.

Similar to the WAVE2 domain native conformation, this could therefore provide a salt-bridge anchoring site for the C-terminus of the TARP ABD. Furthermore, the hydrophobic residues Y765 and G767 of the TARP C-terminus (Figure 6.2.13, cyan) can be seen to locate between 2.5 to 4 Å of a small hydrophobic pocket on the actin surface, positioned around the actin A331 sidechain (Figure 6.2.13, yellow), but also including residues G308-A310 and I329-A331 (Figure 6.2.13, orange residues flanking A331 (yellow)).

The FoldX energy force field (249) was also used, as a software plugin (250) with the YASARA molecular viewer package (251), to subjectively analyse the binding interfaces of top ten models of the two comparable TARP ABD:actin docking calculations. These results are summarised in Figure 6.2.14. The

analysis of the bound state residues for the two models agrees well with the initial observations of hydrophobicity and salt-bridge pairs described above for the unbiased TARP ABD model, and those of WAVE described in the literature.

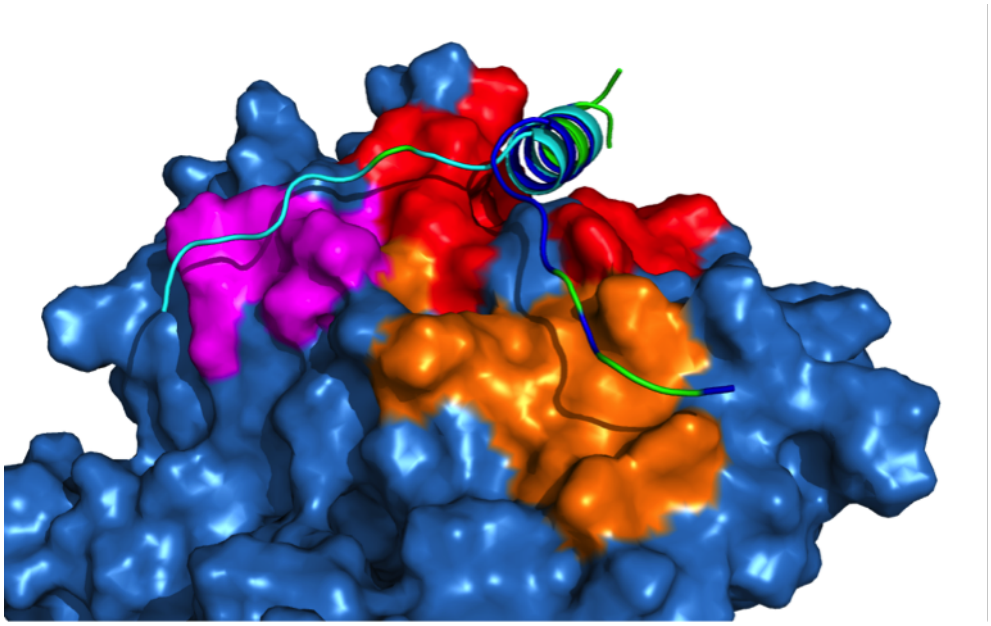

Figure 6.2.14: The structure of the simplified 2A40 actin crystal structure (light blue surface) overlaid with the average structures of the TARP ABD docking calculations (green). Actin residues shown to form a binding interface are coloured according to whether they are involved in both TARP ABD models (red), the WAVE2-adopted conformation model (magenta) or the unbiased TARP ABD model (orange). Residues of the TARP ABD constructs shown to form a binding interface were coloured cyan and dark blue for the WAVE conformation and the unbiased model respectively.

The FoldX algorithm also calculates an effective free energy of binding between two molecular partners. For the top ten models of the two predominant TARP ABD:actin models, these were determined to be -12.01 ± 2.01 and -11.34 ± 2.99 kcal.mol^{-1} for the unbiased and WAVE2-adopted conformation models respectively. This compares to a calculated free energy of +1.437 ± 2.55 kcal.mol^{-1} for the initial extended model that exhibited no C-terminal interaction with actin and obtained a poor overall Rosetta score. These theoretical free energy terms are based upon the total summed energies calculated from the backbone conformations, intermolecular and intramolecular hydrogen bonds, van der Waals contributions and electrostatic interactions. Although these values should not be quantitatively compared to those experimentally determined (252), the similarity between the two binding energies suggests

that both models could be equally favourable and lends further credibility to the idea that the TARP ABD is able to bind actin monomers in multiple ways.

6.3 Discussion

Although previous research had identified the WH2-homologous region of the TARP protein and briefly confirmed its function, no published data is currently available that describes the interaction between actin and the TARP ABD in any biochemical or molecular detail.

Unfortunately, NMR analyses failed to yield conclusive, specific information as to the site of actin interaction within the TARP ABD. Analysis of the bound state ^{1}H-^{15}N HSQC spectrum however did highlight a broad site of interaction from residue A740 at the N-terminus of the WH2 homologous helix to G805 within the C-terminal extended region (Figure 6.2.2). This is in agreement with the known actin-bound structures of WH2 domains which depict a short helical interaction in the actin hydrophobic cleft and stretch of ~10 residues lying across the surface of actin subdomain 2.

Actin is notoriously difficult to crystallise owing to its potential for self association and the structural reorganisation associated with ATP hydrolysis (253). However all elucidated crystallographic structures of actin bound WH2 domains (excluding β-thymosin 4, which possesses a significantly folded domain) typically required substantial truncations of the bound WH2 ligand, as evident in Figure 6.1.3, to enable crystal formation. Although WH2 domains are largely described as heavily disordered, with a propensity to fold upon binding, in the literature (254256) the requirement of short WH2 fragments for actin bound crystallisation suggests that in the bound state, WH2-actin interactions do only occur across a short, but sequential, binding interface. The observations of the TARP ABD residues perturbed upon binding are therefore also in agreement with this idea, and suggests that the identified sequence homology within the TARP ABD to other WH2 domains was an accurate sequence annotation.

The results of the ITC analysis are the first reported thermodynamic characteristics of the TARP ABD:actin interaction. The determined K_d of 102 ± 32.9 nM is in good agreement with those of other WH2 domain:actin interac-

tions and highlights the significance of the primary sequence properties, both in the design of ITC experimentation and in the identification of the determinants related to ligand affinity. Although WH2 domains as a whole should not necessarily be grouped as a single functional entity that only interacts with actin at fixed positions, the correlation between actin affinity and the hydrophobic content of the WH2 helix and that of the C-terminus is striking, and suggests that hydrophobic interactions do indeed play a formative role in WH2:actin complexes as proposed by Chereau et al., (119).

The calculated changes in Gibbs free energy (ΔG = -9.54 kcal.mol^{-1}), enthalpy (ΔH = -16.1 kcal.mol^{-1}) and entropy (-TΔS = 6.56 kcal.mol^{-1}) also offer an insight into the actin binding function of the TARP ABP. As observed for other IDPs, the favourable reaction (-ΔG) is facilitated by an exothermic enthalpic change (-ΔH) that balances the concomitant decrease in entropy (-TΔS) required for folding.

Furthermore, the determination of a TARP ABD:actin K_d allows the retrospective discussion of the NMR experiment in the presence of actin. Given a K_d of 102 nM, with concentrations of 31 μM and 85 μM for the TARP ABD and actin respectively, the expected bound state concentration would be 30.4 μM (99.8% of the total [TARP ABD]). The dominance of the bound state indicates that the observable TARP ABD peaks should not been influenced by those of the unbound state, however the lack of any chemical shift changes is not suggestive of a conformation and environment different to that of the free state. It must therefore be the case that the ITC determined K_d is not appropriate in the absence of ATP. However, a relative difference in the peak intensities suggests binding was occurring, albeit to a lesser extent.

The SRCD investigation into the secondary structure content of the TARP ABD following actin binding provided experimental evidence to suggest that the TARP ABD does indeed fold upon binding. The calculated increase in helical content of 4% suggests than in the bound state the helical content of the TAR ABP rises to a total figure of 9% ± 3.7. With the associated error, this level of structural content could account for the expected level of 11.4% content if the 12 residue helix were indeed to fold as seen for the other WH2 proteins. Unfortunately, due to time constraints, the ~3:1 molar ratio sample was the only sample used for analysis. The calculation to theoretically deduce the complex CD spectrum assumes a 1:1 molar ratio. Although this is not

ideal, it could therefore be the case that excess disordered TARP ABD in the free state is diluting the apparent helical content and that in actuality a higher percentage of α-helix is formed upon binding. Nonetheless, the increase in the helical content of the TARP ABD also validates the creation of preformed structure for use in structural docking models.

A potential criticism to this hypothesis is that there are no means of distinguishing whether the observed changes in secondary structure are actually a consequence of additional α-helix formation within the actin monomer. This possibility can regrettably not be ruled out, and it remains a potential source of error in many SRCD studies of protein:protein complexes. In this instance, it can however be said that no significant structural changes can be seen between the crystal structures of the actin monomers from the WAVE bound state (119) (pdb code: 2A40) and that of native, ligand free, G-actin (257) (pdb code: 3HBT); with a total backbone RMSD between the two of only 0.46 Å (data not shown).

The results from the FlexPep Docking calculations of the TARP ABD and WAVE WH2 domains with actin yielded interesting results. Although the test carried out with the modified WAVE structures were not able to reproduce a bound state model similar to the 2A40 crystal structure, improvements in both energetic scores and backbone RMSDs were observed with backbone conformations approaching that of the native fold (Figure 6.2.10). The work outlining the FlexPepDock methodology (246) describes similar trends in successful calculations, however their improvements were not as dramatic as that seen for the WAVE domain, wherein no features of the true bound state were able to be recreated with starting conformations other than that of the native interaction.

The docking calculations for the TARP ABD provided two potential actin binding modes. The input structures to both models were composed of a preformed WH2-homolgous α-helix to facilitate WH2-like localisation of the TARP ABD within the actin hydrophobic cleft whilst allowing the inherent flexibility of the calculation to find the most energetically favourable conformations.

The resulting TARP ABD:actin models for both the preformed helix and WAVE2-adopted backbone angle input files show equally respectable total Rosetta scores, analogous to those of the native conformation WAVE test scores, even though their C-termini form interactions to the actin surface at

90° to one another, across different actin subdomains. Manual investigation and a FoldX analysis into the possible attractive interactions also bore similar results to each other, and again highlighted the importance of hydrophobic interactions and the potential for salt bridge formation, as had previously been shown for other WH2:actin interactions. The comparable FoldX energies of binding also provide another measure that suggests that the interactions formed in both models are equally favourable.

No currently published WH2:actin crystal structures reflect the possibility of the unbiased docking model of the TARP ABD C-terminus binding actin across the surface of subdomain 3. This does not necessarily exclude the subdomain 3 model from having biological relevance, as isolation of a lowest energy ground state conformation is a well known potential pitfall of studying crystalline samples (258, 259). Without direct structural evidence regarding the actin bound state of the TAPR ABD in solution, the significance of this subdomain 3 model should therefore not be ignored.

Bound state polymorphism, or static "fuzziness", is a characteristic commonly found in IDPs and has been specifically linked to the function of other WH2 domain containing proteins (224, 260). Although the WH2 α-helix is invariably shown to bind in the same orientation, the apparent "fuzziness" of the WH2 domain has been postulated as the cause for its ability to act as a modulator of actin nucleation, polymerisation rates or filament destabilisation depending on the dynamics of the interactions between the WH2 C-terminus and actin subdomain 2 (261).

The finding that both TARP ABD:actin binding models are theoretically probable could therefore agree with observations of other actin binding proteins and the many descriptions of bound state heterogeneity for IDPs in general. Whether or not the subdomain 2 or 3 C-terminal binding sites provide distinct functions however, at this stage, cannot be commented on. It may be equally true that the subdomain 3 model simply provides a dynamic "rest-stop" for the WH2 C-terminus which does not require total molecular dissociation and thus acts to fine tune the dynamics of the WH2 C-terminus.

Chapter 7

Conclusion

As explained in section 1.4, the overall aims of this thesis were to provide structural and dynamic insight into the biophysical properties of the actin binding domain of the *Chlamydia trachomatis* TARP protein and the secreted effector domain of AVR3a11 from *Phytophthora capsici*.

The effector domain of AVR3a11 was shown to form a well ordered four helix bundle in solution, and although this was not a thoroughly novel observation in itself, it does support the characterisation of the fold. It also highlights the potential of the Y131 residue, that was determined to be in a different orientation in the X-ray structure of the same construct, to partake in ligand interactions with the other residues also identified as functionally important. The observations from the dynamic analysis of AVR3a11 portray it as relatively rigid on the very slow timescale of large domain orientations, which reinforces the idea that the core hydrophobic 4-helix bundle is indeed stable in solution.

With respect to the overall description of the proteins dynamics, it was unfortunate that the ^{15}N relaxation data could not be analysed according to the widely used mode free approach. However, the ability to probe the internal motions of the AVR3a11 construct with reduced spectral density analysis, in light of the diffusion model calculated from the X-ray structure suggests that internal motions across the nanosecond timescale or slower are greater than that expected for a rigid protein of the same size and shape.

The AVR3a11 effector of *P. capsici* represents a protein that adheres to the classical interpretation of the structure-function paradigm, whereby a fully formed, definitive structure facilitates the specific intermolecular interactions required for highly selective, high affinity binding events to occur. The observation of multiple native binding partners, both of the host and the parasite, potentially justify this type of structural organisation so as to perhaps ensure that the specific molecular recognition sites for any binding partner are always present and are not reliant upon the downstream consequences of any other binding events. Concomitantly, the significant proportion of internal motions identified throughout the protein might also provide some form of structural plasticity in regards to maintaining these binding sites in the event of any minor conformational rearrangements resulting from the binding of either ligand.

In contrast, although the dynamics of the TARP actin-binding domain (ABD) could not be investigated specifically, the thoroughly novel data and analysis presented in chapters 5 and 6 shows that the domain itself has the biophysical characteristics of a typical disordered protein. Additionally, although the interaction between actin and the TARP ABD could not be investigated at the atomic level; the thermodynamic characterisation, and binding-induced structural rearrangement can be said to be comparable to those of other WH2:actin interactions, for which the TARP ABD has some limited sequential homology. Two potential interaction models of equal in silico significance have also been proposed for the actin:TARP ABD association and although these are not currently confirmed by experimental data, the potential of multiple binding sites, or modes, is also a concept which further agrees with the portrayal of the TARP ABD as a typically disordered protein. The intrinsic disorder inherent to the TARP ABD represents a well characterised method for instilling either high-affinity, low-specificity or low-affinity, high-specificity protein binding events. In the case of the TARP ABD, this is most likely to be high-affinity, low-specificity, as witnessed by the fairly tight Kd and the potential for distinct binding modes across the actin surface. When thought of as part of a modular protein or multimer, within which multiple actin binding domains are present, the benefits of these high affinity, low specificity interactions, coupled with significant conformational freedom, allude to the overall mechanism for co-localisation of the multiple G-actin monomers needed for filament nucleation. Whilst the short WH2-like motif forms a tight interaction with actin, the conformational variability of the bound state C-terminus can be envisaged as conferring a flexible tolerance to the anchor site. The confor-

mational freedom resulting from the disordered nature of TARP could then accommodate the required orientations and proximities for the bound actin monomers to interact with one another as is required for filament nucleation to occur. The results presented in this thesis therefore describe two inherently different proteins. Although the sources of these two proteins are from wholly distinct organisms, as two examples of intracellularly secreted protein effectors which function to modulate host cell processes, they therefore represent two diametrically opposed solutions to arguably very similar evolutionary problems. Nonetheless, the complexities of modulating multifaceted cellular mechanisms, without causing significant perturbation of the native function, so as to either be of immediate detriment to the host cell or used as an indirect measure of infection suggests that both affinity and selectivity are important factors to consider. However, the individual biophysical characteristics of these two proteins can ultimately be described as a function of the native ligands and systems about which they evolved to selectively interact with, and as such this can be used to justify the distinctions between them.

Bibliography

[1] Abedin M, King N (2010) Diverse evolutionary paths to cell adhesion. *Trends Cell Biol* 20:73442.

[2] Kutschera U, Niklas KJ (2005) Endosymbiosis, cell evolution, and speciation. *Theory Biosci* 124:124.

[3] Wagstaff L, Kolahgar G, Piddini E (2013) Competitive cell interactions in cancer: a cellular tug of war. *Trends Cell Biol* 23:1607.

[4] De Kievit TR, Iglewski BH (2000) Bacterial Quorum Sensing in Pathogenic Relationships. *Infect Immun* 68:48394849.

[5] Kohrt HE, Pillai AB, Lowsky R, Strober S (2010) NKT cells, Treg, and their interactions in bone marrow transplantation. *Eur J Immunol* 40:18629.

[6] Holmes DE, Giloteaux L, Williams KH, Wrighton KC, Wilkins MJ, Thompson CA, Roper TJ, Long PE, Lovley DR (2013) Enrichment of specific protozoan populations during in situ bioremediation of uranium-contaminated groundwater. *ISME J* 7:128698.

[7] Lips KR, Brem F, et al. (2006) Emerging infectious disease and the loss of biodiversity in a Neotropical amphibian community. *Proc Natl Acad Sci U S A* 103:316570.

[8] Wahlen BD, Morgan MR, McCurdy AT, Willis RM, Morgan MD, Dye DJ, Bugbee B, Wood BD, Seefeldt LC (2013) Biodiesel from Microalgae, Yeast, and Bacteria: Engine Performance and Exhaust Emissions. *Energy & Fuels* 27:220228.

[9] Datar I, Betti M (2010) Possibilities for an in vitro meat production system. *Innov Food Sci Emerg Technol* 11:1322.

[10] Anderson DM, Frank DW (2012) Five mechanisms of manipulation by bacterial effectors: a ubiquitous theme. *PLoS Pathog* 8:e1002823.

[11] Ciechanover A (1998) The ubiquitin-proteasome pathway: on protein death and cell life. *EMBO J* 17:715160.

[12] Kubori T, Shinzawa N, Kanuka H, Nagai H (2010) Legionella metaeffector exploits host proteasome to temporally regulate cognate effector. *PLoS Pathog* 6:e1001216.

[13] Gruenheid S, DeVinney R, Bladt F, Goosney D, Gelkop S, Gish GD, Pawson T, Finlay BB (2001) Enteropathogenic E. coli Tir binds Nck to initiate actin pedestal formation in host cells. *Nat Cell Biol* 3:8569.

[14] McGhie EJ, Hayward RD, Koronakis V (2001) Cooperation between actin-binding proteins of invasive Salmonella: SipA potentiates SipC nucleation and bundling of actin. *EMBO J* 20:21319.

[15] Smith Portnoy, D. A. GA (1997) How the Listeria monocytogenes ActA protein converts actin polymerization into a motile force. *Trends Microbiol* 5:272276.

[16] Goldberg Theriot, J. A. MB (1995) Shigella flexneri surface protein IcsA is sufficient to direct actin-based motility. *Proc Natl Acad Sci U S A* 92:65726576.

[17] Gouin E, Egile C, Dehoux P, Villiers V, Adams J, Gertler F, Li R, Cossart P (2004) The RickA protein of Rickettsia conorii activates the Arp2/3 complex. *Nature* 427:457461.

[18] Stevens JM, Ulrich RL, Taylor LA, Wood MW, Deshazer D, Stevens MP, Galyov EE (2005) Actin-binding proteins from Burkholderia mallei and Burkholderia thailandensis can functionally compensate for the actin-based motility defect of a Burkholderia pseudomallei bimA mutant. *J Bacteriol* 187:78577862.

[19] Kamoun S (2003) Molecular Genetics of Pathogenic Oomycetes. *Eukaryot Cell* 2:191199.

[20] Phytophthora Database, accessed 08.10.13 Available at: `http://www.phytophthoradb.org/species.php`.

[21] Brasier C (2008) in Proceedings of the Fourth Meeting of the International Union of Forest Research Organizations (IUFRO).

[22] Tyler BM (2002) Molecular basis of recognition between phytophthora pathogens and their hosts. *Annu Rev Phytopathol* 40:13767.

[23] Kasuga T, Kozanitas M, Bui M, Hüberli D, Rizzo DM, Garbelotto M (2012) Phenotypic diversification is associated with host-induced transposon derepression in the sudden oak death pathogen Phytophthora ramorum. *PLoS One* 7:e34728.

[24] Potato and tomato blight, accessed 09.10.13. R Hortic Soc. Available at: `http://apps.rhs.org.uk/advicesearch/profile.aspx?pid=217`.

[25] Elliot M (University of St. Andrews and Science & Advice for Scottish Agriculture. Phytophthora ramorum and P. kernoviae; the most serious threat yet to the UKs native forests and historic gardens.

[26] Brasier C (Forest research) (1999) Phytophthora Pathogens of Trees: Their Rising Profile in Europe.

[27] Erwin DC, Ribeiro OK (1996) Phytophthora diseases worldwide (APS Press).

[28] Agrios GN (2005) *Plant Pathology* (Google eBook) (Academic Press).

[29] Blair JE, Coffey MD, Park S-Y, Geiser DM, Kang S (2008) A multilocus phylogeny for Phytophthora utilizing markers derived from complete genome sequences. *Fungal Genet Biol* 45:266277.

[30] Haas BJ, Kamoun S, et al. (2009) Genome sequence and analysis of the Irish potato famine pathogen Phytophthora infestans. *Nature* 461:393 8.

[31] Morris PF, Ward EWB (1992) Chemoattraction of zoospores of the soybean pathogen, Phytophthora sojae, by isoflavones. *Physiol Mol Plant Pathol* 40:1722.

[32] Judelson HS, Blanco FA (2005) The spores of Phytophthora: weapons of the plant destroyer. *Nat Rev Microbiol* 3:4758.

[33] Tyler BM (2007) Phytophthora sojae: root rot pathogen of soybean and model oomycete. *Mol Plant Pathol* 8:18.

[34] Whisson SC, Boevink PC, et al. (2007) A translocation signal for delivery of oomycete effector proteins into host plant cells. *Nature* 450:1158.

[35] Randall TA, Dwyer RA, et al. (2005) Large-scale gene discovery in the oomycete Phytophthora infestans reveals likely components of phytopathogenicity shared with true fungi. *Mol Plant Microbe Interact* 18:22943.

[36] Jupe J, Stam R, Howden AJ, Morris JA, Zhang R, Hedley PE, Huitema E (2013) Phytophthora capsici-tomato interaction features dramatic shifts in gene expression associated with a hemi-biotrophic lifestyle. *Genome Biol* 14:R63.

[37] Jones JDG, Dangl JL (2006) The plant immune system. *Nature* 444:323 9.

[38] Schulze-Levert P, Farmer E, Zipfel C, Felix G (2005) Plants and animals: a different taste for microbes? *Curr Opin Plant Biol* 8:353360.

[39] Nicaise V, Roux M, Zipfel C (2009) Recent advances in PAMP-triggered immunity against bacteria: pattern recognition receptors watch over and raise the alarm. *Plant Physiol* 150:163847.

[40] Nguyen HP, Chakravarthy S, Velásquez AC, McLane HL, Zeng L, Nakayashiki H, Park D-H, Collmer A, Martin GB (2010) Methods to study PAMP-triggered immunity using tomato and Nicotiana benthamiana. *Mol Plant Microbe Interact* 23:9919.

[41] Bos JIB, Armstrong MR, et al. (2010) Phytophthora infestans effector AVR3a is essential for virulence and manipulates plant immunity by stabilizing host E3 ligase CMPG1. *Proc Natl Acad Sci U S A* 107:990914.

[42] Dong S, Yin W, et al. (2011) Phytophthora sojae avirulence effector Avr3b is a secreted NADH and ADP-ribose pyrophosphorylase that modulates plant immunity. *PLoS Pathog* 7:e1002353.

[43] Marino D, Froidure S, Canonne J, Ben Khaled S, Khafif M, Pouzet C, Jauneau A, Roby D, Rivas S (2013) Arabidopsis ubiquitin ligase MIEL1 mediates degradation of the transcription factor MYB30 weakening plant defence. *Nat Commun* 4:1476.

[44] González-Lamothe R, Tsitsigiannis DI, Ludwig AA, Panicot M, Shirasu K, Jones JDG (2006) The U-box protein CMPG1 is required for efficient activation of defense mechanisms triggered by multiple resistance genes in tobacco and tomato. *Plant Cell* 18:106783.

[45] Kawasaki T, Nam J, Boyes DC, Holt BF, Hubert DA, Wiig A, Dangl JL (2005) A duplicated pair of Arabidopsis RING-finger E3 ligases contribute to the RPM1- and RPS2-mediated hypersensitive response. *Plant J* 44:25870.

[46] Flor H.H. (1942) Inheritance of pathogenicity in Melampsora lini. *Phytopathology* :653669.

[47] Govers F, Bouwmeester K (2008) Effector trafficking: RXLR-dEER as extra gear for delivery into plant cells. *Plant Cell* 20:172830.

[48] Rehmany AP, Gordon A, et al. (2005) Differential recognition of highly divergent downy mildew avirulence gene alleles by RPP1 resistance genes from two Arabidopsis lines. *Plant Cell* 17:183950.

[49] Boutemy LS, King SR, Win J, Hughes RK, Clarke TA, Blumenschein TM, Kamoun S, Banfield MJ (2011) Structures of Phytophthora RXLR effector proteins: a conserved but adaptable fold underpins functional diversity. *J Biol Chem* 286:3583435842.

[50] DeLano WL The PyMOL Molecular Graphics System.

[51] Bush Everett, K. D. RM (2001) Molecular evolution of the Chlamydiaceae. *Int J Syst Evol Microbiol* 51:203220.

[52] Clifton DR, Dooley CA, Grieshaber SS, Carabeo RA, Fields KA, Hackstadt T (2005) Tyrosine phosphorylation of the chlamydial effector protein Tarp is species specific and not required for recruitment of actin. *Infect Immun* 73:38603868.

[53] Brinkman FS, Blanchard JL, et al. (2002) Evidence that plant-like genes in Chlamydia species reflect an ancestral relationship between Chlamydiaceae, cyanobacteria, and the chloroplast. *Genome Res* 12:11591167.

[54] Matsumoto A (1982) Surface projections of Chlamydia psittaci elementary bodies as revealed by freeze-deep-etching. *J Bacteriol* 151:10401042.

[55] Moulder JW (1982) The relation of basic biology to pathogenic potential in the genus Chlamydia. Infection 10 Suppl 1:S108.

[56] Engel J (2004) Tarp and Arp: How Chlamydia induces its own entry. *Proc Natl Acad Sci U S A* 101:99478.

[57] Shahmanesh M (2000) Chlamydia Intercellular Biology Pathogenesis and Immunity. *Sex Transm Infect* 76:324.

[58] Barron AL (1988) Microbiology of Chlamydia:The Chlamydial developmental cycle (CRC press).

[59] Hybiske Stephens, R. S. K (2007) Mechanisms of Chlamydia trachomatis entry into nonphagocytic cells. *Infect Immun* 75:3925 3934.

[60] Elwell CA, Ceesay A, Kim JH, Kalman D, Engel JN (2008) RNA interference screen identifies Abl kinase and PDGFR signaling in Chlamydia trachomatis entry. *PLoS Pathog* 4:e1000021.

[61] Carabeo RA, Grieshaber SS, Fischer E, Hackstadt T (2002) Chlamydia trachomatis induces remodeling of the actin cytoskeleton during attachment and entry into HeLa cells. *Infect Immun* 70:37933803.

[62] Hackstadt T, Fischer ER, Scidmore MA, Rockey DD, Heinzen RA (1997) Origins and functions of the chlamydial inclusion. *Trends Microbiol* 5:288293.

[63] Gérard HC, Freise J, et al. (2002) Chlamydia trachomatis genes whose products are related to energy metabolism are expressed differentially in active vs. persistent infection. *Microbes Infect* 4:1322.

[64] Hybiske K, Stephens RS (2007) Mechanisms of host cell exit by the intracellular bacterium Chlamydia. *Proc Natl Acad Sci U S A* 104:11430 5.

[65] Emerson PM, Bailey RL, Mahdi OS, Walraven GE, Lindsay SW (2000) Transmission ecology of the fly Musca sorbens, a putative vector of trachoma. *Trans R Soc Trop Med Hyg* 94:2832.

[66] WHO, World Health Organization The International Trachoma Initiative. the LS of H&TM (2006) Trachoma control: A guide for programme managers.

[67] WHO — Global incidence and prevalence of selected curable sexually transmitted infections 2008.

[68] CDC. (Center for Prevention Services (U.S.). Division of Sexually Transmitted Diseases). (2008) Sexually Transmitted Diseases Surveillance, 2008.

[69] Bjartling Osser, S., Persson, K. C (2000) The frequency of salpingitis and ectopic pregnancy as epidemiologic markers of Chlamydia trachomatis. *Acta Obs Gynecol Scand* 79:123128.

[70] Ochsendorf FR, Ozdemir K, Rabenau H, Fenner T, Oremek R, Milbradt R, Doerr HW (1999) Chlamydia trachomatis and male infertility: chlamydia-IgA antibodies in seminal plasma are C. trachomatis specific and associated with an inflammatory response. *J Eur Acad Dermatol Venereol* 12:143152.

[71] Papagrigoriadis Rennie, J. A. S (1998) Lymphogranuloma venereum as a cause of rectal strictures. *Postgr Med J* 74:168169.

[72] Yucesan Sriram, S. C (2001) Chlamydia pneumoniae infection of the central nervous system. *Curr Opin Neurol* 14:355359.

[73] Jewett TJ, Fischer ER, Mead DJ, Hackstadt T (2006) Chlamydial TARP is a bacterial nucleator of actin. *Proc Natl Acad Sci U S A* 103:15599 15604.

[74] Clifton DR, Fields KA, Grieshaber SS, Dooley CA, Fischer ER, Mead DJ, Carabeo RA, Hackstadt T (2004) A chlamydial type III translocated protein is tyrosine-phosphorylated at the site of entry and associated with recruitment of actin. *Proc Natl Acad Sci U S A* 101:1016610171.

[75] Jewett TJ, Dooley CA, Mead DJ, Hackstadt T (2008) Chlamydia trachomatis tarp is phosphorylated by src family tyrosine kinases. *Biochem Biophys Res Commun* 371:339344.

[76] Jewett Miller, N. J., Dooley, C. A., Hackstadt, T. TJ (2010) The Conserved TARP actin binding domain is important for Chlamydial invasion. *PLoS Pathog* 7:e1000997

[77] Paunola E, Mattila PK, Lappalainen P (2002) WH2 domain: a small, versatile adapter for actin monomers. *FEBS Lett* 513:9297.

[78] Lane BJ, Mutchler C, Al Khodor S, Grieshaber SS, Carabeo RA (2008) Chlamydial entry involves TARP binding of guanine nucleotide exchange factors. *PLoS Pathog* 4:e1000014.

[79] Carabeo RA, Dooley CA, Grieshaber SS, Hackstadt T (2007) Rac interacts with Abi-1 and WAVE2 to promote an Arp2/3-dependent actin recruitment during chlamydial invasion. *Cell Microbiol* 9:2278 2288.

[80] Mehlitz A, Banhart S, Hess S, Selbach M, Meyer TF (2008) Complex kinase requirements for Chlamydia trachomatis Tarp phosphorylation. *FEMS Microbiol Lett* 289:233240.

[81] Hanahan D (1985) Chapter 6: Techniques for Transformation of E.Coli. In: Glover, D. M. DNA cloning, Volume 1, A practical approach ed Glover D (IRL Press).

[82] Studier FW (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* 41:207234.

[83] Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6:277293.

[84] Wishart DS, Bigam CG, Yao J, Abildgaard F, Dyson HJ, Oldfield E, Markley JL, Sykes BD (1995) 1H, 13C and 15N chemical shift referencing in biomolecular NMR. *J Biomol NMR* 6:135140.

[85] Markley JL, Bax A, Arata Y, Hilbers CW, Kaptein R, Sykes BD, Wright PE, Wüthrich K (1998) Recommendations for the presentation of NMR structures of proteins and nucleic acids. *J Mol Biol* 280:933952.

[86] Vranken WF, Boucher W, et al. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* 59:687 696.

[87] Ulrich EL, Akutsu H, et al. (2008) BioMagResBank. *Nucleic Acids Res* 36:D4028.

[88] Herrmann T, Guntert P, Wüthrich K (2002) Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *J Biomol NMR* 24:171189.

[89] Herrmann T, Guntert P, Wüthrich K (2002) Protein NMR structure determination with automated NOE assignment using the new software

CANDID and the torsion angle dynamics algorithm DYANA. *J Mol Biol* 319:209227.

[90] Volk J, Herrmann T, Wüthrich K (2008) Automated sequence-specific protein NMR assignment using the memetic algorithm MATCH. *J Biomol NMR* 41:127138.

[91] Fiorito F, Herrmann T, Damberger FF, Wüthrich K (2008) Automated amino acid side-chain NMR assignment of proteins using (13)C- and (15)N-resolved 3D [(1)H, (1)H]-NOESY. *J Biomol NMR* 42:2333.

[92] Bartels C, Xia TH, Billeter M, Guntert P, Wüthrich K (1995) The Program Xeasy for Computer-Supported Nmr Spectral-Analysis of Biological Macromolecules. *J Biomol NMR* 6:110.

[93] Keller R (2004) The Computer Aided Resonance Assignment Tutorial (Cantina Verlag).

[94] Guntert P (2004) Automated NMR structure calculation with CYANA. *Methods Mol Biol* 278:353378.

[95] Shen Y, Delaglio F, Cornilescu G, Bax A (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* 44:213223.

[96] Guntert P, Mumenthaler C, Wuthrich K (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J Mol Biol* 273:283298.

[97] Guntert P, Braun W, Wüthrich K (1991) Efficient computation of three-dimensional protein structures in solution from nuclear magnetic resonance data using the program DIANA and the supporting programs CALIBA, HABAS and GLOMSA. *J Mol Biol* 217:517530.

[98] Brünger a T, Adams PD, et al. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 54:90521.

[99] Nederveen AJ, Doreleijers JF, et al. (2005) RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank. *Proteins* 59:66272.

[100] Doreleijers JF, Sousa da Silva AW, Krieger E, Nabuurs SB, Spronk CAEM, Stevens TJ, Vranken WF, Vriend G, Vuister GW (2012) CING: an integrated residue-based structure validation program suite. *J Biomol NMR* 54:26783.

[101] Bhattacharya A, Tejero R, Montelione GT (2007) Evaluating protein structures determined by structural genomics consortia. *Proteins* 66:778795.

[102] Laskowski RA, Macarthur MW, Moss DS, Thornton JM (1993) Procheck - a Program to Check the Stereochemical Quality of Protein Structures. *J Appl Crystallogr* 26:283291.

[103] Vriend G (1990) WHAT IF: a molecular modeling and drug design program. *J Mol Graph* 8:526, 29.

[104] Bieri M, dAuvergne EJ, Gooley PR relaxGUI: a new software for fast and simple NMR relaxation data analysis and calculation of ps-ns and mus motion of proteins. *J Biomol NMR* 50:147155.

[105] Brüschweiler R, Liao X, Wright PE (1995) Long-range motional restrictions in a multidomain zinc-finger protein from anisotropic tumbling. *Science* 268:8869.

[106] Lee LK, Rance M, Chazin WJ, Palmer AG (1997) Rotational diffusion anisotropy of proteins from simultaneous analysis of 15N and 13C alpha nuclear spin relaxation. *J Biomol NMR* 9:28798.

[107] Mandel AM, Akke M, Palmer AG (1995) Backbone dynamics of Escherichia coli ribonuclease HI: correlations with structure and function in an active enzyme. *J Mol Biol* 246:14463.

[108] Palmer AG, Rance M, Wright PE (1991) Intramolecular motions of a zinc finger DNA-binding domain from Xfin characterized by proton- detected natural abundance carbon-13 heteronuclear NMR spectroscopy. *J Am Chem Soc* 113:43714380.

[109] Origin (OriginLab, Northampton, MA).

[110] Pardee JD, Spudich JA (1982) in Structural and Contractile Proteins Part B: The Contractile Apparatus and the Cytoskeleton, Methods in Enzymology., ed Dixie W. Frederiksen LWC (Academic Press), pp 164 181.

[111] Baryshnikova OK, Williams TC, Sykes BD (2008) Internal pH indicators for biomolecular NMR. *J Biomol NMR* 41:57.

[112] Miles A., Wien F, Lees J., Rodger A, Janes R., Wallace B. (2003) Calibration and standardisation of synchrotron radiation circular dichroism and conventional circular dichroism spectrophotometers. *Spectroscopy* 17:653661.

[113] Lees JG, Smith BR, Wien F, Miles AJ, Wallace BA (2004) CDtoolan integrated software package for circular dichroism spectroscopic data processing, analysis, and archiving. *Anal Biochem* 332:285289.

[114] Whitmore L, Wallace BA (2008) Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers* 89:392400.

[115] Provencher SW, Glockner J (1981) Estimation of globular protein secondary structure from circular dichroism. *Biochemistry* 20:337.

[116] Van Stokkum IH, Spoelder HJ, Bloemendal M, van Grondelle R, Groen FC (1990) Estimation of protein secondary structure and error analysis from circular dichroism spectra. *Anal Biochem* 191:1108.

[117] Sreerama N, Woody RW (2000) Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal Biochem* 287:25260.

[118] London N, Raveh B, Cohen E, Fathi G, Schueler-Furman O (2011) Rosetta FlexPepDock web server–high resolution modeling of peptide- protein interactions. *Nucleic Acids Res* 39:W24953.

[119] Chereau D, Kerff F, Graceffa P, Grabarek Z, Langsetmo K, Dominguez R (2005) Actin-bound structures of Wiskott-Aldrich syndrome protein (WASP)-homology domain 2 and the implications for filament assembly. *Proc Natl Acad Sci U S A* 102:166449.

[120] Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen M-Y, Pieper U, Sali A (2007) Comparative protein structure modeling using MODELLER. *Curr Protoc Protein Sci* Chapter 2:Unit 2.9.

[121] Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera–a visualization system for exploratory research and analysis. *J Comput Chem* 25:160512.

[122] Tyburn JM (1998) Variable Temperature Unit User Manual, Version 001 (Bruker SA).

[123] Levitt MH (2008) Spin dynamics : basics of nuclear magnetic resonance (John Wiley & Sons, Chichester, England ; Hoboken, NJ). 2nd Ed.

[124] Evans JNS (1995) Biomolecular NMR spectroscopy (Oxford University Press, Oxford ; New York).

[125] Haken Wolf, H. C. H (2000) The physics of atoms and quanta : introduction to experiments and theory (Springer, Berlin ; New York)6th rev. a.

[126] Guccione Vigni, M. L., Vetri, G. M (1986) The Dynamic Zeeman Effect in the 2-Photon Resonance Condition. *J Phys C-Solid State Phys* 19:1813 1821.

[127] Cavanagh J (1996) Protein NMR spectroscopy : principles and practice (Academic Press, San Diego).

[128] Harris RK (1986) Nuclear magnetic resonance spectroscopy : a physicochemical view (Longman Scientific & Technical ; J. Wiley, Eastern Distribution Center distributor, Burnt Mill, Harlow, Essex, England, Somerset, NJ, USA)Repr. and .

[129] Teng Q (2012) Structural Biology: Practical NMR Applications (Google eBook) (Springer).

[130] Palmer AG, Cavanagh J, Wright PE, Rance M (1991) Sensitivity improvement in proton-detected two-dimensional heteronuclear correlation NMR spectroscopy. *J Magn Reson* 93:151170.

[131] Kay L, Keifer P, Saarinen T (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J Am Chem Soc* 114:1066310665.

[132] Schleucher J, Schwendinger M, Sattler M, Schmidt P, Schedletzky O, Glaser SJ, Sørensen OW, Griesinger C (1994) A general enhancement

scheme in heteronuclear multidimensional NMR employing pulsed field gradients. *J Biomol NMR* 4:3016.

[133] Shaka A., Barker P., Freeman R (1985) Computer-optimized decoupling scheme for wideband applications and low-level operation.

[134] Triple Resonance Backbone Assignment Available at: `http://www.protein-nmr.org.uk/solution-nmr/assignment-theory/triple-resonance-backbone-assignment/`.

[135] Wüthrich K (1983) Sequential individual resonance assignments in the 1H-nmr spectra of polypeptides and proteins. Biopolymers 22:1318.

[136] Figueiredo AMM. (2010) Navigation of the folding landscape of colicin immunity proteins. Dissertation (University of East Anglia).

[137] Bernstein FC, Koetzle TF, Williams GJ, Meyer EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M (1977) The Protein Data Bank: a computer-based archival file for macromolecular structures. *J Mol Biol* 112:53542.

[138] Branden C, Tooze J (1999) Introduction to protein structure (Garland Pub, New York)2nd ed.

[139] Kumar S, Nussinov R (2002) Close-range electrostatic interactions in proteins. *Chembiochem* 3:604617.

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

[141] Lovell SC, Davis IW, Arendall 3rd WB, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC (2003) Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins* 50:437450.

[142] Davis IW, Leaver-Fay A, et al. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* 35:W37583.

[143] Figueiredo AMMP (2010) Navigation of the folding landscape of colicin immunity proteins. PhD thesis (University of East Anglia).

[144] Hendrickx PMS, Gutmanas A, Kleywegt GJ (2013) Vivaldi: visualization and validation of biomacromolecular NMR structures from the PDB. *Proteins* 81:58391.

[145] Bos JIB, Chaparro-Garcia A, Quesada-Ocampo LM, McSpadden Gardener BB, Kamoun S (2009) Distinct amino acids of the Phytophthora infestans effector AVR3a condition activation of R3a hypersensitivity and suppression of cell death. *Mol Plant Microbe Interact* 22:26981.

[146] Carr H, Purcell E (1954) Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments. *Phys Rev* 94:630638.

[147] Anderson WA, Freeman R (1962) Influence of a Second Radiofrequency Field on High-Resolution Nuclear Magnetic Resonance Spectra. *J Chem Phys* 37:85.

[148] Abragam A (1961) The Principles of Nuclear Magnetism (Oxford University Press).

[149] Ottiger M, Bax A (1998) Determination of Relative N-H N , N-C′, Cα-C′, and Cα-Hα Effective Bond Lengths in a Protein by NMR in a Dilute Liquid Crystalline Phase. *J Am Chem Soc* 120:1233412341.

[150] Kroenke CD, Rance M, Palmer AG (1999) Variability of the 15 N Chemical Shift Anisotropy in Escherichia coli Ribonuclease H in Solution. *J Am Chem Soc* 121:1011910125.

[151] Peng JW, Wagner G (1992) Mapping of the spectral densities of nitrogen-hydrogen bond motions in Eglin c using heteronuclear relaxation experiments. *Biochemistry* 31:85718586.

[152] Farrow NA, Zhang O, Szabo A, Torchia DA, Kay LE (1995) Spectral density function mapping using 15N relaxation data exclusively. *J Biomol NMR* 6:15362.

[153] Lipari G, Szabo A (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J Am Chem Soc* 104:45464559.

[154] Lipari G, Szabo A (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. *J Am Chem Soc* 104:45594570.

[155] Tjandra N, Feller SE, Pastor RW, Bax A (1995) Rotational diffusion anisotropy of human ubiquitin from 15N NMR relaxation. *J Am Chem Soc* 117:1256212566.

[156] Colins ES (2001) NMR studies of the structural dynamics and intermolecular interations of colicin E9 and its inhibitor protein. PhD thesis (University of East Anglia).

[157] Clore GM, Szabo A, Bax A, Kay LE, Driscoll P, Gronenborn AM (1990) Deviations from the Simple Two-Parameter Model-Free Approach to the Interpretation of Nitrogen-15 Nuclear Magnetic Relaxation of Proteins. *J Am Chem Soc* 112:49894991.

[158] Clarke J, Hounslow AM, Bycroft M, Fersht AR (1993) Local breathing and global unfolding in hydrogen exchange of barnase and its relationship to protein folding pathways. *Proc Natl Acad Sci U S A* 90:983741.

[159] Huyghues-Despointes BMP, Pace NC, Englander SW, Scholtz JM in Methods in molecular biology (Vol. 68), pp 6992.

[160] Bai Y, Milne JS, Mayne L, Englander SW (1993) Primary structure effects on peptide group hydrogen exchange. *Proteins* 17:7586.

[161] Tjandra N, Wingfield P, Stahl S, Bax A (1996) Anisotropic rotational diffusion of perdeuterated HIV protease from 15N NMR relaxation measurements at two magnetic fields. *J Biomol NMR* 8:27384.

[162] Luginbühl P, Pervushin K V, Iwai H, Wüthrich K (1997) Anisotropic molecular rotational diffusion in 15N spin relaxation studies of protein mobility. *Biochemistry* 36:730512.

[163] Sahu SC, Bhuyan AK, Majumdar A, Udgaonkar JB (2000) Backbone dynamics of barstar: a (15)N NMR relaxation study. *Proteins* 41:460 74.

[164] Figueiredo AM, Whittaker SB-M, Knowling SE, Radford SE, Moore GR (2013) Conformational dynamics is more important than helical propensity for the folding of the all α-helical protein Im7. *Protein Sci* 22:172238.

[165] Garcia de la Torre J, Huertas ML, Carrasco B (2000) HYDRONMR: prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. *J Magn Reson* 147:138146.

[166] Bernadó P, García de la Torre J, Pons M (2002) Interpretation of 15N NMR relaxation data of globular proteins using hydrodynamic calculations with HYDRONMR. *J Biomol NMR* 23:13950.

[167] Chang S-L, Hinck AP, Ishima R (2007) Model-free analysis for large proteins at high magnetic field strengths. *J Biomol NMR* 38:31524.

[168] Win J, Krasileva K V, Kamoun S, Shirasu K, Staskawicz BJ, Banfield MJ (2012) Sequence divergent RXLR effectors share a structural fold conserved across plant pathogenic oomycete species. *PLoS Pathog* 8:e1002400.

[169] Clore GM, Schwieters CD (2006) Concordance of residual dipolar couplings, backbone order parameters and crystallographic B-factors for a small alpha/beta protein: a unified picture of high probability, fast atomic motions in proteins. *J Mol Biol* 355:87986.

[170] Yang J, Tasayco ML, Polenova T (2009) Dynamics of reassembled thioredoxin studied by magic angle spinning NMR: snapshots from different time scales. *J Am Chem Soc* 131:13690702.

[171] Yang L-W, Eyal E, Chennubhotla C, Jee J, Gronenborn AM, Bahar I (2007) Insights into equilibrium dynamics of proteins from comparison of NMR and X-ray data with computational predictions. *Structure* 15:7419.

[172] Reichert D, Zinkevich T, Saalwächter K, Krushelnitsky A (2012) The relation of the X-ray B-factor to protein dynamics: insights from recent dynamic solid-state NMR data. *J Biomol Struct Dyn* 30:61727.

[173] Li D-W, Brüschweiler R (2009) All-atom contact model for understanding protein dynamics from crystallographic B-factors. *Biophys J* 96:307481.

[174] Tompa P, Fersht A (2009) Structure and Function of Intrinsically Disordered Proteins (Chapman and Hall/CRC).

[175] Dunker AK, Obradovic Z (2001) The protein trinity–linking function and disorder. *Nat Biotechnol* 19:8056.

[176] Dunker AK, Oldfield CJ, Meng J, Romero P, Yang JY, Chen JW, Vacic V, Obradovic Z, Uversky VN (2008) The unfoldomics decade: an update on intrinsically disordered proteins. *BMC Genomics* 9 Suppl 2:S1.

[177] Csermely P, Palotai R, Nussinov R (2010) Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends Biochem Sci* 35:53946.

[178] Hsu W-L, Oldfield C, Meng J, Huang F, Xue B, Uversky VN, Romero P, Dunker AK (2012) Intrinsic protein disorder and protein-protein interactions. *Pac Symp Biocomput*:11627.

[179] Musiani F, Ippoliti E, Micheletti C, Carloni P, Ciurli S (2013) Conformational fluctuations of UreG, an intrinsically disordered enzyme. *Biochemistry* 52:294954.

[180] Hahn MW, Wray GA (2002) The g-value paradox. *Evol Dev* 4:7375.

[181] Schad E, Tompa P, Hegyi H (2011) The relationship between proteome size, structural disorder and organism complexity. *Genome Biol* 12:R120.

[182] Uversky VN, Oldfield CJ, Dunker AK (2008) Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu Rev Biophys* 37:21546.

[183] Erickson HP (2009) Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy. *Biol Proced Online* 11:3251.

[184] Fischer H, Polikarpov I, Craievich AF (2004) Average protein density is a molecular-weight-dependent function. *Protein Sci* 13:28258.

[185] Uversky VN (2002) Natively unfolded proteins: a point where biology waits for physics. *Protein Sci* 11:73956.

[186] Marsh JA, Forman-Kay JD (2010) Sequence determinants of compaction in intrinsically disordered proteins. *Biophys J* 98:238390.

[187] Tompa P (2002) Intrinsically unstructured proteins. *Trends Biochem Sci* 27:527533.

[188] Bernadó P, Mylonas E, Petoukhov M V, Blackledge M, Svergun DI (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. *J Am Chem Soc* 129:565664.

[189] Schneidman-Duhovny D, Kim SJ, Sali A (2012) Integrative structural modeling with small angle X-ray scattering profiles. *BMC Struct Biol* 12:17.

[190] Tzeng S-R, Pai M-T, Kalodimos CG (2012) NMR studies of large protein systems. *Methods Mol Biol* 831:13340.

[191] Dyson HJ, Wright PE (2002) Insights into the structure and dynamics of unfolded proteins from nuclear magnetic resonance. *Adv Protein Chem* 62:31140.

[192] Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* 292:195202.

[193] Buchan DWA, Minneci F, Nugent TCO, Bryson K, Jones DT (2013) Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res* 41:W34957.

[194] Uversky VN, Gillespie JR, Fink AL (2000) Why are natively unfolded proteins unstructured under physiologic conditions? Proteins 41:415 427.

[195] Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. *J Mol Biol* 157:105132.

[196] Xue B, Dunbrack RL, Williams RW, Dunker AK, Uversky VN PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochim Biophys Acta* 1804:9961010.

[197] Rose Geselowitz, A.R., Lesser, G.J., Lee, R.H., Zehfus, M.H. GD (1985) Amino acid scale: Average area buried on transfer from standard state to folded protein. *Science* (80-) 229:4.

[198] Abdalla A-M, Bruns CM, Tainer JA, Mannervik B, Stenberg G (2002) Design of a monomeric human glutathione transferase GSTP1, a structurally stable but catalytically inactive protein. *Protein Eng* 15:82734.

[199] Laurent TC, Killander J (1964) A theory of gel filtration and its exeperimental verification. *J Chromatogr A* 14:317330.

[200] Gel filtration: Principles and methods (Amersham pharmacia biotech). 8th Ed.

[201] Blake-Hodek KA, Cassimeris L, Huffaker TC (2010) Regulation of microtubule dynamics by Bim1 and Bik1, the budding yeast members of the EB1 and CLIP-170 families of plus-end tracking proteins. *Mol Biol Cell* 21:201323.

[202] Wallace BA (2009) Protein characterisation by synchrotron radiation circular dichroism spectroscopy. *Q Rev Biophys* 42:31770.

[203] Lopes JLS, Orcia D, Araujo APU, DeMarco R, Wallace BA (2013) Folding factors and partners for the intrinsically disordered protein micro-exon gene 14 (MEG-14). *Biophys J* 104:251220.

[204] Uversky VN, Dunker AK (2010) Understanding protein non-folding. *Biochim Biophys Acta* 1804:123164.

[205] Uversky VN (2009) Intrinsically disordered proteins and their environment: effects of strong denaturants, temperature, pH, counter ions, membranes, binding partners, osmolytes, and macromolecular crowding. *Protein J* 28:30525.

[206] Kjaergaard M, Nørholm A-B, Hendus-Altenburger R, Pedersen SF, Poulsen FM, Kragelund BB (2010) Temperature-dependent structural changes in intrinsically disordered proteins: formation of alpha-helices or loss of polyproline II? *Protein Sci* 19:155564.

[207] Adzhubei AA, Sternberg MJE, Makarov AA (2013) Polyproline-II Helix in Proteins: Structure and Function. *J Mol Biol* 425:21002132.

[208] Panchal Bhavesh, N. S., Hosur, R. V. SC (2001) Improved 3D triple resonance experiments, HNN and HN(C)N, for HN and 15N sequential correlations in (13C, 15N) labeled proteins: application to unfolded proteins. *J Biomol NMR* 20:135147.

[209] Cierpicki T, Otlewski J (2001) Amide proton temperature coefficients as hydrogen bond indicators in proteins. *J Biomol NMR* 21:249261.

[210] Cook BW, Shaw GS (2012) Architecture of the catalytic HPN motif is conserved in all E2 conjugating enzymes. *Biochem J* 445:16774.

[211] Tamiola K, Acar B, Mulder FAA (2010) Sequence-specific random coil chemical shifts of intrinsically disordered proteins. *J Am Chem Soc* 132:180003.

[212] Marsh JA, Singh VK, Jia Z, Forman-Kay JD (2006) Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation. *Protein Sci* 15:2795804.

[213] Wishart D, Sykes B (1994) The 13C Chemical-Shift Index: A simple method for the identification of protein secondary structure using 13C chemical-shift data. *J Biomol NMR* 4.

[214] Traficante DD (1991) Relaxation. Can T2, be longer than T1? *Concepts Magn Reson* 3:171177.

[215] R Dominguez (2007) in Actin-Monomer-Binding Proteins (Landes Bioscience and Springer Science and Business Media), pp 107115.

[216] Csizmók V, Szollosi E, Friedrich P, Tompa P (2006) A novel two- dimensional electrophoresis technique for the identification of intrinsically unstructured proteins. *Mol Cell Proteomics* 5:26573.

[217] Ganem B, Henion JD (2003) Going gently into flight: analyzing noncovalent interactions by mass spectrometry. *Bioorg Med Chem* 11:311314.

[218] Light-Wahl KJ, Schwartz BL, Smith RD (1994) Observation of the Noncovalent Quaternary Associations of Proteins by Electrospray Ionization Mass Spectrometry. *J Am Chem Soc* 116:52715278.

[219] Schwalbe H, Fiebig KM, Buck M, Jones JA, Grimshaw SB, Spencer A, Glaser SJ, Smith LJ, Dobson CM (1997) Structural and dynamical properties of a denatured protein. Heteronuclear 3D NMR experiments and theoretical simulations of lysozyme in 8 M urea. *Biochemistry* 36:897791.

[220] Uversky VN (2011) Intrinsically disordered proteins from A to Z. *Int J Biochem Cell Biol* 43:10901103.

[221] Janin J, Sternberg MJE (2013) Protein flexibility, not disorder, is intrinsic to molecular recognition. *F1000 Biol Rep* 5:2.

[222] Uversky VN, Dunker AK (2013) The case for intrinsically disordered proteins playing contributory roles in molecular recognition without a stable 3D structure. *F1000Prime Rep* 5.

[223] Mukhopadhyay R, Hoh JH (2001) AFM force measurements on microtubule-associated proteins: the projection domain exerts a long-range repulsive force. *FEBS Lett* 505:3748.

[224] Tompa P, Fuxreiter M (2008) Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem Sci* 33:28.

[225] Sugase Dyson, H. J., Wright, P. E. K (2007) Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* 447:10211025.

[226] Shoemaker BA, Portman JJ, Wolynes PG (2000) Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc Natl Acad Sci U S A* 97:886873.

[227] Dyson HJ, Wright PE (2002) Coupling of folding and binding for unstructured proteins. *Curr Opin Struct Biol* 12:5460.

[228] Alberts B (2002) Molecular biology of the cell (Garland Science, New York). 4th Ed.

[229] May R, Machesky L (2001) Phagocytosis and the actin cytoskeleton. *J Cell Sci* 114:10611077.

[230] Otterbein Graceffa, P., Dominguez, R. LR (2001) The crystal structure of uncomplexed actin in the ADP state. *Science* (80-) 293:708711.

[231] Stryer L, Berg JM, Tymoczko JL (2002) Biochemistry (W.H.Freeman & Co Ltd).

[232] Blanchoin L, Pollard TD (2002) Hydrolysis of ATP by polymerized actin depends on the bound divalent cation but not profilin. *Biochemistry* 41:597602.

[233] Bryan J, Coluccio LM (1985) Kinetic analysis of F-actin depolymerization in the presence of platelet gelsolin and gelsolin-actin complexes. *J Cell Biol* 101:123644.

[234] Winder SJ, Ayscough KR (2005) Actin-binding proteins. *J Cell Sci* 118:6514.

[235] Mofrad Kamm, Roger D. MRK (2006) Cytoskeletal mechanics : models and measurements (Cambridge University Press, Cambridge ; New York).

[236] Goley ED, Welch MD (2006) The ARP2/3 complex: an actin nucleator comes of age. *Nat Rev Mol Cell Biol* 7:71326.

[237] Pavlov D, Muhlrad A, Cooper J, Wear M, Reisler E (2007) Actin filament severing by cofilin. *J Mol Biol* 365:13508.

[238] Derry JM, Ochs HD, Francke U (1994) Isolation of a novel gene mutated in Wiskott-Aldrich syndrome. Cell 79:following 922.

[239] Ducka AM, Joel P, Popowicz GM, Trybus KM, Schleicher M, Noegel AA, Huber R, Holak TA, Sitar T (2010) Structures of actin-bound Wiskott-Aldrich syndrome protein homology 2 (WH2) domains of Spire and the implication for filament nucleation. *Proc Natl Acad Sci U S A* 107:1175762.

[240] Irobi E, Aguda AH, Larsson M, Guerin C, Yin HL, Burtnick LD, Blanchoin L, Robinson RC (2004) Structural basis of actin sequestration by thymosin-beta4: implications for WH2 proteins. *EMBO J* 23:3599608.

[241] Hertzog M, van Heijenoort C, et al. (2004) The beta-thymosin/WH2 domain; structural basis for the switch from inhibition to promotion of actin assembly. *Cell* 117:61123.

[242] Notredame Higgins, D. G., Heringa, J. C (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* 302:205217.

[243] Poirot O, OToole E, Notredame C (2003) Tcoffee@igs: A web server for computing, evaluating and combining multiple sequence alignments. *Nucleic Acids Res* 31:35033506.

[244] Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:17927.

[245] Cowieson NP, Miles AJ, Robin G, Forwood JK, Kobe B, Martin JL, Wallace BA (2008) Evaluating protein:protein complex formation using synchrotron radiation circular dichroism spectroscopy. *Proteins* 70:11426.

[246] Raveh B, London N, Schueler-Furman O (2010) Sub-angstrom modeling of complexes between flexible peptides and globular proteins. *Proteins* 78:202940.

[247] Das R, Baker D (2008) Macromolecular modeling with rosetta. *Annu Rev Biochem* 77:36382.

[248] Bradley P, Misura KMS, Baker D (2005) Toward high-resolution de novo structure prediction for small proteins. *Science* 309:186871.

[249] Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L (2005) The FoldX web server: an online force field. *Nucleic Acids Res* 33:W382 8.

[250] Van Durme J, Delgado J, Stricher F, Serrano L, Schymkowitz J, Rousseau F (2011) A graphical interface for the FoldX forcefield. *Bioinformatics* 27:17112.

[251] Krieger E, Koraimann G, Vriend G (2002) Increasing the precision of comparative models with YASARA NOVA–a self-parameterizing force field. *Proteins* 47:393402.

[252] Guerois R, Nielsen JE, Serrano L (2002) Predicting Changes in the Stability of Proteins and Protein Complexes: A Study of More Than 1000 Mutations. *J Mol Biol* 320:369387.

[253] Graceffa P, Dominguez R (2003) Crystal structure of monomeric actin in the ATP state. Structural basis of nucleotide-dependent actin dynamics. *J Biol Chem* 278:3417280.

[254] Safer D, Sosnick TR, Elzinga M (1997) Thymosin beta 4 binds actin in an extended conformation and contacts both the barbed and pointed ends. *Biochemistry* 36:580616.

[255] Domanski M, Hertzog M, Coutant J, Gutsche-Perelroizen I, Bontems F, Carlier M-F, Guittet E, van Heijenoort C (2004) Coupling of folding and binding of thymosin beta4 upon interaction with monomeric actin monitored by nuclear magnetic resonance. *J Biol Chem* 279:2363745.

[256] Tompa P, Fuxreiter M, Oldfield CJ, Simon I, Dunker AK, Uversky VN (2009) Close encounters of the third kind: disordered domains and the interactions of proteins. *Bioessays* 31:32835.

[257] Wang H, Robinson RC, Burtnick LD (2010) The structure of native G-actin. Cytoskeleton (Hoboken) 67:45665.

[258] Halle B (2002) Flexibility and packing in proteins. *Proc Natl Acad Sci* 99:12741279.

[259] Putnam CD, Hammel M, Hura GL, Tainer JA (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q Rev Biophys* 40:191285.

[260] Fuxreiter M, Tompa P (2012) Fuzziness: Structural Disorder in Protein Complexes (Google eBook) (Springer).

[261] Renault L, Bugyi B, Carlier M-F (2008) Spire and Cordon-bleu: multifunctional regulators of actin dynamics. *Trends Cell Biol* 18:494 504.

Appendix A

$TARP_{726\text{-}825}$ NMR resonance assignment

Table 1: The NMR assignments for the $TARP_{726\text{-}825}$ construct main conformation, comprising the actin binding domain of the full length *C. trachomatis* LGV2 TARP protein (CT456). Construct residues 1-5 are remnants of the cleavage site from removal of the GST purification tag.

Residue number			$TARP_{726\text{-}825}$ chemical shifts				
Construct	Full length	Residue	HN	N	CO	Cα	Cβ
1	-	Gly	-	-	-	-	-
2	-	Pro	-	-	177.02	63.22	32.32
3	-	Leu	8.58	122.09	177.88	55.46	42.15
4	-	Gly	8.44	110.23	174.23	45.33	-
5	-	Ser	8.20	115.60	174.30	58.39	64.03
6	726	Asp	8.46	122.10	176.02	54.36	41.16
7	727	Asp	8.30	121.19	176.76	54.35	41.17
8	728	Ser	8.39	116.74	175.45	59.39	63.82
9	729	Gly	8.49	110.81	174.30	45.51	-
10	730	Ser	8.13	115.72	174.61	58.37	63.91
11	731	Val	8.19	121.64	176.31	62.33	32.74
12	732	Ser	8.45	119.65	174.73	58.32	63.93
13	733	Ser	8.46	118.33	174.77	58.52	63.93
14	734	Ser	8.45	117.91	174.89	58.66	63.88
15	735	Glu	8.45	122.67	176.94	57.06	30.13

Residue number			$TARP_{726-825}$ chemical shifts				
Construct	Full length	Residue	HN	N	CO	Cα	Cβ
16	736	Ser	8.24	116.09	174.64	59.02	63.73
17	737	Asp	8.29	122.20	176.82	54.51	40.96
18	738	Lys	8.19	121.66	176.83	56.91	32.58
19	739	Asn	8.36	118.80	175.10	53.49	38.93
20	740	Ala	8.09	124.03	177.79	52.85	19.22
21	741	Ser	8.27	115.02	174.64	58.39	63.81
22	742	Val	8.09	121.34	176.64	62.43	32.68
23	743	Gly	8.44	111.99	173.99	45.32	-
24	744	Asn	8.37	118.81	174.91	53.33	39.10
25	745	Asp	8.42	120.38	176.26	54.31	41.27
26	746	Gly	8.16	108.91	172.05	44.96	-
27	747	Pro	-	-	177.01	63.42	32.06
28	748	Ala	8.37	123.49	178.20	52.71	19.02
29	749	Met	8.28	119.50	176.54	55.76	32.47
30	750	Lys	8.23	121.39	176.50	56.97	32.94
31	751	Asp	8.24	120.81	176.69	54.68	41.04
32	752	Ile	8.01	121.02	176.84	62.06	38.49
33	753	Leu	8.24	123.84	177.91	55.73	42.00
34	754	Ser	8.08	115.66	174.61	58.99	63.64
35	755	Ala	8.12	125.34	177.92	52.92	19.16
36	756	Val	7.92	118.56	176.34	62.74	32.48
37	757	Arg	8.26	124.45	176.10	56.07	30.70
38	758	Lys	8.27	122.65	176.20	56.46	33.14
39	759	His	8.37	120.50	174.95	56.15	30.20
40	760	Leu	8.17	123.50	176.65	55.15	42.53
41	761	Asp	8.35	121.28	175.79	54.43	41.13
42	762	Val	7.93	120.06	175.40	62.10	33.01
43	763	Val	8.11	124.21	175.47	61.98	32.93
44	764	Tyr	8.50	126.16	174.23	55.58	38.36
45	765	Pro	-	-	177.35	63.57	31.99
46	766	Gly	8.00	108.62	173.89	45.18	-
47	767	Asp	8.21	120.20	176.41	54.39	41.31
48	768	Asn	8.52	119.37	176.00	53.39	38.84
49	769	Gly	8.50	109.13	174.83	45.76	-
50	770	Gly	8.26	108.65	174.23	45.30	-

Residue number			$TARP_{726\text{-}825}$ chemical shifts				
Construct	Full length	Residue	HN	N	CO	Cα	Cβ
51	771	Ser	8.23	115.50	175.07	58.24	63.95
52	772	Thr	8.32	115.93	174.57	61.99	69.67
53	773	Glu	8.38	122.93	176.49	56.58	30.55
54	774	Gly	8.25	110.16	171.81	44.67	-
55	775	Pro	-	-	177.20	63.14	32.12
56	776	Leu	8.33	122.06	177.57	55.35	42.21
57	777	Gln	8.33	121.36	175.74	55.75	29.46
58	778	Ala	8.34	125.13	177.50	52.73	19.21
59	779	Asn	8.42	117.58	175.26	53.41	38.68
60	780	Gln	8.36	120.75	176.07	56.02	29.64
61	781	Thr	8.29	115.83	174.51	61.94	69.86
62	782	Leu	8.32	124.80	177.71	55.45	42.39
63	783	Gly	8.32	109.33	173.70	45.19	-
64	784	Asp	8.20	120.41	176.10	54.45	41.23
65	785	Ile	8.07	120.88	176.24	61.11	38.70
66	786	Val	8.26	125.41	176.03	62.43	32.60
67	787	Gln	8.48	124.86	175.45	55.77	29.75
68	788	Asp	8.41	122.09	176.26	54.48	41.10
69	789	Met	8.37	120.91	176.45	55.71	32.96
70	790	Glu	8.52	121.85	176.97	56.96	30.18
71	791	Thr	8.24	114.88	174.95	61.99	69.77
72	792	Thr	8.19	116.08	175.21	62.22	69.87
73	793	Gly	8.46	111.27	174.43	45.49	-
74	794	Thr	8.11	113.38	174.77	61.76	70.01
75	795	Ser	8.43	118.24	174.61	58.45	63.82
76	796	Gln	8.49	122.31	175.96	55.96	29.46
77	797	Glu	8.44	121.87	176.63	56.81	30.37
78	798	Thr	8.24	116.16	174.30	62.29	69.79
79	799	Val	8.19	123.84	175.93	62.27	32.75
80	800	Val	8.29	124.88	175.87	62.18	32.68
81	801	Ser	8.43	121.23	173.02	56.01	63.53
82	802	Pro	-	-	176.53	63.38	31.88
83	803	Trp	8.05	120.73	176.29	57.39	29.24
84	804	Lys	7.95	124.38	176.32	56.21	33.07
85	805	Gly	7.51	108.99	173.78	45.15	-

Residue number			$TARP_{726-825}$ chemical shifts				
Construct	Full length	Residue	HN	N	CO	Cα	Cβ
86	806	Ser	8.19	115.33	174.89	58.33	64.02
87	807	Thr	8.28	115.40	173.88	61.69	69.88
88	808	Ser	7.99	123.35	178.58	60.11	64.83
89	809	Ser	-	-	-	-	-
90	810	Thr	-	-	-	-	-
91	811	Gly	-	-	-	-	-
92	812	Ser	-			-	-
93	813	Ala	-	-	-	-	-
94	814	Gly	-	-	-	-	-
95	815	Gly	-	-	-	-	-
96	816	Ser	-	-	-	-	-
97	817	Gly	-	-	-	-	-
98	818	Ser	-	-	-	-	-
99	819	Val	-	-	-	-	-
100	820	Gln	-	-	-	-	-
101	821	Thr	-	-	-	-	-
102	822	Leu	-	-	-	-	-
103	823	Leu	-	-	-	-	-
104	824	Pro	-	-	-	-	-
105	825	Ser	-	-	-	-	-