# Role of Paxillin and Polyadenylate Binding Protein-1 Complex in mRNA Trafficking During Cell Migration

Thesis Submitted for the Degree of

Doctor of Philosophy

At the University of Leicester

by

Stuart Roy Parnham, MSc (Nottingham Trent)

Department of Biochemistry

University of Leicester

January 2010

**Acknowledgements**

I would like to begin with thanking my supervisors: Dr Jim Norman for always being able to put a positive attitude on a negative result and for being able to make me see things in that way, for this I will always be thankful even if it was for such a brief time. Dr Igor Barsukov for introducing me to NMR spectroscopy and for spending, what seemed like, endless hours of training in the fundamentals of NMR spectroscopy. Your calm attitude and 'Zen-like' demeanour in the face of angry opposition I will never forget, you were a friend and mentor and for this I thank you. Professor Gordon Roberts who finally ended up as my supervisor, thank you for being a great arbitrator and without whom I would not be in a position to submit this thesis. I would also like to thank Dr Mark Pfuhl, who was my research committee and the Biotechnology and Biological Sciences Research Council for funding this studentship.

I would like to offer a special thank you to Dr Fred Muskett whose help and advice on NMR spectroscopy, protein structure assignment and structural calculation helped me to advance when I was stalled. Your advice on the NMR chapters was greatly appreciated. Dr Alison Woods for being there for support and for supplying a majority of the protein constructs used in this research. Dr Mark Carr who was willing to read and critic the NMR chapters, for that I thank you.

My thanks go to all at the Henry Wellcome Laboratories for Structural Biology at the University of Leicester. To Paul, Lorna, Kirsty, Cat, Vaclav, Phil and Ojay, you have all been good friends and have made this research project such a pleasure to do.

Finally and as always my greatest thanks go to my wife Sarah and son Lewis who's kindest, support and understanding have never wavered throughout this project. I could not and would not have done any of this without them; you are the air that I breathe. To my parents, thank you for believing in me, my millstone is now gone.

Finally to all my friends and family that I have not mentioned, thank you.

Stuart Roy Parnham, Leicester, July 2009.

**Stuart Roy Parnham**

**Title:** Role of Paxillin and Polyadenylate Binding Protein-1 Complex in mRNA Trafficking During Cell Migration

## Abstract

Cellular migration is dependent upon the efficient formation of focal adhesions at the point where the cytoplasmic components engage the extracellular matrix via the integrin proteins. The leading edge of the migratory cell is also a site where proteins such as actin are synthesised. In order for these events to take place mRNA transcripts must be transported from the nucleus to the leading edge by a protein or protein:protein complexes that are capable of nucleo-cytoplasmic shuttling. For example paxillin is a scaffold protein that is a central component of focal adhesions and is capable of nucleo-cytoplasmic shuttling. Polyadenylate binding protein-1 was found to be an abundant co-immunoprecipitant of paxillin in lamellipodium formations.

Using overlapping PABP-1 and paxillin constructs it was found that one of paxillins N-terminal LD domains (LD1) interacted with PABP-1 RRM (RNA recognition motif) 1 and 2. The NMR derived structures of PABP-1 RRMs 1 and 2 in their unbound forms are presented within this report. Previous reports have indicated the presence of a paxillin binding sub-domain within the PABP-1 RRM 1 domain. NMR titration data, using a synthetic paxillin LD1 peptide, revealed a binding interaction site within PABP-1 RRM 2 and not within the PABP-1 RRM 1 domain.

Also reported here are the structural details of the PABP-1 RRM 2/paxillin LD1 complex. The binding interaction appears to be in the fast exchange regime with an estimated Kd of ~211μM. Experimental evidence shows the binding to be electrostatically driven and confined to the four stranded antiparallel β-pleated sheet. The interaction site is shared with the polyadenylated tail of nascent mRNA transcripts. NMR titration data indicates a competition for this site to be biased toward mRNA. This linked with other experimental data, presented here, and would suggest a more complicated picture of binding to include multiple sites of contact.

**Contents**

## List of Figures

**Figure** **Page Number**

**Figure** **Page Number**

## List of Tables



## List of Equations

## Abbreviations

| | |
|---|---|
| **ARIA** | Ambiguous Restraints for Iterative Assignment |
| **CANDID** | Combined automated NOE assignment and structure |
| **CCPN** | Collaborative computer project for NMR |
| **FPLC** | Fast protein liquid chromatography |
| **GST** | Glutathione-S-transferase |
| **HSQC** | Heteronuclear single quantum coherence |
| **JNK** | Jun N-terminal Kinase |
| **LB** | Luria Bertani |
| **MAPK** | Mitogen-activated protein kinase |
| **NH** | Backbone amide group |
| **NOE** | Nuclear Overhauser effect |
| **NOESY** | Nuclear Overhauser effect spectroscopy |
| **PABP-1** | Polyadenylate binding protein 1 |
| **Poly(A)** | Polyadenylate |
| **r.m.s.d** | Root mean squared deviation |
| **TALOS** | Torsion angle likelihood obtained from shift and sequence similarity |
| **TOCSY** | Total correlation spectroscopy |
| **WATERGATE** | Water suppression by gradient-tailored excitation |
| **Φ (Phi)** | Amide nitrogen alpha carbon dihedral angle |
| **Ψ (Psi)** | Alpha carbon carboxyl carbon dihedral angle |

## 1.0 Introduction

### 1.1 Cell Migration

Cell migration plays a critical role in numerous biological processes which initiate from the very beginning of life. During embryogenesis, cellular migrations are crucial to the development of the organism; ranging from gastrulation to the development of the nervous system. In an adult organism, cell migration plays a dominant role in normal physiology as well as pathology. An inflammatory response, for example, invokes the migration of leukocytes to the locale of the insult; it is here that they mediate phagocytic and immune responses (8). Migration of fibroblasts and vascular endothelial cells is critical for wound healing. Cellular migration plays a crucial role in cancer cell metastasis where local tissue is invaded by malignant cancer cells breaking away from the primary tumour site. This movement is of great significance in cancer mortality. *In vivo* the cells break away from the main malignant mass and move through the extracellular matrix (ECM) by deforming or degrading the ECM and surrounding tissues and travel via the lymphatic or cardio vascular systems to set up secondary tumour sites (9,10). Tumour cells have been classified into two classes; those that adopt an elongated conformation and those that are rounded, both of which are inter-convertible. However, it has been observed that particularly motile cancer cells are able to adopt a mixture of conformations based upon their intrinsic environments by extending protrusions into the ECM (11). Elongated movement (mesenchymal), such as that seen in melanoma cells, is driven by activation of the GTPase Rac and its association with the Rac guanine exchange factor (GEF) DOCK 3. Rounded cellular movements (amoeboid) make use of Rho-kinase signalling which inactivates Rac thereby suppressing mesenchymal movement (12).

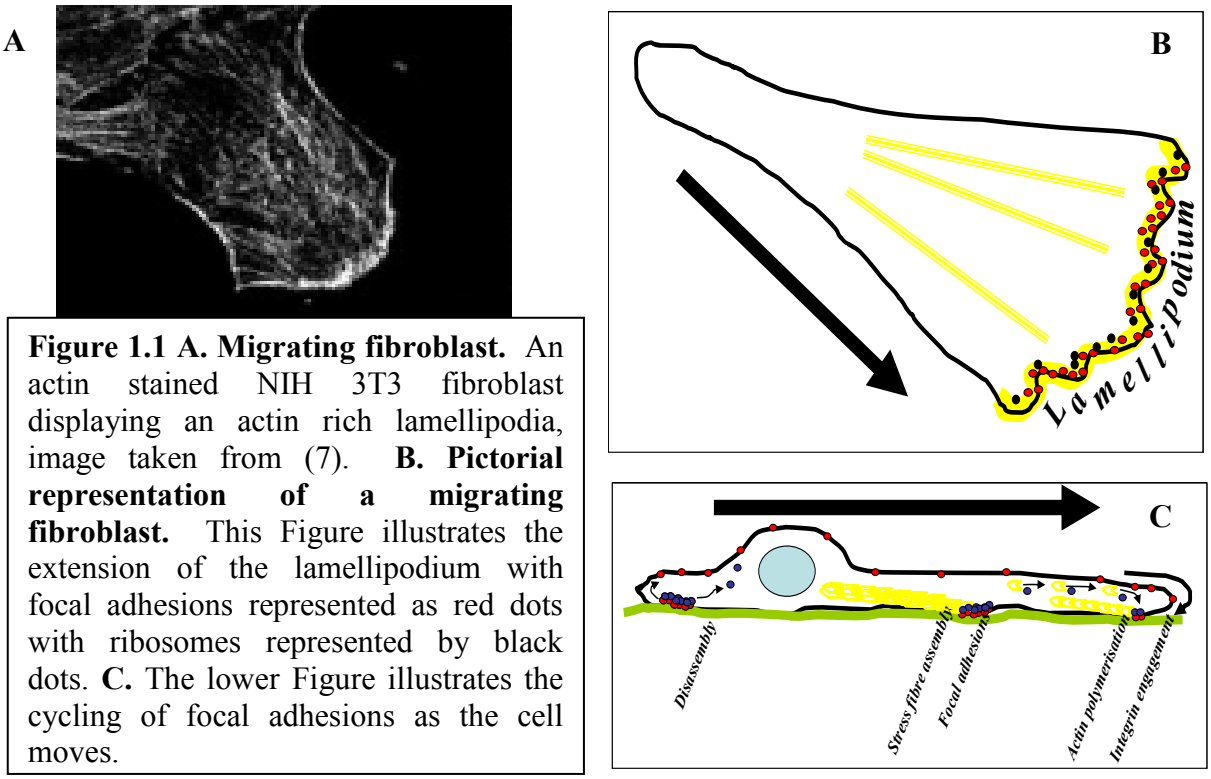



**Figure 1.1 A. Migrating fibroblast.** An actin stained NIH 3T3 fibroblast displaying an actin rich lamellipodia, image taken from (7). **B. Pictorial representation of a migrating fibroblast.** This Figure illustrates the extension of the lamellipodium with focal adhesions represented as red dots with ribosomes represented by black dots. **C.** The lower Figure illustrates the cycling of focal adhesions as the cell moves.

A cell's initial response to a migratory stimulus is to organise key components of its cytoskeleton and extend protrusions with regards to the extra-cellular environment. Protrusions can be large broad 'sheet-like' structures called lamellipodia or 'spike-like' structures called filopodia both of which are driven by actin polymerisation. These structures are stabilised by adhering to the ECM via transmembrane receptors known as integrins. Cells are able to vary their adhesive properties by selectively expressing different integrins thereby modulating their specificity for external ligands (13). The ECM is linked to the actin cytoskeleton via the integrins through structures known as focal adhesions (FAs), (14). As the cells leading edge extends forward it must create new adhesion sites, produce contractile forces and orchestrate the detachment and disassembly of FAs at the rear of the cell. Each of these events must be carefully coordinated to generate optimum forward movement whilst maintaining the structural integrity of the actin cytoskeleton.

Actin filaments are intrinsically polarised by rapid growth at their 'plus' ends and slow growth at their 'minus' ends; it is this polarisation that is the driving force in membrane protrusion (14). Organisation of the filaments is dependent upon the type of cell and protrusion. The architecture of a migrating fibroblasts lamellipodium requires the actin

filaments to form a ‘dendritic’ network with branches forming off the parent filament. Actin polymerisation can be stimulated in many ways such as increasing the rate at which monomers are added, decreasing the rate of depolymerisation, and facilitating new filaments by branching thus creating new plus ends and increasing the need for actin monomers (10). Actin polymerisation is facilitated by the hydrolysis of ATP this creates a difference between the ATP-bound barbed end and the ADP bound pointed end. A balance between polymerisation and depolymerisation is attained thereby maintaining actin filament length however, addition of pure actin appears too slow to account for the speeds of cellular migration that have been documented. Actin binding proteins such as cofilin, profilin and WASP homology 2 (WH2) have been shown to bind directly to the actin filaments facilitating the rapid polymerisation and depolymerisation of actin (15).

Initiation of adhesion formation is linked with actin assembly in the lamellipodia which requires an activation of the Rho, Rac, Cdc42 and Arp2/3 pathways. Adhesions grow in size as they mature under the main body of the cell. One of the main triggers for maturation is an increase in the tensile force within the actin filaments (16). As the cell moves forward the focal adhesions under the main body progressively move rearward as the mature adhesions reduce in size. Disassembly occurs in the perinuclear region and is brought about by the decrease in tensile strength. This decrease is likely due to the inhibition of myosin and RhoA (17). However, detachment can be accomplished by actomyosin stress fiber contraction, which in turn requires Rho activity. It is critical that the destabilisation of adhesions occurs prior to detachment in order to prevent membrane damage from occurring and patches of the membrane being left behind on the ECM. Microtubules act in concert with the many occurring events to release the tensile forces through calpain-driven proteolysis and thereby destabilising adhesion components; this requires focal adhesion kinase activity. It has been theorised that microtubules may lend some assistance in the stimulation of adhesion formation, although indirectly, since microtubule polymerisation activates Rac activity (18). This is achieved through the addition of a microtubule tip binding protein (EB1) which recruits adenomatous polyposis coli (APC) to the tips of the microtubules. APC has been shown to interact with Rac-GEF and thereby concentrating Rac activity to the focal adhesions. APC has been shown to act as an antagonist to Rho which aids focal adhesion turnover (19).

## 1.2 Focal adhesions

Focal adhesion formation is critical not only for cellular migration but also to cell adhesion and extracellular matrix reorganisation (20). The integrins are the key cell surface receptors that mediate the cell's interaction with the extracellular matrix and thereby initiate adhesion. These receptors are heterodimeric transmembrane proteins that facilitate the link between the cell's actin cytoskeleton and the extracellular matrix. Integrins function as mechanotransducers and transform the mechanical forces exerted by the extracellular matrix and the cytoskeleton (3). The extracellular integrin domain binds extracellular ligands, while the short cytoplasmic tail of integrins are able to recruit components and signalling molecules to the focal adhesion site through carefully regulated phosphorylations (21); which facilitate the integrin's indirect connection to the actin cytoskeleton. Following adhesion, to the extracellular matrix, the integrins cluster which in turn promotes the localisation of intracellular signalling molecules (22). The clustering of the integrins allows their β domains to activate non-receptor kinases such as focal adhesion kinase (FAK) which increases local concentrations of tyrosine phosphorylated kinases (23). The integrin adhesome can be broken down further into three basic categories of proteins, which are recruited to cell matrix adhesion sites these are integrin binding proteins, adaptor proteins (that lack intrinsic enzymatic activity) and kinases/phosphatases (24). Talin and FAK are examples of proteins that are able to bind directly to the cytoplasmic tail of integrins. This is of course an over simplification of events that make up the integrin adhesome. This network consists, at least to date, of some 156 components that are linked and modified by ~690 interactions (25). The detailed description of these interactions is beyond the scope of this report and so it would be more efficient to give an overview of the functional make up of the adhesome. The focal adhesion network includes 25 adaptor proteins, 24 cytoskeletal proteins, 9 actin binding proteins, 10 serine/threonine protein kinases, 3 serine/threonine phosphatases, 9 tyrosine phosphatases, 8 tyrosine kinases, 8 GAPs, 8 Gefs, 5 GTPases and 32 other types of components (25).

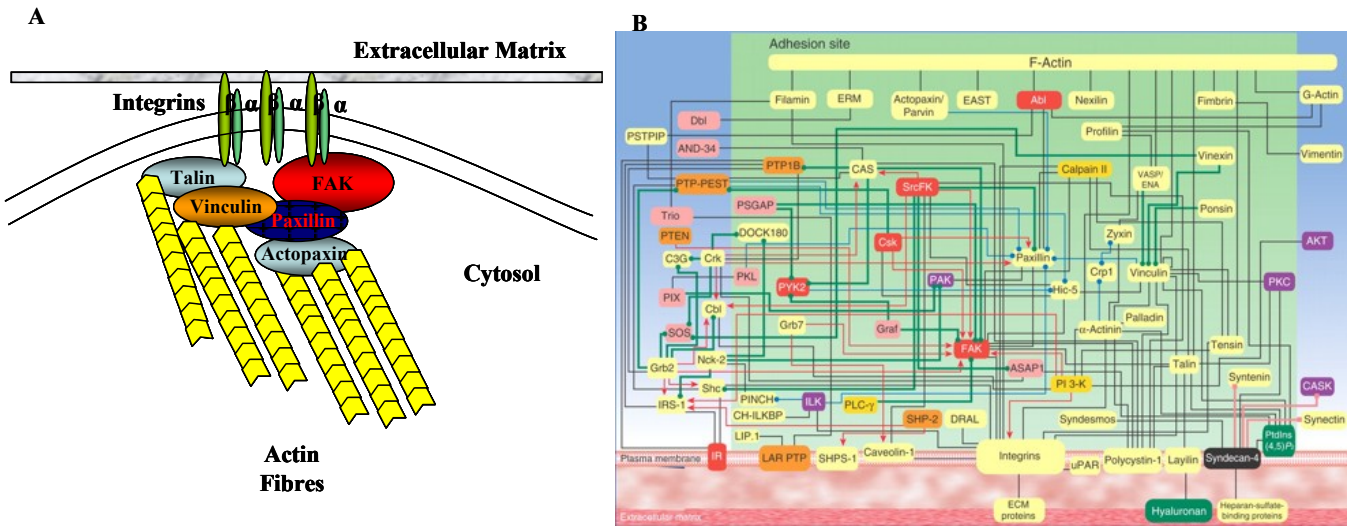



**Figure 1.2 A) Simplified diagram for the major components of focal adhesions.** Shown above are the critical proteins and interactions that comprise a focal adhesion. These proteins and kinases provide a link between the short cytoplasmic tails of integrins and the actin cytoskeleton thereby making an intimate connection between the cell and the extracellular matrix. **B) Components of cell-matrix adhesions.** The actual adhesome comprises of ~156 components with ~690 interactions during cell migration, picture taken from (2).

FAK has been shown to be an important mediator of cell growth, cell proliferation, cell survival and cell migration. In each of these cases it has been seen that FAK is dysfunctional in cancer cells (26). FAK has been found to be significantly elevated in 100% of colonic lesions, 88% of breast tumours and has proven to be an early indicator of breast tumorigenesis (20). It has been seen that the linked activities of Src and FAK control changes in cell to cell and cell to matrix adhesions, cellular migration and invasion. Both Src and FAK have been associated with more aggressively invasive tumour phenotypes (27). FAK interacts with the cytoplasmic tales of β integrins via its N-terminal domain. In the meantime the carboxyl terminal domain contains binding sites for other focal adhesion proteins, such as talin and paxillin, which further aid in the recruitment of FAK to adhesion complexes (28). Talin and paxillin act together with FAK to facilitate a link between membrane bound integrins and the actin cytoskeleton; this is discussed in section 1.5.

Talin is an abundant cytoplasmic protein that helps complete the link between the β-integrins and the actin cytoskeleton as well as performing a pivotal role in the formation of focal

adhesions. Talin possesses binding sites for the β-integrin subunits, actin filaments and vinculin which provides an alternative mode of actin binding. Talin is composed of a 50kDa head domain which contains the primary integrin binding site. The remainder of the protein contains a 220kDa rod domain that contains a low-affinity integrin binding site, and is discussed in detail in section 1.4, as well as binding sites for vinculin and actin (29).

Vinculin is an adaptor protein that plays a key role in the regulation of focal adhesions stabilising focal adhesions and suppressing cell migration. Regulation is released by changes in concentration of inositol phospholipids. Activated vinculin is seen at the proximal edge of gliding or dissolving focal adhesions in migratory cells. In contrast, very little active vinculin was witnessed in the vicinity of stable or protruding adhesions (30). It is comprised of three major domains; an N-terminal head domain, a flexible proline-rich hinge region and a C-terminal tail domain. The intra molecular interactions between the head and tail domains constrain vinculin to an inactive form thereby forcing it to locate exclusively to the cytoplasm. Upon its recruitment to the focal adhesions its structure is switched to its active conformation by its interaction with talin whereby its domains undergo a global conformational change, (section 1.4.2). This process is crucial to its further interaction with talin, α-actinin, ponsin, actin, phosphatidylinositol (4,5)-bisphosphate and paxillin (31-33).

Paxillin is an adaptor protein that is central and critical to focal adhesion complexes providing links to FAK, integrins, talin via vinculin, actin and microtubules (34,35). Paxillin is a multidomain protein capable of nuclear-cytoplasmic shuttling that becomes localised at the leading edge of migratory cells (36). The N-terminal region contains five LD domains that have been found to be conserved amongst species. These leucine rich regions act as binding sites for focal adhesion proteins such as actopaxin, vinculin and FAK with their binding sites often overlapping ensuring recruitment of the proteins to the focal adhesions (37). The C-terminal region contains four double zinc fingers that facilitate paxillin's interaction with actin and microtubules (34). These LIM domains serve as binding sites for several structural and regulatory proteins including tubulin and tyrosine phosphatase (PTP-PEST) each providing important roles for the control of focal adhesion dynamics (38,39).

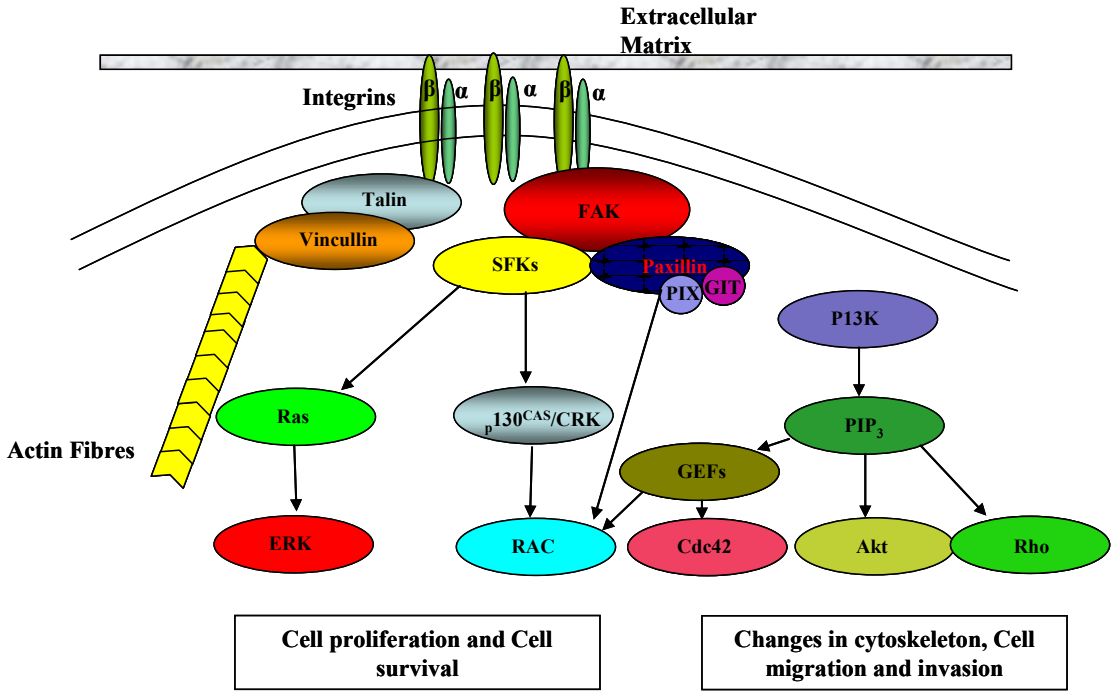



**Figure 1.3 Depiction of out to in signalling through the focal adhesions.** FAK is a key component of the focal adhesion complex and is recruited by integrins during signalling. Through the recruitment of SFKs, paxillin and P13K, FAK is able to affect cell proliferation, cell surviva. This facilitates a link between integrins to the actin cytoskeleton via paxillin and its association with actopaxin and changes within the cytoskeleton gene expression cell migration and invasion via P13K (3).

Some of the more critical 'outside in' signalling pathways are illustrated in Figure 1.3. It can be seen from the diagram that FAK and paxillin are key components that facilitate the signalling cascade through phosphorylations. FAK activates the Src family of kinases (SFKs) and P13K which provides links for pathways leading to the restructuring of the cytoskeleton, gene expression, cell migration and tumour invasion (3). It is through the signalling down stream of P13K that affects the activation of Akt and the small GTPases Rac, Rho and Cdc42, that enables the restructuring of the actin cytoskeleton in migratory cells. FAK is able to bind to the Pix/GIT/paxillin complexes thereby promoting cell proliferation and cell survival. Recent studies have shown that upon FAK depletion in metastatic cells, phosphotyrosine-containing proteins switch from focal adhesions to invadopodia via both spatial and temporal regulation of c-Src activity (26). While a majority of the migratory cell signalling is facilitated through FAK, the focal adhesion complex also provides a link to the actin cytoskeleton via paxillin's interaction with vinculin and actopaxin. Paxillin phosphorylation

contributes to the regulation of the Rho family of GTPases and therefore to the coordination of the cytoskeleton (40). Talin provides a more structural function with regards to cell migration by its ability to bind directly to actin and indirectly via its interaction with vinculin.

### 1.3 Integrins

Each integrin subunit consists of a large extracellular domain, a short transmembrane domain and a short cytoplasmic tail. The extracellular domain of integrins often binds to an extracellular matrix component, whilst the cytoplasmic domain is linked indirectly to actin filaments. Ligand binding to this extracellular domain brings about conformational changes leading to the clustering of integrins (22). This clustering is essential as it initiates intracellular signalling via protein tyrosine phosphorylation and activation of small GTPases and has been shown to induce changes in phospholipid biosynthesis (41). Invasion of prostate cancer cells can be ascribed to the reorganisation and clustering of α2 integrin subunits, resulting in the activation of the associated FAK/Src/paxillin/Rac/JNK pathway (42).

Cells are able to vary their adhesive properties by selectively expressing different integrins and modulating their specificity for external ligands (13). Integrins are heterodimers comprised of non-covalently linked 18α and 8β subunits which form 24 known αβ glycoprotein units, dependent on cell type and cellular function. Two major subfamilies of integrins are used and are defined by $\beta_1$ and $\alpha_V$ subunits. The $\beta_1$ subunit pairs with at least 12 different α subunits in order to adhere to a variety of extracellular matrix components (43). From a structural perspective, the recombinant αA-domain of the α-subunit and has been shown to fold independently and has been extensively studied; the first structure was with a $Mg^{2+}$ ion bound (1). The αA-domain is not unlike a GTPase domain in which a catalytic site at the apex is replaced by a conserved metal-ion-dependent adhesion site (MIDAS) occupied by a $Mg^{2+}$ ion (1). The domain was found to adopt a classical α, β Rossmann fold with five parallel β-strands and one single anti-parallel β-strand at its core. These β-strands are surrounded on all sides by a series of seven α-helices, Figure 1.4 (44). The 'open' form is distinguished from the 'closed' form by the inward movement of the N-terminal α1 helix resulting in a two-turn downward shift of the C-terminal α7 helix (1).

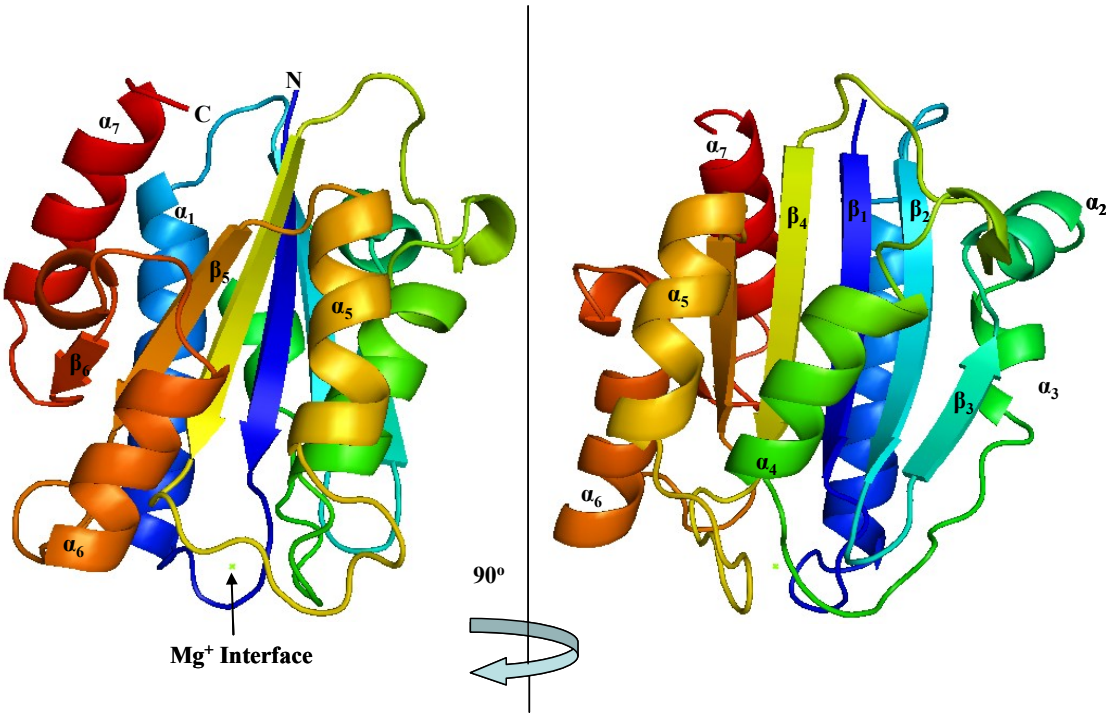



**Figure 1.4 A-domain from the α-subunit of the integrin CRD.** At the core of the structure lies five parallel β-strands ($\beta_1$, $\beta_2$, $\beta_4$, $\beta_5$, $\beta_6$) and one single anti-parallel β-strand ($\beta_3$). A total of seven α-helices adorn the periphery of the β-sheet core. The $Mg^{2+}$ ion was found bound at the apex of the structure (PDB entry 1m1u).

The crystal structure of integrin αVβ3 is without a αA domain. The αV subunit consists of 4 domains; an N-terminal seven-bladed β-propeller, an Ig-like thigh domain and two large β-sandwich domains (CALF1 and 2), Figure 1.5. The β3 subunit has eight domains including an N-terminal cysteine rich Plexin-Semaphorin-Integrin (PSI) domain (1). The propeller and βA domains assemble into a 'head' structure which facilitates the formation of the αβ heterodimer. Stability of the heterodimer maybe due to glycosylation sites within the propeller region of the αV subunit (45). The integrin head structure sits on top of the α and β chains described as 'legs' formed of the thigh and CALF domains of the αV subunit and the PSI domain of the β3 subunit. A crystal structure of the αVβ3 domain in complex with a high affinity pentapeptide cilengitide (RGD), shows that RGD inserts into a crevice between the propeller and the βA domain. The RGD side chains exclusively contact the propeller and βA domain thus drawing these two domains closer together (46). The crystal structure of integrin αIIbβ3 illustrates an arrangement of domains similar to that seen for integrin αVβ3 (47). Negative stain electron microscopy with class averaging found three of the αIIbβ3 ectodomain constructs to be in a bent conformation. This conformation has been described as a low affinity integrin state which acts as a starting point for integrin extension. In the

extended conformation the α and β legs straighten at the knees and extend away from the headpiece (47). During cell migration the open head conformation engages extracellular ligands and the integrins along with their focal adhesion complexes are cycled under the cell toward the rear (16). The normal force exerted on the integrin as the rear of the cell pulls away from the ligand on the extracellular matrix stabilises the closed headpiece conformation with low affinity for extracellular ligands (16,47). The integrin can then return to the low affinity bent conformation completing the integrin adhesion/detachment cycle (47).

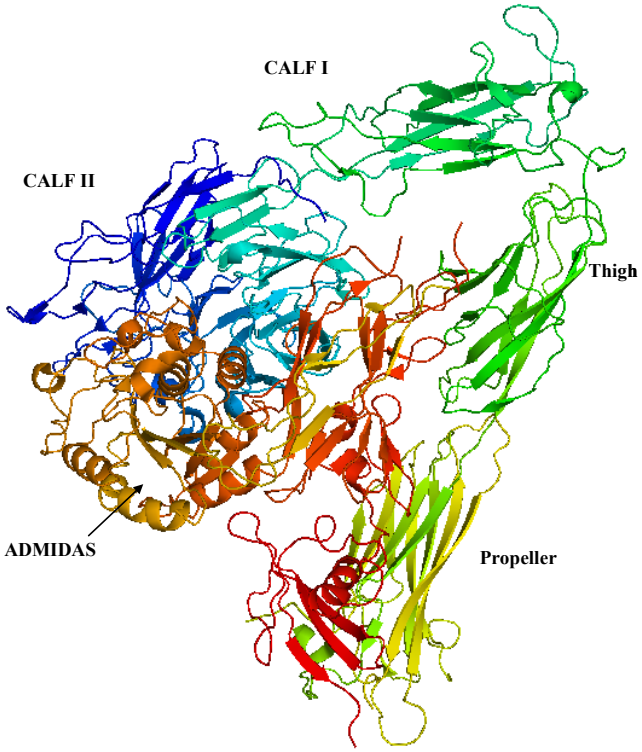



**Figure 1.5 Crystal structure of the αVβ3 integrin ectodomain complex.** Illustrated is the seven bladed β-propeller region in yellow, the two calf domains are presented as β-sheet sandwiches, in green, and the Ig like thigh domain, also in green. The ADMIDAS domain is occupied by a $Ca^{+}$. This stabilises the unliganded state of the integrin (PDB entry 1M1X) (1).

## 1.4 Talin interactions

### 1.4.1 Talin

Talin plays an important role in focal adhesion formation and cellular migration by making vital connections to the short cytoplasmic tails of integrin β subunits. Talin is also able to bind directly to the actin cytoskeleton as well as indirectly via its interactions with other actin binding proteins, such as vinculin, thus creating a stable bridge effect between the

extracellular matrix and intracellular cytoplasmic components (48). This bridge is facilitated through a specific ionic interaction between talin and the membrane-proximal helix of the β-tail of integrins. Upon binding the integrin α/β salt bridge is disrupted thereby releasing the integrin heterodimer from its inactive state (49).

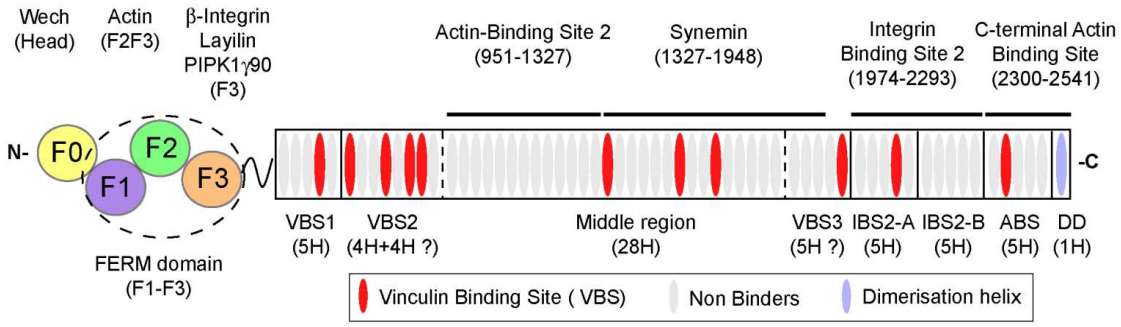



**Figure 1.6 Talin domain structure.** The 50kDa talin head contains a FERM domain comprised of three sub-domains which provides a binding site for β-integrin. The rod domain comprises 62 predicted α-helices with 11 vinculin binding sites depicted in red. A second integrin binding site can be found in the rod domain IBS-2A and IBS2-B. The diagram was taken from (4).

Talin consists of a 50kDa head domain, which contains the principal integrin binding site, and a 220kDa rod domain (50). The head domain contains a FERM domain that is further divided into three sub-domains or lobes (F1-F3). Tryptic digests and the consequential binding studies have shown that the F3 lobe (residues 300-400) is capable of direct interactions with the integrin β-subunit (48). The crystal structure of this F2/F3 fragment illustrates that the F2 sub-domain consists of four α-helices and two short $3_{10}$ helices that are located in the loop connecting α-helices 2 and 3, Figure 1.7A. In contrast, the F3 sub-domain consists of two anti-parallel β sheets the first of four and the second of three. These are packed into a β-sandwich which encloses the hydrophobic core with a single α-helix located at the C-terminus (51). The C-terminus of talin is predicted to contain 62 helices that are organised into bundles to form a rod like structure. The N-terminal region of the rod contains the first of 11 vinculin binding sites (VBS), (29). The C-terminal portion of the rod domain is comprised of a series of 5 helix bundles followed by a single C-terminal helix that forms an anti-parallel homodimer (52). This domain has recently been studied by NMR and crystallographic techniques and is referred to as a THATCH domain (talin/H1P1R/Sla2p actin tethering C-terminal homology) (29). Small angle X-ray scattering, combined with the crystallographic structure of the region and NMR data have shown this dimer to bind three actin monomers of the same actin filament but surprisingly it does not facilitate F-actin cross-linking, (29,52).

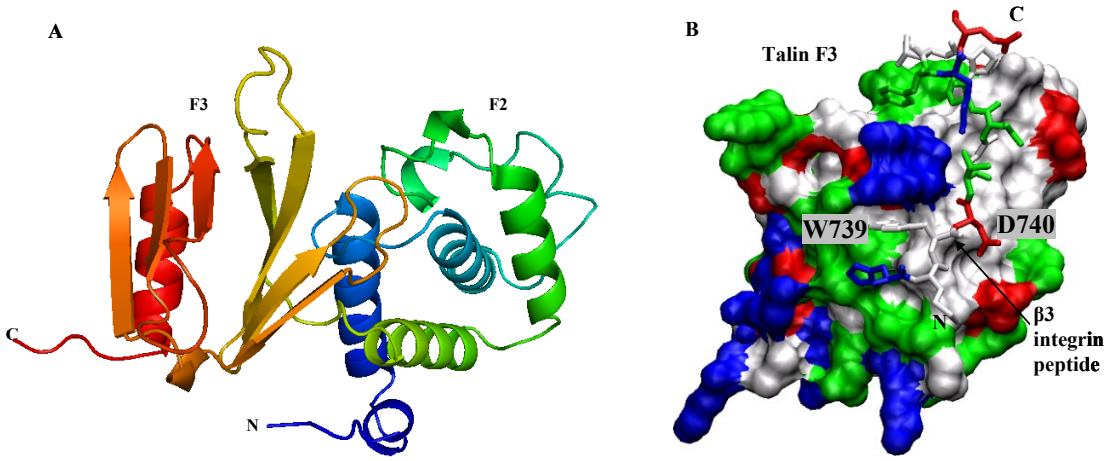



**Figure 1.7 Crystal structures of the Talin FERM domain (F2/F3 lobes) in their free and bound forms.** **A)** Crystal structure of the Talin FERM domain residues 195-400 (PDB data file 1MIX) illustrating the F2 and F3 lobes. **B)** Crystal structure of the F2/F3 lobes of the talin head in complex with the β integrin tail, integrin residues 735-743, these residues are linked to the N-terminal region of the talin FERM domain to form a chimera (PDB data file 1MIZ). The crystal structure shows the β integrin tail in close proximity to the loop connecting H3 and H4. It is worth noting that NMR solution data has illustrated a more intimate complex where the talin F3 lobe demonstrates chemical shift perturbations centred on the β5 strand.

Initial binding studies employed NMR and a full length $\beta_3$ integrin tail. The largest spectral changes were seen in residues 732-750 of the β3 integrin. Binding occurs through a hydrophobic dominant area and is centred on the FERM domain β5 strand but also affects residues of the β6 strand, the C-terminal portion of helix 5 and the β4-β5 loop (53). Interestingly, a second integrin binding site was found in the C-terminal rod domain of talin (54). A crystal structure of this site, residues 1974-2293, shows that it is comprised of a tandem pair of five helix bundles. The upper bundle is named IBS2-A and the lower bundle IBS2-B (Figure 1.8), (3). Pull down assays confirmed that this region interacts with $\beta_3$ integrin tails and that it is critical that the entire module of this five helix bundle be intact for binding to occur (4). It has been suggested that the head domain converts the integrins into the high affinity state, while the rod domain contributes to the integrin clustering via this five helix bundle domain (54).

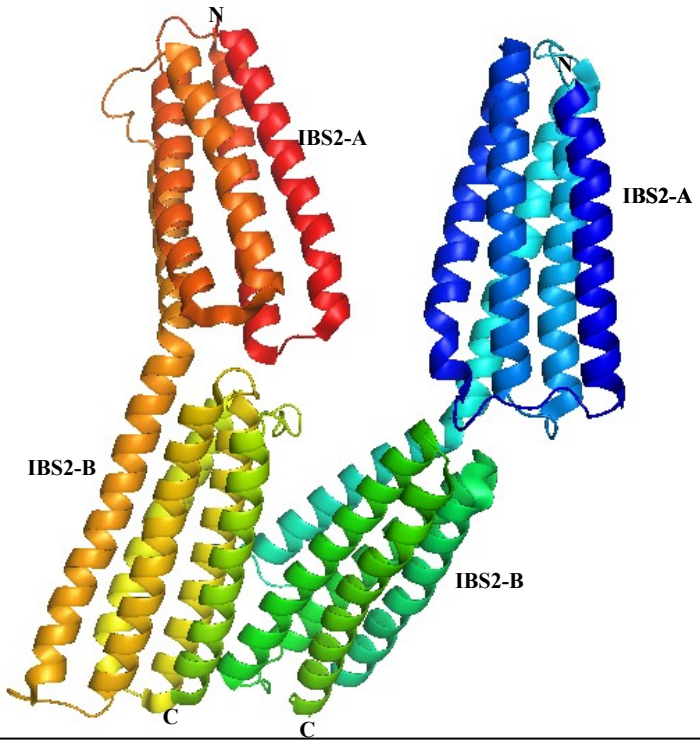



**Figure 1.8 Crystal structure of the talin rod IBS2 domain.** This Figure illustrates two rod domains in close proximity. IBS2-A domain is comprised of a 4 helix bundle (H47-50) residues 1974-2139, IBS2-B is composed of another 4 helix bundle (H53-56). Both parts of the domain are linked by an extended helix running down the spine of the domain. This extended helix was described as two helices making the two domains into five helix bundles (4), PDB entry 3dyj.

### 1.4.2 Talin-vinculin complex

The identification of talin's second integrin binding site illustrates its importance in order for functional focal adhesions to form around clustered integrins (54). Further to this is the identification of 11 vinculin binding sites (VBS), of which three functional sites have been investigated, located within the talin rod domain. Each VBS is capable of binding to the same location within the vinculin head domain (Figure 1.9A, B and C) (33,55,56). The activity of VBS's is differentially regulated by the force exerted by actomyosin contraction.

Vinculin is a 116kDa actin binding protein that consists of a globular head linked to a tail domain by a proline-rich region (30). A majority of the ligand-binding sites are masked by an intra-molecular interaction between the head and tail domains. A popular theory is that the vinculin molecule attains equilibrium between its active and inactive states. Structural studies on vinculin illustrate how the head/tail interaction inhibits the ligand-binding sites by steric

and allosteric mechanisms (57). On recruitment to cell to cell and cell to matrix adhesions, vinculin becomes activated and mediates numerous protein to protein interactions. These interactions regulate the links between F-actin and the adhesion/integrin protein families (58). The structure reveals a five domain auto-inhibited conformation in which the tail is grabbed in a pincer like manner by domains D1-D3. The vinculin tail makes two contacts with the head domain and one contact with the neck (58). Domain D1 packs against the tail to form the major interface with a binding affinity of 50-80nM. The tail domain buries around 1300Å of the head domain D1 (29,58) by binding to helices 1 and 4 of the D1 domain. The second interface is on the opposite side to the D1 interface where the D3 domain is brought into close proximity with the vinculin tail (58). The third contact occurs between the bottom of the vinculin tail bundle and the top of the D4 domain, this interface includes main chain hydrogen bonding and a well defined salt bridge between Glu 775 and Arg 978 where the second domain is largely polar (58). Domain D1 of vinculin contains binding sites for talin and α-actinin and is comprised of seven α-helices that are organised into two 4 helix bundles joined by a long common helix (Figure 1.9) similar to that seen in the talin IBS2 domains. The proline rich region between domain D4 and the vinculin tail domain provides an interaction site for VASP, vinexin, ponsin and the Arp2/3 complex (30). In its autoinhibited state the proline rich domain of vinculin is obscured (30). A popular theory is that activation of vinculin occurs when two or more of its binding partners are in close proximity.

The head and tail domains of talin are able to interact with one another which inhibit vinculins interaction. It has been seen that Y377 of the F3 head domain docks into a hydrophobic pocket at one end of a five-helix C-terminal bundle. The flexible loop of the F3 domain interacts with a cluster of acidic residues located in the middle of helix 4. A competition exists between the β-integrin tail and talins rod domain for this F3 binding site (59,60). Talins affinity for vinculin is increased by PIP2 which inhibits talin's head and tail domains from interacting (59). Talin's multi functional rod domain also provides an interaction site for vinculin which spans residues 482-655. This particular fragment is part of a five helix bundle where helices 2 to 5 form a right handed up and down four helix conformation. Helix 2 adopts an elongated conformation allowing it to connect to helix 1. A crystal structure of the vinculin head domain in complex with the talin rod (talin residues 482-636) reveals that the talin contacting helix (residues 605-636) is buried within the complex. The four helices of the vinculin head surround residues 605-636 from talin resulting in formation of a five helix bundle (33,55,56). Physiological forces that are exerted on the talin rod domain by its

interaction with integrin and actin have been shown to cause a stretching effect within this domain. Recent studies have shown this stretching to increase vinculin binding thereby linking its recruitment to applied migrational forces (61).

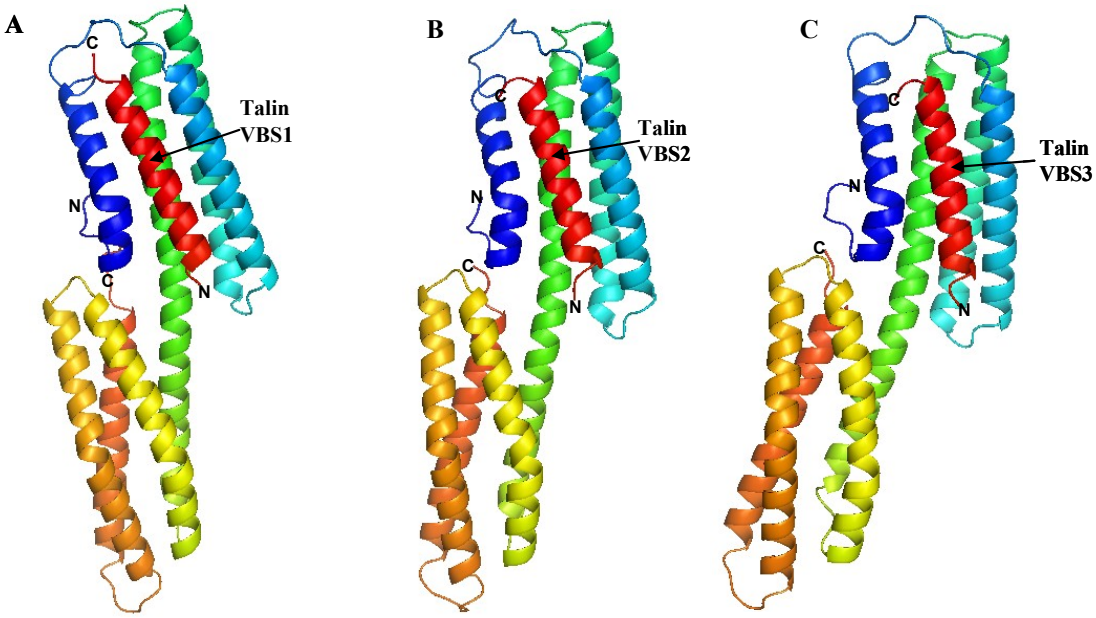
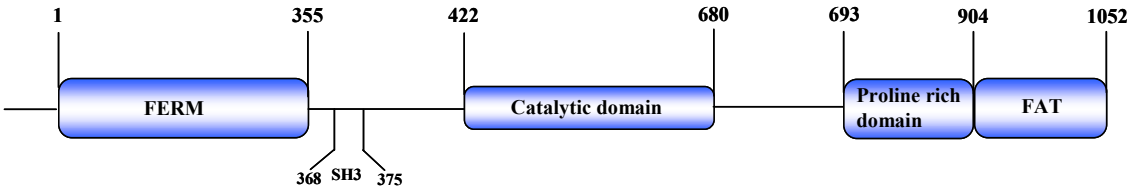



**Figure 1.9 Crystal structure of the vinculin head domain (1-258) in complex with talin rod VBS 1, 2 and 3 peptides.** **A)** Illustrates the crystal structure of the vinculin head domain in complex with the talin VBS1 peptide from the talin rod, residues 605-628, (PDB data file 1TO1). **B)** Illustrates the crystal structure of the vinculin head domain in complex with the second VBS site in the talin rod, residues 849-879, (PDB data file 1U6H). **C)** Illustrates the crystal structure of the vinculin head domain in complex with a third VBS domain from the talin rod, residues 1947-1970, (PDB data file 1XWJ).

## 1.5 Focal Adhesion Kinase




**FAK – Focal Adhesion Kinase**

**FAT – Focal Adhesion Targeting**

**FERM – Four point one-Ezrin-Radixin Moesin homology**

**FRNK – FAK-related non-kinase**

**Figure 1.10 Schematic representation of focal adhesion kinase.** The N-terminal region contains the FERM domain which provides for the integrin interaction of FAK. C-terminal to this separated by a flexible linker region is the kinase region of FAK followed by the proline rich region. The C-terminus of FAK contains the FAT domain which provides a binding interface for the focal adhesion proteins.

### 1.5.1 FAK

Focal adhesion kinase (FAK) is a non-receptor cytoplasmic tyrosine kinase that plays a key role in several different cellular processes through its scaffolding function (28). FAK consists of an N-terminal domain that interacts with membrane proteins such as platelet-derived growth factor receptors, epidermal growth factor receptors and integrins. This domain shows homology with ezrin, radixin and moesin families and is therefore termed the FERM domain, (62). The FERM domain acts in an inhibitory fashion with the centrally-located kinase domain. FERM domains are found in various cell migration signalling proteins and mediate interactions at the leading edge of the migratory cell via their binding partners. The FERM domain of FAK is connected to the kinase domain via a 60-residue linker region. This linker region is particularly important as it contains a phosphorylation site of Y397 which is important for FAK functionality. Initially it was thought that this region was largely disordered. However, crystal structures have shown that the first ten residues of this linker region are bound to the cleft formed between the F1 and F3 lobes of the FERM domain. Larger truncations of FAK revealed this ordered region extends to L375 which encompasses the Src SH3 interaction site (63). The SH3 site, residues 368-375, interacts with the F3 lobe

while the phosphorylation site (Y397) packs against the F1 lobe (64). Following autophosphorylation of Y397, the linker region disengages from the F1 lobe which allows the recruitment of Src family kinases to the SH3 domain. The subsequent phosphorylation of Y576 and Y577 by Src family kinases enables the formation of an active signalling complex (65). The free F3 lobe is then allowed to interact with the cytoplasmic tails of integrin receptors (5).

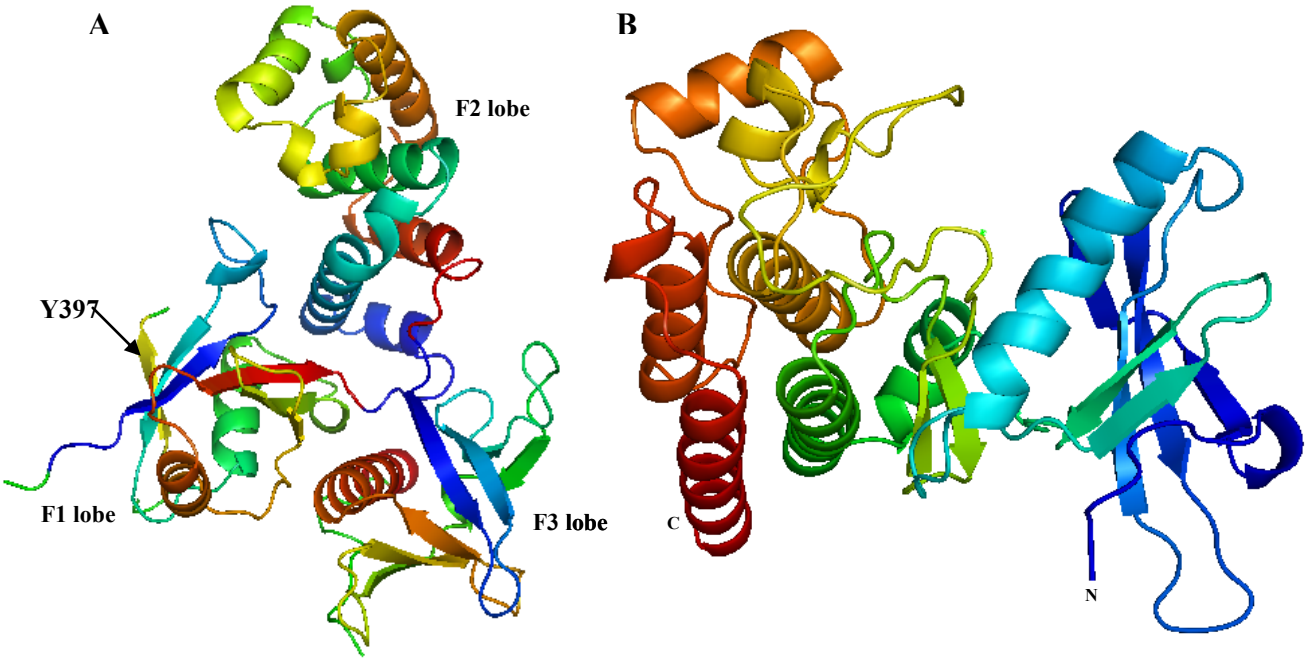



**Figure 1.11 FAK N-terminal FERM domain. A) Inactive form of the FERM domain.** The FERM domain presents three distinctive lobes. The F3 lobe contains a protein interaction groove similar to that seen in the talin FERM domain. The linker region between the FERM domain and the kinase domain masks this protein interaction groove suggesting a regulatory function (PDB file 1mp8). **B) Active form of FAK's catalytic site.** The active phosphorylated FAK kinase adopts a conformation that is immune to FERM inhibition due to Y397 phosphorylation and subsequent recruitment of Src family kinases (PDB file 2J0L) (5).

The C-terminal domain has three proline rich regions and also a focal adhesion targeting (FAT) domain (66). This carboxyl FAT domain contains binding sites for paxillin and talin which cooperate with FAK to link the integrin receptors to the actin cytoskeleton. The FAT structure shows a well ordered domain spanning residues 921-1046 and is organised into a right-handed up-down four helix bundle. The helices are arranged in a straight anti-parallel conformation connected by short turns. This gives the bundle a highly compact and symmetrical appearance with an almost square cross section, Figure 1.12, (67).

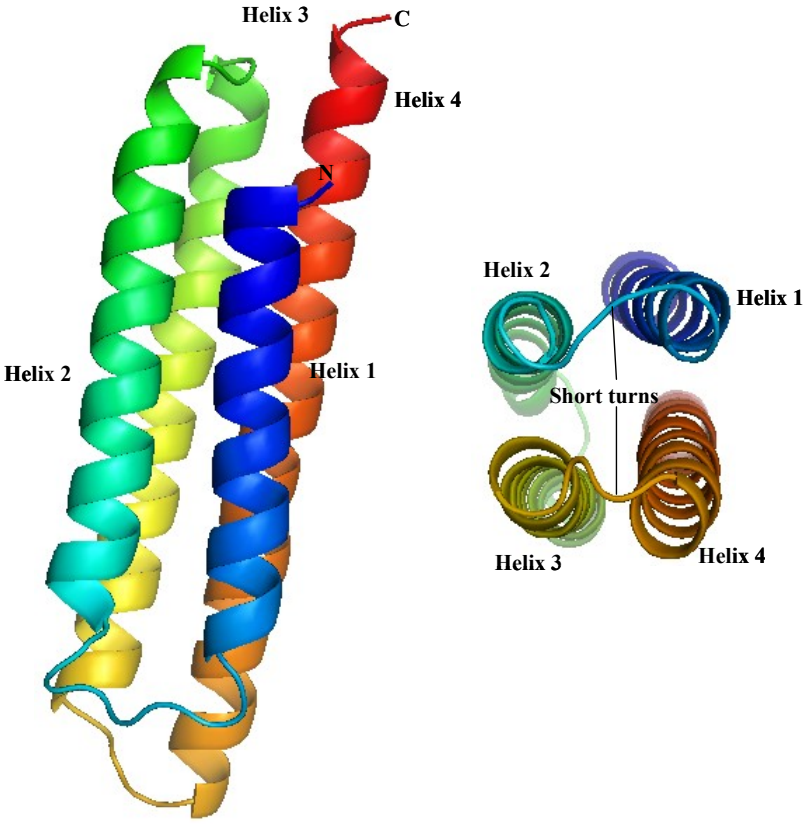



**Figure 1.12 Four helix bundle of the FAT domain.** The well ordered FAT domain of focal adhesion kinase presents a four helix bundle arranged in an anti-parallel conformation. The helices are close together and linked by short turns which help provide a square appearance to the cross-sectional area of the bundle (PDB 1qvx).

### 1.5.2 FAK interactions

Activation of FAK requires the autophosphorylation of residue Y397; this has been shown to be temporally related to the clustering of the integrin receptors. Upon phosphorylation of Y397 the Arp2/3 complex is unable to interact with FAK thus removing FAKs influence on actin polymerisation (68). The clustering of integrin receptors induces a conformational change in the FAK that then alters the interaction of the FERM domain with the centrally located kinase domain of the FAK, Figure 1.11b (69). Tyrosine 397 provides a high affinity binding site for Src, following the phosphorylation of cellular Src and its recruitment cellular Src is then available to phosphorylate Y576/Y577 in the kinase domain of FAK, (23).

The FAT domain contains two hydrophobic patches (HP 1 and 2) each of which is formed at the interface of the two helices (Figure 1.13). It can be seen that several hydrophobic residues have their side chains exposed to the solvent, thereby extending the hydrophobic core of the bundle (67). HP 1 is comprised of the groove between α-helices 2 and 3, while HP 2 is located on the opposite side of the bundle and is formed by the N-terminal region of α-helix 1 and the C-terminal region of α-helix 4 (67). Closer inspection of the amino acid sequence has found that a paxillin binding site (PBS) is located within the HP 2 site. Modelling paxillin LD 2 in its helical conformation into the HP 1 site revealed an excellent fit in which five hydrophobic side chains (valines and leucines) from LD2 fit into the hydrophobic groove forming a five helix bundle (70). In addition to these residues three acidic residues, including D146, were shown to make critical salt bridges to the FAT domain. The paxillin LD 4 domain was found to be of a similar fit and mode of binding thus forming a final conformation of a six helix bundle (70). It has recently been found that the phosphorylation of S273 in paxillin LD4 leads to the destabilisation of the α-helix which leads to decreased binding affinity of FAK (71).

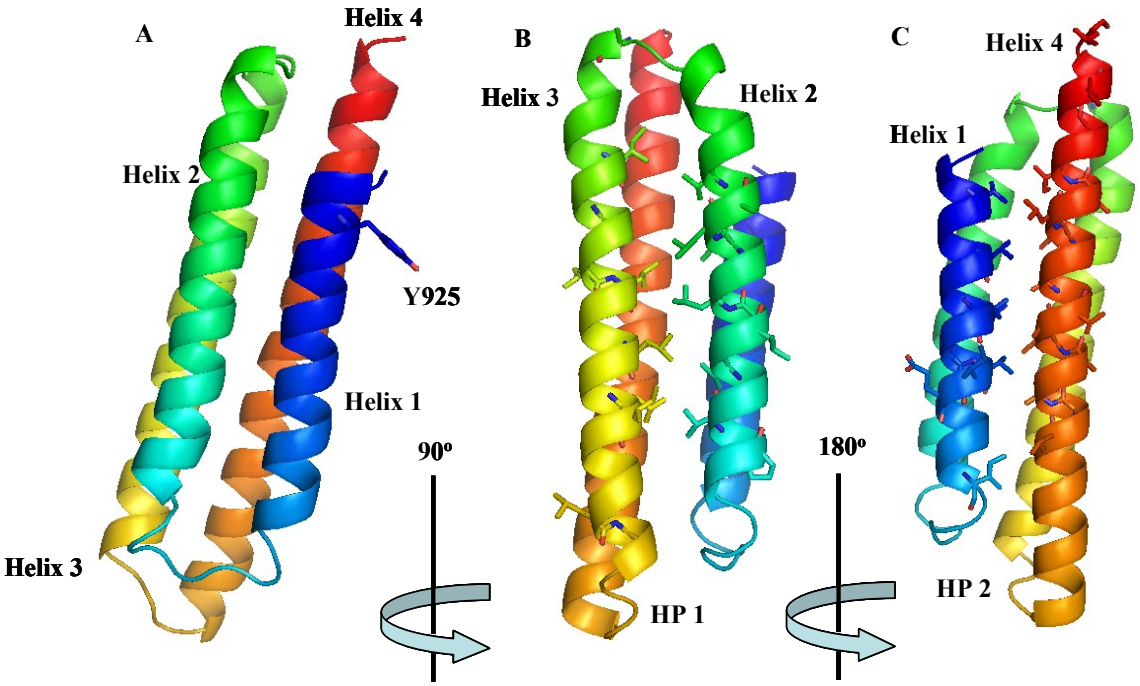



**Figure 1.13 FAK FAT domain interaction site with paxillin LD2 and LD4.** The modification of Y925 shown here **A)** allows for the minimal paxillin binding region of FAK to be exposed. **B)** Paxillin LD2 helix fits into the hydrophobic groove between helices 2 and 3 of the FAT domain of FAK (hydrophobic patch, HP, 1) (PDB file 1ow8). **C)** Paxillin LD 4 helix binds in a similar fashion between helices 1 and 4 of the FAT domain of FAK or the HP 2 site (PDB file 1ow7).

FAK and Src function in a manner that enables them to activate one another. Src activity is increased by the opening of the closed auto-inhibited conformation of Src thus resulting in the increased auto-phosphorylation event (28) and the phosphorylation of FAK Y576/Y577 required for increased FAK activity.

Regulation of FAK occurs through several mechanisms such as a novel inhibitory protein known as FAK inhibitory protein (FIP200) and the FAK related non-kinase (FRNK). The N-terminal domain of FIP200 has been shown to associate with the kinase domain of FAK inhibiting its kinase activity and therefore inhibiting cell migration and proliferation (72). FRNK was found to be an inhibitor of FAK by interacting with the C-terminal proline rich regions and the FAT domain. Due to the low levels of FRNK within migratory cells it was determined that its inhibitory effect is important only at the early stages of cell adhesion to the extracellular matrix (73).

### 1.6 Paxillin

Paxillin is a 559 amino acid protein with a molecular weight of 68kDa and was first described as a phosphotyrosine protein in Src transformed cells, (74). The first paxillin isoform to be identified was termed paxillin-α, two other variants have been identified as paxillin-β and paxillin-γ, (75). Paxillin α is abundant in most tissue types while studies have shown paxillin variants β and γ to be limited to cancerous tissues such as HeLa S3 epithelial carcinoma, K562 chronic myelogenous leukaemia cells, SW480 colorectal adenocarcinoma cells, A549 lung carcinoma cells and G361 melanoma cells (73). Both variants β and γ displayed an insert down-stream of the LD4 domain, between lysine 277 and phenylalanine 278. Paxillin-β has a 34 amino acid residue insert, whereas paxillin-γ has a 48 amino acid residue insert. Both inserts show little homology to one another and therefore are likely to have different modes of action. For example, paxillin-β displays a reduced affinity for vinculin but is able to bind FAK with similar affinity to that of paxillin-α. In contrast, paxillin-γ is unable to bind FAK, but retains its affinity for vinculin (34).

#### 1.6.1 Integrin/Paxillin interactions

Paxillin plays a central role in focal adhesion complexes through its LD protein-protein interactions. Paxillin is able to associate with actin through its interaction with vinculin which is able to bind α-actin and F-actin (76). Actopaxin acts in a similar fashion to that of vinculin and provides a link to F-actin, spectrin and fimbrin (82). Actopaxin is composed of a pair of calponin-homology domains which are responsible for its interaction with the cytoskeletal components (34).

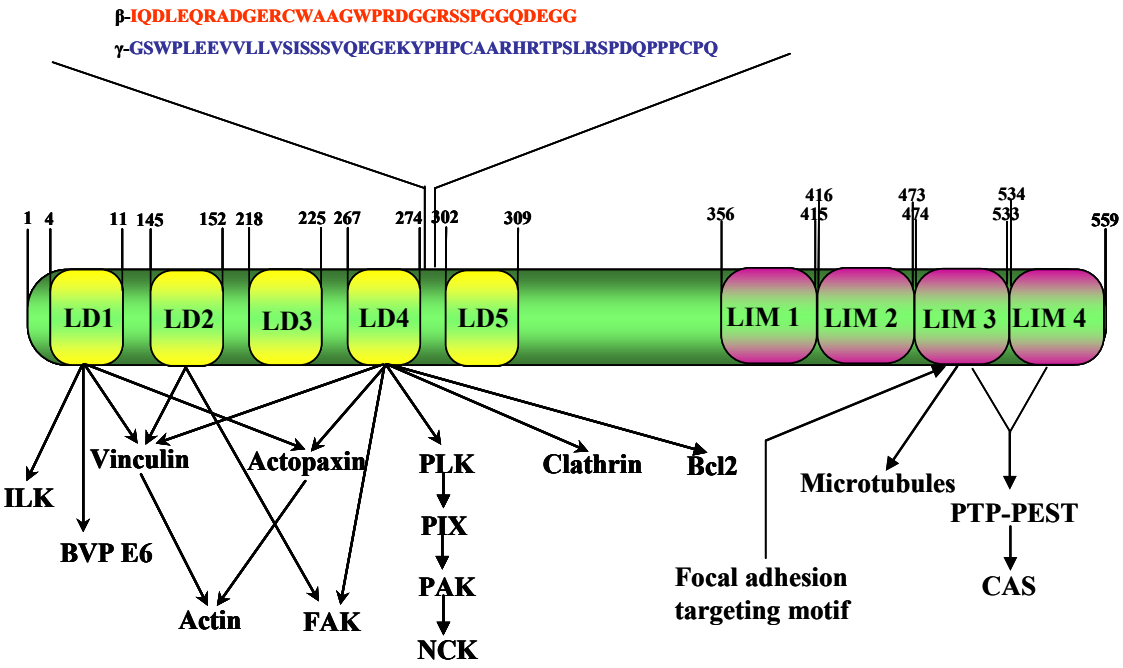



**Figure 1.14 Graphical representation of the 68kDa protein α-Paxillin.** Many of paxillins binding partners are able to bind multiple LD domains as well as form important downstream signalling connections. α-paxillin is the most abundant in migratory cells while β and γ forms are present in cancerous cells.

Initially it was thought that paxillin was able to bind to integrin $\alpha_1\beta_3$ cytodomains, but the data were obtained using synthetic peptides and have not since been substantiated (77). However, paxillin can be recruited to $\alpha_4\beta_1$ integrins by binding to the cytodomain of the $\alpha_4$ subunit (78). The $\alpha_4$-integrin-paxillin complex inhibits the formation of a stable lamellipodium at the leading edge of the migratory cell. When bound to the $\alpha_4$ integrin subunit, paxillin's LD4 domain is available to recruit the ADP-ribosylation factor GTPase-activating protein (Arf-GAP) (78). This leads to a dramatic decrease in Arf activity and thus reduces Rac activity. The $\alpha_4$ integrin tail is phosphorylated at the leading edge of the migratory cell and

dephosphorylated at the sides and rear, (79). This leads to localised $\alpha_4$ integrin-paxillin interactions and promotes directional cell migration through the inhibition of Rac at the sides and rear of the cell, (80). Although pair wise interaction between the α4 integrin tails and paxillin is sufficient for normal Rac1 regulation it is the ternary complex that is essential for focused Cdc42 activity at the lamellipodial leading edge and therefore for directed cell movement (78).

### 1.6.2 Paxillin LD Domains

The most extensively researched and characterised domains within paxillin are the LD motifs, which are so termed as they contain leucine and aspartate residues. These domains were first identified as part of a study aimed at identifying the binding sites for vinculin and FAK (81). Subsequent sequence alignments revealed the binding sites for vinculin and FAK to share a leucine-rich motif that was found to be repeated within the N-terminus of paxillin. The consensus sequence 'LDXLLXXL' was proposed to be involved in multiple protein-protein interactions, (35,81). The N-terminus of paxillin contains four highly conserved LD domains. However a fifth degenerate LD domain (LD3) was originally, omitted due to the lack of a conserved LD start, but has since been included (82). The LD domains of paxillin were originally modelled and suggested to fold as amphipathic α-helices with the leucines providing a hydrophobic interface on one side of the helix (83). This structural model has been confirmed experimentally as part of related studies, such as the FAT domain of FAK in complex with Paxillin LD2 (67).

There are a number of tyrosine and serine/threonine phosphorylation sites that are interspersed between the LD domains as well as potential SH3 binding domains (84). A phosphorylation modification site lies on the first turn of α-helix 1, residue Y118 is present with its side chain solvent exposed. Deletion mutagenesis studies implicated this site as being the minimal paxillin binding region following phosphorylation of this residue. Paxillin's interaction with FAK through its FAT domain is mediated by the LD2 and 4 domains of paxillin. Secondary structure predictions of these two leucine rich motifs have shown them to form α helices. Binding studies for two of the individual paxillin LD domains returned a Kd of ~4μM; larger fragments encompassing both the LD domains were found to bind 5-10 fold more tightly (67).

The LD domains clearly provide specific interaction sites to facilitate the interaction of paxillin with other proteins. LD1 links with actopaxin (85), ILK (86), vinculin (87) and the papillomavirus protein E6 (82). The LD domains also present an overlapping set of protein-protein interaction sites for example, LD2 also provides an interface for vinculin, FAK and PYK2 (87). Similarly, the LD4 domain provides an interface for actopaxin, FAK, PYK2 and Arf-GAPs (34,35). The Arf-GAP protein PKL (protein kinase linker) binds directly to LD4 and provides a link to a protein complex containing the PAK-interacting guanine nucleotide exchange factor (PIX), PAK and the adaptor protein Nck. These components act as mediators to Rho GTPase signalling (34). PAK and PIX both localise with paxillin at the base of the lamellipodia in Cdc42 and Rac mediated focal complexes. This association suggests that paxillin is required for the localisation of PAK and PIX to the paxillin-PKL-PIX-PAK-Nck complex (34,35).

### 1.6.3 Paxillin LIM domains

Many critical focal adhesion proteins influence signalling pathways that control gene expression (88). Focal adhesion proteins are capable of nuclear-cytoplasmic shuttling, and many of the focal adhesion components enhance this behaviour through very distinctive LIM domains (Lin11-Isl1-Mec3); consisting of double zinc finger structures with very specific binding partners (35,89). Zyxin was one of the first focal adhesion proteins found to undergo nuclear-cytoplasmic shuttling (90). Zyxin contains three C-terminal LIM domains and a leucine-rich nuclear export signal. The N-terminus has been shown to associate with the actin cytoskeleton influencing its assembly and organisation during cell migration (91). A binding partner of Zyxin is a cysteine-rich protein (CRP), these proteins are predominately located at focal adhesions as well as along actin filaments and the nucleus (92). CRPs are easily identified by the presence of two LIM domains which are followed by a glycine-rich repeat. CRP family members serve as scaffolds for the assembly of transcriptional complexes within the nucleus (88). Paxillin contains four C-terminal LIM domains that are arranged in tandem (93). Structural studies of the LIM domains of CRP have shown that each of the individual zinc fingers consisted of two anti-parallel β-strands which are separated by a tight β-turn (94). The two zinc fingers of each LIM domain pack tightly together due to hydrophobic interactions. At the C-terminus of the LIM domain is a short α-helix. The C-terminal LIM domain, Figure 1.15, is important in anchoring paxillin to the plasma membrane, and LIM 3 has been shown to be essential in maintaining the integrity of paxillins anchorage to the

leading edge of the migratory cell, with LIM 2 playing a supporting role (81). Furthermore, it has been shown that during cell migration events it is the phosphorylation of serine and threonine residues of LIM domains 2 and 3 that leads to the localisation of paxillin to focal adhesions (34), although the molecular significance of this, and the binding partners of paxillins LIM domains in regard of recruitment to focal adhesions remains unclear(35).

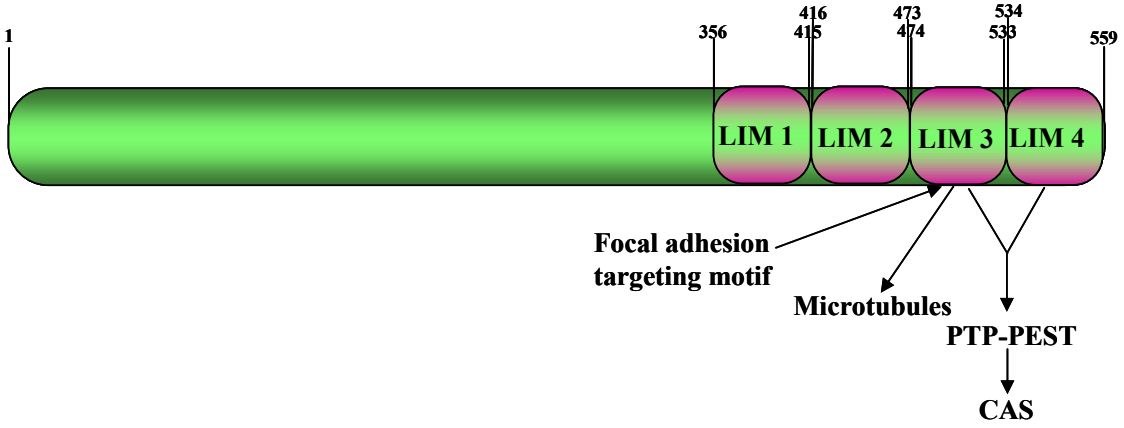



**Figure 1.15 Paxillin double zinc finger LIM domains.** The LIM domains are arranged in tandem with a structural conformation of two anti-parallel β-strands. Specific interactions target paxillin to the leading edge of the migratory cells via the cell actin cytoskeleton and microtubules. The interaction between paxillin and PTP-PEST allows for Cas to be recruited to the focal adhesion complexes which eventually leads to focal adhesion complex disassembly.

Interestingly, LIM domains 3 and 4 are known to provide an interface for PTP-PEST (95) and it is the binding of PTP-PEST to LIM3/4 that is responsible for its transient targeting to the focal adhesions. PTP-PEST activity requires a phosphorylation event at Y31 and Y118 within paxillins N-terminus as well as ARF GAP PKL/GIT2 binding to paxillins LD4 domain. PTP-PEST is then able to regulate adhesion induced Rac signalling (96). This in turn facilitates the dephosphorylation of p130Cas which is important for the disassembly of focal adhesions towards the rear of the migratory cell (97). In addition to its well established influence with the actin filaments and focal adhesions, paxillin has also been shown to be bound to the microtubular cytoskeleton. A study employing a yeast two-hybrid screen showed that paxillin's LIM 2 and 3 domains interacted directly with α-tubulin (98). This association was confirmed by co-precipitation experiments which also indicated that paxillin can interact with γ-tubulin. It is likely that this interaction with γ-tubulin recruits paxillin to the microtubule organising centre (MTOC) in lymphoblast cells; although the potential

significance of this interaction and the role that paxillin plays at the MTOC is not known (99). Microinjection of LIM 2 and LIM 3 domains of paxillin into fibroblasts were found to localise specifically to focal adhesion sites replacing full length paxillin (38). This led to a decrease in microtubule catastrophes at this site. In adhesion free cytoplasm microtubules can grow for 4.9μm, on average, without catastrophes whereas at adhesion sites catastrophes occur after only 0.7μm (38). According to the recent model paxillin serves as a docking site for the catastrophe factor, possibly through paxillins phosphotyrosine-containing domain within its amino-terminal domain. When a microtubule approaches an adhesion site a paxillin associated factor can trigger the microtubule to undergo a catastrophic event and depolymerise (38).

### 1.6.4 Paxillin binding site (PBS) in paxillin binding proteins

The first binding site for paxillin in an interacting protein was identified within the C-terminal region of vinculin (100). The binding interface on FAK was found to show a similar amino acid sequence (101); thus these similar interfaces were termed paxillin binding sites (PBS). The sequence alignment and the subsequent consensus sequence of paxillin binding partners are shown in Figure 1.16. Alignment of these sequences demonstrate that the PBS motifs are poorly conserved at the amino acid level compared to the LD domains of paxillin (36). However, structural predictions of these sites indicated that each of these PBS sequences forms an α-helical structure flanked by regions of random coil (102). It is suggested that it is the heterogeneity between the PBS motifs that forms the basis for their selective binding to particular paxillin LD domains (83). Woods *et al* 2005 identified two PBS like sequences within the N-terminal region of polyadenylate binding protein 1 (PABP-1); these are discussed in more detail within paragraph 1.9.

| | | **Putative PBS sequence** | |
|---|---|---|---|
| **Vinculin** | ***1025*** | **KASDEVTRLAKEVA** | ***1038*** |
| **Actopaxin** | ***273*** | **KLNLEVTELETQFA** | ***290*** |
| **PABP1*(PBS1)*** | ***17*** | **DLHPDVTEAMLYEK** | ***30*** |
| **p95PKL*(PBS2)*** | ***763*** | **RIHVAVTEMAALFP** | ***786*** |
| **PABP1*(PBS2)*** | ***345*** | **EATKAVTEMNGRIV** | ***358*** |

**Figure 1.16 Sequence alignment of paxillin binding proteins.** Three of the main paxillin binding contributors contain areas of amino acid similarity within their paxillin interacting interfaces. These sequences produce the not so well conserved PBS consensus sequence.

### 1.6.5 Paxillin Phosphorylation

A large number of stimuli are known to induce tyrosine phosphorylation of paxillin (103). FAK along with its relatives are major components in the regulatory mechanisms of tyrosine phosphorylation of paxillin, (104,105). This regulation in turn mediates a cycle of adhesion strengthening and weakening thereby delaying cellular retraction (106). A recent study has shown that the phosphorylation of paxillin at positions Y31 and Y118 substantially increases paxillin's affinity for FAK (107). Members of the Src family of kinases are also known to phosphorylate paxillin; indeed paxillin was originally purified from Src transformed fibroblast cells. Src activity is increased in the absence of its regulatory kinase Csk (C-terminal Src kinase), and this leads to increase the tyrosine phosphorylation of paxillin (35,108). Equally important is the regulation of the dephosphorylation of paxillin in signalling pathways (103). PTP-PEST was identified as a 125kDa protein with an N-terminal catalytic domain and a long proline rich C-terminal domain housing the PEST sequences, (109). Both paxillin and Hic-5 have been found to bind to this proline rich region of PTP-PEST (110). PTP-PEST is a cytoplasmic PTP that does not localise at the focal adhesion complexes which only permits its interaction with free paxillin, (103). PTP-PEST is able to regulate cell migration and spreading by decreasing Rac1 activity which may occur through the dephosphorylation of GIT2 thereby allowing the PTP-PEST to destabilise the GIT2/PIX/PAK complex (94).

The function of tyrosine phosphorylation is to create binding sites for SH2 domain containing signalling proteins. Tyrosine residues 31, 118 and 182 in paxillin are found embedded in high affinity binding sites for the adaptor protein Crk. It is the phosphorylation of these sites that result in the recruitment of Crk and Crkl into complex with paxillin (77). Crk and Crkl are then able to associate via their SH3 domains with other signalling molecules such as DOCK180, (103). Paxillin has been shown to experience increased phosphorylation on serine residues when fibroblasts attach to fibronectin or when macrophages adhere to vitronectin (111). The phosphorylation events of paxillin in adhering cells are predominately mediated by PKC. The activation of PKC can also lead to ERK activation inducing the phosphorylation of paxillin, (112,113).

### 1.6.6 Focal adhesion components containing LIM domains

Two other LIM domain-containing focal adhesion proteins that are capable of nucleocytoplasmic shuttling are members of the paxillin family, paxillin and hydrogen peroxide inducible clone-5 (Hic-5). Both paxillin and Hic-5 possess four C-terminal LIM domains and several N-terminal LD domains. Within focal adhesion sites they recruit and are recruited by signalling proteins such as focal adhesion kinase and play an important role in cell migration events, (84).

Hic-5 shows 57% homology to paxillin and contains an N and C-terminal organisation of LD and LIM domains respectively (114). Hic-5 is found at focal adhesion sites where it is mobilised and accumulates on the actin stress fibers during cell migration, (115). Hic-5 facilitates the accumulation of NADP oxidase at the focal complexes via its recruitment of TRAF4 (85). NADP oxidase activity results in $H_2O_2$ production, which oxidises and inactivates PTP-PEST in the cytoplasm. This inactivation of PTP-PEST gives rise to increased Rac activation which in turn enhances cell motility (116). Extracellular signals such as those of oxidative stress are able to induce Hic-5 translocation from the cytoplasm to the nucleus in order to down regulate $H_2O_2$ production, (116).

Abelson non-receptor tyrosine kinase (Abl) is another protein commonly located within focal adhesion complexes and within the nucleus. This multifunctional protein contains multiple nuclear localisation and export motifs and has binding sites for both F-actin and DNA. Abl is recruited to cytoplasmic sites of integrin contact with the extracellular matrix where it facilitates phosphorylation of focal adhesion constituents such as paxillin, (117). This protein has been found to control actin assembly and cell surface ruffling when it is complexed with the protein Abelson-interactor-1, (118). Taken together these various interactions between paxillin and signalling kinases, actin binding proteins, integrin and tubulin support a role for paxillin as a multi-functional adapter. It can mediate the recruitment of substantial signalling moieties whilst trafficking between both actin and microtubule cytoskeletons. A mechanism for the regulation of cell motility that involves the nuclear-cytoplasmic shuttling of paxillin exists (85). Paxillin is able to interact with PABP-1 thus promoting the export of PABP-1

bound mRNA's (7). It is postulated that paxillin acts as a chaperone that directs the targeting of specific mRNA's to nascent focal adhesions (7).

### 1.7 Cell polarity and mRNA localisation

It is known that the focal adhesion structures perform a vital role as part of a complex array of proteins that act to maintain the polarity of the migratory cell. The establishment and maintenance of polarity during directional migration is mediated by both positive and negative feedback mechanisms. These mechanisms include integrins, phosphoinositides, cytoplasmic adaptor proteins and Rho family GTPases (21). One of the ways that a cell actively maintains its polarity is the transport of mRNAs to certain intracellular locations. Injection of localising mRNA into the cytoplasm of a cell revealed that the mRNA need not originate in the nucleus in order for it to localise properly in the cytoplasm (119). For example β-actin and all seven components of the Arp2/3 actin nucleation promoting complex are associated with cellular dynamics and motility at the cell's periphery (85). High concentrations of β-actin mRNA have been found at the leading edge of locomoting fibroblasts (120). In order to facilitate the transport of mRNAs to intracellular locations, protein complexes specifically capable of selectively binding mRNA transcripts are required. Paxillin and other adaptor proteins capable of interacting with cytoskeletal filaments and nucleo-cytoplasmic shuttling are ideal candidates for this purpose. Examination of spreading initiation centres (SICs) found a class of proteins, that specifically act to transport mRNAs as determined by nucleic acid staining, bound to talin and paxillin immunoprecipitants (121). The presence of these nucleo-cytoplasmic shuttling scaffold proteins, such as zyxin and CRP, at the leading edge of migratory cells and within the nucleus indicates a possible mode of action for mRNA delivery to cytoplasmic locations. rRNA probes confirmed the presence of ribosomes within the endoplasmic reticulum but also at the leading edge of the lamellipodia (121).

Asymmetric β-actin localisation is stimulated by extracellular signals that act via the Rho GTPase pathway. β-actin mRNA contains a localisation signal (zip code) that is located within the 3′ untranscribed region (UTR). It has been seen that Zip-code binding protein-1 (ZBP-1) recognises this localisation signal and remains with the nascent mRNA during its translocation from the nucleus to the cytoplasm (122). Within hippocampal cells ZBP-1 movement relies exclusively on microtubules, suggesting that β-actin mRNA can

simultaneously bind both microtubule based motors and myosins. It has been seen that the inhibition of dynein results in the loss of ASH-1 mRNA localisation (119). Interestingly it has been witnessed that mRNA transcripts which lack this localisation signal are also capable of the same movements but with reduced frequency and distance. Cytoskeletal dependent localisation of mRNA requires the mRNA to be anchored to the filaments and it is through depolymerisation of the actin filament, for example, that releases ZBP-1 from its cytoskeletal association (119).

A class of receptors called karyopherins are responsible for the nucleocytoplasmic transport of numerous proteins and certain classes of RNA. The small GTPase Ran controls the assembly of the karyopherin-cargo complex (123). When Ran-GDP binds to an importin in the cytoplasm, it triggers a release of the cargo similarly when Ran-GTP binds an exportin it facilitates the assembly of the exportin cargo complex (124). It has been seen that mRNA export into the cytoplasm can be via the Crm-1 or exportin-5 pathways. Indeed these pathways are also used to export micro-RNA (miRNA) transcripts in conjunction with a scaffold protein. An example of this is exportin-5 that binds a miRNA transcript with its scaffold protein ILF-3 (125). In order for RNA to be transported via this pathway it must not contain a 5′ overhang. The common route of mRNA transport is via the Crm-1 pathway which again requires the RNA transcript to be part of a transport complex, for example miRNA with drosha/DGCR8 (126) and mRNA with PABP-1/paxillin (127). At least one of the cargo proteins must contain a nuclear export signal such as that found in paxillin, which has been seen to export via the Crm-1 pathway (127).

### 1.8 PABP-1 in mRNA translational control

Regulation of gene expression is more complicated in eukaryotic cells than in bacteria and is reliant upon the spatial and temporal separation of transcription and translation. For example, eukaryotic transcription is carried out from within the nucleus. Three types of RNA polymerases are employed; RNA polymerase I is located within the nucleoli where it transcribes the ribosomal RNA of 18S, 5.8S and 28S; ribosomal RNA are synthesised by RNA polymerase III which is found in the nucleoplasm; RNA polymerase II is also located in the nucleoplasm and is responsible for the precursors of mRNA (128,129). Following transcription, the 5’ triphosphate end of nascent RNA is conjugated with GTP, which is then methylated by s-adenosyl-methionine. This is commonly known as the CAP. The cap

provides a contribution to the stability of mRNAs by protecting the 5' end from phosphatases and nucleases. In addition to the 5'-cap most eukaryotic mRNAs contain a polyadenylate tail at their 3' end. Primary transcripts are cleaved by a specific endonuclease that recognises the sequence AUAAA. Following cleavage by the endonuclease a polyadenylate polymerase adds approximately 250 adenine residues to the 3' end of the transcripts using ATP as a donor in this reaction (130,131). The long polyadenylate tail collaborates with the 5'-cap to protect the mRNA from digestion by nucleases and provides stability during its transport to the cytoplasm to engage with the translational machinery.

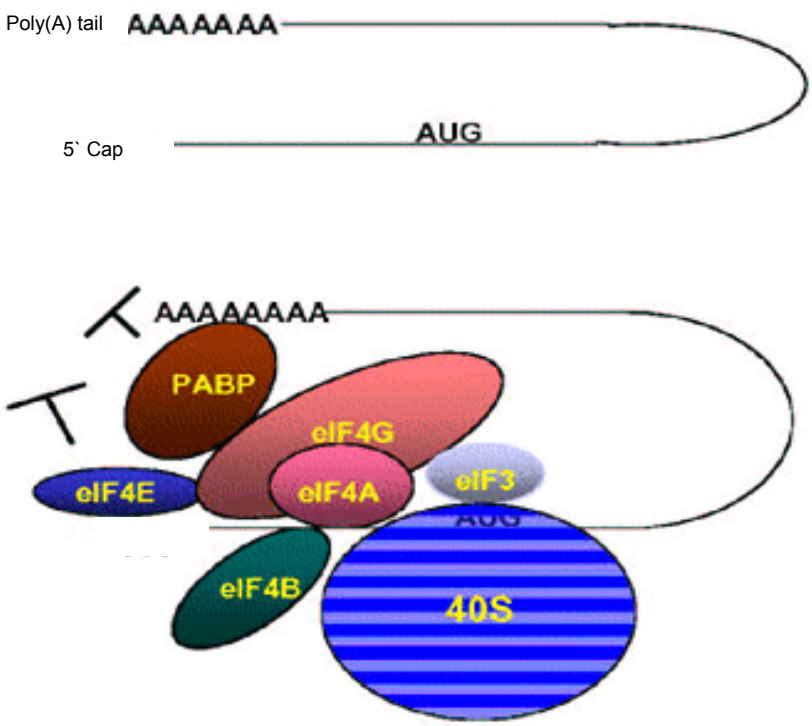



**Figure 1.17 Eukaryotic initiation factor complex which facilitates communication between the 5′-CAP and 3′-polyadenylated tail.** Formation of the initiation complex begins with the recognition of the 5′-CAP by eIF4E. eIF4G acts as a scaffolding protein providing a bridge between eIF4E and the other eIF proteins. As can be expected for such a central protein it is also a site of regulation. PABP binds to eIF4G via its RRM 1 domain and brings the 3′-polyadenylated tail into close proximity with the 5′-CAP, thus facilitating communication and inducing the 'closed loop' formation of initiation.

From the point of mRNA translation and throughout its nuclear export, the poly(A) tail can be found in complex with multiple copies of a 71 kDa polyadenylate binding protein (PABP-1) (132). Each PABP-1 molecule binds to approximately 27 adenylate nucleotides thus forming

a head to tail repetitive pattern of PABP-1 that gives the poly(A) tail protection from RNase cleavage (133).

In addition to this role in promoting mRNA stability during transport PABP-1 plays an important role in translation initiation. Eukaryotic translation requires the assembly of a large multi-protein complex that promotes the circularisation of mRNA and thus allows the 5' CAP and 3' poly(A) tail to communicate (134,135). The assembly of this circular complex is initiated by the recognition of the 5' CAP by the eukaryotic initiation factor eIF4E. eIF4E then provides a binding site for the scaffold protein eIF4G which directs the 43S subunit of the ribosome to the 5' CAP. eIF4G is then recruited by its proximity to eIF4E and this provides a platform for multiple protein-protein interactions including PABP-1, that lends to the circularisation and translational enhancement, Figure 1.17, (136).

### 1.8.1 Domain structure of PABP-1

Published PABP sequences contain four highly conserved RNA recognition motifs (RRMs), Figure 1.18, arranged in tandem (137). These are followed by a variable proline rich C terminus that is not as well conserved. Numerous studies have investigated the modular organisation of PABP and have shown that it is not essential for successful RNA binding to have all of the RRM domains present. For example, it was demonstrated that a polypeptide containing the two N-terminal RRMs (RRM1/2) binds specifically to polyadenylated RNA with an affinity close to that of wild type PABP (138,139). Recent studies have shown that the RRM1/2 tandem domain also supports the interaction with eIF4G and polyadenylate binding protein interacting protein 1 (PAIP-1), albeit at a reduced affinity relative to wild type (140).

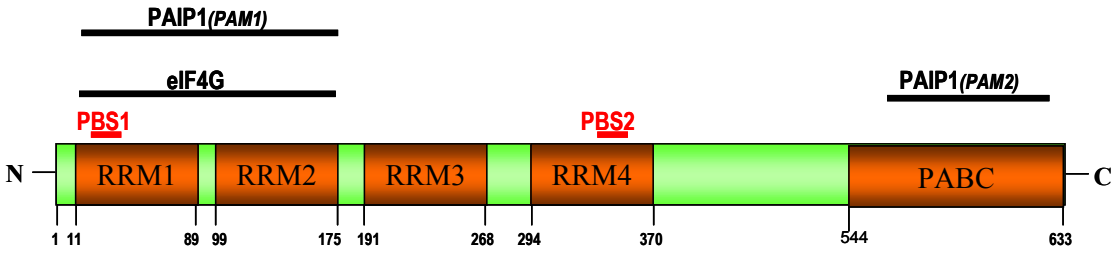



**Figure 1.18 Schematic diagram of PABP-1.** The diagram illustrates the RNA recognition motifs (RRMs) arranged in tandem at the N-terminal region of PABP-1. Indicated within RRMs 1 and 4 are the identified paxillin binding sub-domains that share similarities to those found in actopaxin. Also shown are the areas of binding for eIF4G and PAIP1.

### 1.8.2 PABP-1 RRM1/2 Structural Characteristics

A three dimensional crystal structure of PABP-1 RRM1/2 in complex with an 11-nucleotide poly(A) tail was determined, Figure 1.19 (6). The two structurally similar RRM domains consist of four β-strands and two α-helices that are connected by a nine residue domain linker (90-98). The two domains form a trough that binds poly (A) RNA to give an anti-parallel protein/RNA conformation. In general the floor of the trough is formed by two adjacent anti-parallel β-sheets with the β-strands arranged in space in the order of S2-S3-S1-S4 (RRM 1)-S2′-S3′-S1′-S4′(RRM 2). The sides of the RNA-binding trough are formed by the S2-S3 loop of RRM 1, the inter-domain linker, S2′-S3′ loop of RRM 2 and part of the inter-domain linker between RRM 2 and RRM 3. During the course of their study Deo *et al* found that in the absence of RNA the inter-domain linker between RRM 1 and 2 was unstructured, but was found to adopt an α-helix in the presence of RNA (6).

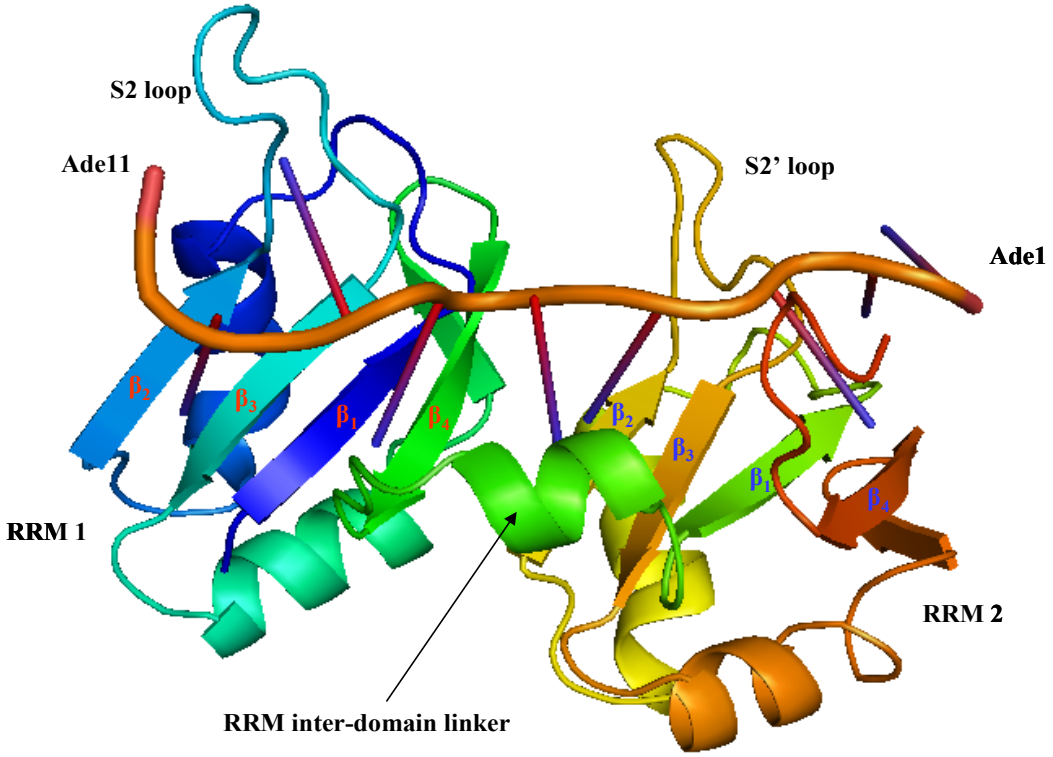



**Figure 1.19 Crystal structure of the PABP-1 RRM1/2 domain in complex with a 11 ribonucleotide poly(A) tail.** The diagram illustrates the RNA binding trough that is formed by the two domains. The RNA (orange) can be seen to be bound to the floor of the trough to form an anti-parallel conformation of protein and RNA. The floor of the trough is formed by the two β-sheets. The walls of the trough are formed by the S2 loop of RRM 1, the RRM inter-domain linker, S2′ loop of RRM 2 and part of the inter-domain linker between RRM 2 and RRM 3 (PDB data file 1CVJ).

PABP-1 RRM1/2 makes significant contacts with the RNA backbone which participates in a number of electrostatic interactions. These interactions include K104-Ade2, Y140-Ade3, Y14-Ade6, Y54-Ade8, Y56-Ade8 and R89-Ade8. The 2´hydroxyl groups interact with the protein suggesting that the interaction is weak (6).

The aliphatic region of K104 makes Van der waals contacts with the 3´face of Ade2 while at the same time N105 accepts a hydrogen bond from the N6 atom of RNA, Figure 1.20A. The plane of F142 binds the 5´ face of Ade4 which in turn stacks with Ade5, Figure 1.20B. Below the 3´ face of Ade5 are the interaction sites for R94 and M85 which generate a fourth stacking layer. It is through a mixture of stacking interactions and the hydrogen bond donation from N6 that adenine is selected and guanine excluded. The interaction of N100 distinguishes purines from pyrimidines and S127 aids in PABP-1 specificity for adenine.

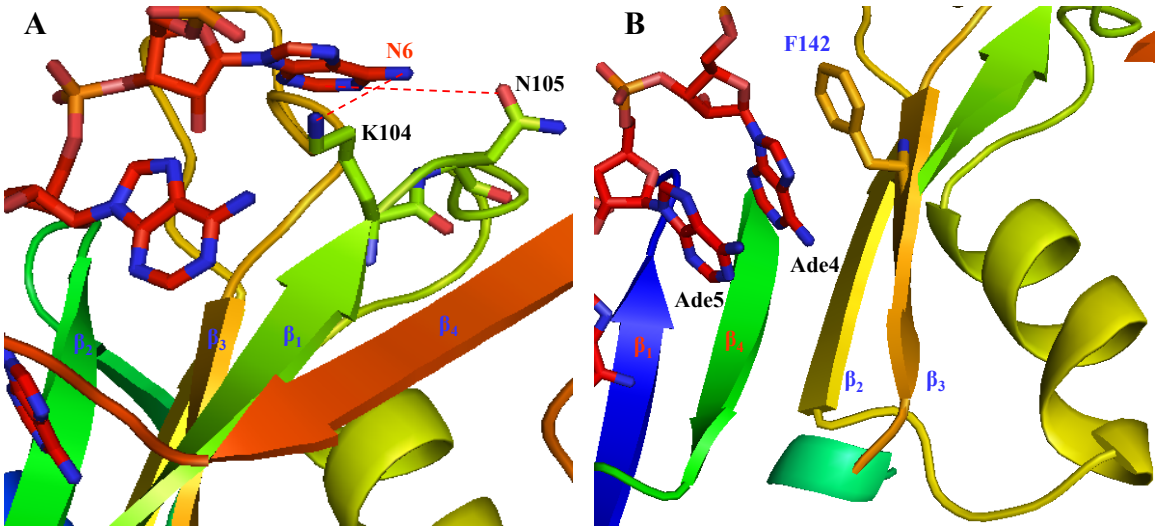



**Figure 1.20 PABP-1 RRM 2 RNA interactions. A.** K104 engages Ade 2 in electrostatic interactions while N105 of PABP-1 RRM 2 is able to form a hydrogen bond with N6 of Ade2. **B.** F142 with its aromatic ring solvent exposed on the $\beta_3$ strand of PABP-1 RRM 2 is able to form stacking interactions with Ade 4 of the poly (A) tail of mRNA, which in turn forms another stacking interaction with Ade 5.

In conjunction with the stacking interactions, the RNA ribose moieties appear to be sandwiched between the aromatic and aliphatic side chains of PABP-1 RRM1/2. Ade3 was found to be sandwiched between F102 and R179 and is specified by the interaction between

K104 and Ade2, Figure 1.21A. Similarly Ade6 is sandwiched between Y14 and R94 and is specified by the interactions of W86 and D88, Figure 1.21B (6).

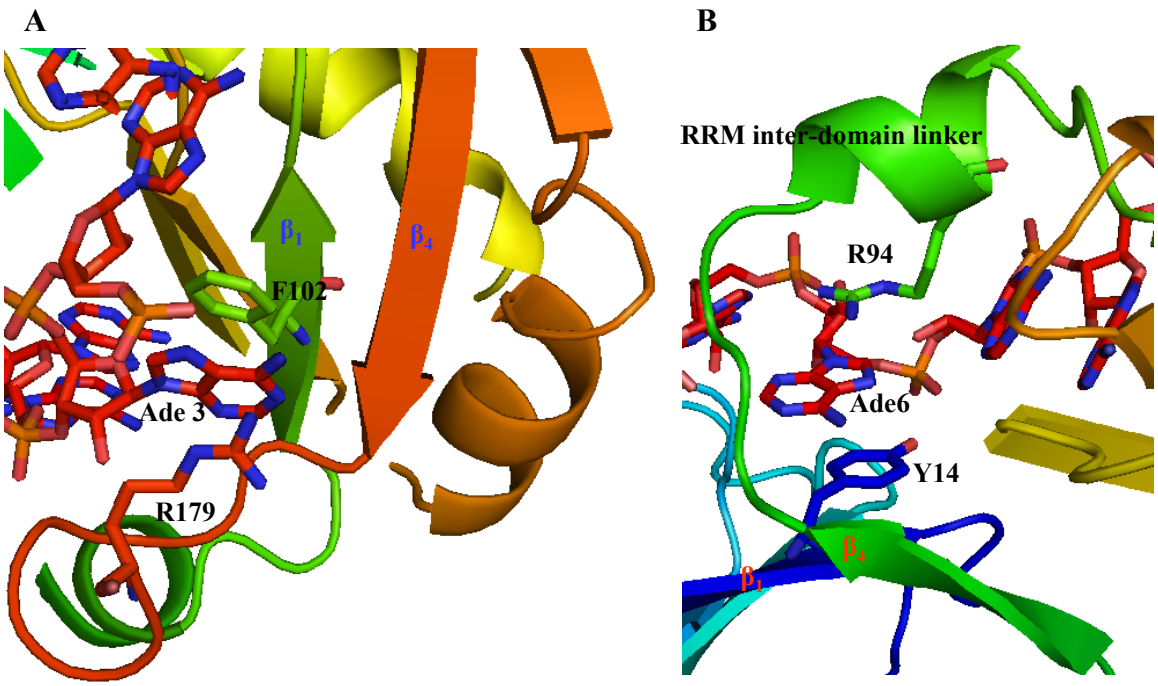



**Figure 1.21 PABP-1 RRM 1 and 2 domain sandwich interactions with the poly(A) tail of mRNA. A.** F102 located on the $\beta_1$ strand of PABP-1 RRM 2 forms a stacking interaction with Ade 3 of the mRNA and sandwiches it between R179 located within the PABP-1 RRM 2 and 3 linker region. **B.** Y14 located on the $\beta_1$ strand of PABP-1 RRM 1 forms a stacking interaction with Ade 6 of the mRNA and sandwiches, in a similar fashion, between R94 located within the small helical region of the inter-domain linker between PABP-1 RRM 1 and 2.

Polypyrimidine tract binding protein (PTB) is an RNA binding protein that functions as a splice repressor and acts in a similar manner to PABP-1, in that the RNA interaction site lies across the β-sheet, with its RRM domains maintaining a similar structural topology. Unlike PABP-1, PTB shows specificity for poly (CU) sequences allowing it to compete with the essential splicing factor U2AF for the 3′ splice sites (141). Interestingly the RRM 1 domain of PTB interacts weakly with RNA, perhaps because the RRM 1 and 2 domains of PTB lack the hydrophobic interactions between the aromatic residues and the RNA bases that are key to the PABP-1/RNA interactions. PTB RRM3/4 domains provide a different functionality to that seen in PABP-1 (currently unknown) in that they bring two distantly located pyrimidine tracts to within ~30Å and thus reduce the amount of RNA looping. It is this looping that prevents the assembly of spliceosomal components (141).

### 1.8.3 The characteristics of the PABP-1 C-terminus

The C-terminal portion of PABP-1 contains a 75 residue conserved region (PABC) which is homologous to a subset of the HECT-family E3 ubiquitin ligases (142). It has been seen that the adenomatous polyposis coli (APC) tumour suppressor specifically interacts with the C-terminal domain of PABP-1. Neither Ran-GTP nor importin-β was detected but subsequent treatment with RNase disrupted the interaction. These results would indicate that the interaction is mediated through the RNA itself providing an insight to PABP-1/RNA transcript specificity (143). In mammals, the PABC domain of PABP-1 enables the recruitment of several different translation factors to the mRNA poly-(A) tail, including PAIP-2 and Rna15p (144). A study employed NMR spectroscopy to investigate the PABC portion of PABP-1 and the residues involved in interactions with PAIP-2. NMR spectra obtained found that the 75 residue C-terminal domain is almost three quarters α-helical, containing a total of 5 α-helices, with a hydrophobic core and compact globular structure (145). The α-helices are arranged in an arrowhead formation (Figure 1.22) with helix 1 forming the tip, and helices 2 and 4 forming the sides. Helix 3 crosses helices 2 and 4, forming a salt bridge between amino acids K35 and E42. Helix 5 is a long C-terminal helix that constitutes the shaft of the arrow (145). Following a titration with Paip-2, it was found that residues K35, V68 and M39, of the PABP C-terminus display the largest shifts upon Paip-2 binding, and that the molecular surface around the residues is disturbed by Paip-2 binding thus revealing a putative peptide binding site.

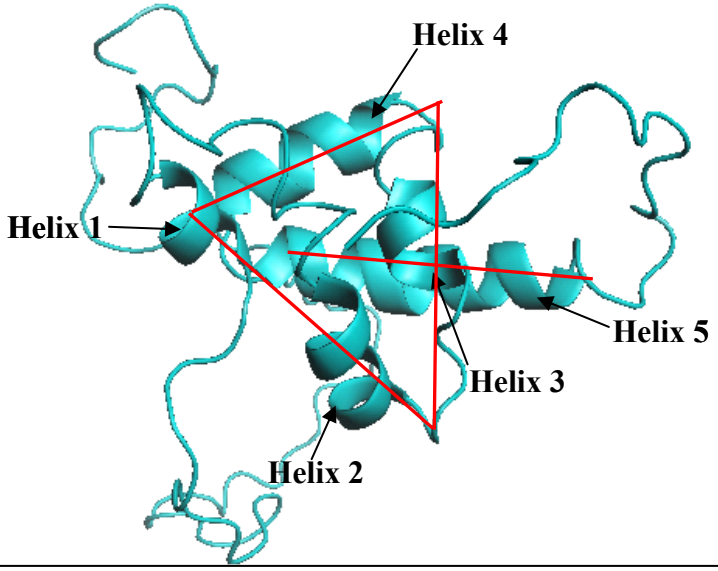



**Figure 1.22 C-terminus region of PABP-1 displaying an arrowhead formation.** Helix 1 forms the tip of the arrowhead with helices 2 and 4 forming the sides. Helix 3 crosses helices 2 and 4 and is held in position by a salt bridge between residues K35 and E42. Helix 5 forms the shaft of the arrowhead formation. At the centre of the 5 helices is a well maintained hydrophobic core giving the overall structure a globular appearance (PDB file 1I2T).

### 1.9 Paxillin/PABP-1 interaction during cell migration events

Biochemical studies focused upon identifying proteins that co-immunoprecipitate with paxillin were carried out (7). Pull-down assays using both paxillin α and β isoforms were used to isolate co-immunoprecipitants in conjunction with yeast two hybrid screens. In order to identify proteins that form stable complexes with paxillin; magnetic beads conjugated with anti-paxillin antibodies were used to separate the stable complexes. These complexes were separated by SDS-PAGE gel electrophoresis and the bands excised. These were then digested using trypsin and the resulting peptides analysed by matrix assisted laser desorption ionisation (MALDI) mass spectrometry. PABP-1 was identified as an abundant co-immunoprecipitant of paxillin. Saturation binding analysis using enzyme linked immunosorbent assays determined this association to have a $K_d$ of ~10nM.

Treatment of migratory cells with leptomycin B resulted in the inhibition of the exportin 1 nuclear export pathway. Using florescence labelled paxillin and PABP-1 it was possible to observe the accumulation of these complexes within the nucleus of the cell thereby identifying the nuclear export pathway used. Further to this cell fractionation studies found that the bulk of paxillin/PABP-1 complexes were present within the dense endoplasmic reticulum. Using similar florescence labelled protein it was found that paxillin β/PABP-1 complexes to be located at the leading edge of lamellipodia in migratory cells. These data suggest that paxillin is involved in the delivery of mRNA to the leading edge of the migratory cell (7).

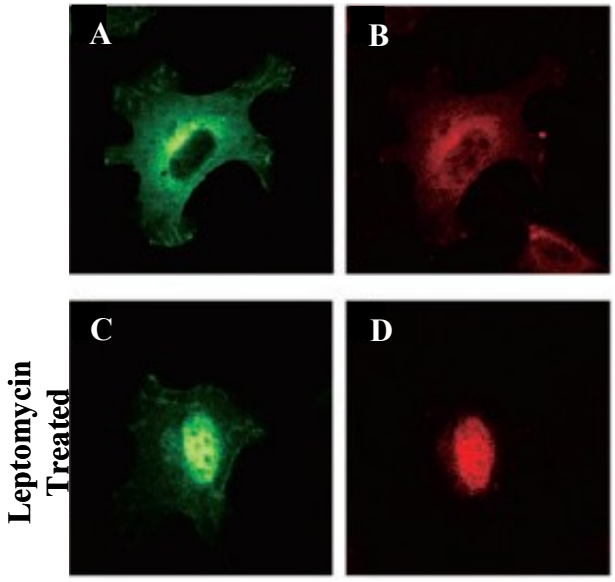



**Figure 1.23 NIH 3T3 Fibroblast cells transfected with EGFP tagged paxillin α. A)** EGFP paxillin α was fixed and stained, it can be seen that paxillin α is predominately located at the leading edge of the migratory cell. **B)** The cells were then stained with anti-PABP-1 followed by a Texas-red conjugated secondary antibody; this clearly illustrates the co-localisation of PABP-1 at the leading edge of the cell. **C and D)** Following treatment with leptomycin the majority of paxillin and PABP-1 are retained in the nucleus of the fibroblast and cellular migration is severely diminished. Image taken from (7).

As mentioned earlier Woods *et al* 2005 identified PBS like sequences within the N-terminal region of PABP-1; the first (PBS1) within the RRM1 domain and the second (PBS2) within the RRM 4 domain. Two mutations were introduced into each of the PBS sites of PABP-1 the first within the RRM1 D21R and T22N; and the second within the RRM4 domain A349R and T351N. Florescence labelled mutants revealed that the intracellular distribution of PABP-1 $PBS1^{RN}$ showed very little difference from the wild type PABP-1. Approximately

40% of cellular PABP-1 PBS2$^{RN}$ and PABP-1 PBS1/2$^{RN}$ was located within the nucleus (127). Wound healing assays indicated that mutation of either PBS1 or PBS2 was insufficient to negate the binding of His-PABP-1. Cell migration was reduced by 40% with the PBS2 mutant whilst the PBS1 mutant, surprisingly, had little to no effect on migrational movements. The question remained, was the altered nucleocytoplasmic distribution of these mutants due to their inability to associate with paxillin? To answer this RNAi was used to suppress the cellular levels of paxillin. This resulted in the accumulation of PABP-1 within the nucleus, which was shown to be capable of shuttling upon addition of paxillin (127).

Initial binding experiments (7) found paxillin LD4 domain to be involved with the paxillin/PABP-1 interaction; paxillin constructs without this domain yielded no significant binding interactions. These binding data led to the theory that the paxillin LD4 was binding to the PABP-1 RRM 4 domain, given the reduction of migrational movements upon mutation of the PBS2 site. However, initial NMR titration data (not presented) revealed no chemical shift changes in the uniformly $^{15}$N labelled PABP-1 RRM 4 sample upon the addition of paxillin LD4. The titration was repeated with GST tagged paxillin LD constructs only to reveal identical results (not shown). As will be discussed in later chapters the paxillin/PABP-1 interaction maybe more complicated than was first thought. Indeed the position of the PBS 1 site within the PABP-1 RRM1 domain is located on α-helix 1 far removed from the poly(A) mRNA interface. Disruption of this may lead to destabilisation of the α-helix but would not disturb the mRNA interface. It was curious that this mutation did not result in the reduction of cellular migration. It may have been the case that the disruption of the α-helix in PABP-1 RRM1 was sufficient to render any possible internal interaction with PABP-1 RRM4 thus preventing total disruption of the paxillin/PABP-1 interaction. The success of the initial PABP-1 RRM1/2 NMR titration trials and the current crystal structure of this double domain in complex with poly (A) mRNA make this a tantalising area of research.

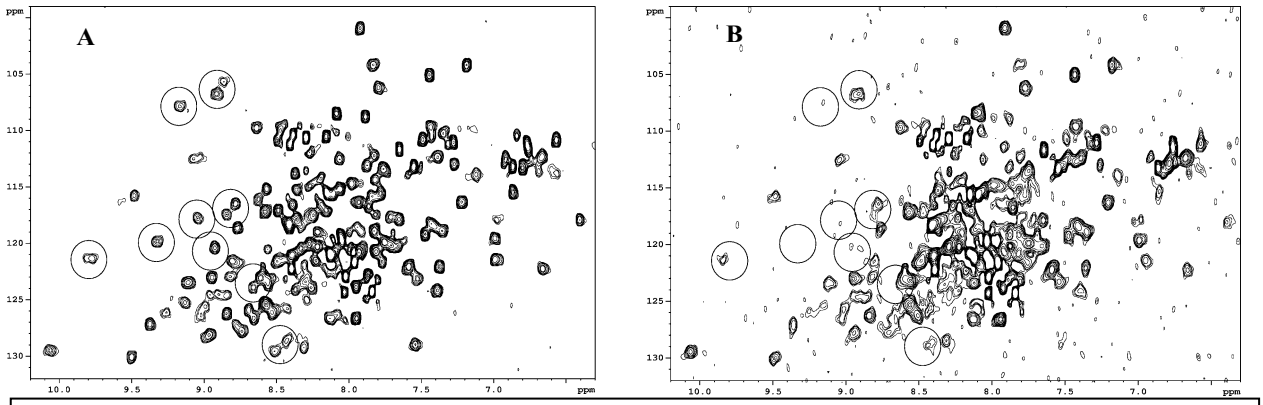



**Figure 1.24 $^{15}$N resolved HSQC spectra of PABP-1 RRM1/2 domains. A)** Illustrates a 6 Histidine tagged PABP-1 RRM1/2 domain in its free form. **B)** Illustrates a 6 Histidine tagged PABP-1 RRM1/2 domain following titration of paxillin LD1/2 domain at a 1:1 ratio. The circled HSQC signals illustrate the disappearance and movement of HSQC signals upon addition of the paxillin ligand. The spectra clearly demonstrates the feasibility of using NMR to further investigate this critical area of protein:protein and protein:RNA interactions.

### 1.10 Aims of the project

- To characterise the molecular interaction between paxillin and PABP-1.
  - Use biochemical assays to identify the regions of paxillin that bind to PABP-1 (Chapter 3).
- To use NMR techniques to elucidate the structure of the PABP-1 RRM 2/paxillin LD1 complex.
  - Identify the possible competition between paxillin LD1 and poly (A) mRNA for the binding interface.
- To use NMR spectroscopy to elucidate the structure of the PABP-1 RRM 1 and characterise the predicted PBS domain within PABP-1 RRM 1.
- To determine any possible inter-domain interactions between PABP-1 RRM 1 and 2 domains.
  - Examine the effects of PABP-1 RRM 1/2 tandem domain on paxillin binding.

## Chapter 2 Materials and methods

### 2.1 DNA Methods

PABP-1 RRM 1 PCR primers were designed in accordance with the published nucleotide sequence accession number NW_002568. An NCO I restriction enzyme site was included into the forward primer of RRM 1; similarly a Kpn I site was designed into the reverse primer. The primers were designed to have similar GC content and melting points.

**Table 2.1 Table of PCR primers for PABP-1 RRM 1**

| Primer | Primer Sequence | Melting temp | GC content |
|---|---|---|---|
| RRM1 Forward | 5′-CATGCCATGGGTATGAACCCCAGTGCCCCC -3′ | 73°C | 63% |
| RRM1 Reverse | 5′-TGGTCTCAGCGTGATTAAGGTACCCTCGAGCGG-3′ | 73°C | 58% |

Lyophilised primers were re-hydrolysed to give a 200µM concentration; each primer was aliquoted to prevent contamination. A working stock concentration of 10µM was required for the PCR reaction.

#### 2.1.1 PCR Reaction Protocol

Roche expand high fidelity PCR kit was used as described in table 3.2 with a primer concentration of 300nM and a dNTP mix concentration of 10µM. Reaction mixture A and mixture B were combined and overlaid with 25µl of mineral oil. The PCR conditions used are as follows:

One cycle of: Initial denaturation 95°C 4 minutes

Followed by 30 cycles of:

| | | |
|---|---|---|
| Denaturation | 94°C | 1 minute |
| Annealing | 55°C | 1 minute |
| Extension | 72°C | 1 minute |

A further cycle of 72°C for 4 minutes was employed to use any left over primers.

### 2.1.2 Legation

The cloning vector used was pET-M11, supplied by Dr Alison Woods. To enable successful ligation into the vector the PCR product must have an over hang of nucleotides i.e. 'sticky ends'. In order to facilitate this PCR product and pET-M11 vector must undergo restriction enzyme digest with NCO I and Kpn I restriction enzymes. Both the PCR product and vector were cut using the reaction protocol set out in table 2.2 and incubated for 1 hour at 37°C shaking.

**Table 2.2 Restriction Enzyme Protocol.**

| Sample | DNA Product | $dH_2O$ | 10X NEB 1 buffer * | 100X BSA | Kpn-I | NCO-I |
|---|---|---|---|---|---|---|
| Neg control | None | 5µl | 2µl | 2µl | 1µl | 1µl |
| PCR product | 10µl | 5µl | 2µl | 2µl | 1µl | 1µl |
| pET-M11 | 10µl | 5µl | 2µl | 2µl | 1µl | 1µl |

* NEB 1 buffer, supplied with Kpn-1 restriction enzymes, by New England Biolabs is universally adaptable to both restriction enzymes.

The DNA products were purified using a Gene Clean kit supplied by Bio101 Inc. Table 2.3 presents the consistently successful protocol for the T4 DNA ligase reaction.

**Table 2.3 Ligation Protocol**

| Sample | Vector | Insert | $dH_2O$ | 2x Ligation buffer | Quick T4 DNA ligase |
|---|---|---|---|---|---|
| Neg cntl | 1µl | None | 8µl | 10µl | 1µl |
| 1 | 1µl | 1µl | 7µl | 10µl | 1µl |
| 2 | 1µl | 3µ | 5µl | 10µl | 1µl |
| 3 | 1µl | 5µl | 3µl | 10µl | 1µl |

The solution was vortex mixed for 5 seconds and incubated at room temperature for 5 minutes followed by a further incubation of 5minutes on ice. The resultant DNA plasmid solution was stored at -20$^{o}$C until DNA sequencing has been carried out to confirm correct ligation of the target sequence.

### 2.1.3 PCR screen

Competent DH5α cells were transformed using the newly ligated pETM-11 vectors containing the DNA insert. The transformation protocol described in below was used to transform the competent DH5α cells and the plates incubated overnight at 37$^{o}$C. The plates were then checked for successful transformation and positive colonies confirmed by PCR screen.

Each colony was removed from the plate and resuspended in nuclease free water. A PCR was carried out using the DNA primers to confirm the presence of the insert. Upon successful identification of the insert 10µl of the resuspended colony was used to inoculate 10ml of enriched LB-broth. These were incubated overnight at 37$^{o}$C in order to make minipreps of the cloned DNA insert and vector for DNA sequencing. DNA minipreps were made using a DNA miniprep kit supplied by Promega and using their protocol.

## 2.2 Transformation protocol

Transformation of competent *E.Coli* cells required the incubation of DNA with the competent bacterial cells. 10ng of recombinant DNA was introduced to BL21 DE3 competent cells and allowed to combine on ice for 20 minutes. The DNA solution was then heat shocked for 30 seconds at 40$^{o}$C and allowed to equilibrate at room temperature for 10 minutes. The solution was added to 1ml of pre-warmed (37$^{o}$C) enriched LB and incubated at 37$^{o}$C for one hour.

The transformed cells were plated out on to kanamycin (40mg/l) enriched agar plates by centrifuging the cell suspension for 10 seconds at low speed to pellet the transformed cells the LB decanted the pellet was resuspended with 150µl of fresh LB. The Kanamycin enriched plates were incubated at 37$^{o}$C overnight, the plates were then checked the following day for growth, positive growth plates were sealed with parafilm and stored at 4$^{o}$C.

## 2.3 Protein expression

In order for the target protein to be labelled with $^{15}N$ and $^{13}C$ it was necessary to express it in M9 minimal media (appendix I). Initially one colony was removed from the newly transformed plate and resuspended in 100µl LB kanamycin (40mg/l) this was then incubated at 37$^{o}$C shaking until an optical density of 0.7 was achieved at an absorbance of 595nm. This was then used to inoculate 1ml of $^{15}N$ enriched M9 media and incubated at 37$^{o}$C until an O.D of 0.7 was reached. Once again this was used to further inoculate 5ml of $^{15}N$ enriched M9 media and incubated at 37$^{o}$C until an O.D of 0.7 was obtained. The 5ml cell suspension was then decanted into the pre-warmed 37$^{o}$C $^{15}N$ enriched M9 media, and incubated at 37$^{o}$C shaking until an optical density of ~0.65-0.8 was reached. The cells were then induced with 1mM IPTG per 500ml suspension, the cell suspension was induced overnight at 37$^{o}$C shaking. Induction temperature trials were carried out in order to obtain optimum protein yields, as summarised in table 2.4.

**Table 2.4 lists protein constructs expressed and their associated cell lines, purification method, yields and induction temperatures.**

| Protein Construct | Competent Cells | Purification Column | Average Protein Yield | Induction Temperature |
|---|---|---|---|---|
| Full length PABP-1 | BL21 DE3 | Nickel | 0.21 mg/ml | 37$^{o}$C |
| PABP-1 RRM1/2 | BL21 DE3 | Nickel | 0.79 mg/ml | 15$^{o}$C |
| PABP-1 RRM3/4 | BL21 DE3 | Nickel | 0.97 mg/ml | 37$^{o}$C |
| PABP-1 RRM 1 | BL21 DE3 | Nickel | 1.4 mg/ml | 15$^{o}$C |
| PABP-1 RRM 2 | BL21 DE3 | Nickel | 2.32 mg/ml | 37$^{o}$C |
| PABP-1 RRM 4 | BL21 DE3 | Nickel | 1.96 mg/ml | 37$^{o}$C |
| PABP-1 RRM1/2 R21, N23 PBS mutant | BL21 DE3 | Nickel | 1.2 mg/ml | 37$^{o}$C |
| | | | | |
| Full length Paxillin | BL21 | GST | 0.75 mg/ml | 30$^{o}$C |
| Paxillin 55-313 | BL21 | GST | 0.16 mg/ml | 15$^{o}$C |
| Paxillin N-terminus | BL21 & DE3 | GST | 0.43 mg/ml | 37$^{o}$C |
| Paxillin LD1/2 | BL21 | Nickel | 0.89 mg/ml | 15$^{o}$C |

| | | | | |
|---|---|---|---|---|
| Paxillin LD1/2 pETM-20 | BL21 DE3 | Nickel | 1.2 mg/ml | 15°C |
| Paxillin LD3/4 | BL21 | Nickel | 0.98 mg/ml | 37°C |
| Paxillin LD 1 | BL21 DE3 | GST | 0.58 mg/ml | 37°C |
| Paxillin LD 2 | BL21 | GST | 0.46 mg/ml | 37°C |
| Paxillin LD 5 | BL21 | GST | 0.36 mg/ml | 37°C |

### 2.3.1 Protein purification

The induced cell suspension was harvested by centrifugation at 5000 rpm for 8 minutes; and the resulting pellet was resuspended in 10ml phosphate buffer A (appendix I) per flask. The cell suspension was treated with 10µl DNAse I, the cells were disrupted by French press and the resultant cytosol centrifuged at 18,000 rpm for 30 minutes. The cytosol was passed down a 5ml nickel charged metal chelate column, the column was then placed into an FPLC system and eluted on an increasing gradient of phosphate buffer B.

## 2.4 Histidine tag removal

The N-terminal six histidine tag facilitated the purification of the protein and was left attached to the protein during the course of the ELISA binding assays. However, for the purpose of determination of the native protein structure it was necessary to remove this N-terminal tag. The pETM-11 plasmid employs a Tev protease cleavage site to facilitate efficient removal of the His-tag.

Prior to the removal of the His-tag it was necessary to change the purified protein buffer from phosphate FPLC elution buffer to Tev protease buffer. A PD10 column was used to facilitate the buffer exchange, and 0.1mM DTT was added to the eluted protein solution. The final protein concentration was determined by measuring the absorbance at 280nm and using a theoretical extinction coefficient as determined Swiss-Prot database tools.

### 2.4.1 Tev Protease time course

To determine the efficiency of the AcTev protease a time course be carried out, Figure 2.1. This was performed on a protein sample diluted to 2mg/ml at a volume of 25µl, AcTev protease (supplied by Invitrogen) concentrations of 1, 2, 4 and 10units and left to incubate

overnight at room temperature. The time course was repeated on undiluted protein solution using 2 and 4units on a volume of 35µl at a concentration of 7.8mg/ml. It was decided that a AcTev protease concentration of 2 units was the optimum able to cleave 0.25mg of target protein overnight at room temperature.

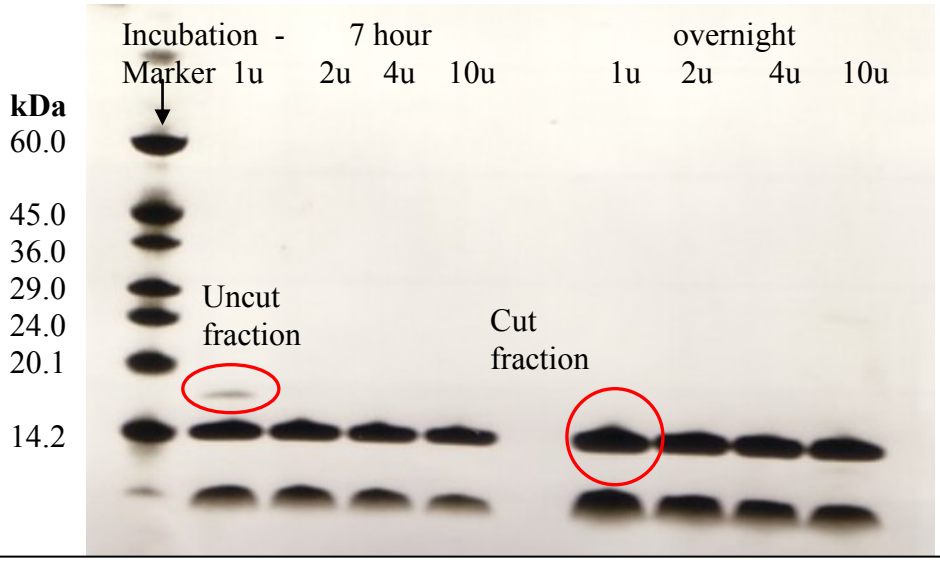



**Figure 2.1 AcTev Protease time course for a protein solution.** The Figure illustrates the a final protease time course performed on 0.25g of protein with varying units of protease over a set period of time.

The cleaved His-tags were removed by passing the AcTev protease/protein solution through a 5ml nickel charged metal chelate column, this was carried out after the buffer was changed to phosphate buffer.

## 2.5 Enzyme Linked ImmunoSorbent Assay (ELISA)

Each well of a Nunc Maxisorp 96 well plate was labelled with 100ng of full length His-tagged PABP-1. The plate was then incubated overnight at 4$^{o}$C.

Full length GST-tagged paxillin was to be used as a positive control and purified GST as a negative control with a GST-tagged paxillin construct as the test subject. A serial dilution of the protein to be analysed was used.

The wells of the plate were initially washed twice with 1xPBS 0.1% Tween pH 7.4. The wells were then loaded with the GST labelled protein ligands, according to table 2.5 below, and incubated at 4$^{o}$C for one hour.

**Table 2.5 ELISA serial dilution ligand loading.**

| | GST negative control | | Full length Paxillin | | Paxillin 55-313 | |
|---|---|---|---|---|---|---|
| Well No. | 1 | 2 | 3 | 4 | 5 | 6 |
| A | 90nM | 90nM | 90nM | 90nM | 90nM | 90nM |
| B | 45nM | 45nM | 45nM | 45nM | 45nM | 45nM |
| C | 22.5nM | 22.5nM | 22.5nM | 22.5nM | 22.5nM | 22.5nM |
| D | 11.25nM | 11.25nM | 11.25nM | 11.25nM | 11.25nM | 11.25nM |
| E | 5.6nM | 5.6nM | 5.6nM | 5.6nM | 5.6nM | 5.6nM |
| F | 2.8nM | 2.8nM | 2.8nM | 2.8nM | 2.8nM | 2.8nM |
| G | 1.4nM | 1.4nM | 1.4nM | 1.4nM | 1.4nM | 1.4nM |
| H | 1xPBS | 1xPBS | 1xPBS | 1xPBS | 1xPBS | 1xPBS |

Anti GST-HRP primary antibody (supplied by Dr Alison Woods) was used in order to avoid the need of an additional secondary antibody step. The ELISA plate was washed three times with 1xPBS 0.1% Tween pH 7.4. Anti GST-HRP antibody was diluted 1:1000 with 1xPBS 0.1% Tween pH 7.4 1% BSA and 50µl loaded into each well. The plate was then incubated at 4$^{o}$C for 45 minutes.

Excess antibody was removed by washing the plate four times with 1XPBS 0.1% Tween pH 7.4. In order to elucidate a colorimetric change the antibody was treated with 15mg ONPG dissolved in 12ml citrate phosphate buffer and 1.25µl hydrogen peroxide. 50µl of this solution was loaded into each well and allowed to develop over 10 minutes at room temperature. Once the colour change had reached a moderate intensity the reaction was stopped by the addition of 50µl 4M $H_2SO_4$.

The absorbance of the reaction was measured using a plate reading spectrometer at a wavelength of 490nm. After the initial ELISA was performed only one change was made to the protocol which was the starting concentration of the paxillin ligands. A starting concentration of 300nM and subsequent serial dilution from that point was used for each ELISA.

## 2.6 NMR Spectroscopy

NMR spectra were typically obtained from 350µl protein samples of ~1.2mM PABP-1 RRMs 1 and 2 in 20mM $Na_2HPO_4$, 100mM NaCl, 2mM DTT pH6.5 but also containing either 90% $H_2O$/10% $D_2O$ or 100% $D_2O$ where appropriate and at a temperature of 30$^oC$. NMR 2D and 3D datasets for the sequential assignment of PABP-1 RRM 2 protein backbone were collected on a Bruker 600 MHz spectrometer, protein side-chain datasets were performed using a Bruker 600 MHz spectrometer with cryoplatform. All 3D NOESY experiments were carried out using a Bruker 800 MHz spectrometer with cryoplatform. Table 2.6 lists the NMR experiments employed for the structural determination of PABP-1 RRM domains in their unbound states, the 3D datasets were collected over a 90 hour period.

**Table 2.6 Table of NMR datasets commonly employed in amino acid protein assignments.**

| | NMR Experiment | References |
|---|---|---|
| Backbone Assignments | $^{15}$N HSQC | (146) |
| | CBCACONH | (147) |
| | HNCACB | (147,148) |
| | HBHACONH | (149) |
| | HBHANH | (149) |
| | HNCO | (150) |
| | HNCACO | (151) |
| Side-chain Assignments | $^{13}$C HSQC | (152) |
| | [H]CCH-TOCSY | (153,154) |
| | H[C]CH-TOCSY | (153,154) |
| | $^{13}$C NOESY HSQC | (155) |
| | $^{13}$C NOESY (aromatics) | (155) |
| | $^{15}$N NOESY | (156,157) |

The details of the experimental NMR datasets will be discussed in more detail within each protein domain chapter. Experimental parameters differed from protein to protein and were optimised to each particular protein construct (appendix 2). The WATERGATE method was employed in order to suppress the water signal for the duration of the experiment where needed (158).

Each NMR dataset was Fourier transformed using Bruker TOPSPIN edition 1.2, it can be assumed that linear prediction was not used to enhance spectral resolution unless stated in the respective protein chapters. Amino acid assignment was performed using CCPN (Computer Collaborative Project for NMR) assignment software (159).

## 2.7 Amino acid assignments

### 2.7.1 Amide backbone assignment

The principal process behind the sequential assignment of backbone resonances was first described by Wuthrich and co-workers (160). The initial premise makes use of $H^N$ $H^\alpha$ scalar couplings in order to establish sets of spin systems based upon a protein of given residues with distinct backbone based spin systems.

Amino acid assignment was performed using CCPN (Computer Collaborative Project for NMR) assignment software (159). The software assigns a resonance to each signal in a spin system. A processed and UCSF converted $^{1}H$ $^{15}N$ HSQC spectrum was imported into the project. The $H^N$ signals of the HSQC were picked, the additional signal pairs from the glutamines were not included, and a resonance assigned to each spin system.

The HSQC signals are used to navigate and identify their reciprocal signals in the HNCACB and CBCA(CO)NH. The Cα and Cβ signals, within the 3D experiments, were confirmed by their random coil chemical shift index (161). The intra-residue signals were identified by the HNCACB experiment and the inter-residue (i-1) signals by the CBCA(CO)NH. Each of the Cα and Cβ signals were propagated to their relative $H^N$ signals in the HSQC spectrum via its assigned resonance. The Cα and Cβ signals in both spectra are labelled as i or i-1 respectively. Sequence specific assignment of these spin systems is carried out by the identification of the amino acid and its neighbour (i-1). The chemical shift index of amino

acids can result in ambiguous assignment of amide backbone residues. For this reason unambiguous residues such as alanine, glycine, serine and threonine are used as start points for the ordering of the spin systems within the protein sequence. Using this regime it is relatively simple to identify and assign regions of the amide backbone.

In order to simplify and expedite the identification and assignment of backbone residues an addition to the CCPN software was created. A python macro, written by Dr Igor Barsukov, for simulated annealing was used to provide an energy term based on the i and i-1 scalar coupling. To ensure correct assignment of backbone residues, chemical shifts for both Cα and Cβ inter and intra residues were confirmed by comparing to published random coil NMR chemical shifts of common amino acids (161). Any high energy bonds can be attributed to poor centring of the picked signals and incorrect labelling of inter and intra Cα and Cβ signals. Carbonyl carbons are assigned using the HNCO and HN(CA)CO spectra, inter-residue connectivity can also be determined using these spectra. The Cα and Cβ protons are assigned using the HBHA(CO)NH and HBHANH, this is accomplished by navigating to the spectral planes from the chemical shift data from the assigned backbone residues.

### 2.7.2 Amino acid side-chain assignments

The assignment of the amino acid side chains was not as straight forward as that of the backbone. A $^{1}H$ $^{13}C$ HSQC was used to assign the methyl regions of the side chains. Carbon edited TOCSY and proton edited TOCSY were collected and used to find the methyl components of the side chains.

Using the Cα and Cβ chemical shifts, obtained from the backbone assignments, the additional residue peaks can be found in each TOCSY experiment. From these cross peaks it is possible to identify the proton and carbon chemical shift coordinates. These can then be used to identify the correct side chain atom in the carbon edited HSQC spectrum. Once this has been identified it is possible to select that signal and navigate from there to the carbon and proton TOCSY spectra. Confirmation of the correct side chain atom can be confirmed from the TOCSY spectra as it will contain the cross peak signals for the other atoms within that side chain spin system.

### 2.7.3 NOE assignments

An NOE or Nuclear Overhauser Effect spectrum will give signals from pairs of hydrogen atoms that close together, less than 5Å, in space even though they may not be close together in sequence determined by the cross-relaxation between the two spin systems. These spectra are important in NMR as they derive the distance constraints that will give the data for the secondary and tertiary structures. For the purpose of protein structure determination, within this project, three types of NOESY spectra are collected $^{13}C$ edited NOESY-HSQC for the aliphatic region and one for the aromatic region and a $^{15}N$ edited NOESY spectrum. The NOE experiments that were carried out on PABP-1 RRM 1 and 2 were performed on 90% $H_2O$ and 10% $D_2O$.

NOE correlations are not restricted to sequential interactions but also occur between non-sequential residues. NOE data for a residue is laid out in strips and the NOE signals picked for each side chain and amide atom. With the various planes of the spectra laid out in strips the intra-residue NOE signals are assigned and the inter-residue signals were left unassigned. Structural calculations with the CYANA/CANDID calculation protocol (162) will give rise to the assignment of inter-residues and produce an output so that the inter-residue NOE assignments can be checked, structural refinements are discussed as part of the NMR chapters (chapters 3 and 4). The aromatic protons of phenylalanine, tyrosine, histidine and tryptophan residues are reliant on intra-residue NOE correlations for their assignments; this is due to the fact that they are not scalar-coupled to the rest of the side chain. NOE signals to the aromatic region can become overlapped dependent on the amount of aromatic residues with that protein to resolve these signals a 2D NOESY spectrum can be performed on a 100% $D_2O$ unlabelled sample, chapter 3.

## Chapter 3 PABP-1 RRM 2

### 3.1 Introduction

PABP-1 RRM 2 is the second domain of four tandemly arranged RNA recognition motifs. Recently, the crystal structure of the PABP-1 RRM1/2 in complex with RNA has been determined (142). RRM 2 plays a critical role in the stabilisation of the PABP-1-poly (A) mRNA complex which makes the identification of a paxillin interaction site even more interesting with regards to mRNA transport. The structure of the PABP-1 RRM 2 domain in its unbound state has not been determined. Structural comparisons between the NMR derived structure in its unbound state and the crystal structure of RRM 2 in its bound state were carried out to determine whether there is a conformational change on binding mRNA. Structural features of the domain are presented and discussed. The structure of this domain in its unbound state is important if our understanding of the function of the PABP-1 RRM 2-paxillin LD1 complex is to be determined.

### 3.2 Results and Discussion

#### 3.2.1 Enzyme Linked ImmunoSorbent Assay

In order to better characterise the interaction of paxillin with PABP-1 it was decided that a map of paxillin/PABP-1 interactions was needed. This map of interactions would also serve to focus structural efforts. As previously described mutations within the PBS 1 and 2 sites of PABP-1, in RRM1 and RRM4 respectively, were found to abolish paxillin binding, thereby compromising cell migration (127). However, the structural effects of these mutations were unknown. It maybe possible that the mutations within the PBS sites destabilised the structure of PABP-1 RRM1 or 4 thereby negating paxillin binding or PABP-1 binding to polyadenylated tails of mRNA transcripts. A structural approach based upon biochemical

interaction site mapping would lead to more accurate creation of mutants that would inhibit paxillin binding but not mRNA binding.

The interactions between various paxillin and PABP-1 polypeptides was therefore analysed using an enzyme linked immunosorbent assay (ELISA). Paxillin LD domain constructs and PABP-1 RRM domain constructs were supplied by Dr Alison Woods. Paxillin constructs were generated to include a GST N-terminal tag and PABP-1 constructs with a six histidine N-terminal tag; a complete list of these constructs is given in table 2.4.

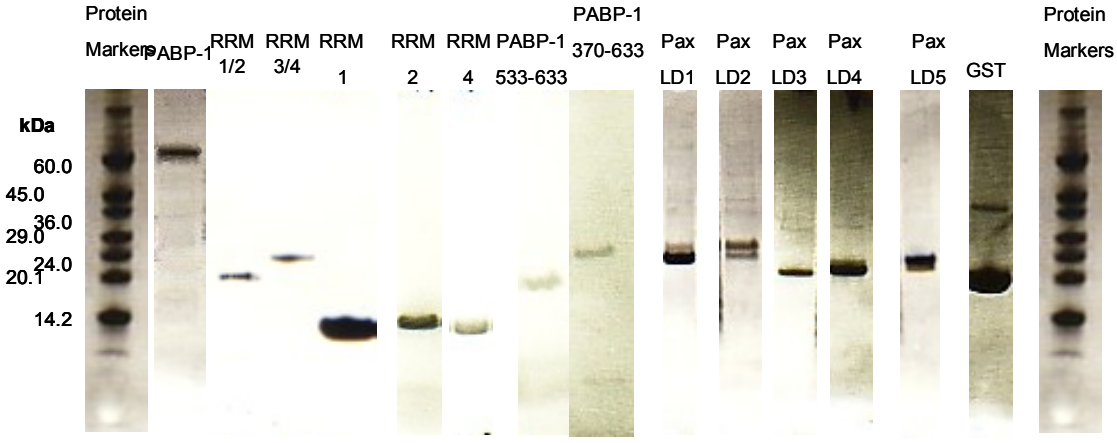



**Figure 3.1 illustrates strips of SDS-PAGE gels used to identify target proteins.** The protein bands displayed for the truncated PABP-1 proteins all contain a six histidine residue tag at their N-terminus. All paxillin proteins were GST-fusion proteins; as the GST was to be used as a negative control this has been included within this figure.

Table 2.4 (page 56) presents the cells lines and purification method used for each of these protein constructs. PABP-1 constructs were purified using a nickel column and yielded a high protein concentration (Figure 3.1); full length PABP-1 contained two lower molecular weight contaminants which were removed following ion exchange column purification (Figure 3.1). Paxillin constructs were purified by glutathione-agarose column chromatography. GST-tagged protein constructs were suspended in batch overnight at 4$^{o}$C with the glutathione-agarose, which was then washed thoroughly before elution of bound protein. This method produced a good protein yield although SDS-PAGE gels showed the presence of additional bands. Gel filtration was employed as a further purification step, but resulted in an almost total loss of target protein. After great discussion it was decided to proceed following an additional ion exchange purification step. Although the GST-paxillin constructs still contained additional bands, the biochemical results remained constant and

reproducible. Figure 3.1 is composed of strips taken from different SDS-PAGE gels and combined for ease of illustration.

For the ELISA experiments, PABP-1 constructs were bound to the wells of a 96 well microtiter plate using a fixed optimised quantity of protein (100ng); GST-paxillin constructs were then added in varying concentrations. An anti GST-HRP primary antibody was used to detect any paxillin/PABP-1 interaction by colorimetric change upon addition of ONPG; this was measured at a wavelength of 490nm (method description can be found on page 56). The ELISA experimental protocol was optimised with full length PABP-1 and full length paxillin as positive controls and GST as a negative control. The assay was initially performed with a maximum paxillin protein concentration of 90nM, but this was optimised to 300nM to produce the desired amount of detectable signal. Binding at each protein concentration was assayed in triplicate, and each set of experiments was performed in triplicate; the results are presented as a mean ± standard error.

### 3.2.2 PABP-1 full length and double RRM domain analysis

Previous studies (7,127) showed that full length PABP-1 binds to full length paxillin with a $K_d$ of ~10nM. Therefore, binding of paxillin to PABP-1 was used as a positive control for all ELISA experiments. Full length GST-paxillin was expressed and purified using a glutathione column but this resulted in an impure product. After repeated attempts to improve expression levels and improve product purity it was decided to use an existing GST-paxillin stock that had been expressed and purified by Dr Alison Woods. The results presented in Figure 3.2 are the mean of the absorbance readings at 490nm. Initially ELISA experiments with full length paxillin and PABP-1 were carried out to determine the paxillin concentration that resulted in maximum binding. The higher the protein concentration the more the results varied making it difficult to reproduce. For this reason it was decided that a maximum GST-paxillin concentration of 300nM would provide a reproducible binding result that could be used to identify paxillin/PABP-1 interaction sites.

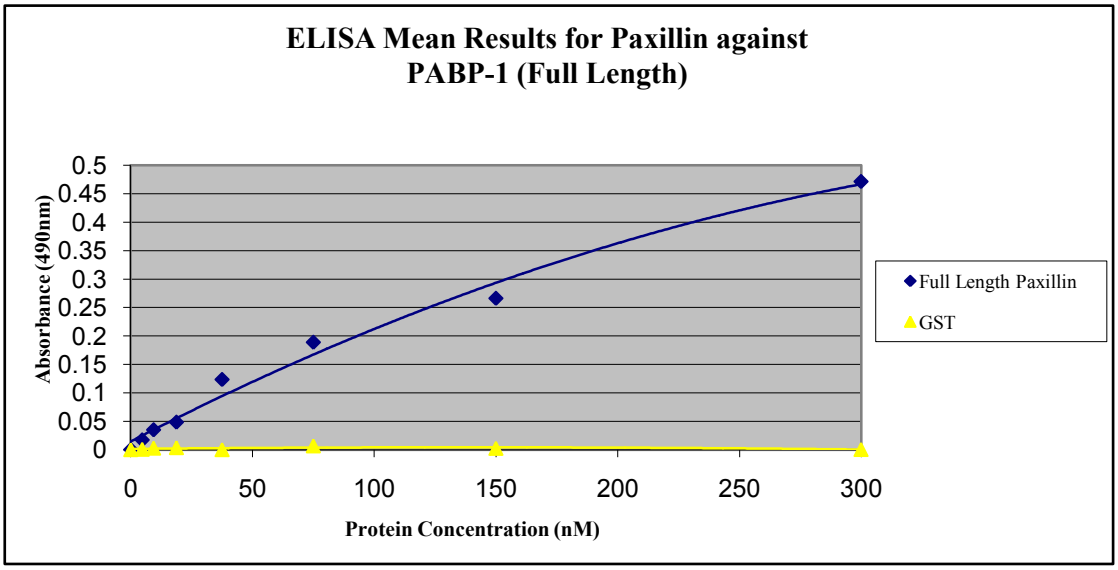



**Figure 3.2** Graph of mean results obtained from triplicate ELISA assays for full length paxillin against full length PABP-1. As can be seen, saturable binding with the GST-paxillin has not been reached. However, full length paxillin against full length PABP-1 was employed as a positive control and GST as a negative control for further subsequent assays.

PABP-1 tandem RRM domains 1/2 and 3/4 were both found to bind full length paxillin (Figure 3.3A). Given the difficulties in expressing and purifying the paxillin 55-313 construct it was decided that further binding studies would generate a better understanding if they were carried out using individual PABP-1 RRM and paxillin LD domains. The interaction between paxillin and PABP-1 was thought to be centred on the N-terminal domains of each protein. To test this hypothesis, binding of GST-paxillin to the C-terminal region of PABP-1 (constructs 533-633 and 370-633) was analysed. Both constructs were found to bind full length paxillin. A positive binding event was witnessed for paxillin LD5 to the PABP-1 370-633 construct,

but not to the PABP-1 533-633 construct. This result indicates a possible paxillin binding site may exist between residues 370 and 533 of the C-terminal region of PABP-1.

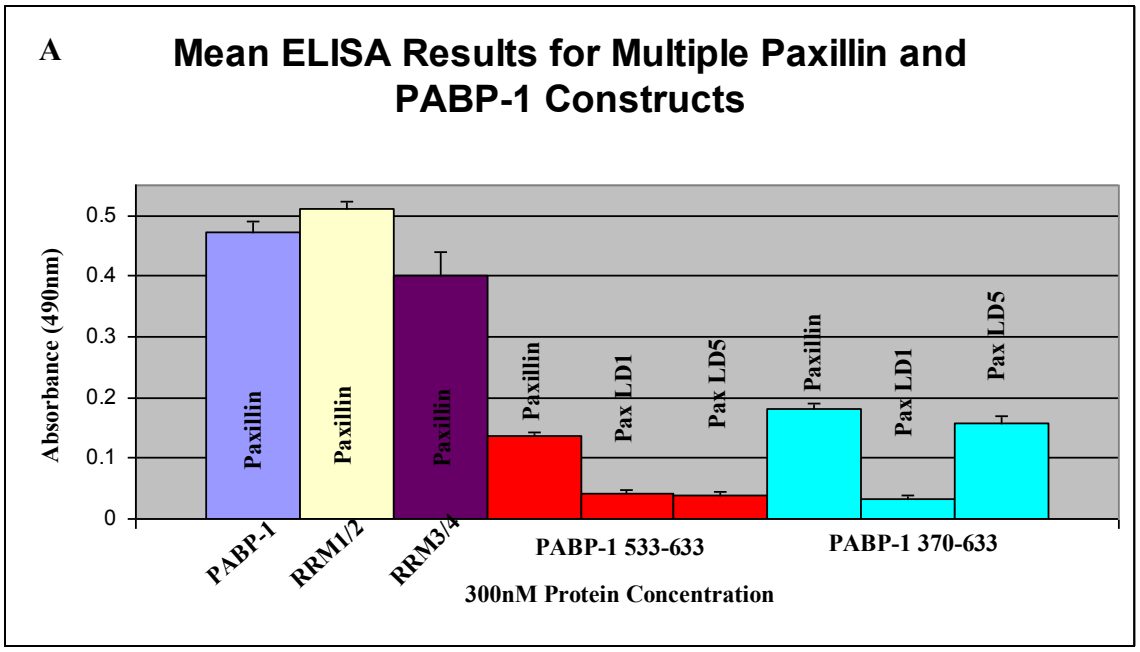
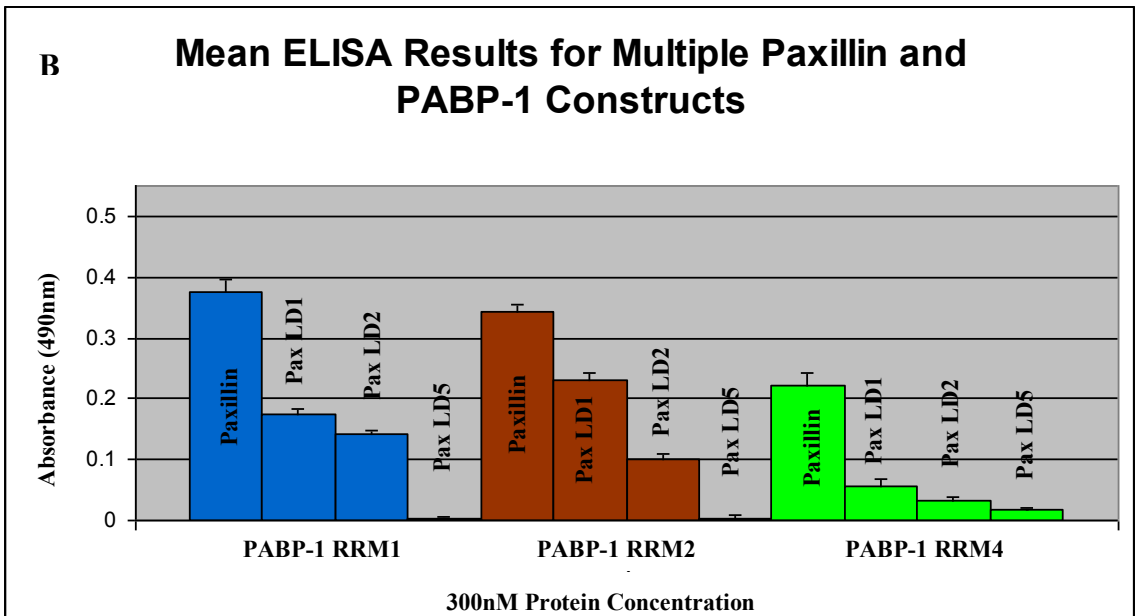



**Figure 3.3 A and B Illustrating the combined biochemical analysis results for the individual PABP-1 domains against paxillin LD domains.** The Figure illustrates binding of paxillin constructs (300nM) to PABP-1 constructs. The results shown here represent the calculated mean from the triplicate assays. The error bars represent the standard error of the mean calculated from the standard deviations.



The individual PABP-1 RRM domains were analysed in order to establish an accurate map of paxillin LD domain binding sites within the N-terminal region of PABP-1. The PABP-1 RRM1 domain bound both paxillin LD1 and LD2 domains, but not the LD5 domain (Figure 3.3B). Similar results were obtained for the RRM2 domain, although LD1 appeared to bind more strongly than LD2. By contrast, although the RRM4 domain bound to full-length paxillin, it showed little or no binding to any of the three paxillin LD domains tested.

The tandem PABP-1 RRM1/2 domain construct was also analysed for paxillin LD1 binding and gave a similar result to that obtained with the PABP-1 RRM2 construct. However, the PABP-1 RRM4 domain gave a positive binding result with full length paxillin (Figure 3.3B). Given the heterogeneous nature of some of the paxillin constructs these results can only be used as a guide. The fact that the PABP-1 RRM4 domain bound to full length paxillin but not to any of the N-terminal paxillin constructs could indicate that there may be another possible site of interaction within the C-terminal region of paxillin.

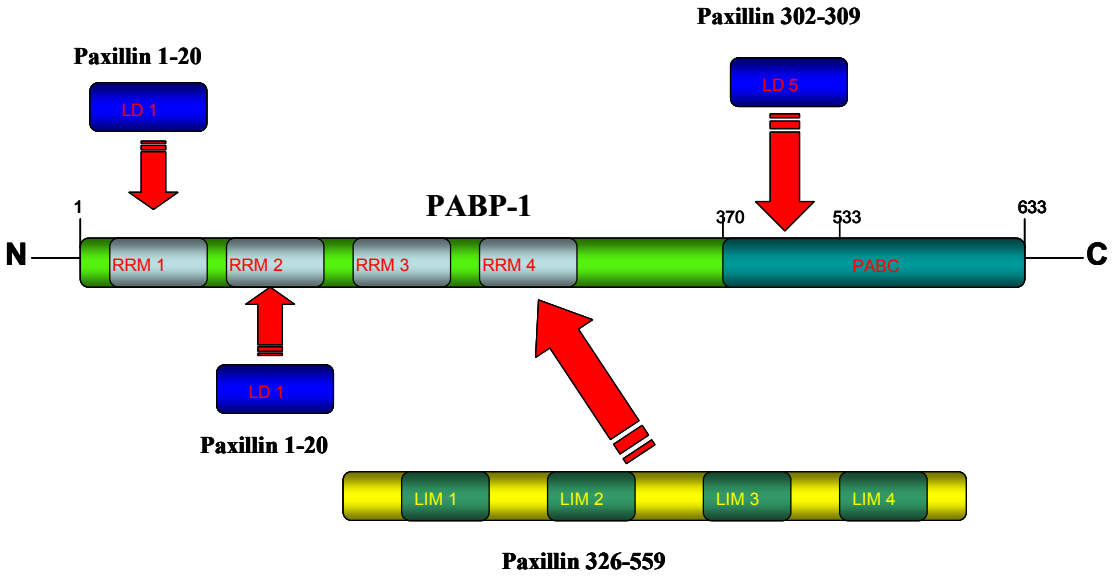



**Figure 3.4 Illustrates the Protein-Protein binding map for the Paxillin/PABP-1 complex.** The diagram summarises the various constructs of PABP-1 and their potential interactions within paxillin. The diagram aims to present a possible interaction map of the two proteins by incorporating the biochemical data presented above. Many interactions were ambiguous and as a result were omitted from this early model.

Initial expectations were that the PABP-1/paxillin interactions would be limited to the RRM and LD domains. However, the ELISA results suggest a more complicated picture. It would appear from the results that the binding interaction may involve the C-termini of both proteins as well as the RRM and LD domains (Figure 3.4).

To summarise, it was found that paxillin LD1 bound to the PABP-1 RRM1/2 double domain but also to PABP-1 RRM 1 and RRM 2. Although paxillin LD1 appears to bind to both RRMs 1 and 2, it might be that it binds primarily to PABP-1 RRM 2 but that the site overlaps with PABP-1 RRM 1 via the inter-domain linker. In the introduction, the functionality of the RRM1/2 inter-domain linker in mRNA binding was discussed. The flexibility of this linker may well be involved in the paxillin LD1 interaction which would suggest that paxillin and the poly (A) tail of mRNAs may compete for binding to PABP-1. This is a provocative theory and so it was decided to follow this line of investigation.

### 3.2.3 PABP-1 RRM 2 expression and binding data

Homology sequence modelling found the PABP-1 RRM 2 domain to span from G99 to S175 inclusive, and construct boundaries were therefore set at D90 to E182 in order to fully encompass this domain. The PABP-1 RRM 2 constructs in pETM-11 vectors, which were designed to attach an N-terminal six histidine tag, were prepared and supplied by Dr Alison Woods. Expression induction temperatures were as described in Chapter 3 and the protein appeared to be stable for long periods of time. Figure 3.5 illustrates an SDS-PAGE gel of PABP-1 RRM 2 displaying the high level of protein purification. This Figure is an example of the typical level of purity following protein expression and purification by a nickel charged metal chelate column. Lanes 5 and 6 both show strong protein bands for the RRM 2 protein at around 14.2 kDa protein marker, but they also show the presence of two minor additional proteins at ~36 and ~60 kDa respectively. These contaminants were removed following the cleavage of the N-terminal tag and a subsequent purification step. The molecular weight of PABP-1 RRM 2 without its N-terminal tag is 11.3 kDa. Typical levels of protein yields following cleavage of the N-terminal 6-histidine tag and associated buffer exchanges were ~22 mg of protein per litre of M9 media.

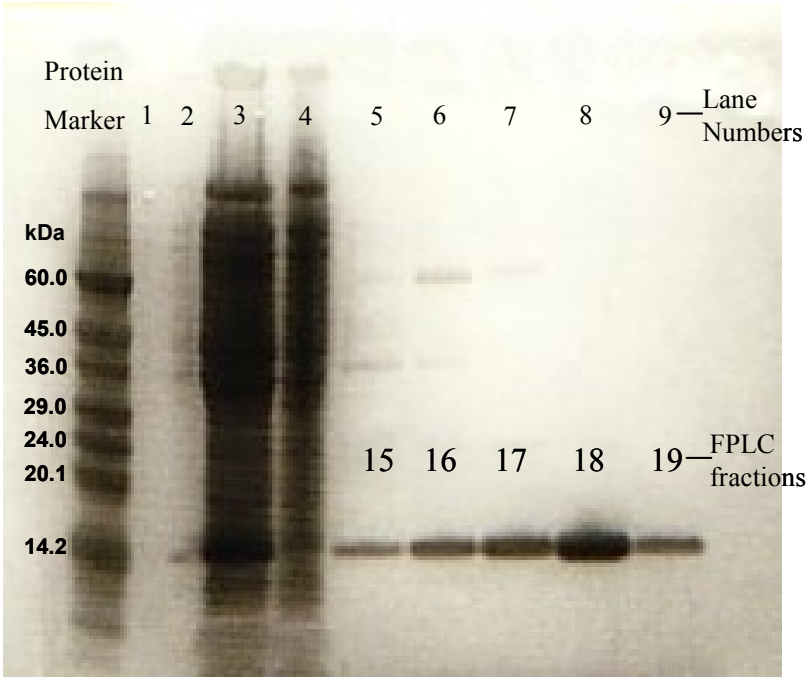



**Figure 3.5 SDS-PAGE gel of PABP-1 RRM 2.** **Lane 3** displays whole cell lysate fraction of RRM 2 following induction and mechanical cell disruption. **Lane 4** Cell lysate following centrifugation to remove cellular debris. **Lanes 5-9** Purified RRM 2 FPLC fractions.

As already discussed enzyme linked immunosorbent assays (ELISA) were employed in an attempt to determine a map of the binding interactions between paxillin LD domains and PABP-1 RRM domains. However, the labile nature of paxillin meant that determination of binding affinity was unreliable. Figure 3.6 illustrates a graph depicting the mean of the ELISA data for paxillin LD1 and LD2 against PABP-1 RRM2. Full length PABP-1 and full length paxillin were used as a positive control (although they produced inconsistent data between experiments) while GST was used as a negative control. Figure 3.6 indicates that

there maybe positive binding between paxillin LD1 and PABP-1 RRM 2. However, paxillin LD2 bound less well, and at the same level detected with all PABP-1 RRM constructs tested.

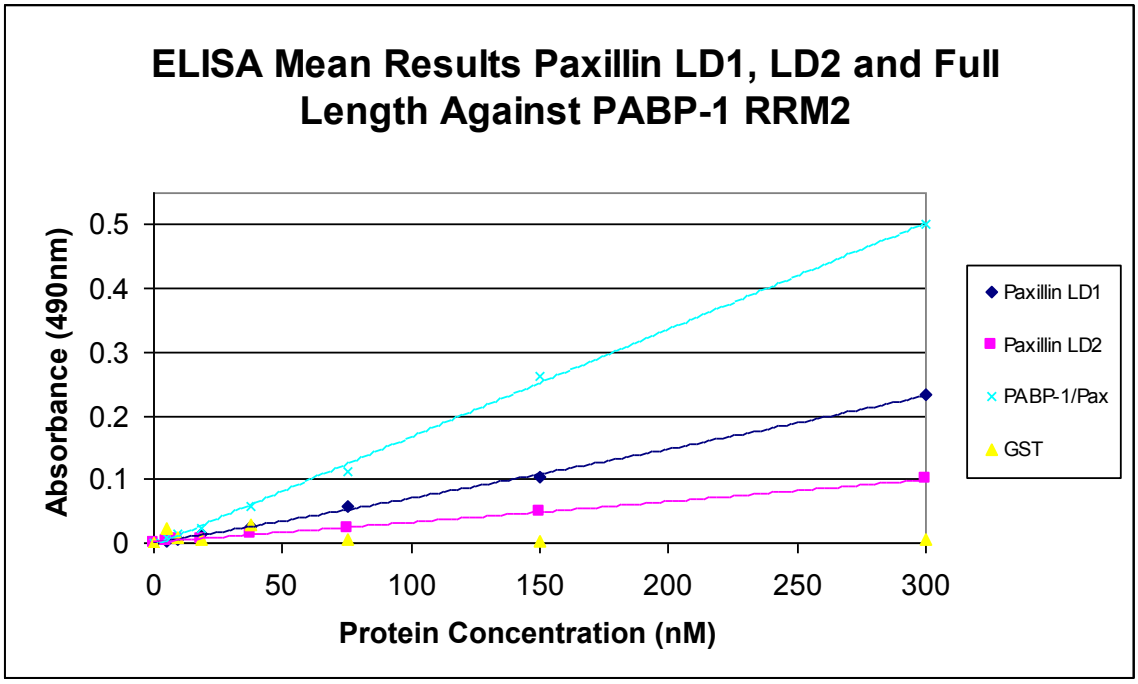



**Figure 3.6 Graph of mean ELISA data for paxillin LD1 and LD2 against PABP-1 RRM 2.** It would appear that a possible positive binding interaction can be seen for paxillin LD1. The weak interaction between paxillin LD2 and PABP-1 RRM 2 was considered a negative result as it was present in all PABP-1 RRM domains analysed.

### 3.2.4 PABP-1 RRM 2 sequence specific assignments

Sequence specific assignments were carried out using CCPN software using the method described in Chapter 2. Figure 3.7 illustrates an $^{15}N$ resolved HSQC spectrum displaying the backbone NH signals and side chain $NH_2$ group signals. The HNCACB and CBCA(CO)NH spectra (Figure 3.8a) show excellent signal intensity and very few artefacts. Both spectra were collected over a 72 hour period and their high quality was due to the size of the protein

and the relatively high protein concentration. Indeed, it can be seen in this example that the signals in the HNCACB spectrum show both i and i-1 signals. The i-1 sequential assignments were confirmed using the CBCA(CO)NH spectrum. Sequential assignment of the carbonyl carbons were facilitated by collecting HNCO and HN(CA)CO spectra. The connectivity of the amide sequential assignments was confirmed using the HNCO and HN(CA)CO spectral data. The HNCO spectrum produced i-1 while the HN(CA)CO provided data to the intra-residue. All of the expected signals were observed even in the less sensitive experiments.

Backbone NH, Cα and Cβ residue assignments were 98% complete. Cα and Cβ protons were assigned using HBHA(CO)NH and HBHANH spectra (Figure 3.8b). Both experiments gave a good signal to noise ratio with no spectral overlap. As Hα and Hβ were of opposite phase in the HBHANH spectrum assignment was relatively straight forward. Inter-residue assignments to the Hα and Hβ protons were carried out using strip plots such as those seen in Figure 3.8b.

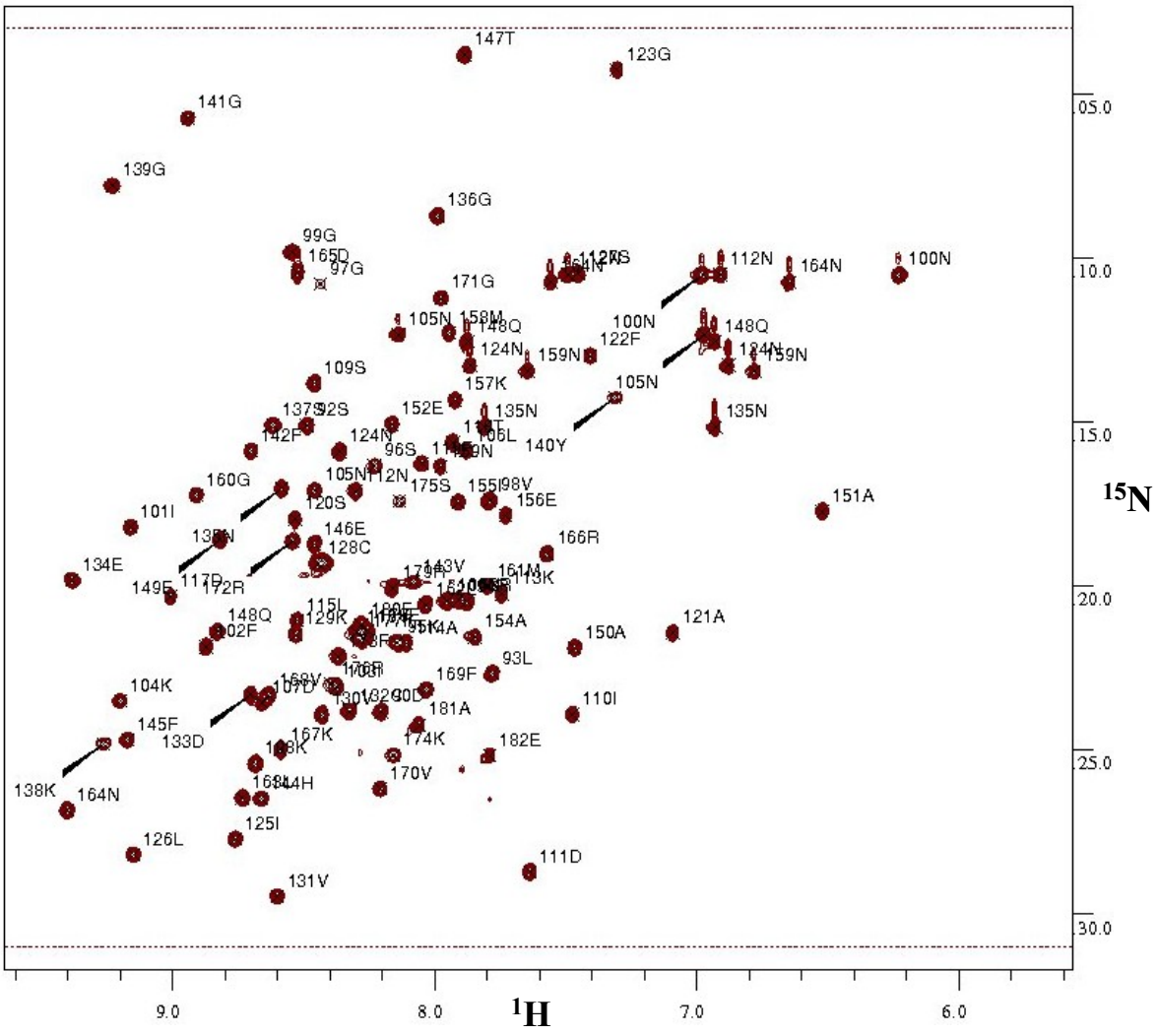



**Figure 3.7 $^{1}$H $^{15}$N edited HSQC spectrum for PABP-1 RRM 2.** Backbone NH assignments and side chain $NH_2$ groups. Sequential numbering is that of human PABP-1 SwissProt accession number P11940 with additional N-terminal 88Gly and 89Ala as remnants of the cloning process included.

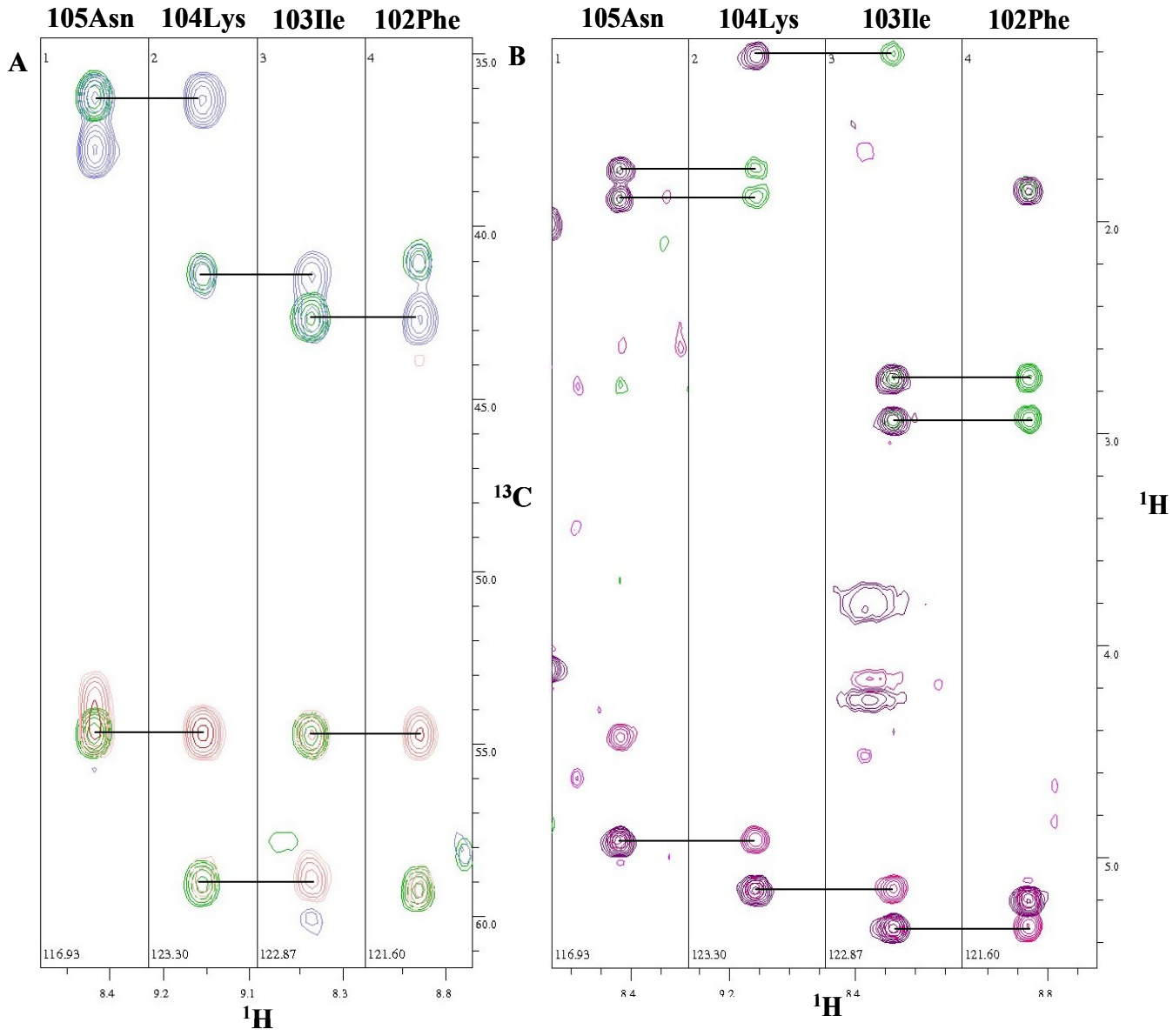



**Figure 3.8 Selected strips from the triple-resonance spectra illustrating sequential connectivities for residues 102-105 of PABP-1 RRM 2.** **A.** Superposition of CBCA(CO)NH and HNCACB spectra. Positive contour level of CBCA(CO)NH are shown in dark green ($CB_{i-1}N_iHN_i$ and $CA_{i-1}N_iHN_i$ correlations); positive contour level of HNCACB are shown in red ($CA_iN_iHN_i$ correlations), negative contour level of HNCACB are shown in blue ($CB_iN_iHN_i$ correlations). Sequential connectivities are represented by horizontal lines. **B.** Superposition of HBHA(CO)NH and HBHANH spectra. Positive contour level of HBHA(CO)NH are shown in purple ($HB_{i-1}N_iHN_i$ and $HA_{i-1}N_iHN_i$ correlations); positive contour level of HBHANH are shown in pink ($HA_iN_iHN_i$ correlations), negative contour level of HBHANH are shown in dark green ($HB_iN_iHN_i$ correlations). Sequential connectivities are

### 3.2.5 Side-chain assignments

All side chain assignments were performed in a progressive manner building on to the already assigned amide backbone and Cα and Cβ assignments. Figure 3.9 illustrates a $^{1}H$ $^{13}C$ resolved HSQC spectrum displaying expanded $C_{\alpha}$ and methyl regions. Spectral overlap was experienced within the centre region of the spectrum. Signals to the side chains located within this region were resolved using the [H]CCH-TOCSY and H[C]CH-TOCSY spectra, described assignment protocol chapter 2. Figure 3.10B illustrates selected strips from the $^{1}H$ $^{13}C$ resolved TOCSY spectrum for I103. Strip plots were used to identify the other side chain carbon and proton resonances within that residue. The corresponding signal in the $^{13}C$-HSQC spectrum was then identified and another strip plot created. The $H\gamma_1$ strip is a good example were the two protons are overlapped and could not be readily resolved and therefore were assigned as ambiguous.

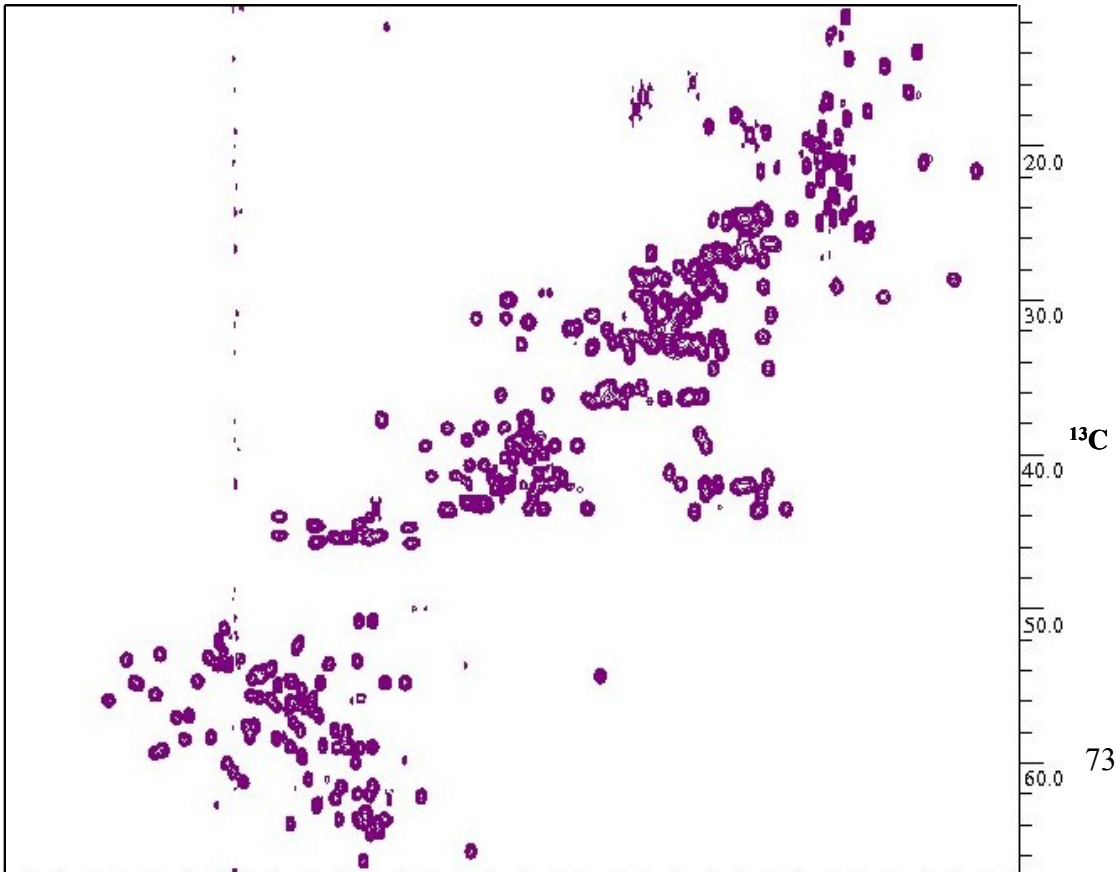


**Figure 3.9 $^{1}H$ $^{13}C$ HSQC spectrum for PABP-1 RRM 2.** The spectrum is well resolved with minimal spectral overlap of signals. Side chain residues were assigned in this spectrum via $^{1}H$-TOCSY and $^{13}C$-TOCSY spectra.

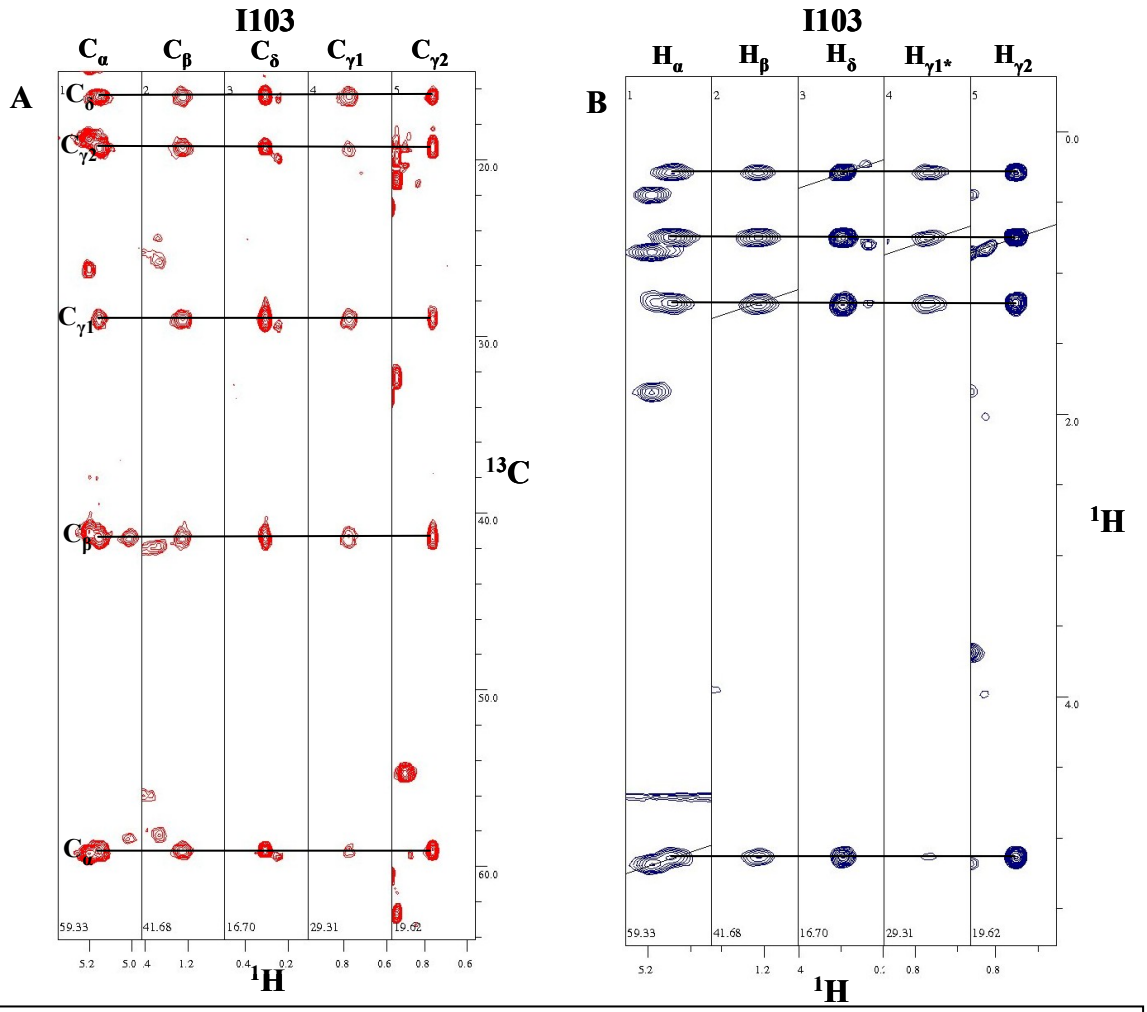



**Figure 3.10 Selected strips from the $^{13}C$ and $^{1}H$ resolved TOCSY spectra illustrating intra-residue side chain connectivities for I103. A.** Side-chain carbon atom positions, shown as red signals, for I103. The carbon signals can be seen within each spectral plane. **B.** Side-chain protons, shown as blue signals, for I103. Each proton signal can be seen in each strip-plot.

Assignment of the aromatic side-chains proved to be more difficult. The PABP-1 RRM 2 domain contains seven phenylalanines which resulted in spectral overlap of the aromatic $^{1}H$

$^{13}$C HSQC and $^{13}$C NOESY spectra. In order to resolve these overlapped signals a high resolution 2D $^{1}$H TOCSY was performed in $D_2O$ and correlated with a $^{13}$C-HSQC of the aromatic region, Figure 3.11. The Cβ Hβ chemical shift of each phenylalanine were previously determined, the Hβ chemical shift is visible within the Hδ aromatic 2D $^{1}$H TOCSY and so could be used to identify that particular residue. Although the use of these experiments was successful in resolving the overlapped signals it was found that not all of the Cζ Hζ atoms could be identified.

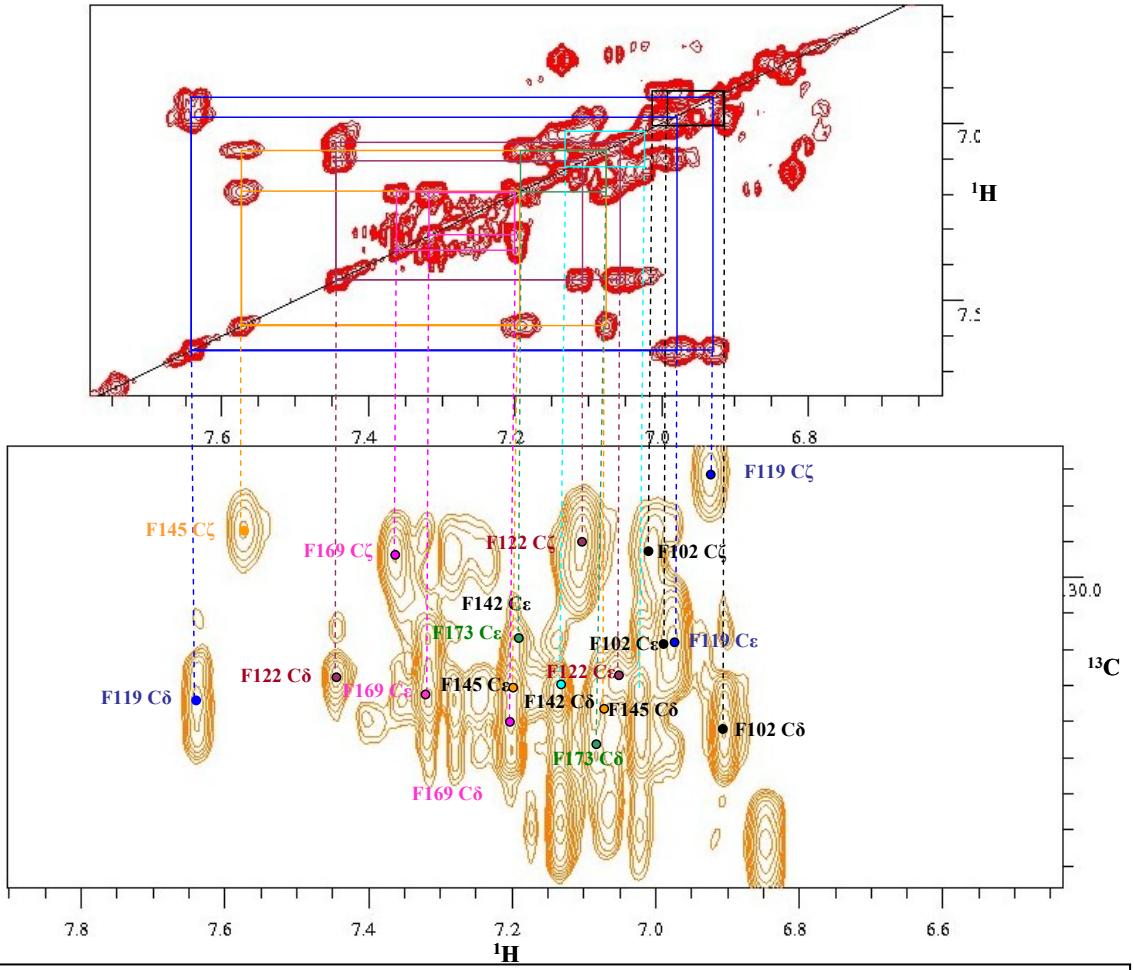



**Figure 3.11 Aromatic $^{1}$H $^{1}$H-TOCSY and $^{1}$H $^{13}$C HSQC spectra to resolve the overlapped phenylalanine signals.** The top spectrum illustrates a 2D-TOCSY spectrum of the aromatic region illustrating the $H_\delta$, $H_\varepsilon$ and $H_\zeta$ spin systems. The identified spin systems are correlated to the $^{13}$C HSQC aromatic spectrum (bottom) by dotted lines. This spectrum illustrates the carbon atom assignments for the identified ring spin systems. Not all Cζ atoms were identified.

### 3.2.6 PABP-1 RRM 2 NMR structure calculation

Initial structural calculations were performed using CYANA/CANDID protocol version 2.1 for the determination of unique assignments of NOE derived data. Dihedral constraints generated by TALOS were also included in the calculations. Initial calculations derived a folded structure with 86% of NOE signals assigned using a chemical shift tolerance for $^{1}H$ 0.03 and 0.3 for both $^{15}N$ and $^{13}C$. Unassigned NOE signals were examined to ensure there were no systematic errors. They were then examined to determine whether they were from another residue that had not been assigned by the calculation protocol. The NOE data was then refined by referring to the structure and used to generate constraint tables for ARIA1.2 where the final structural water refinements took place. Structural refinements took into account any raised NOE energies that were highlighted as part of the ARIA output by cross referring to the group of signals or angle constraints and correcting the violating constraint. Table 3.1 lists the final restraints generated from 20 lowest energy water refined structures derived from 100 random start structures.

**Table 3.1 Solution Structure Determination of PABP-1 RRM 2**

| Experimental restraints | |
|---|---|
| Restraints | |
| Unique | 2171 |
| Intraresidue | 722 |
| Sequential | 476 |
| Short range (l< [i-j] < 5) | 328 |
| Long range ([i-j] > 4) | 645 |
| φ/ψ dihedral angles[a] | 156 |
| Energies (Kcal $mol^{-1}$)[b] | |
| Total | -3508.85 ± 114.81 |
| Van Der Waals | -850.50 ± 15.40 |
| NOE | 96.60 ± 0.0027 |
| RMS deviations[b] | |
| NOEs (Å) (no violations > 0.5 Å) | 0.028 ± 0.0085 |
| Dihedral restraints ($^{o}$) | 0.58±0.011 |
| Bonds (Å) | 0.0038 ± 0.00015 |
| Angles ($^{o}$) | 0.51 ± 0.019 |
| Impropers ($^{o}$) | 1.66 ± 0.13 |
| Ramachandran map analysis[c] | |
| Allowed regions | 84.4% |
| Additional allowed regions | 14.5% |
| Generously allowed regions | 0.5% |
| Disallowed regions | 0.6% |
| Pairwise rms difference (Å) | |
| Residues 90-182 (heavy atom) | 1.151 |
| $2^{nd}$ Structure Residues 90-182 (backbone) | 0.503 |

[a] From chemical shifts using Talos.
[b] Calculated in Aria 1.2 for the 20 lowest energy structures refined in water.
[c] Obtained using PROCHECK-NMR.

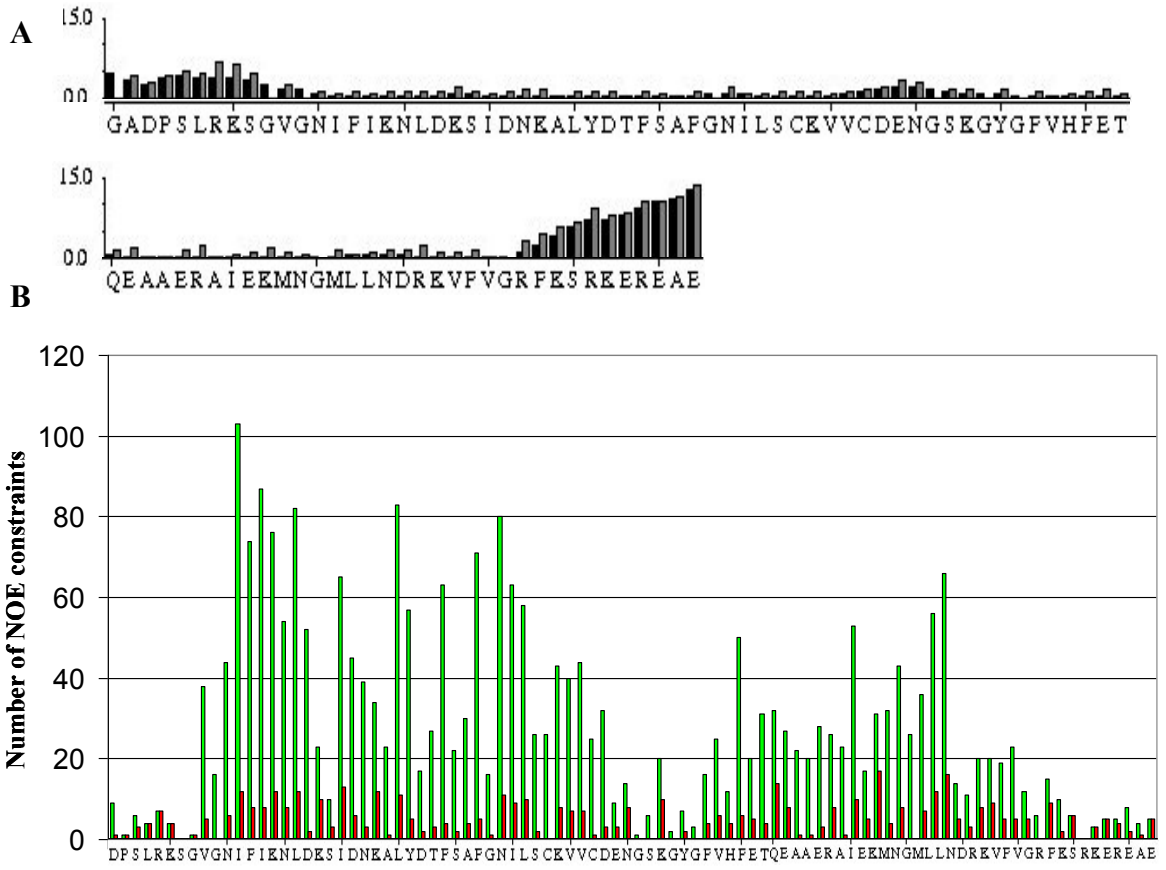



**Figure 3.12 Statistics on a residue per residue basis for the 20 lowest energy conformers representing the NMR structure of PABP-1 RRM 2. A.** Depicts the average local RMSD for the backbone in black and the side chain in grey. **B.** Illustrates the number of NOE constraints per residue with the intra-residue NOE constraints shown by red bars.

The final water refined lowest energy conformers shown in Figure 3.13, were analysed by Procheck-NMR (163). Figure 3.13 show the global fold to be $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$, which is seen in other RRM domain containing proteins. Figure 3.12(B) illustrates the RMSD analysis on a residue by residue basis for the amide backbone (black) and the side-chain (grey). High local RMSD values can be seen at the N and C-termini ranging from 5-15Å owing to limited NOE data to those regions. A rise in RMSD values can be seen between V130 and G139 this indicates that this region maybe mobile. Figure 3.12(B) illustrates the number of NOE constraints on a residue by residue basis. It can be seen that low numbers of NOE constraints is accompanied by a rise in RMSD values as can be seen in Figures 3.12(A) and 3.12(B) for the N and C-termini.

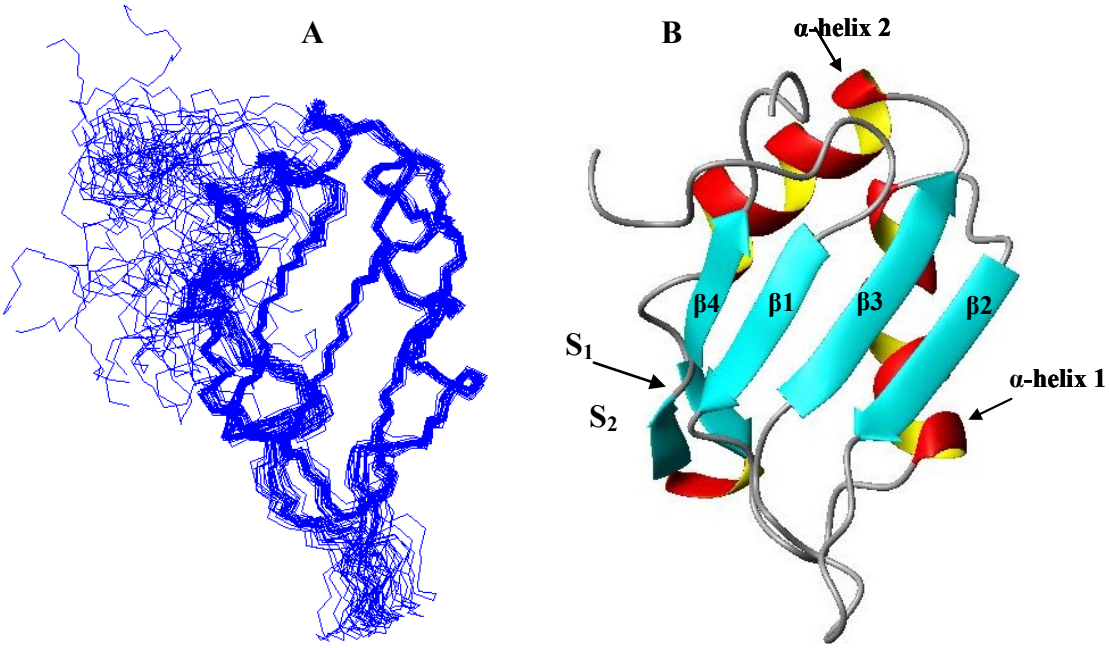



**Figure 3.13 Superposition of 20 lowest energy converged structures for PABP-1 RRM 2. A.** Overlay of structures following a water refinement step from ARIA. **B.** Secondary structure depiction yielding a four stranded anti-parallel β-sheet with two α-helices crossing over the β-sheet on one side. A small anti-parallel β-sheet is seen as part of the loop between the C-terminus of helix-2 and the N-terminus of the $\beta_4$-strand.

### 3.2.7 Description of secondary structure

Proteins with regions of secondary structure display distinctive patterns of inter-residue NOEs. For example an α-helix is characterised by strong HN(i)/HN(i-1) and HN(i)/HA(i),

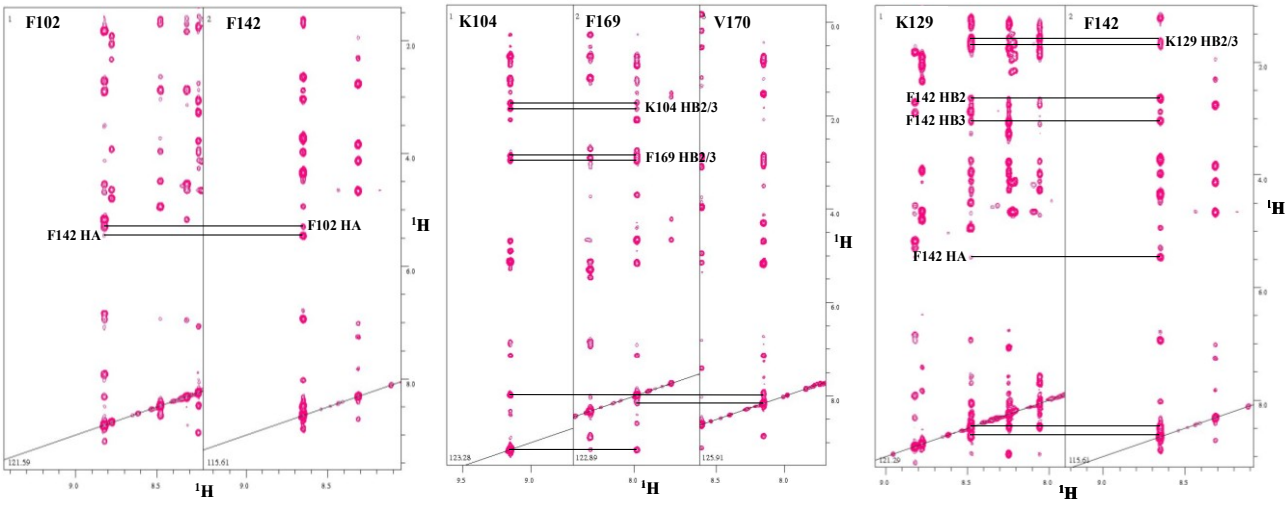
weak/medium HN(i)/HA(i-1), weak/medium HN(i)/HA(i-3), HN(i)/HA(i-4) and HA(i)/HB(i-3), (160). A $3_{10}$ helix shows a similar pattern but with only HN(i)/HA(1-2), HN(i)/HA(i-3) and HA(i)/HB(i-2). A β-sheet will typically give an NOE pattern between HN/H$\alpha_{(i,j)}$ and H$\alpha$/HN$_{(i,j)}$ on adjacent strands.

After structure calculations the secondary structure NOE data was checked against the expected NOE patterns. Long range NOE signals that are indicative of an anti-parallel β-sheet were identified between residues I101-K104, L126-K129, Y140-F146 and R166-V170. The NOE data from these residues has been summarised in Figure 3.14(D) with supporting $^{15}$N-NOESY-HSQC strip plots. These constraints determine the order and alignment of the β-strands. NOE data from β strands 1, 2 and 4 gave NOE signals to Y140 thus extending the β-strand by one residue in comparison to the other 4 residue strands. N159 presents NOE signals from the C-terminus of helix 2 to F169 on the $\beta_4$-strand suggesting that these two residues are close in space and that the C-terminus of helix 2 is within 6Å of the $\beta_4$-strand.

NOE's to the residues in helix 1 give a typical NOE pattern of an α-helix. Figures 3.15a and b both display the NOE pattern described above for an α-helix, with strong NOE signals to HN(i)/HA(i) and weak/medium signals to HN(i)/HA(i-3) HN(i)/HA(i-4) and HA(i)/HB(i-3). A commonly witnessed aspect of this helix is that three residues on the C-terminus of helix 1 gave an NOE pattern of HN(i)/HA(i-2) and HA(i)/HB(i-2) indicative of a 3.10 helix, see Figure 3.15a. This distinctive feature of helix 1 was not seen in the helix 2. The C-terminal region of helix 1 has two phenylalanines both of which may cause the helix to extend through hydrophobic interactions with the hydrophobic core. In contrast to this, helix 2 has no aromatic residues or other residues that may disrupt the α-helix.

NOE data provides supporting evidence that a small β-sheet is formed between the C-terminus of helix-2 and the $\beta_4$-strand of the main β-sheet (Figure 3.13). NOE evidence for loop between helix2 and the $\beta_4$-strand suggests a formation of a second small β-sheet M161-L163 of $S_1$ and N164 and D165 of $S_2$. The turn itself is a tight turn giving a NOE pattern similar to that of a $3_{10}$ helix HA(i)/HB(i-2) suggesting that the two sides of the loop are in close proximity with one another. The NOE data suggests that L163 on the $S_1$-strand may form a hydrogen bond with R166 on the $\beta_4$-strand causing the sides of the loop to pinch together. Figure 3.13b illustrates that the $S_2$-strand is separated from the $\beta_4$ strand of the main β-sheet by a short random coil region. The presence of this second small β-sheet appears to

be a common feature in RRM binding proteins, although there is no tangible evidence to suggest that it is involved with RNA binding.



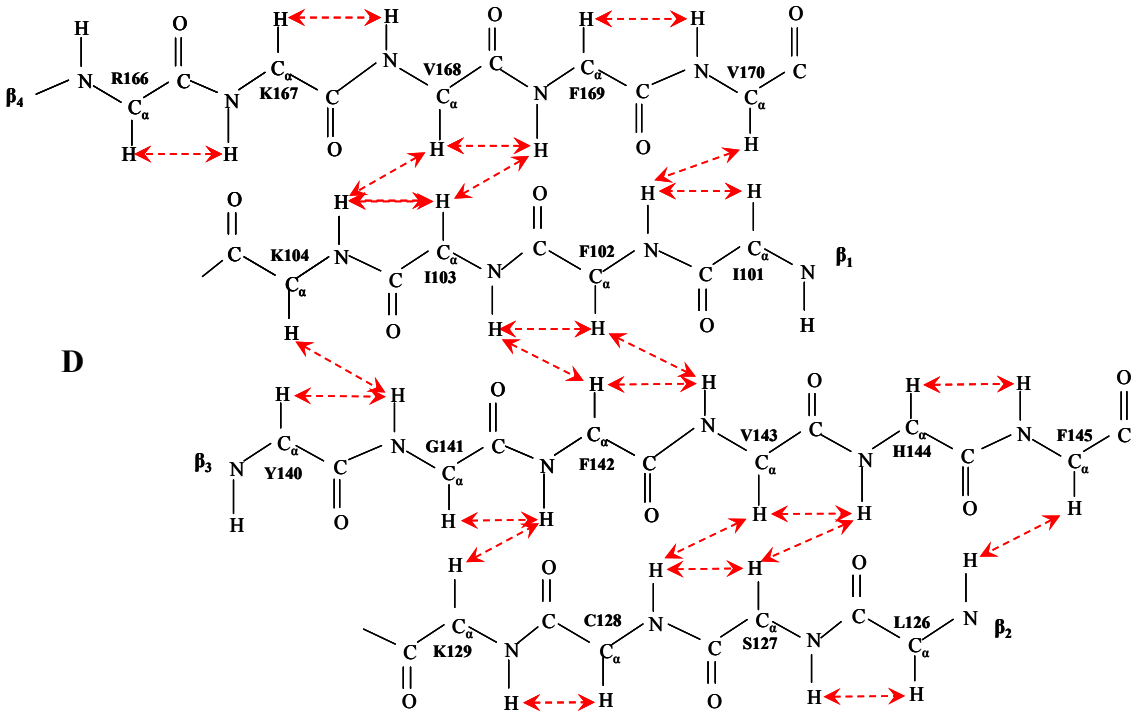



**Figure 3.14 Selected strips from $^{15}$N edited NOESY spectrum connectivities for the $\beta_1$, $\beta_2$, $\beta_3$ and $\beta_4$ strands for PABP-1 RRM 2. A.** $^{15}$N edited NOESY-HSQC spectra with NOE connectivities for residues F102 and F142 of $\beta_1$ and $\beta_3$ strands respectively. **B.** $^{15}$N edited NOESY HSQC spectra with NOE connectivities for residues K104 on $\beta_1$ strand, F169 and V170 of the $\beta_4$ strand with the cross connectivities which form part of the anti-parallel β-sheet. **C.** $^{15}$N edited NOESY HSQC spectra with NOE connectivities for residues K129 on the $\beta_2$ strand and F142 on the $\beta_4$ strand. **D.** Schematic illustration of the four stranded anti-parallel β-sheet of PABP-1 RRM 2, uniquely assigned NOEs that determined the ordering and alignment of the β strands are displayed as arrows.

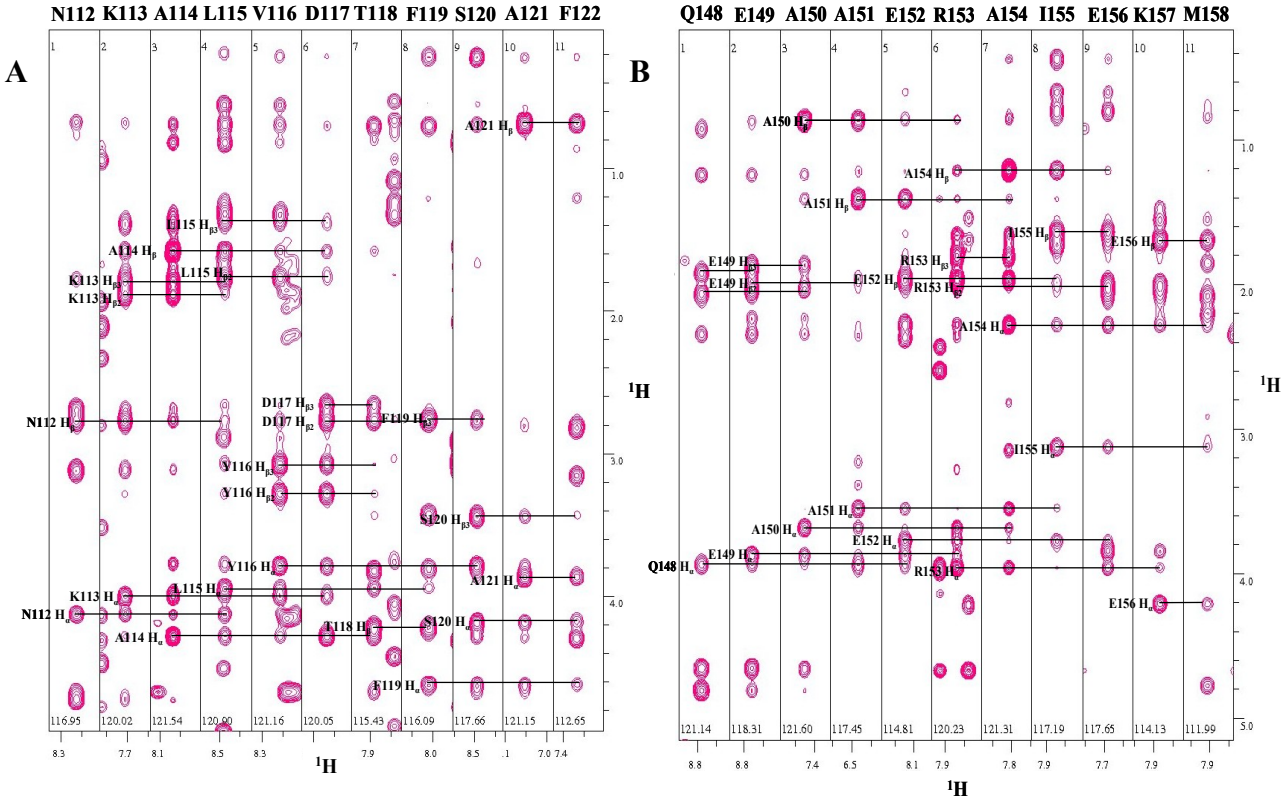
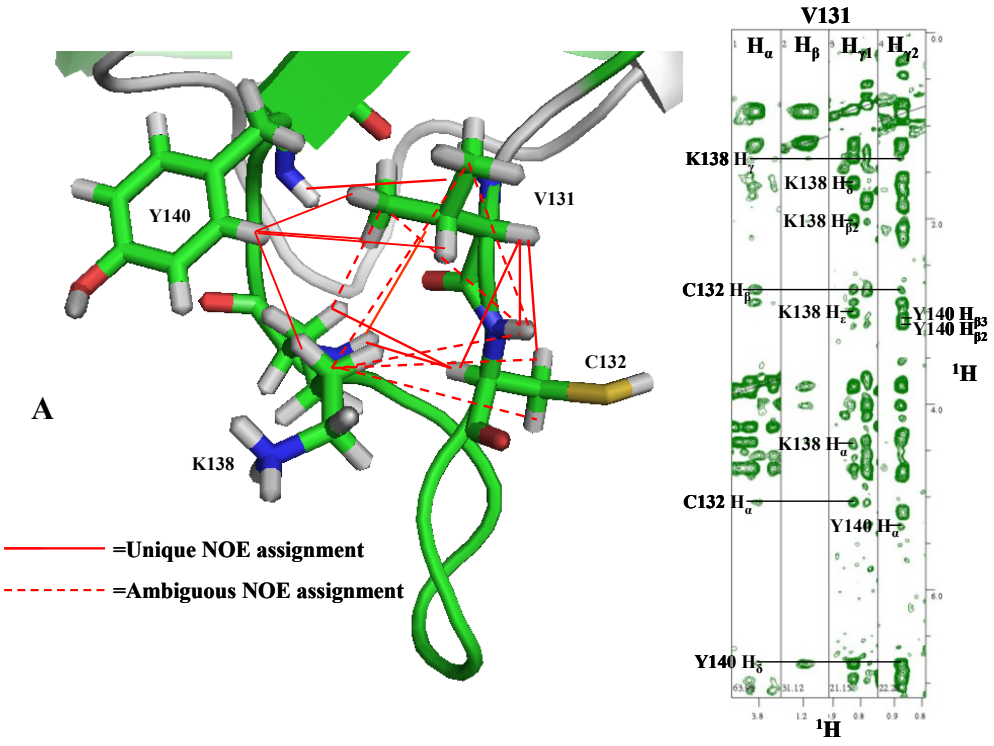



**Figure 3.15 Selected strips from $^{15}$N edited NOESY-HSQC spectrum NOE connectivities for the helical regions of PABP-1 RRM 2. A.** Strip plot of a $^{15}$N edited NOESY-HSQC illustrating an NOE pattern that is indicative of an α-helix. The strip plot illustrates $HN_{i\pm3}$ for residues N112-F119 and $HN_{i\pm2}$ for S120-F122, weak/medium NOE for A114 HA and V116 HA HB of helical region 1 of PABP-1 RRM 2. **B.** Strip plot of a $^{15}$N edited NOESY HSQC spectrum illustrating an α-helical NOE pattern. The strip plot illustrates $HN_{i-3}$ for residues E149-A154 a similar pattern maybe present but many of the signals are overlapped. Weak/medium NOE signals were also seen for E149 HA and HB, R153 HA, I155 HA and A154 HB of the α-helix 2.

One feature of PABP-1 RRM 2 is the loop (loop3) between the $\beta_2$ and $\beta_3$-strands. This loop is shows medium range NOE's giving it a β-strand appearance (similar to that of the S-sheet) Figure 3.16(A). The turn of loop3 itself demonstrates NOE's to i-2 indicating that it is a tight turn, Figure 3.16(B). The loop appears to be conserved within the RRM domains of RNA interacting proteins; other loops within the PABP-1 RRM 2 domain do not show the same characteristics. The loop structure maintains its tight architecture by making polar contacts within itself. Residues N135 and D133 show multiple NOE connections and polar contacts may form, Figure 3.16(B). These contacts result in bringing the sides of the loop together making it a tight turn at the end. E134 projects outward from the turn of loop3 making it accessible to RNA ligands. Although the study by Deo *et al* 1999 did not identify E134 as an active component in PABPs interaction with the poly (A) mRNA tail, the RNA titration performed later in this chapter shows a weak interaction for E134 with the RNA ligand. It can be seen in Figure 3.16(A) that Y140 is bent away from the β-strand ($\beta_3$) as NOE signals were observed between residues in loop3 and Y140. Y140 is of particular biological importance as it is central to the stacking interactions with RNA. Multiple NOE signals were seen from K138 and the residues of loop3 indicating that its orientation is across the loop. NOE signals from the side chain regions of K138 to the aromatic ring of Y140 and K138s orientation lends further evidence to supports this feature.

The structure of PABP-1 RRM 2 shows the two helices packed against the hydrophobic core that is presented on the rear of the β strands. The C-terminal hydrophobic residue F122 of helix 1 projects inward toward A154 on helix 2, NOE signals, Figure 3.17(A), were detected between F122 and A150, A151 and A154 indicating the presence of hydrophobic interactions between the two helices. The hydrophobic interaction between A152 and A154 of helix 2 with F122 and I125 of helix 1 facilitates the close proximity of the two helices. This is further supported by the possible polar contacts between N124 and E146. The hydrophobic core has a central phenylalanine (F145) located on the $\beta_3$-strand. Figure 3.17(B) illustrates the NOE signals across the β-sheet and to α-helix 2 from F145. The α-helices pack against the hydrophobic residues, located on the rear of the β-sheet, to form an integral part of the hydrophobic core itself. Hydrophobic residues (V168 and V170) on the $\beta_4$-strand appear somewhat removed from the hydrophobic cluster but are able to make hydrophobic interactions with I155 on helix 2 and F119 on helix 1. NOE data from the aromatic ring of F119 to V168 and V170 indicate that it is bent toward the N-terminus of the helix and toward the $\beta_4$-strand.


A
Y140
V131
C132
K138
=Unique NOE assignment
=Ambiguous NOE assignment
V131
Hα
Hβ
Hγ1
Hγ2
K138 Hγ
K138 Hδ
K138 Hβ2
C132 Hβ
K138 Hε
Y140 Hβ3
Y140 Hβ2
K138 Hα
C132 Hα
Y140 Hα
Y140 Hδ
1H
1H

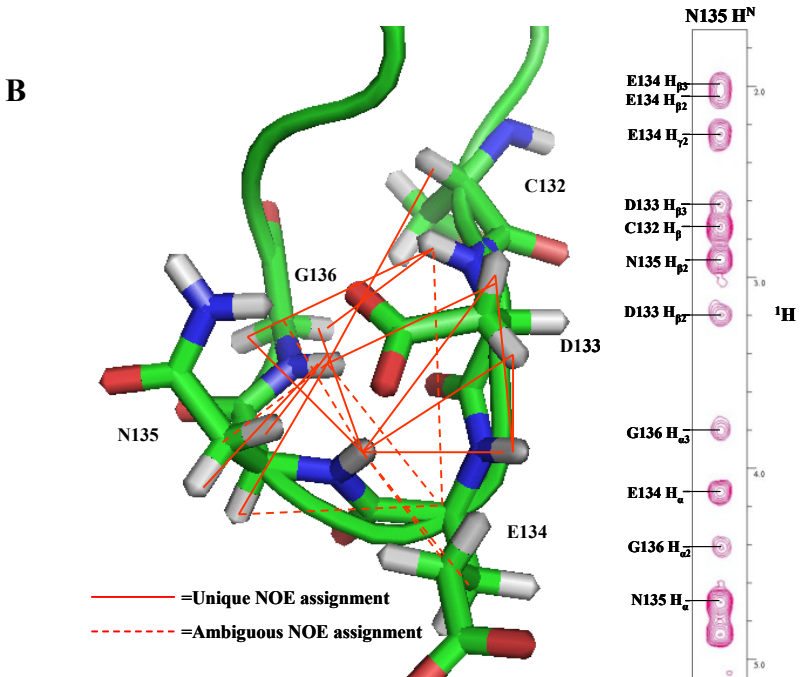

B
C132
G136
D133
N135
E134
=Unique NOE assignment
=Ambiguous NOE assignment
N135 HN
E134 Hβ3
E134 Hβ2
E134 Hγ2
D133 Hβ3
C132 Hβ
N135 Hβ2
D133 Hβ2
1H
G136 Hα3
E134 Hα
G136 Hα2
N135 Hα
1H

**Figure 3.16 Illustrates the side-chain and structural characteristics for loop 3 of PABP-1 RRM 2.** These Figures depict the NOE connections between residues of loop 3. **A.** NOE signals of the residues at the base of the loop and NOE's detected to Y140 indicates that it is bent away from the β-strand. **B.** Short range NOE's were witnessed in the tight turn of loop3. Many of these NOE's originate from the NH of N135 that appears to act as an anchor.

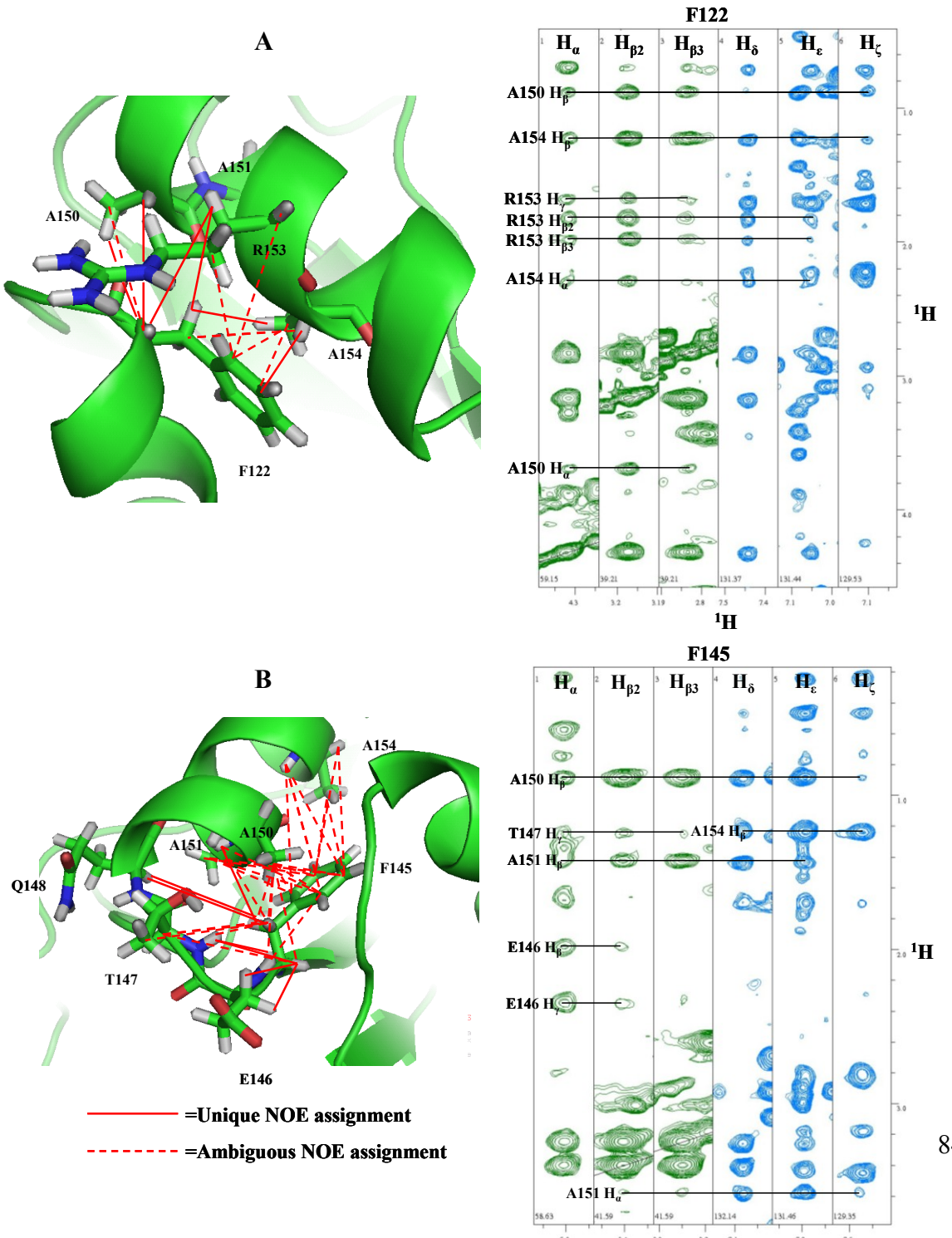


**Figure 3.17 NOE constraints between the two helices of PABP-1 RRM 2. A.** Hydrophobic interactions exist between the two helices which move them into close proximity to one another. NOE data presented in the selected strips support the connectivities demonstrated on the structure. **B.** The C-terminus of helix one and the N-terminus of helix two appear close together forming an arrow-head like appearance. Central to this is F145 making numerous NOE connectivities as illustrated by the selected strips of the $^{13}$C-NOESY spectra for the aliphatic (green signals) and aromatic regions of F145 (blue signals).

The hydrophobic core that is formed between the two helices and the β-sheet is well maintained however, there are still hydrophobic residues that are present on the other side of the solvent exposed face of the β-sheet. The NOE data suggests the presence of possible hydrogen bonds and Van der Waals interactions that exist to help maintain their solvent exposure. Figure 3.18(A) illustrates NOE signals from V131 located toward the C-terminus of the $\beta_2$-strand. NOE data suggests that V131 plays a key role by forming Van der Waals interactions with Y140 and K138. As mentioned earlier NOE data shows that the aromatic ring of Y140 is bent away from the β-sheet suggesting possible Van der Waals interactions with K104 on $\beta_1$, calculated Van der Waals interactions were based on the 20 lowest energy water refined structures and outputted from MolMol (164).

Three phenylalanines are present on the surface of the β-sheet with their aromatic rings exposed. As mentioned in chapter 2 the obtained NOE data was used to calculate possible hydrogen bonds and Van der Waals interactions; hydrogen bonds were not measured directly but calculated from the 20 lowest energy water refined structures. Figure 3.18(B) illustrates the NOE signal data seen for the centrally located F102 which is extensively involved in RNA binding (6). The position of the aromatic ring suggests that a Van der Waals interaction may exist between Hδ protons of residues of F102 and F142. NOE signals across the β-sheet indicate that key hydrogen bonds between K129 and F142, G171 and F102, F169 and K104 form. To reinforce this interaction between the strands Van der Waals interactions were predicted between K104 on the $\beta_1$ strand and V168 on $\beta_4$, K104 also presents a possible Van der Waals interaction to the aromatic ring on F102 and Y140.

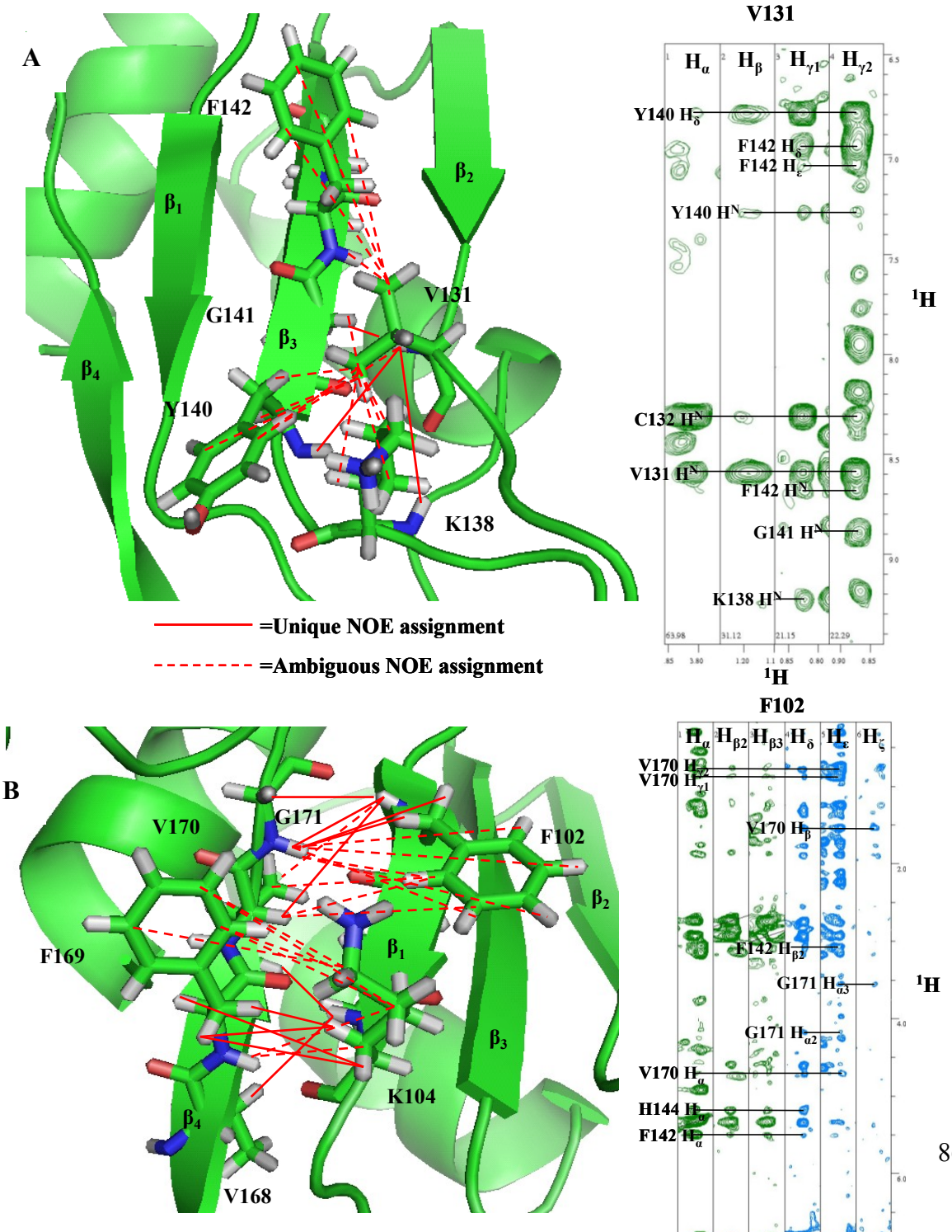

A
F142
β1
β2
G141
β3
β4
Y140
V131
K138
=Unique NOE assignment
=Ambiguous NOE assignment
V131
Hα
Hβ
Hγ1
Hγ2
Y140 Hδ
F142 Hδ
F142 Hε
Y140 HN
C132 HN
V131 HN
F142 HN
G141 HN
K138 HN
1H
1H
B
V170
G171
F102
β2
F169
β1
β3
K104
β4
V168
F102
Hα
Hβ2
Hβ3
Hδ
Hε
Hζ
V170 Hγ2
V170 Hγ1
V170 Hβ
F142 Hβ2
G171 Hα3
G171 Hα2
V170 Hα
H144 Hα
F142 Hα
1H

**Figure 3.18 illustrates supporting NOE constraints for the β-sheet of PABP-1 RRM 2. A.** NOE constraints for V131, it can be seen that this residue presents key NOE signals which suggests K138 is in close proximity with Y140 suggesting possible Van der Waals interactions between Y140 and K138. **B.** Demonstrates the key phenylalanine (F102) which plays an important role in maintaining the hydrophobicity of the β-sheet. As can be seen NOE signals were witnessed to F169 and F142 as well as Y140 allowing for interactions with the poly(A) tail of mRNA.

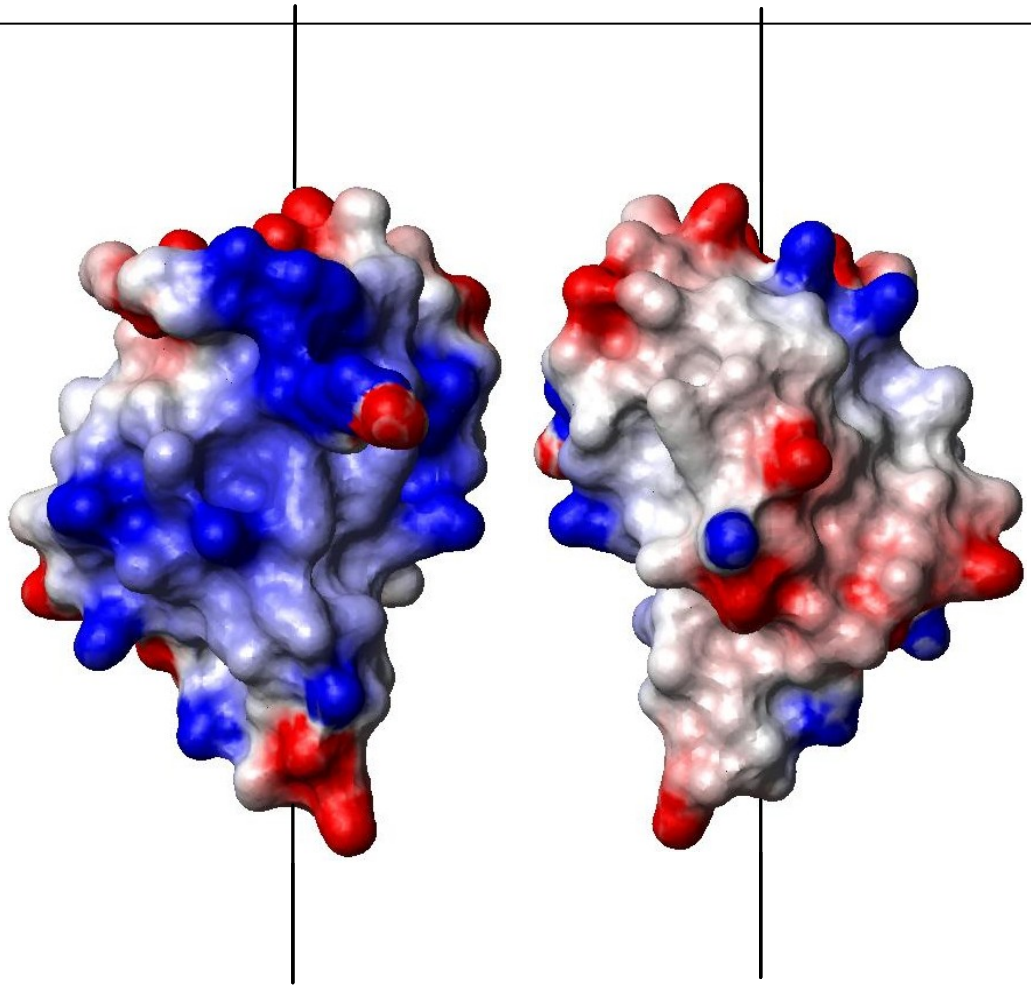


**Figure 3.19 Electrostatic surface representation of PABP-1 RRM 2.** The two Figures illustrate a surface view of PABP-1 RRM 2 coloured according to its electrostatic potential. The left Figure shows an area of significant positive charge across the surface of the β-sheet with the lightly coloured phenylalanines interspersed throughout the sheet. The right hand Figure, rotated 90$^{o}$, illustrates the α-helical side of PABP-1 RRM 2 with negatively charged residues in red and neutral residues in white.

Figure 3.19 shows a depiction of the electrostatic surface charge for PABP-1 RRM 2 the charge was determined using molmol based on the mean of 20 lowest energy water refined conformers. As can be seen the left Figure shows the positive charge that is present across the β-sheet. Dispersed across the β-sheet are the aromatic rings of F102, Y140, F142, H144 and F169, and have all been shown to provide stacking interactions with the purine rings of the polyadenylated tail of PABP-1 RRM 2 (6). In contrast to this positively charged surface is a largely neutral surface on the α-helical side of PABP-1 RRM 2. Helix 2 presents two positively charged residues (R151 and K156) and three negatively charged residues (E49, 152 and 156) that project outward away from the rest of the structure. A distinct positive patch is seen on the electrostatic surface of PABP-1 RRM 2, see Figure 3.19, for helix 2. These may possibly make an important contribution to further protein-protein interactions such as PABP-1s interaction with the eIF complex.

### 3.2.9 Comparison of NMR and Crystal structures

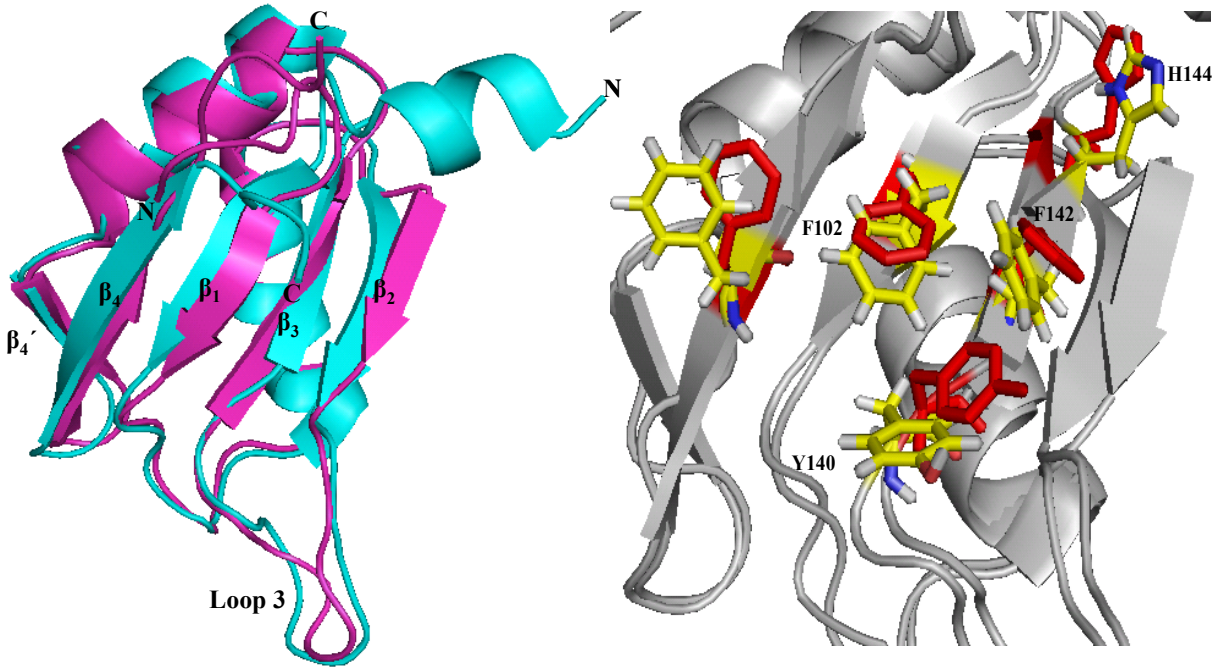



**Figure 3.20 illustrates an overlay of the mean NMR derived structure of PABP-1 RRM 2 with the crystal structure of PABP-1 RRM 2 in complex with RNA (6).** The NMR derived structure (pink) and the crystal structure (light blue) overlay with very few differences, the main differences exist in the helices. **B.** Overlay of the aromatic rings on the solvent exposed face of PABP-1 RRM 2, crystal structure in complex with poly (A) RNA (red), NMR structure unbound (yellow).

The crystallographic structure of PABP-1 RRM 1/2 in complex with a polyadenylated tail of mRNA has been solved (6). Figure 3.20 shows a direct comparison between the NMR structure of PABP-1 RRM 2 in its unbound state with the crystal structure of PABP-1 RRM in complex with RNA. The two structures overlay very well with very few deviations indicating that the NMR structure displays very little conformational changes between the free and bound conformations. The notable differences lay with the aromatic residues of the β-sheet. F142 illustrates that in the unbound state it is rotated in the opposite direction, NOE data to the aromatic ring supports this orientation of the aromatic ring. Deo *et al* 1999 showed that the linker region between PABP-1 RRM 1 and 2, residues 90-98, was random coil in the absence of RNA. This would account for the differences witnessed in the N-terminus of NMR derived structure produced in this project which was in the absence of RNA. NOE data to both the helices of the NMR structure indicate that they are half a turn longer than those seen in the crystal structure. The differences in the β-strands maybe due to free and the bound state of PABP-1 RRM 2 however, the deviations seen are not particularly large.

A direct comparison between the PABP-1 RRM 2 and the RRM domains of hnRNPA1, HUD and Nucleolin is presented in Figure 3.21. As can be seen all structures demonstrate identical secondary structure characteristics in the form of $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$. Each of the RRM domains also display the $S_1$ $S_2$-strands located at the C-terminus of α-helix 2. Similarly, the RRM 2 domains of HUD and nucleolin both contain a loop3 similar to that seen in PABP-1 RRM 2. Interestingly, this loop is not as exaggerated in the RRM 1 domains of hnRNPA1, HUD and nucleolin.

In addition to the secondary structure similarities the RRM 2 domains of nucleolin and HUD show similar modes of RNA interaction. For example, nucleolins interaction with RNA involves stacking interactions between the aromatic rings of Y140-$\beta_3$ and Y169-$\beta_4$ and hydrophobic interactions with two exposed Leucine residues, L103-$\beta_1$ and L128-$\beta_2$ (165). These interactions provide similar anti-parallel conformation of protein and RNA as seen for PABP-1 RRM 2. HUD indicates that hydrophobic and aromatic interactions between the exposed aromatic rings of F170-$\beta_3$ and F202-$\beta_4$ and RNA allow for a similar conformation of RNA and protein as seen for PABP-1 RRM 2 and nucleolin (166). Deo *et al* 1999 reported stacking interactions within the PABP-1 RRM 1 and 2 domains. The RRM 1 domain of hnRNPA1 displays similar stacking interactions as those seen in the RRM 1 domain of PABP-1 between the aromatic rings of Y13-$\beta_1$ and F56-$\beta_3$ (167).

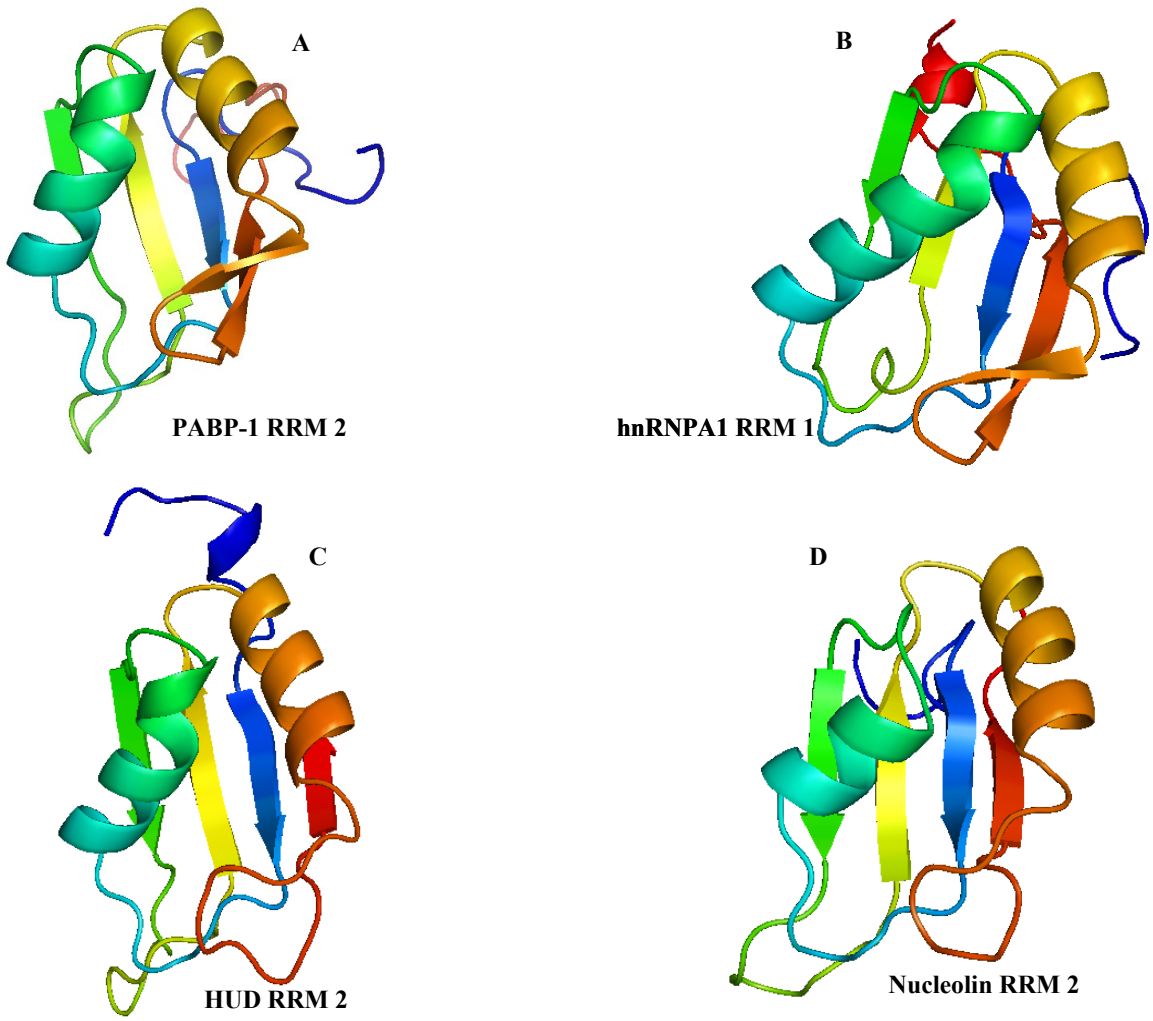



**Figure 3.21 Comparison of the RNA binding domains of PABP-1 RRM 2, hnRNPA1, HUD and Nucleolin.** This Figure illustrates the similarity between the RRM domains of different proteins. As was found for the structure of PABP-1 RRM 2 the RRM fold consists of $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$, this fold was also seen for the three protein examples. The structure of PABP-1 RRM 2 shows a high degree of similarity to the RRM domains of hnRNPA1, HUD and Nucleolin.

### 3.3 PABP-1 RRM 2 titration series

Biochemical results identified paxillin LD1 as a potential binding partner for PABP-1 RRM 2. With the structure of PABP-1 RRM 2 in its unbound state now known, a better understanding of their relationship can be obtained by NMR titration assays. Following a titration event by recording NMR spectra will enable the interacting residues to be identified by their chemical shift changes caused by the difference in the magnetic environment of the free and bound states. For these assays a synthetic paxillin LD 1 peptide was manufactured, supplied by Dr John Keyte, University of Nottingham. A synthetic peptide was used for the titration experiments as it was of high purity and had no additional residues attached as a result of protease cleavage. The peptide itself was fourteen residues in length and consists of the following amino acid sequence DD**LDALLADL**ESTT, the colored residues indicate the LD domain.

#### 3.3.1 Paxillin LD1 titration

A $^{15}$N labelled PABP-1 RRM 2 protein sample was prepared and its pH measured, it was important to make every effort to maintain the pH in order to prevent unwanted changes in HSQC signals due to an alteration in pH. An $^{15}$N HSQC spectrum was recorded at each of the following ratios of protein to peptide concentrations 1:1, 1:1.5, 1:2, 1:3, 1:4 and 1:6. The pH of the protein solution was adjusted back to the starting pH, in this case pH 6.5, after each addition of the peptide.

The chemical shift perturbations that were witnessed, Figure 3.22, appeared to be residue specific as no further resonances changed upon increased molar excess of paxillin LD1 peptide. The resonances that were perturbed during the titration continued to move as the concentration of peptide was increased. This indicates that the observed resonances were the population weighted average of the resonances in the free and bound states that is, the system was in fast exchange on the NMR timescale which is consistent with weak binding. The titration did produce changes in signals for a number of residues that were located at the flexible termini of PABP-1 RRM 2 for example G99 and G171, these may contact the peptide transiently due to the flexible nature of the termini. Upon inspection of the affected residues and their position on the NMR elucidated PABP-1 RRM 2 structure it was found that a binding interface exists across the β-sheet of PABP-1 RRM 2. Indeed, the interaction site shares a number of residues that are involved in poly (A) RNA binding specifically K138, Y140 and F142.

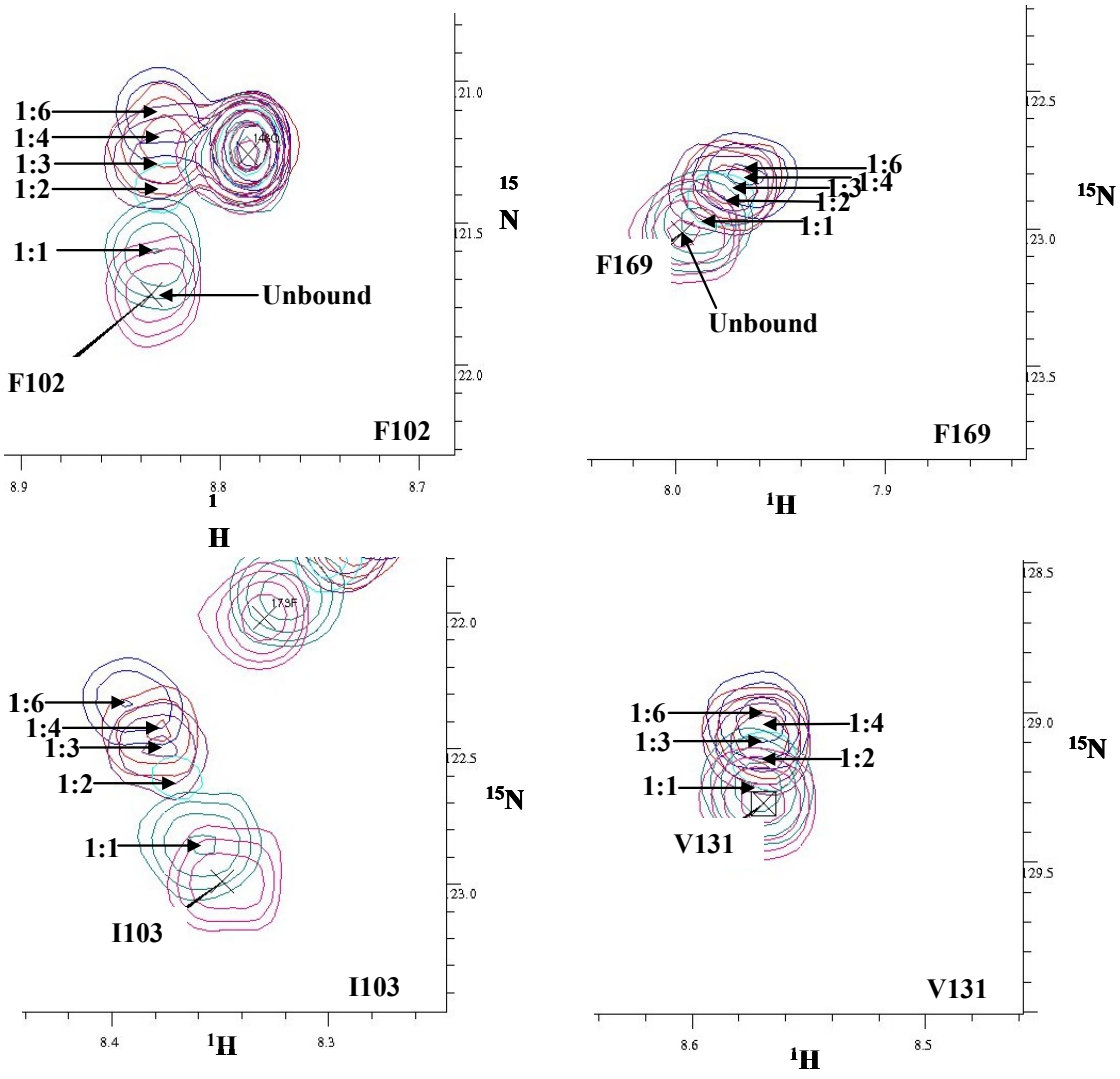



**Figure 3.22 Selected residues with chemical shift perturbations $^{1}H$ $^{15}N$ HSQC spectrum of a 15N labelled PABP-1 RRM 2 titration with paxillin LD1 synthetic peptide. A.** The titration was performed at a molar excess of 1:1 (blue), 1:2(teal), 1:3(purple), 1:4(brown) and 1:6(dark blue) paxillin LD1 peptide. The start position of the titration is labelled with the residue type and number (pink signal). This titration was

### 3.3.2 PABP-1 RRM 2 charge dependent binding

Analysis of PABP-1 RRM 2 residues that show chemical shift perturbations indicated that there were charged residues that complement those on the peptide and maybe involved in the formation of salt bridges. Analysis of these affected residues with the elucidated PABP-1 RRM 2 structure revealed that the interaction site is located across the solvent exposed face of the β-sheet. The presence of charged residues on the β-sheet of PABP-1 RRM 2 would suggest that the interaction of paxillin LD1 maybe electrostatically driven. Trials were performed on PABP-1 RRM 2 to establish the lowest NaCl concentration the protein can be placed in before precipitation occurs; this was found to be ~20mM.

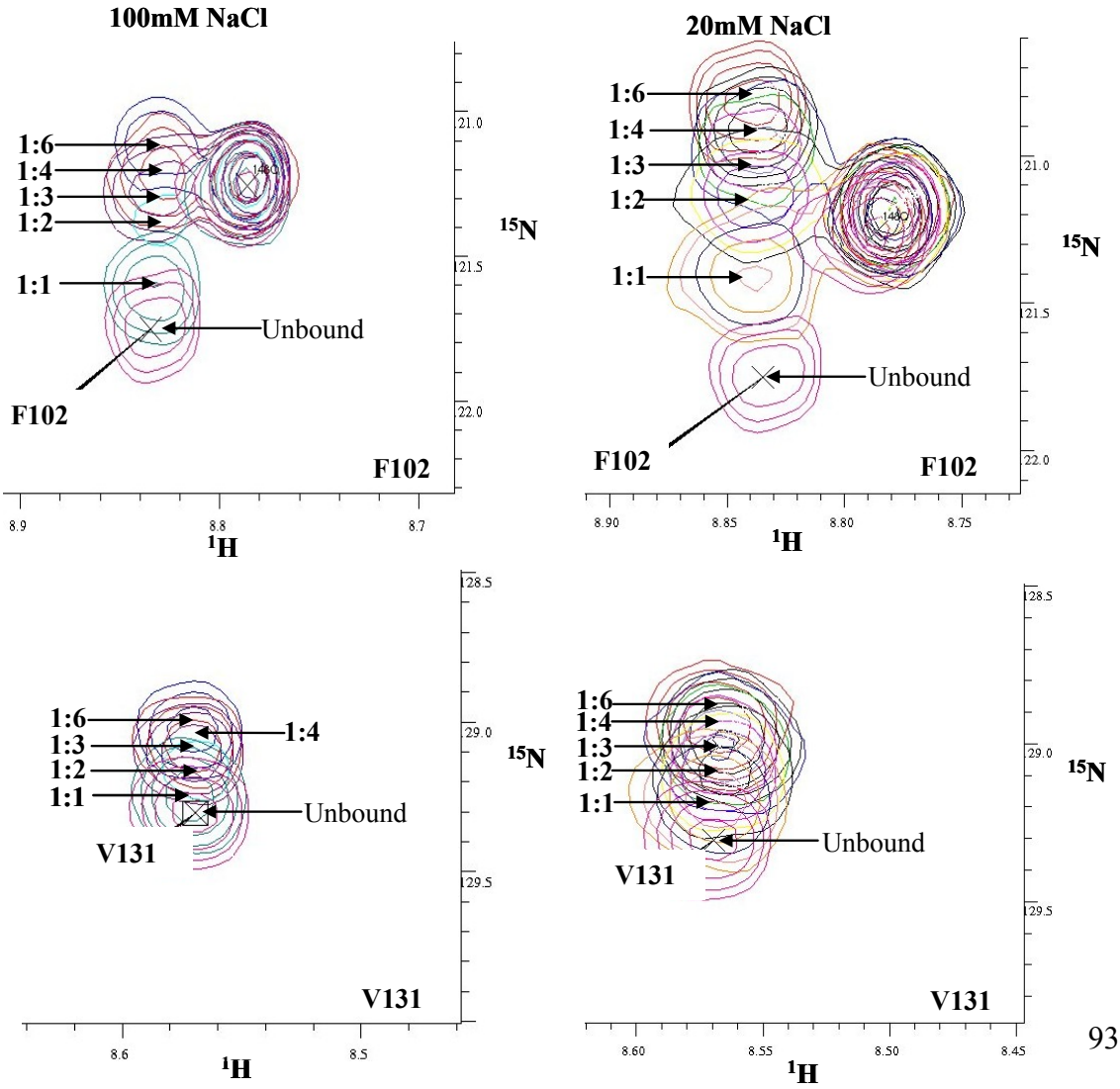
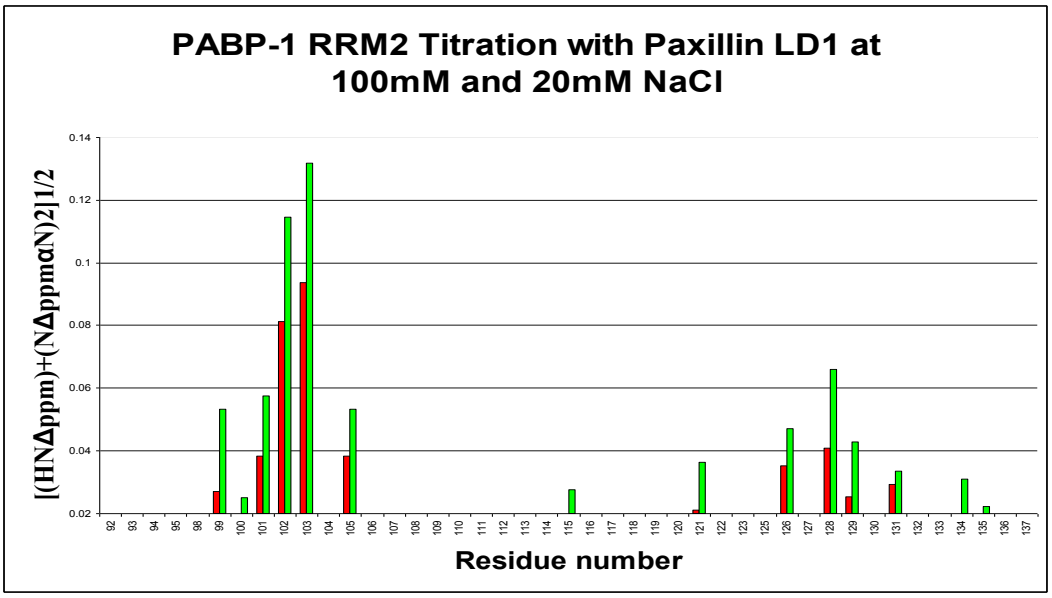


**Figure 3.23 $^{1}H$ $^{15}N$ HSQC spectrum of a $^{15}N$ labelled PABP-1 RRM 2 titration with paxillin LD1 synthetic peptide at 100mM and 20mM NaCl.** Six step titration series was repeated at 100mM NaCl (Figures on the left), and 20mM NaCl concentration (Figures on the right), the start position for each titration is labelled with the associated residue number.

The titration in 3.3.1 was carefully reproduced at 100mM NaCl and again at 20mM NaCl. Figure 3.23 illustrates the results of these titrations performed with identical conditions. An increase in chemical shift perturbations was observed at 20mM NaCl for each of the affected residues of the original titration series. These could be a direct result of changes in Kd or of changes in chemical shift perturbations. These chemical shift perturbations were calculated on a residue by residue basis and mapped graphically, Figure 3.24. The number of residues that display chemical shift differences from that of the free protein does not alter between NaCl concentrations moreover the amount of shift witnessed is greater at 20mM NaCl. The increase in chemical shift perturbations may indicate that the interaction is indeed electrostatically driven and by reducing the NaCl content increases the number of salt bridges formed between the positive charges of the protein and the negatively charged residues of the peptide thereby increasing the strength of the binding interaction.

$$\min \Delta\delta = \min \left[ \left( {}^{HN}\Delta_{ppm} \right) + \left( {}^{N}\Delta_{ppm}\alpha_{N} \right)^{2} \right]^{1/2} \qquad \textbf{3.1}$$

The chemical shifts shown in Figures 3.24 a and b and Figure 3.25 a have been scaled relative to their ppm scales. For example nitrogen ranges from ~99-135 and proton ranges from ~10-6 ppm therefore the measurements have been scaled down to minimise any over emphasis that may occur due to the difference in spectral width ranges. A scaling factor was introduced to account for these chemical shift differences in backbone $^{15}N$ relative to $^{1}H$, a $\alpha_N$ value of 0.14 was applied to all residues with exception to Glycine where $\alpha_N$ value of 0.2 was applied (168). The glycine residues are scaled differently due to their restricted nitrogen chemical shift range in comparison to the other amino acid residues. In equation 3.1 the $^{HN}\Delta_{ppm}$ and $^{N}\Delta_{ppm}$ values are the chemical shift changes observed in both the proton and nitrogen dimensions throughout the titration.



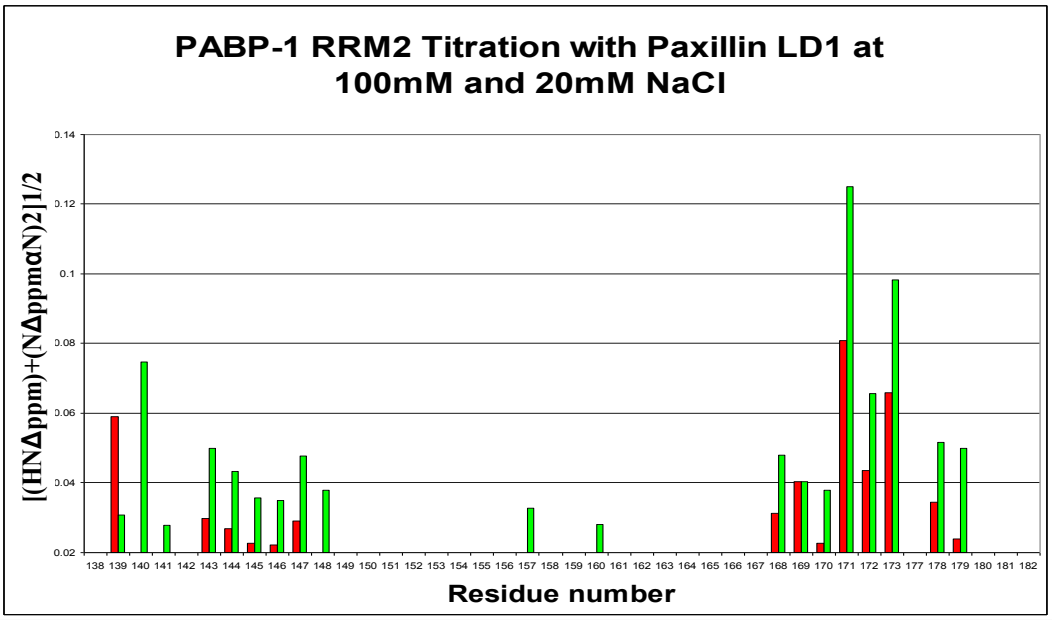




**Figure 3.24 Minimum chemical shift perturbations for paxillin LD1 titrated against PABP-1 RRM 2 at low and high NaCl concentrations.** A graphical representation of a titration at 100mM NaCl shown in **RED** and 20mM NaCl shown in **GREEN**. The nitrogen and proton chemical shifts have been calculated using the minimum chemical shift protocol: min $\Delta\delta$ = min $[(^{HN}\Delta_{ppm})+(^{N}\Delta_{ppm}\alpha_N)^2]^{1/2}$ where the $^{HN}\Delta_{ppm}$ and $^{N}\Delta_{ppm}$ are the $^{1}H_{HN}$ and $^{15}N_{NH}$ chemical shift changes respectively. A scaling factor was introduced to account for spectral

The chemical shift perturbation data for PABP-1 RRM 2/paxillin LD1 complex illustrates a specific binding event with a defined binding interface on the β-sheet. This data when combined with the filtered NOE data, described in paragraph 3.3.3, presented residues that were specific to paxillin LD1 binding. Figure 3.25A illustrates the binding of three of these residues which display well characterised chemical shift perturbations. Unfortunately it was not possible to reach the point of saturation in the titration series due to precipitation thereby making the determination of Kd difficult. The titration results illustrate the binding of paxillin LD1 to the RNA interaction site of PABP-1 RRM2 domain. When the on/off rate of exchange is faster than that of the NMR experiment then an average is detected by the system it can be said that this is in fast exchange. With each molar edition of paxillin it can be seen that the $^{1}H$ $^{15}N$ signals of the selected residues appear to move so the interaction is in the fast exchange regime. The position of the averaged resonance ($\nu_{obs}$) is weighted by the fractional populations of the free ($P_f$) and bound ($P_b$) states.

$$\nu_{obs} = P_f\nu_f + P_b\nu_b \qquad \textbf{3.2}$$

The bound population can be calculated using equation 3.3.

$$Pb = \frac{Mt + Lt + Kd - \sqrt{(Mt + Lt + Kd)^2 - 4MtLt}}{2Lt} \qquad \textbf{3.3}$$

For a titration in which the ligand (Lt) is added to the protein (Mt) the plotted difference between the observed and the free chemical shifts (Δν) against the ligand concentration then the chemical shift reaches a constant value at saturating concentrations of the ligand (Figure 3.25B residue I110). For weakly bound complexes, such as this, not every protein molecule will be bound when the concentrations of protein and ligand are equal. Therefore the observed protein resonance keeps moving until every protein molecule is bound ($P_b$ = 1). Figure 3.25b illustrates a fitted curve using a Kd value of $2.11 \times 10^{-3}$ M.

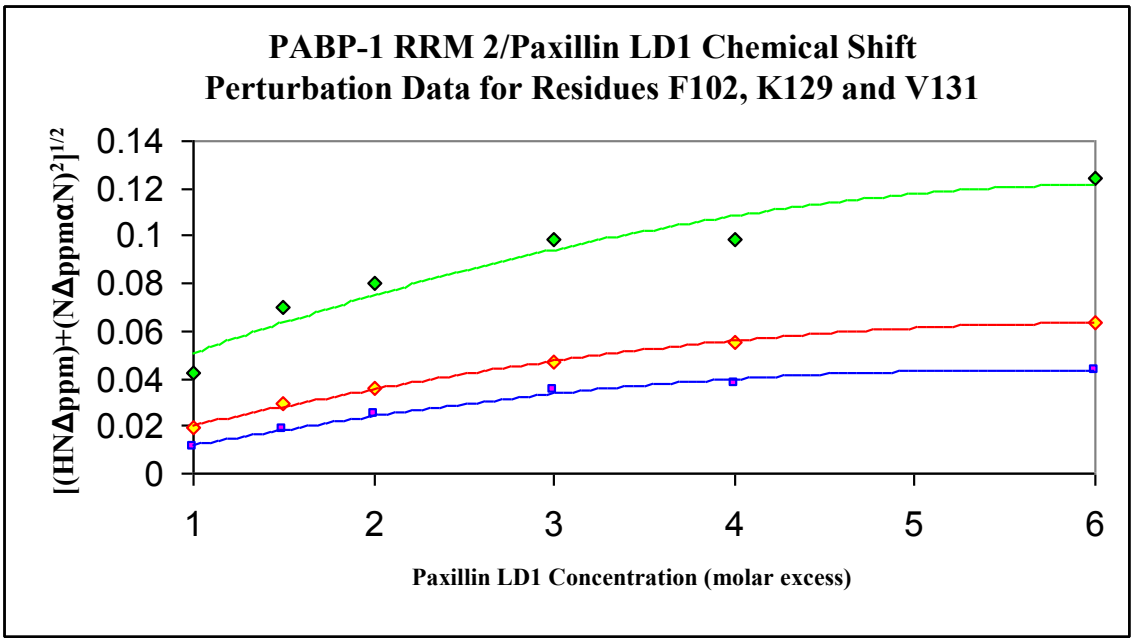
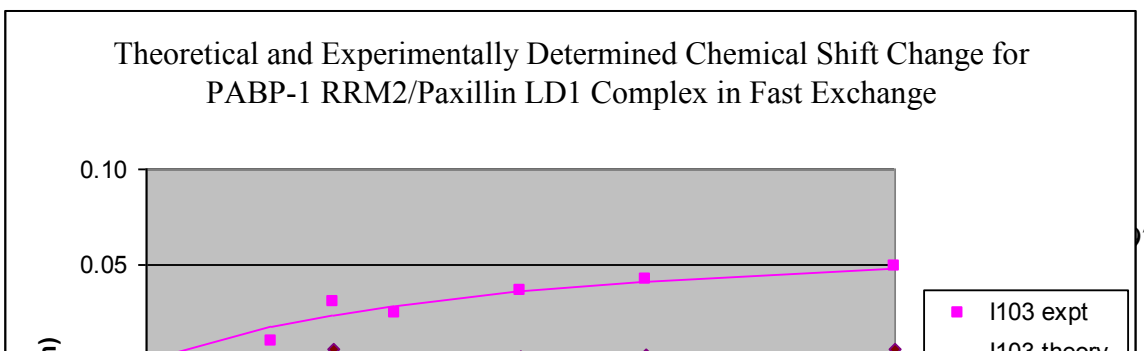



**Figure 3.25 Minimum chemical shift perturbations for paxillin LD1 titrated against PABP-1 RRM 2 residues F102, K129 and V131.** **A.** Graphical representation of a titration at 20mM NaCl shown in F102 (green), K129 (red) and V131 (blue). The nitrogen and proton chemical shifts have been calculated using the minimum chemical shift protocol: min $\Delta\delta$ = min $[(^{HN}\Delta_{ppm})+(^{N}\Delta_{ppm}\alpha_N)^2]^{1/2}$ where the $^{HN}\Delta_{ppm}$ and $^{N}\Delta_{ppm}$ are the $^{1}H_{HN}$ and $^{15}N_{NH}$ chemical shift changes respectively. A scaling factor was introduced to account for spectral width differences in backbone $^{15}N$ relative to $^{1}H$, a $\alpha_N$ value of 0.14 was applied to all residues with exception to Glycine where $\alpha_N$ value of 0.2 was applied. **B.** Change in protein chemical shift with change in ligand concentration. The curves were calculated using Mt=1mM, Lt=0-6mM, $\Delta\delta$=±0.15ppm and Kd=2.11x10$^{-3}$ M. Theoretical values were used to determine the fit of a binding curve to the data.



### 3.3.3 Filtered NOE experiments

A filtered NOE experiment was adapted from a published pulse sequence (169). Gradients were adapted to provide adequate water suppression and filters for a $H_2O$ based sample (Dr Igor Barsukov unpublished). Figure 3.26 illustrates the effectiveness of the filter, due to the weakness of the binding of paxillin LD1 a six molar excess was used for the filtered experiments. NOE signals were seen from previously identified contacting residues to the unlabeled paxillin LD1 peptide. Table 3.2 lists the contacting residues that display NOE connections to the peptide.

**Table 3.2 PABP-1 RRM 2 amino acid residues showing NOE signals to the paxillin LD1 peptide**

| PABP-1 RRM2 Residues | PABP-1 RRM 2 protein to peptide NOEs $^1$H-$^1$H NOEs |
|---|---|
| 95K | Hδ Hε Hγ |
| 102F | Hβ Hδ Hε |
| 129K | Hβ Hδ Hε Hγ |
| 131V | Hγ1 Hγ2 |
| 138K | Hδ Hε Hγ |
| 140Y | Hβ2 Hβ3 Hδ |
| 142F | Hδ1 |
| 173F | Hβ Hδ1 |

| | |
|---|---|
| 177K | Hδ Hε Hγ |

The paxillin LD1 synthetic peptide amino acid sequence displays a high degree of symmetry, DDLDALLADLESTT, making assignment of specific NOE signals difficult. To resolve this issue 2D TOCSY and NOESY experiments were performed on the peptide in its bound and unbound states. It was possible to only make ambiguous NOE assignments from a specific protein residue to either all of the leucine or aspartate residues of the peptide. It was possible to identify a specific NOE to the glutamate of the peptide thereby tethering the peptide to a specific protein residue in this case K177.

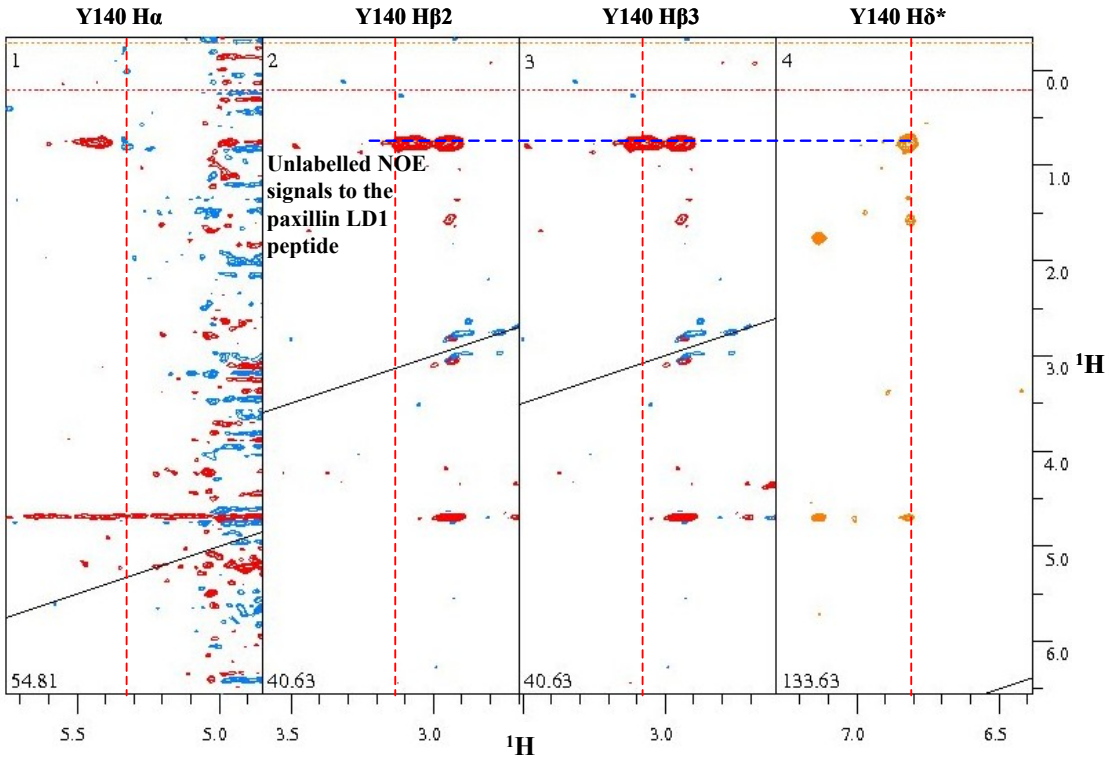



**Figure 3.26 $^{15}$N $^{13}$C double filtered experiment for PABP-1 RRM 2 in complex with paxillin LD1.** Paxillin LD1 was titrated in at 6 molar excess with filters applied to both the $^{15}$N and $^{13}$C channels. Spectral strips 1-3 show the aliphatic region of Y140 with NOE connectivity from the protein to the peptide in strip plot 3. The aromatic region of Y140 in strip plot 4 displays NOE connectivities to a Hδ leucine of paxillin LD1, given the number of leucines within this peptide unique assignments were not possible.

### 3.3.4 PABP-1 RRM 2/paxillin LD1 docking

The filtered NOE signals obtained on the complex in 20mM NaCl were ambiguous. Unique NOE assignments to D11 and T14 of the peptide made it possible to tether the ligand at one end. The peptide was docked to the PABP-1 RRM 2 using HADDOCK (High Ambiguity Driven protein-protein Docking), (170). The filtered NOE data presented in table 3.2 was used to determine AIR (Ambiguous Interaction Restraints) constraints of the active site in order to define the binding interface by mapping it onto the established PABP-1 RRM 2 structure PDB file (Figure 3.23). It was found that this produced a very small interface and so chemical shift perturbation data from the 20mM NaCl titration series, see Figure 3.27, was used to extend this interface. The passive site was then determined as those residues that indicated less significant chemical shift perturbations and were solvent accessible (>50%) neighbours of the residues of the active site.

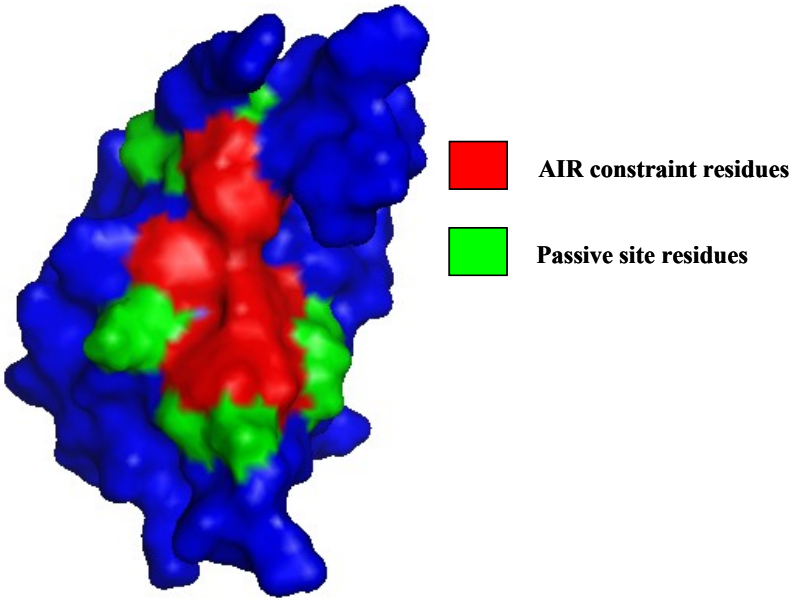



**Figure 3.27 Coloured surface plot of PABP-1 RRM 2 with active and passive sites.** AIR constraints were generated using information obtained from filtered NMR NOESY spectra and chemical shift perturbation data obtained from paxillin LD1 titration. Passive site residues were based on, the neighbouring residues of the AIR constraint active site, chemical shift perturbation data and >50% solvent accessibility.

The HADDOCK protocol generates a number of clustered structures that are grouped mainly through which conformations energetically favourable, other factors such as the number of structures are included within that particular cluster, the Buried Surface Area measurement and per-residue interaction energy analyses all must be considered when analysing the cluster that best fits the data. In the case of this complex, consideration was given to which cluster of

structures resulted in a favourable energies and surface area measurements but also which cluster best satisfied the filtered NOE data and chemical shift perturbation data. Figure 3.28 illustrates a secondary structure of the PABP-1 RRM2/paxillin LD1 complex and is the most energetically favourable model. The leucine residues of paxillin LD1 interact with the solvent exposed hydrophobic residues on the β-sheet of PABP-1 RRM2 forming a hydrophobic core. A number of charged aspartate residues on the peptide are able to form a number of salt bridges thereby promoting paxillins interaction with this site.

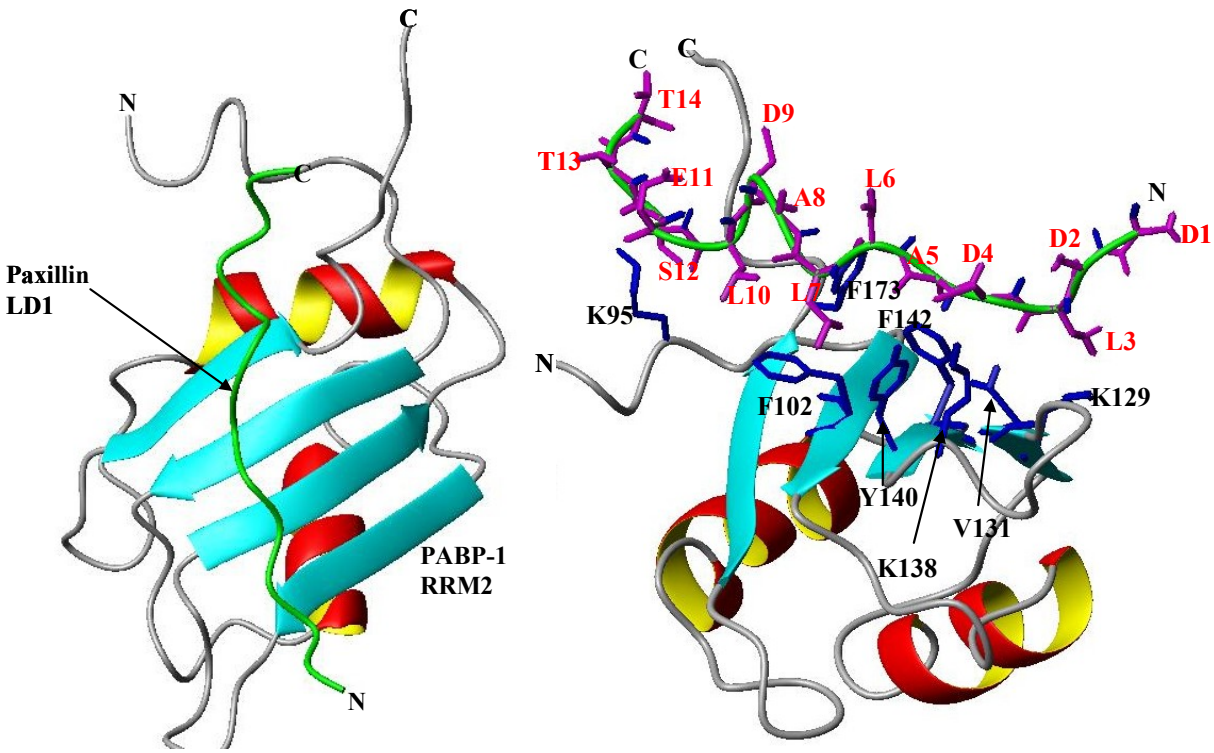



**Figure 3.28 Structure of PABP-1 RRM 2 in complex with paxillin LD1.** **A.** Paxillin LD1 synthetic peptide docked across the β-sheet of PABP-1 RRM 2. **B.** PABP-1 RRM 2 residues demonstrating NOE signals to the paxillin LD1 peptide are shown as blue side-chains. The position of paxillin LD1 side-chains (in purple) satisfies NOE and chemical shift perturbation data.

### 3.3.5 Poly(A) mRNA competition assay

The results on the PABP-1 RRM 2/paxillin LD1 complex have accurately identified the binding interface for paxillin LD1. This site is located across the β sheet which has been previously identified as the RNA binding interface. The predominant question is whether or not paxillin LD1 shares the site with the RNA or whether it competes for the site? To answer this, a manufactured 6'mer polyadenylate RNA oligonucleotide was created, supplied by Dharmacon Inc. Multiple molecules of PABP-1 bind to the polyadenylate tail of nascent mRNA to form a repeating unit, for this reason only a short polyadenylate tail was ordered to ensure only a single molecule of PABP-1 RRM 2 binds.

The weak binding nature of paxillin LD1 to PABP-1 RRM 2 made the investigation of a possible competition between poly(A) mRNA and paxillin LD1 difficult. All titrations were performed on the same day at pH 6.5 and a temperature of 303K. A $^{1}H$ $^{15}N$ HSQC titration was performed on PABP-1 RRM 2. The first titration was performed for poly(A) mRNA (6'mer oligonucleotide) against RRM 2 at 1:0.5, 1:1 and 1:1.5 molar equivalents of RNA (Figure 3.29). This titration was extended to 2 molar equivalents of RNA but the residue specific HSQC signals did not display any further shifts indicating that saturation had occurred at around 1.5mM poly(A) mRNA. The second titration was performed for paxillin LD1 against RRM 2 at 1:1, 1:4 and 1:6 molar equivalents of LD1 peptide. With the data generated for these two individual complexes an established titration pattern for both RNA and LD1 is known. However, due to the weak binding nature of paxillin LD1 it was not possible to reach a point of saturation therefore a concentration of 6 molar excess was chosen which was known not to cause protein precipitation.

The competition assay was first performed on the PABP-1 RRM 2-paxillin LD1 complex at 1:6 molar equivalents of LD1. RNA was titrated against the PABP-1 RRM 2-paxillin LD1 complex at a ratio of 1:0.25(poly(A)mRNA), 1:0.5, 1:1 and 1:1.5 the results of which can be seen in Figure 3.30. The next titration series was performed on PABP-1 RRM 2-poly(A) RNA complex at 1:1.5 molar equivalents of RNA to form the initial complex. Paxillin LD1 was titrated in at 1:1, 1:4 and 1:6 molar equivalents, the results of which can be seen on Figure 3.31.

When poly(A) mRNA was titrated into the established PABP-1 RRM 2-paxillin LD1 complex significant chemical shift movement away from the PABP-1-paxillin complex was seen,

Figure 3.30. Further additions of the RNA resulted in a majority of the PABP-1 RRM2-paxillin LD1 complexes being replaced by PABP-1 RRM 2-poly(A) mRNA complexes.

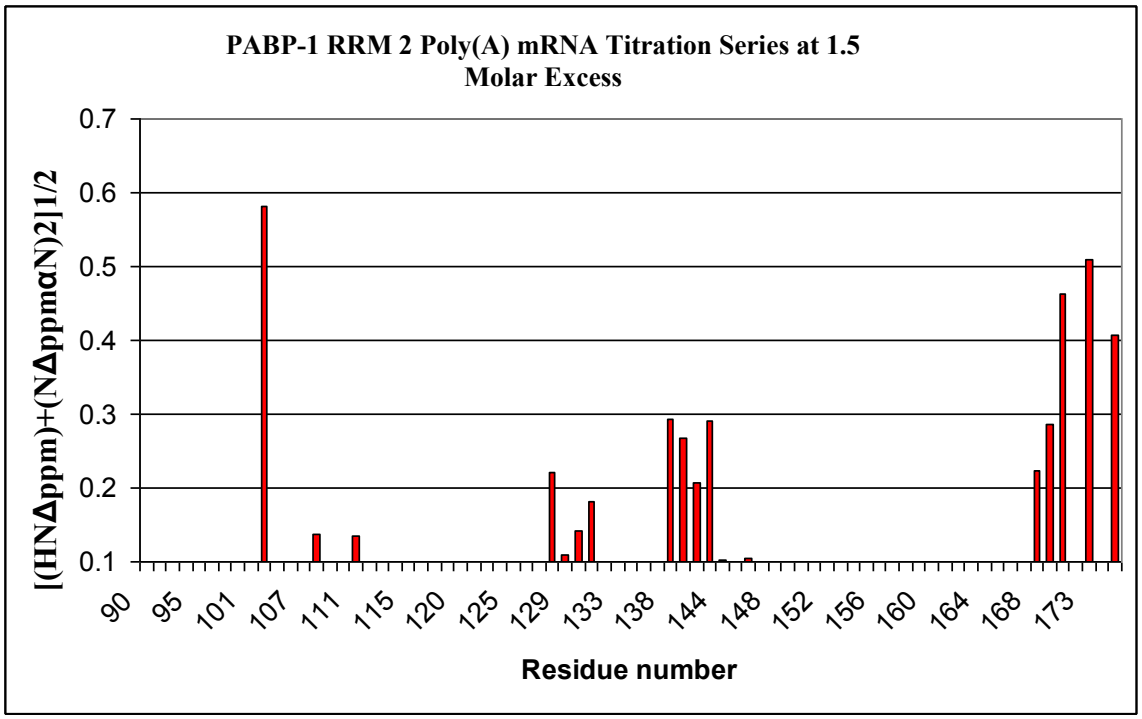



**Figure 3.29 Minimum chemical shift perturbations for Poly(A) mRNA titrated against PABP-1 RRM 2.** A graphical representation of a titration at 1.5 molar excess of poly(A) mRNA shown in **RED**. The nitrogen and proton chemical shifts have been calculated using the minimum chemical shift protocol: min $\Delta\delta$ = min $[(^{HN}\Delta_{ppm})+(^{N}\Delta_{ppm}\alpha_N)^2]^{1/2}$ where the $^{HN}\Delta_{ppm}$ and $^{N}\Delta_{ppm}$ are the $^{1}H_{HN}$ and $^{15}N_{NH}$ chemical shift changes respectively. A scaling factor was introduced to account for spectral width differences in backbone $^{15}N$ relative to $^{1}H$, a $\alpha_N$ value of 0.14 was applied to all residues with exception to Glycine where $\alpha_N$ value of 0.2 was applied.

The competition assay was repeated this time starting with a preformed PABP-1 RRM 2-poly(A) RNA complex it required 6 molar equivalents of the paxillin LD1 peptide in order for a significant disruption of the complex to occur. In contrast to the $^{15}N$ HSQC (Figure 3.30) the $^{15}N$ HSQC (Figure 3.31) for this assay revealed a mixture of PABP-1 RRM 2-paxillin LD1 and PABP-1 RRM 2-poly (A) RNA complexes present. These data would suggest that although poly(A) mRNA has a higher affinity for this binding site paxillin LD1 is still able to interact across the β-sheet of PABP-1 RRM 2, albeit in a weak almost transient way. Of course it can not be excluded that the shifts observed upon addition of the peptide to the protein-RNA complex reflect a secondary binding site for the peptide.

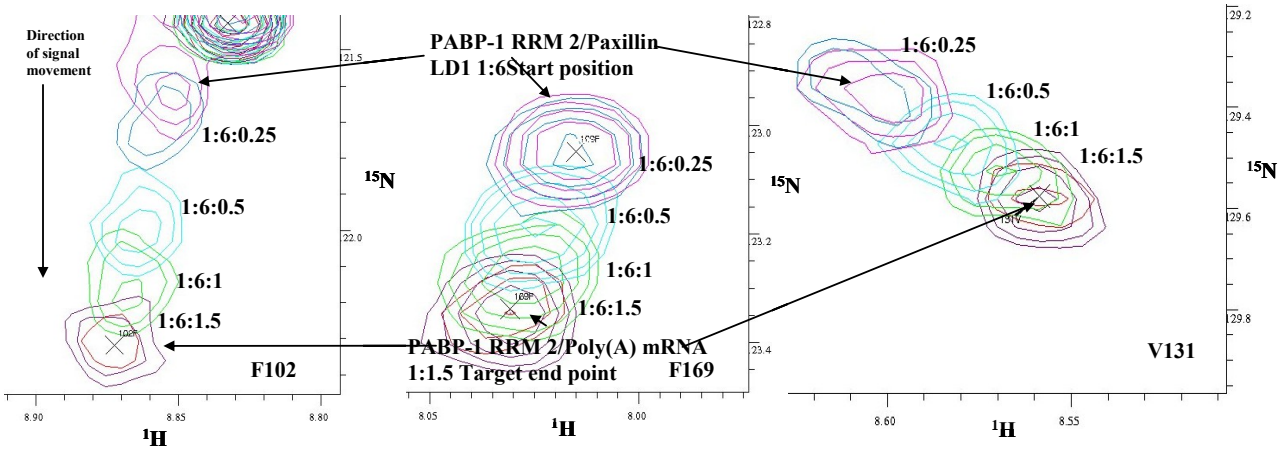



**Figure 3.30 above $^{1}H$ $^{15}N$ edited HSQC PABP-1 RRM 2/paxillin LD1 complex with poly (A) mRNA tail titrated in at1:6:0.25, 1:6:0.5, 1:6:1 and 1:6:1.5.** RRM2/LD1 complex was the starting point at a protein:peptide concentration of 1:6 and is marked, first point RNA titration was 0.25 molar equivalents. Second point RNA titration was 0.5 molar equivalents, third point RNA titration was 1.0 molar equivalents, and a final point of 1.5 molar equivalents. Expanded regions illustrating F102, F169 and V131 of PABP-1 RRM 2, titrated RNA in complex. Start and target end points show additional signals these correspond to overlaid spectra of individual protein:peptide and protein:poly(A)mRNA titrations to ensure accuarate referencing of spectra.

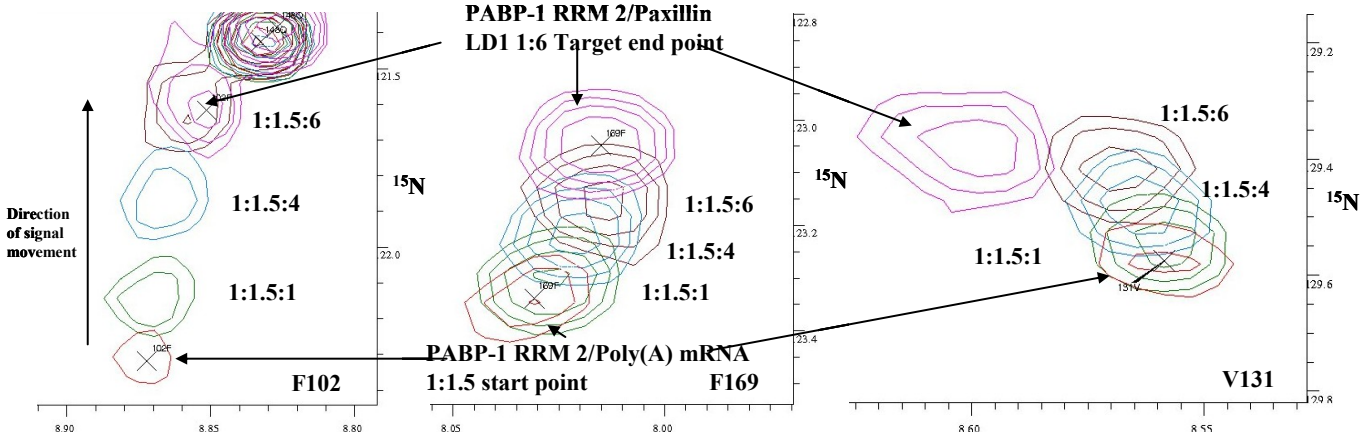


**Figure 3.31 $^{1}H$ $^{15}N$ edited HSQC spectra for PABP-1 RRM 2/poly (A) mRNA tail with paxillin LD1 titrated in at 1:1, 1:4 and 1:6 molar equivalents.** **A.** RRM2/RNA starting point is 1:1.5 molar equivalents. The first LD 1 titration point was 1:1.5:1 molar equivalents, second titration point was 1:1.5:4 molar equivalents. The third LD1 titration point was 1:1.5:6 molar equivalents. Expanded regions for residues F102, F169 and V131 of PABP-1 RRM 2, F102 and F169 were found to bind mRNA directly (6) whilst V131 appears to be unique to paxillin LD1 interaction. Start and target end points show additional signals these correspond to overlaid spectra of individual protein:peptide and protein:poly(A)mRNA titrations to ensure accurate referencing of spectra.

### 3.4 Conclusion

The elucidated structure gave a typical RRM domain conformation of $\beta_1$ $\alpha_1$ $\beta_2$ $\beta_3$ $\alpha_2$ $\beta_4$. This fold is similar to those seen in Nucleolin, HNRNP and HUD. The two α-helices are packed against the four stranded anti-parallel β-sheet. A hydrophobic core is located between the α-helices and the β-sheet. F145 is centrally located within the hydrophobic core providing multiple NOE signals to the β-strands and the α-helices. Hydrophobic interactions between the helices facilitate their close proximity allowing for possible polar interactions between the helices and the β-sheet. NOE data indicates that several positive residues extend outward away from the helix 2, in general helix 2 displays a positive charge when compared to helix 1. The NMR derived unbound structure revealed the F102, F142 and F169 are present on the β-sheet. The experimental data demonstrated that their aromatic rings extend outward away from the face of the β-sheet providing a hydrophobic interface for possible ligand binding. The NOE data was used to predict hydrogen and Van der waals interactions, based upon residue distance, which further support the structure. PABP-1 RRM 2 structure presents a mobile loop between the $\beta_2$ and $\beta_3$-strands. This loop also appears in other RNA binding proteins making it a common feature for RRM domains. Dispersed across the surface of the β-sheet and β-hairpins are positive charged residues. It is the hydrophobic and charged residues that are involved in the stacking interactions in RNA binding to PABP-1. These residues also compliment the largely hydrophobic and electronegative characteristics of

paxillins LD 1 domain unlike those residues of the proposed binding site. Indeed closer inspection of the PABP-1 RRM2 amino acid sequence revealed that it bears no similarity to the PBS motif identified in other paxillin binding partners. It is not to say that the proposed PBS motif be discounted it merely does not play a functional role within this particular protein domain. Moreover this finding further complicates the identification of paxillin's binding partners.

The biochemical data presented at the beginning of this chapter have shown the PABP-1/paxillin interaction to be much more complicated. The evidence indicated that an interaction site existed for paxillin within PABP-1 RRM2 domain but also within the PAPB-1 RRM1. The absence of a conserved PBS motif within PABP-1 RRM2 is surprising and further complicates the characterisation of paxillins binding partners with the knowledge that paxillin is able to bind to sites that are devoid of this motif. The C-terminal region of PABP-1 was able to bind both LD2 and LD5 domains of paxillin. Similarly the C-terminal region of paxillin gave a positive indication of binding to PABP-1's RRM4 domain.

NMR titration results with a synthetic paxillin LD1 peptide revealed residue specific chemical shift perturbations. The further addition of paxillin LD1 peptide resulted in increased chemical shift perturbations but did not induce changes in the HSQC signals from any additional residues. A titration performed at low NaCl concentrations on PABP-1 RRM 2 resulted in increased chemical shift perturbations. These data confirmed that the complex maybe electrostatically driven. Filtered NOE experiments gave NOE data from the protein to the peptide this data was supported by the assignment of the protein in its bound state. Further NOE data was collected on the peptide itself allowing for the successful docking of the peptide to PABP-1 RRM 2 using HADDOCK. The binding interaction appears to be in the fast exchange regime with a calculated Kd ~2.11mM. The weak binding nature of this complex would indicate the necessity of further sites of paxillin interaction along the length of PABP-1.

Some of the residues involved in the binding of paxillin LD1 were also involved in the binding of the poly (A) mRNA tail. This finding naturally presents the question of competition for the binding interface or a cooperative binding. The NMR competition assay revealed that both paxillin LD1 and the poly (A) mRNA tail compete for the binding interface on PABP-1 RRM 2. Although the RNA binds to PABP-1 RRM 2 with greater affinity,

paxillin LD1 was able to push the RNA oligonucleotide out of the binding interface upon addition of six molar equivalents of the peptide paxillin LD1. Similarly the RNA was able to completely push the peptide out of its binding site with only a 1:1.5 molar excess of RNA. The PABP-1/mRNA crystal structure presented by Deo *et al*, (142), illustrated the specificity of PABP-1 for the poly(A) tail of mRNA. The competition data would indicate that the competition for this active binding site is biased toward polyadenylated mRNA. However it would be interesting to examine the effect on cell migration by mutating of the specific residues involved with paxillin binding. The weak interaction of this complex, as previously mentioned, would indicate the presence of additional binding sites therefore if this particular site was a primary interaction site then cell migration would be greatly reduced.

# Chapter 4 PABP-1 RRM 1

## 4.1 Introduction

PABP-1 RRM 1 provides a binding interface for the poly-adenylated tail of nascent mRNA and is an active proponent involved in PABPs interaction with eIF4G. The mRNA tail was shown to specifically interact with the RRM 1 domain residues Y14, Y54, Y56 and R89. Residues M85 of PABP-1 RRM 1 and R94 from an inter-domain linker of PABP-1 RRM1/2 and are involved with stacking interactions with Ade-5 of an 11'mer poly(A) mRNA. Woods *et al* 2005 identified one such domain, bearing similarity with the PBS domain of actopaxin, within the N-terminus of PABP-1 RRM 1 residues D17-K30. Woods *et al* 2005 performed mutational assays on this region, D21-R21 and T23-N23. Their results indicated that a reduction in cell migration occurred when this mutant was introduced into migratory fibroblasts (Chapter 1).

The ELISA results in Chapter 3 identified a binding site for paxillin LD1 peptide within the PABP-1 RRM 1 domain, although the binding affinity was less than that seen in the RRM2/LD1 interaction. Given the roles played by the RRM 1 domain in its interaction with eIF4G and mRNA, it is considered a biologically important domain. To further investigate this interaction RRM1 was expressed and labelled with both $^{15}N$ and $^{13}C$.

## 4.2 Results and discussion

### 4.2.1 PABP-1 RRM 1 construct

PCR primers were designed to encompass the RRM 1 domain. The GC-content of both the forward and reverse PCR primers was kept similar to one another. Figure 4.1 is a representation of the RRM1 PCR primers; the primers contain an NCO-I restriction enzyme site in the forward primer and a KPN-I site in the reverse primer upstream from the 'stop' codon.

Forward primer

NCO-I Start

5'-CAT GCC ATG GGT ATG AAC CCC AGT GCC CCC AGC TAC CCC ATG GCC TCG CTC TAC GTG GGG-3'

Reverse primer

KPN-I Stop

5'- TGG TCT CAG CGT GAT TAA GGT ACC CTC GAG CGG-3'

**Figure 4.1 Representation of the PABP-1 RRM 1 PCR primer design.** The forward primer was extended in order to facilitate the removal of a second NCO-I site located downstream from the start codon. The engineered restriction enzyme sites are shown in red and the start and stop codons identified.

PABP-1 RRM1 nucleotide sequence was analysed using the New England Biolab restriction enzyme cutter (www.neb.com). The results showed that an NCO-I site was located downstream of the start codon and would have resulted in the removal of the first 10 N-terminal amino acid residues. The forward primer shown in Figure 4.1 was extended to include this NCO-I site and introduce a point mutation thus removing the site but maintaining the nucleotide sequence of the amino acid.

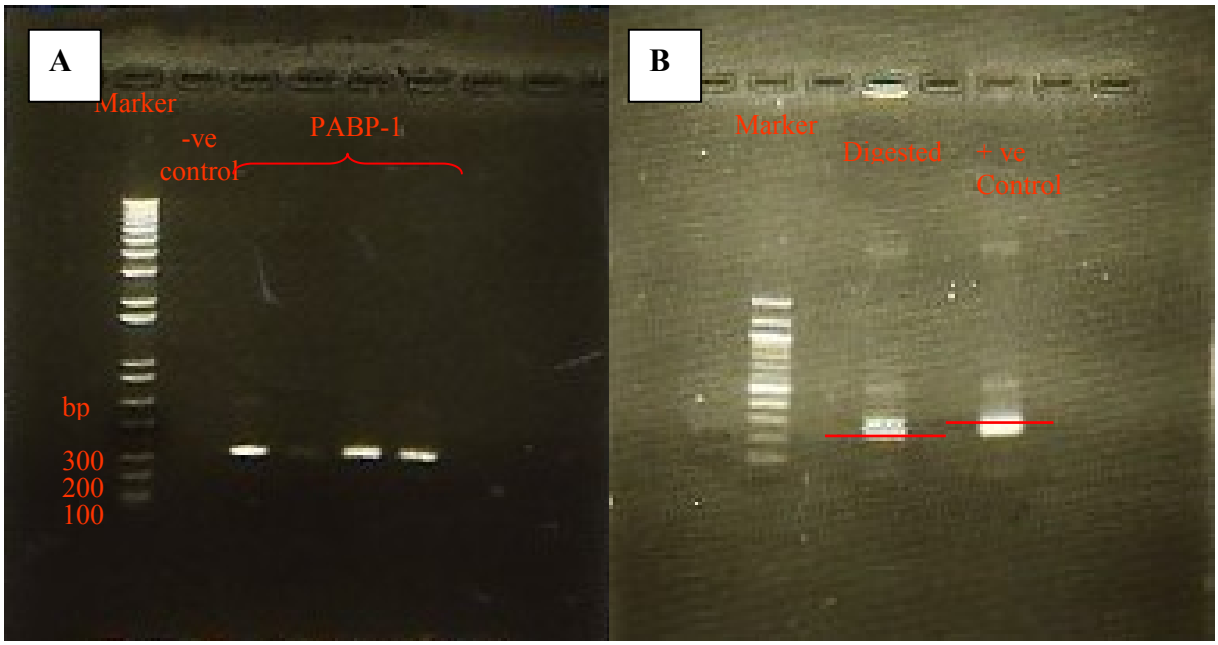



**Figure 4.2 2% agarose gel of the PCR and enzyme digest products.** **A.** PABP-1 RRM 1 PCR products (303bp) were purified by 2% agarose gel and the products treated with NCO-I/KPN-I restriction enzymes. **B.** Reduction in fragment size following the enzyme digest to give the PCR product adhesive ends prior to ligation.

### 4.2.2 Protein expression

Competent BL21 DE3 cells were transformed using the RRM1 constructs and expressed. Cell cultures were induced at an optical density of 0.7 and a temperature of 37$^{o}$C. Protein expression concentrations were very low at this temperature even at varying inductions times. Figures 4.3a, b and c shows the FPLC traces for PABP-1 RRM 1 expression trial using 500ml enriched media cultures at varying induction temperatures. The FPLC traces clearly demonstrate the improvement of protein expression at lower induction temperatures. An optimum induction temperature of 15$^{o}$C was found to provide the maximum protein yield for PABP-1 RRM 1 in M9 minimal media.

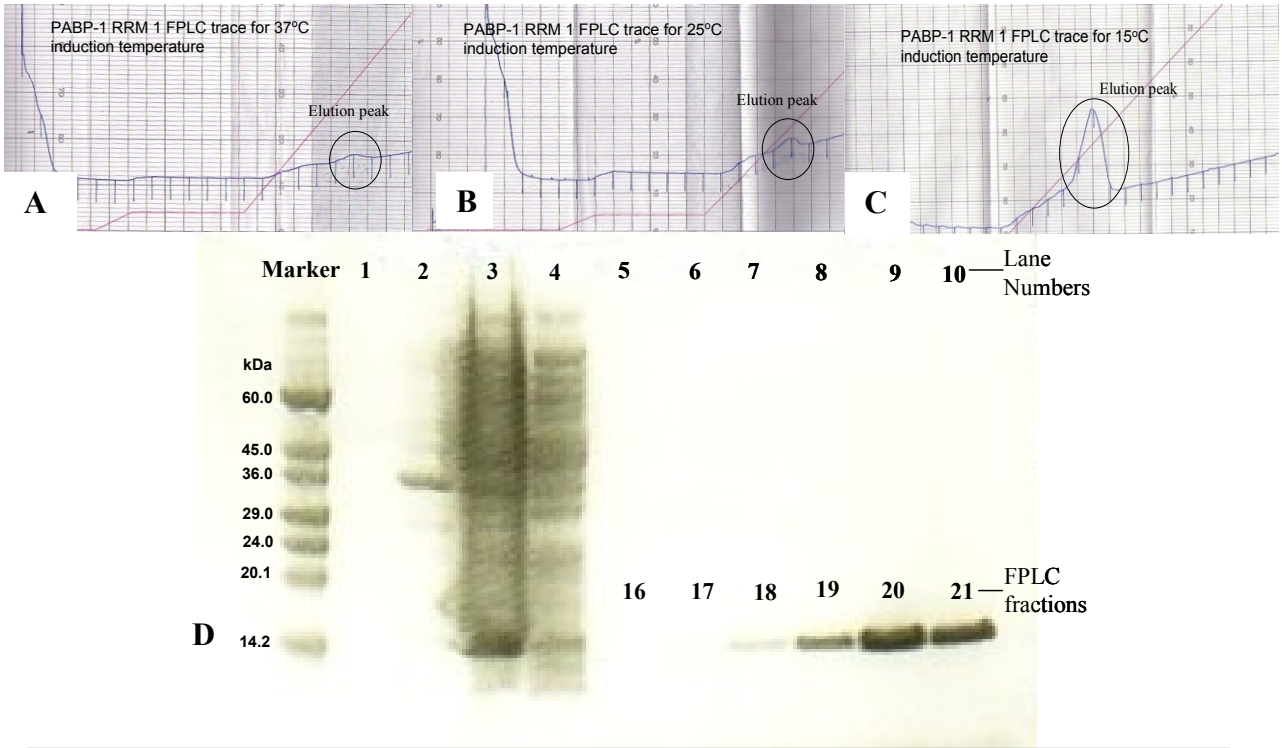



**Figure 4.3 Temperature dependent protein expression trials for PABP-1 RRM 1. A.** FPLC trace for an overnight induction at 37°C in 0.5l of M9 media. **B.** FPLC trace for an overnight induction at 25°C in 0.5l of M9 media. **C.** FPLC trace for an overnight induction at 15°C in 0.5l of M9 media. **D.** PABP-1 RRM 1 protein bands following an overnight induction at 15°C. Lane 2 the uninduced whole cell fraction, lane 3 lysate fraction following cell lysis and lane 4 lysate flow-through following purification. Lanes 5 through 10 PABP-1 RRM 1 FPLC fractions.

### 4.2.3 PABP-1 RRM 1 sequence specific assignments

Sequence specific assignments were carried out using CCPN software. Figure 4.4 illustrates an $^{15}N$ resolved HSQC spectrum displaying the backbone NH signals and side chain $NH_2$ group signals. The HNCACB and CBCA(CO)NH spectra (Figure 4.5A) show excellent signal intensity and very few artefacts. Both spectra were collected over a 72 hour period and their high quality was due to the size of the protein and the relatively high protein concentration. Indeed, it can be seen in this example that the signals in the HNCACB spectrum show both i and i-1 signals. The i-1 sequential assignments were confirmed using the CBCA(CO)NH spectrum. Sequential assignment of the carbonyl carbons were facilitated

by collecting HNCO and HN(CA)CO spectra. The connectivity of the amide sequential assignments was confirmed using the HNCO and HN(CA)CO spectral data (Figure 4.5B). The HNCO spectrum produced i-1 while the HN(CA)CO provided data to the intra-residue. All of the expected signals were observed even in the less sensitive experiments.

Backbone NH, Cα and Cβ residue assignments were 96% complete with no detectable signals to the three C-terminal residues. Cα and Cβ protons were assigned using HBHA(CO)NH and HBHANH spectra. Both experiments gave a good signal to noise ratio with no spectral overlap. As Hα and Hβ were of opposite phase in the HBHANH spectrum, assignment was relatively straight forward.

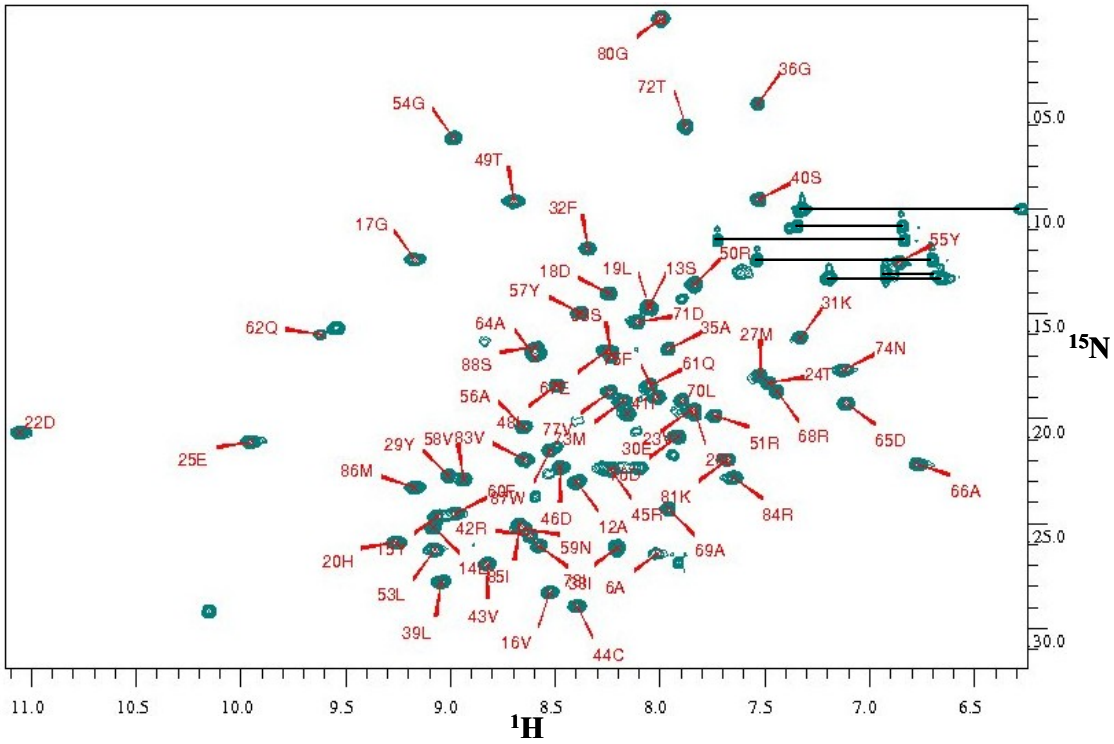



**Figure 4.4 $^{1}H$ $^{15}N$ HSQC spectrum of PABP-1 RRM 1.** The spectrum shows a pure sample with no proteolysis, signals show good dispersion that is indicative of a folded protein. Sequential numbering is that of human PABP-1 SwissProt accession number P11940.

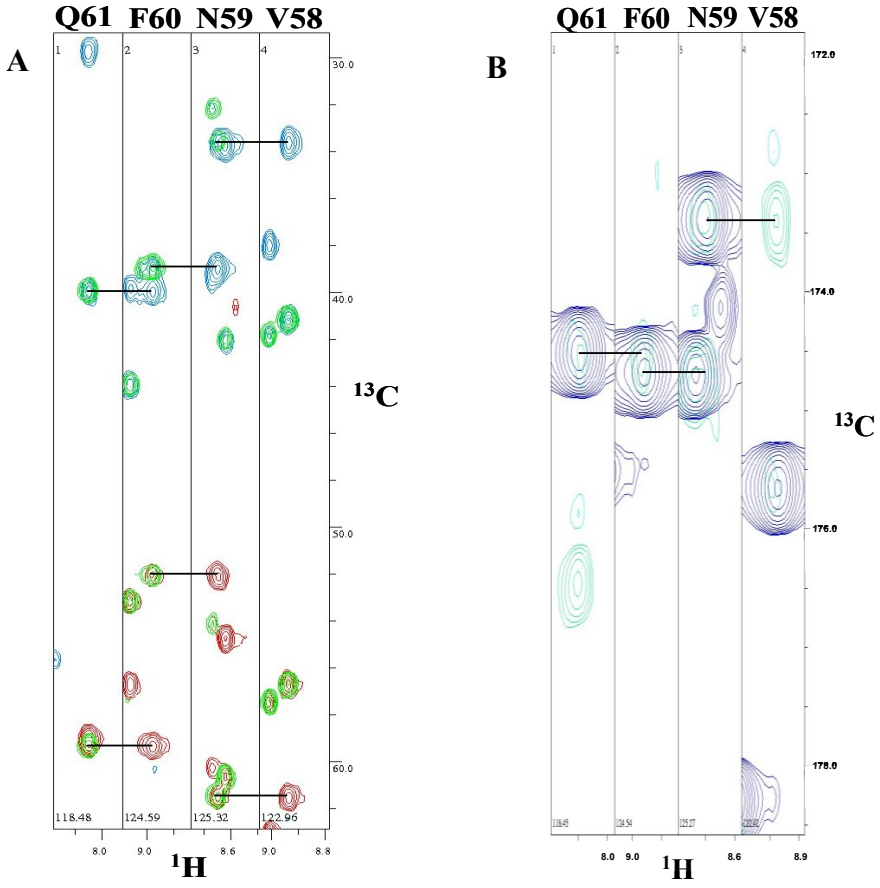



**Figure 4.5 Selected strips from the triple-resonance spectra illustrating sequential connectivities for residues 102-105 of PABP-1 RRM 1.** **A.** Superposition of CBCA(CO)NH and HNCACB spectra. Positive contour level of CBCA(CO)NH are shown in dark green ($CB_{i-1}N_iHN_i$ and $CA_{i-1}N_iHN_i$ correlations); positive contour level of HNCACB are shown in red ($CA_iN_iHN_i$ correlations), negative contour level of HNCACB are shown in blue ($CB_iN_iHN_i$ correlations). Sequential connectivities are represented by horizontal lines. **B.** Superposition of HNCO and HNCACO spectra. Positive contour level of HNCO shown in blue ($CO_{i-1}N_iHN_i$ correlations); positive contour level of HNCACO are shown in light green ($CA_iCO_iN_iHN_i$ correlations). Sequential connectivities are represented by horizontal lines.

### 4.2.4 Side-chain assignments

All side chain assignments were performed in a logical and progressive manner building on to the already assigned amide backbone. Assignments to the $C_\beta$ and its accompanying protons were carried out in a similar manner as PABP-1 RRM 2 (chapter 3) thus providing a bridge

between the amide backbone and side chain assignments. Figure 4.6 illustrates a $^{1}H$ $^{13}C$ resolved HSQC spectrum displaying expanded $C_{\alpha}$ and methyl regions. Signals to the side chains located within overlapped regions were resolved using the [H]CCH-TOCSY and H[C]CH-TOCSY spectra. Figure 4.7B illustrates selected strips from the $^{1}H$ resolved TOCSY spectrum for I48. The $H\gamma_1$ strip is a good example were the two protons are resolved and their assignments given.

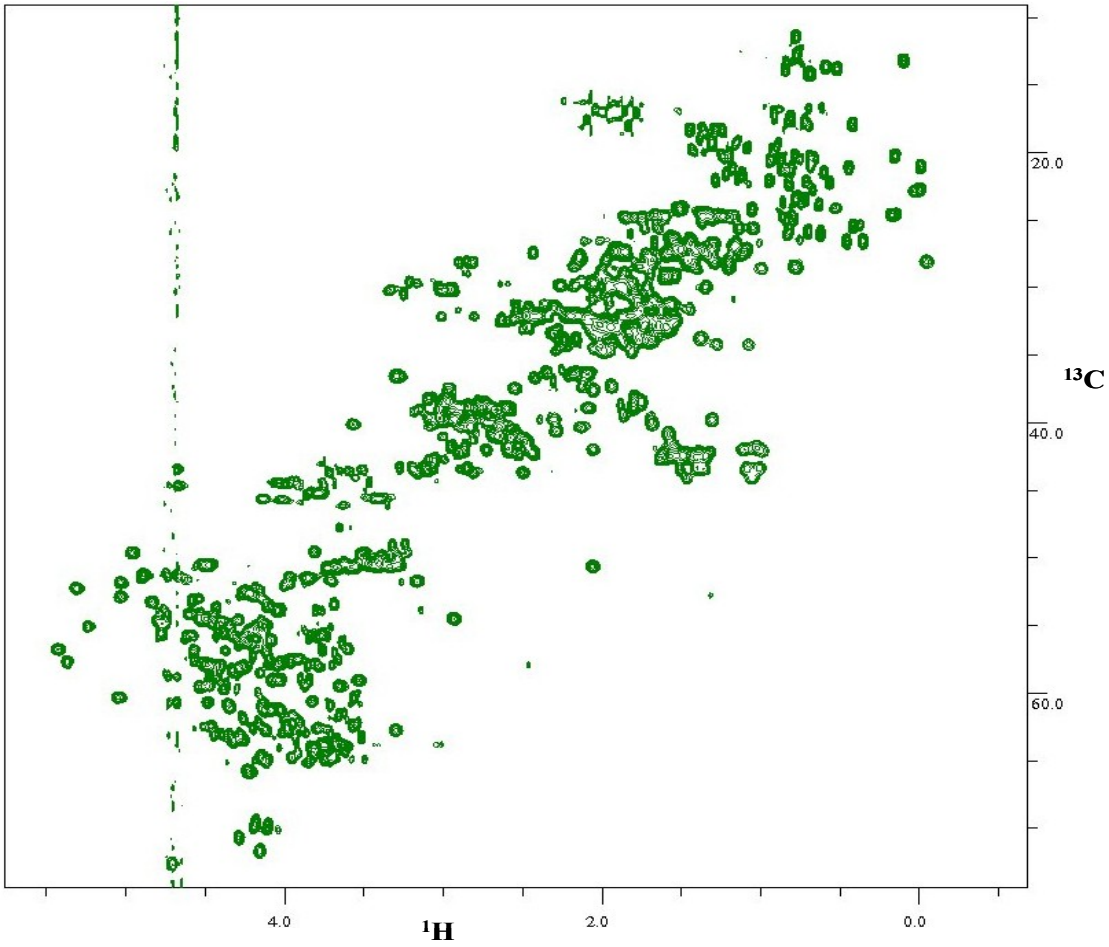



**Figure 4.6 $^{1}H$ $^{13}C$ HSQC spectrum side chain assignments for PABP-1 RRM 1.** The spectrum is well resolved with minimal spectral overlap of signals. Side chain residues were assigned in this spectrum via $^{1}H$-TOCSY and $^{13}C$-TOCSY spectra.

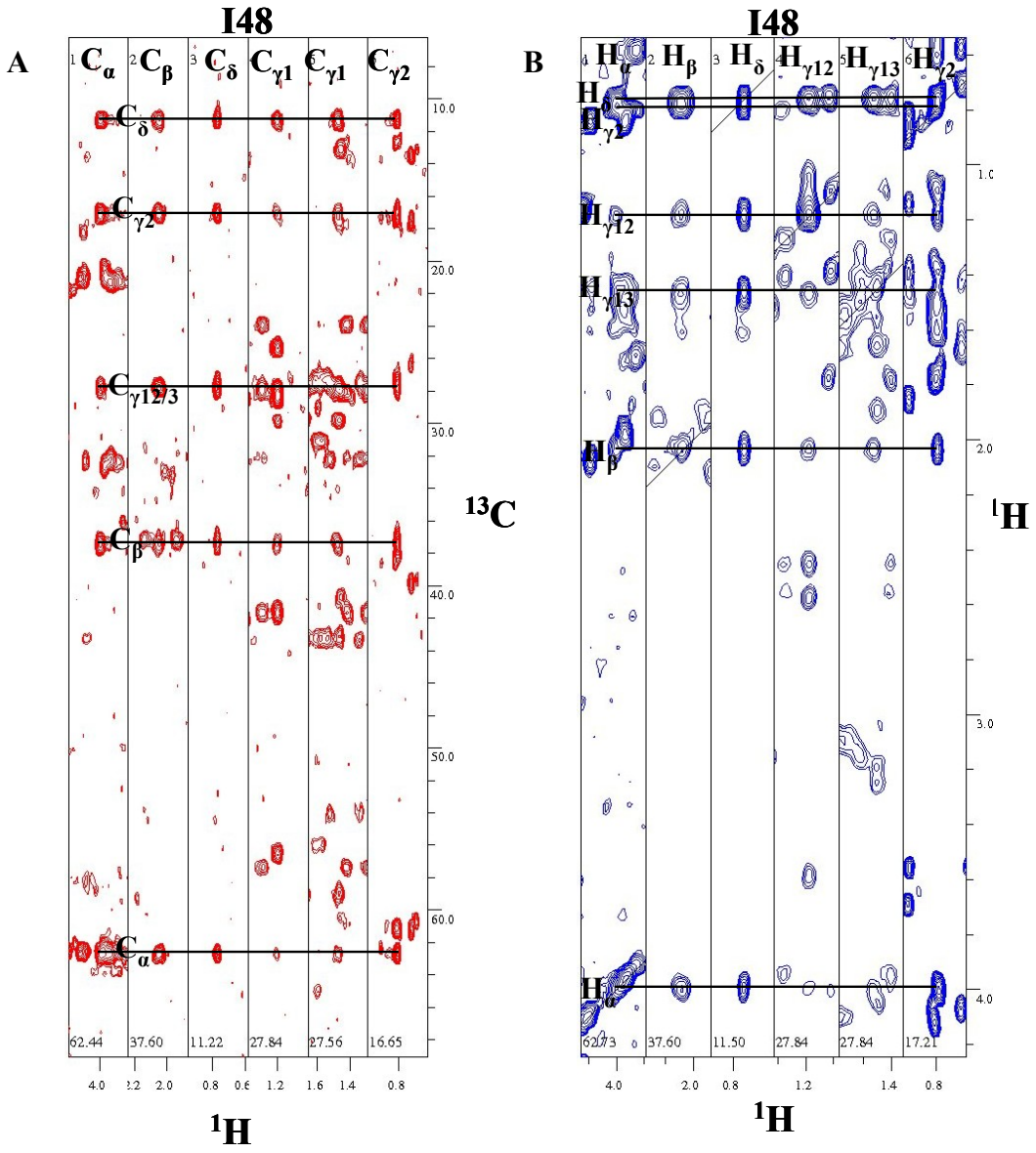



**Figure 4.7 Selected strips from the $^{13}$C and $^{1}$H resolved TOCSY spectra, intra-residue side chain connectivities for I48 of PABP-1 RRM 1. A.** Side-chain carbon atom positions, shown as red signals, for I48. The carbon signals can be seen within each plane. **B.** Side-chain carbon protons, shown as blue signals, for I48, each resolved proton signal can be seen in each strip.

### 4.3 PABP-1 RRM 1 structure

Initial structural calculations were performed using CYANA/CANDID protocol version 2.1 for the determination of unique assignments of NOE derived data. Dihedral constraints generated by TALOS were also included in the calculations. Initial calculations derived a folded structure with 89% of NOE signals assigned using a chemical shift tolerance for $^{1}H$ 0.03 and 0.3 for both $^{15}N$ and $^{13}C$. Unassigned NOE signals were examined to ensure there were no systematic errors. They were then examined to determine whether they were from another residue that had not been assigned by the calculation protocol. The NOE data was then refined by referring to the structure and used to generate constraint tables for ARIA1.2 where the final structural water refinements took place. Structural refinements took into account any raised NOE energies that were highlighted as part of the ARIA output by cross referring to the group of signals or angle constraints and correcting the violating constraint. Table 7.1 lists the final restraints generated from 20 lowest energy water refined structures derived from 100 random start structures.

The final water refined lowest energy conformers, Figure 4.9, were analysed by Procheck-NMR (163). Figures 4.8 and 4.9 show the global fold to be $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$, this type of fold is seen in other RRM domain proteins which will be discussed later in the chapter. Figure 4.8(A) illustrates the RMSD analysis on a residue by residue basis for the amide backbone (black) and the side-chain (grey). High local RMSD values can be seen at the N and C-termini ranging from 2-5Å as expected. A rise in RMSD values can be seen between V43 and S52 this would indicate a certain amount of dynamic motion within the loop 3 region. Figure 4.8(B) illustrates the number of NOE constraints on a residue by residue basis. It can be seen that low NOE constraints is accompanied by a rise in RMSD values as can be seen for the N and C-termini

**Table 4.1 Solution Structure Determination of PABP-1 RRM 1**

| Experimental restraints | |
|---|---|
| Restraints | |
| Unique NOEs | 1817 |
| Intraresidue | 771 |
| Sequential | 367 |
| Medium range (l< [i-j] < 4) | 183 |
| Long range ([i-j] > 5) | 496 |
| φ/ψ dihedral angles[a] | 108 |
| Energies (Kcal $mol^{-1}$)[b] | Mean ± S.D |
| Total | -2948.58 ± 72.97 |
| Van Der Waals | -741.88 ± 8.99 |
| NOE | 70.20 ± 0.0023 |
| RMS deviations[b] | |
| NOEs (Å) (no violations > 0.5 Å) | 0.028 ± 0.0023 |
| Dihedral restraints ($^o$) (no violations > $5^o$ ) | 0.62 |
| Bonds (Å) | 0.0043 ± 0.00016 |
| Angles ($^o$) | 0.54 ± 0.021 |
| Impropers ($^o$) | 1.44 ± 0.093 |
| Ramachandran map analysis[c] | |
| Allowed regions | 86.3% |
| Additional allowed regions | 13.5% |
| Generously allowed regions | 0.2% |
| Disallowed regions | 0.0% |
| Pairwise rms difference (Å)[d] | |
| Residues 11-88 (heavy atom) | 1.100 |
| $2^{nd}$ Structure residues 11-88 (backbone) | 0.503 |

[a] From chemical shifts using Talos.
[b] Calculated in Aria 1.2 for the 20 lowest energy structures refined in water.
[c] Obtained using PROCHECK-NMR.

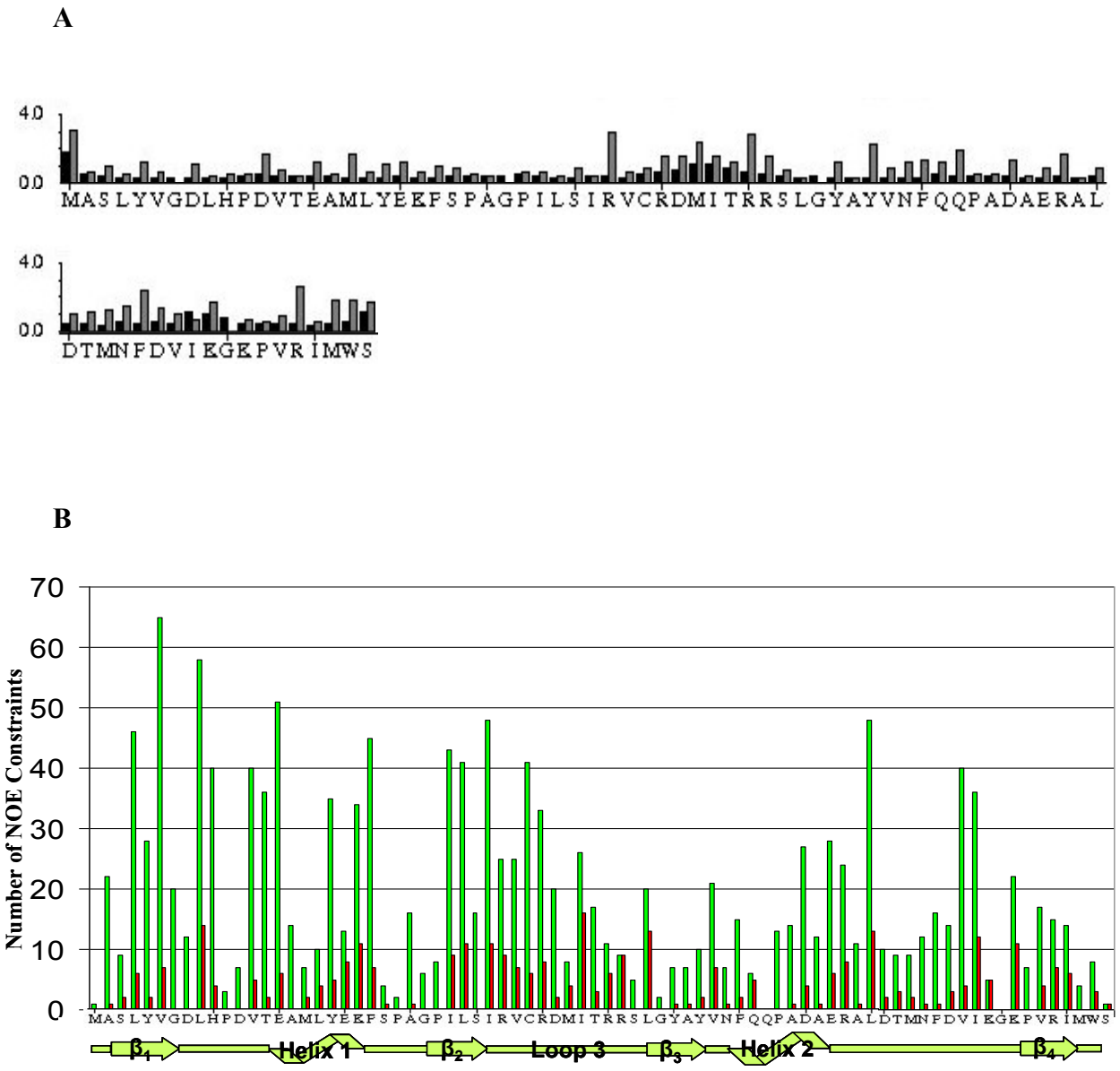



**Figure 4.8 Statistics on a residue per residue basis for 20 lowest energy conformers representing PABP-1 RRM 1. A.** Average local RMSD for the backbone and side-chain residues. **B.** Total number of NOE restraints shown as green bars and the intra-residue NOE restraints shown as green bars. These analysis were performed using Procheck NMR.

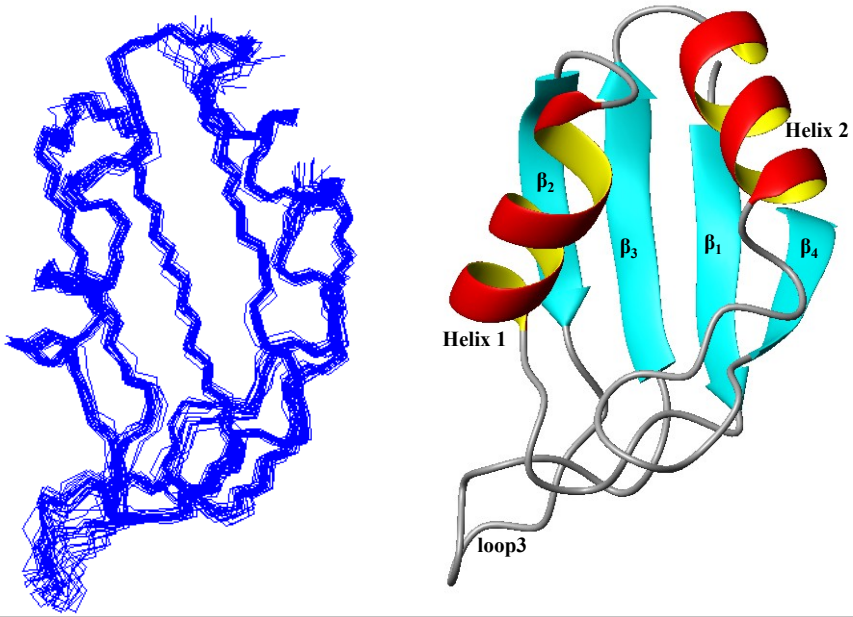



**Figure 4.9 Superposition of the 20 lowest energy converged structures for PABP-1 RRM 1. A.** A good convergence of structures however; the loop 3 region demonstrates a less stringent superposition. **B.** Cartoon representation of the water refined secondary structure for PABP-1 RRM 1. The structure depicts two α-helices crossing over a four stranded anti-parallel β-sheet resulting in a $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ fold similar to that seen in the PABP-1 RRM 2 structure.

### 4.3.1 Description of Secondary structure

Long range NOE signals that are indicative of an anti-parallel β-sheet were identified between residues S13-G17, L39-R42, A56-F60 and R84-M86. Figure 4.10 illustrates an example of the NOE connectivities that were detected by a $^{15}N$ edited NOESY-HSQC spectrum across the β-sheet. A β-sheet will typically give an NOE pattern between HN/H$\alpha_{(i,j)}$ and H$\alpha$/HN$_{(i,j)}$ on adjacent strands. These constraints determine the order and alignment of the β-strands. NOE data between L14 on the $\beta_1$-strand and M86 and Y57 on the $\beta_4$ and $\beta_3$-strands respectively indicates that a kink in the sheet maybe present this is further supported by the $^{13}C$ NOE data.

NOE's to the residues in helix 1 give a typical NOE pattern for an α-helix. Figures 4.10(A) and (B) both display the NOE pattern for an α-helix, with strong NOE signals to HN(i)/HA$_{(i)}$ and weak/medium signals to HN(i)/HA$_{(i-3)}$ HN(i)/HA$_{(i-4)}$ and HA(i)/HB$_{(i-3)}$ as indicated. A commonly witnessed aspect of this helix is that three residues on the C-terminus of the helical

region 1 gave an NOE pattern of HN(i)/$HA_{(i-2)}$ and HA(i)/$HB_{(i-2)}$ indicative of a 3.10 helix, Figure 4.10(A). This distinctive feature of helix 1 was not seen in helix 2. Helix 1 shows a significant bend at the C-terminal of the helix caused by P34. P34 is positioned at the C-terminus of the helix and causes an expansion of the last turn of the α-helix. In contrast to this helix 2 has no aromatic residues or other residues that may disrupt the α-helix.

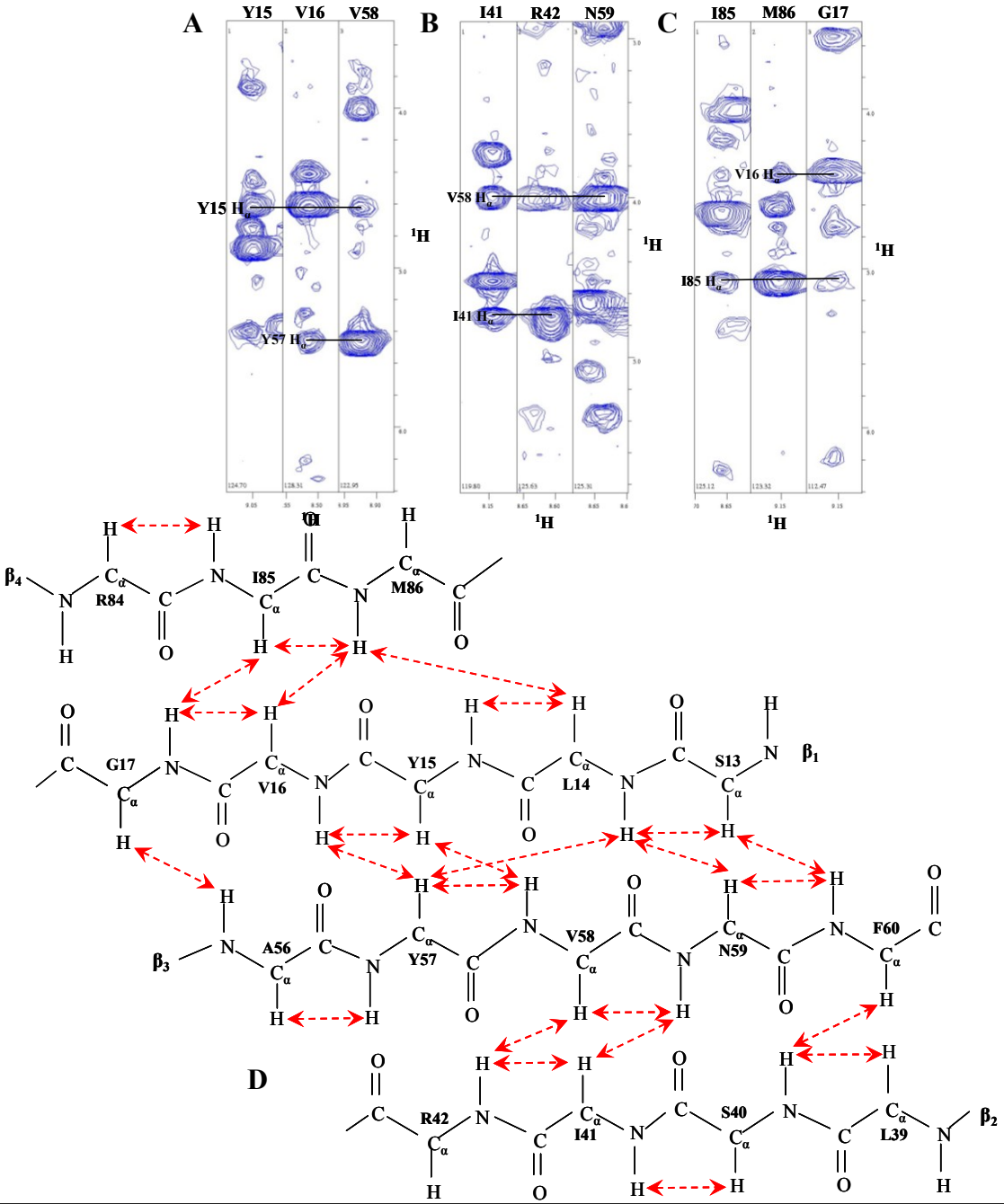



**Figure 4.10 Selected strips from $^{15}N$ edited NOESY spectrum the $\beta_1$, $\beta_2$, $\beta_3$ and $\beta_4$ strands for PABP-1 RRM 1. A.** $^{15}N$ edited NOESY-HSQC spectra with NOE connectivities for residues Y15, V16 and V58 of $\beta_1$ and $\beta_3$ strands respectively. **B.** NOE connectivities for residues I41 and R42 on $\beta_2$ strand, and N59 on the $\beta_3$ strand with the cross connectivities which form the anti-parallel β-sheet. **C.** NOE connectivities for residues I85 and M86 on the $\beta_3$ strand and G17 on the $\beta_1$ strand. **D.** Four stranded anti-parallel β-sheet of PABP-1 RRM 1, uniquely assigned NOEs that were used to elucidate the ordering and alignment of the β strands are displayed as arrows.

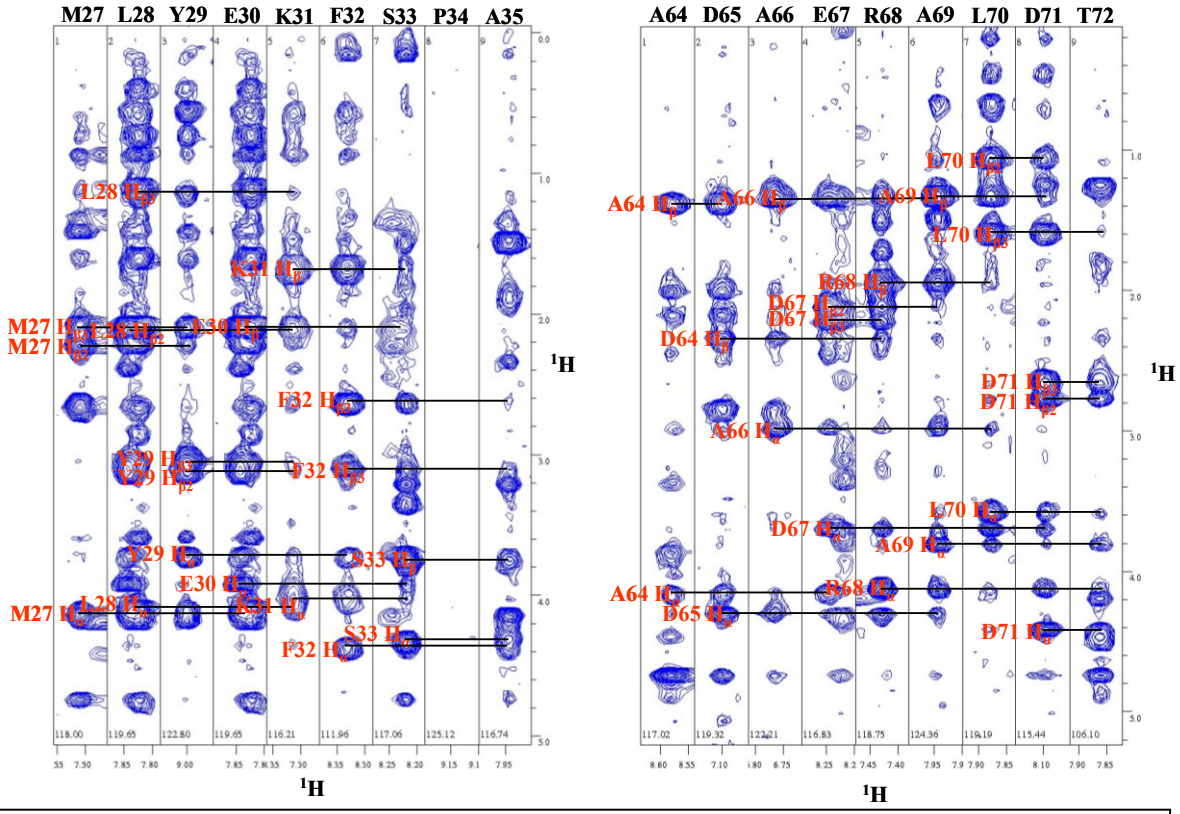



**Figure 4.11 Selected strips from $^{15}$N edited NOESY HSQC spectrum with NOE connectivities for the helical regions of PABP-1 RRM 1. A.** Strip plot of a $^{15}$N edited NOESY HSQC with an NOE pattern that is indicative of an α-helix; strong $HN_i$ $HA_i$, weak/medium $HN_i$ $HA_{i-3}$, $HN_i$ $HA_{i-4}$ and $HA_i$ $HB_{i-3}$ for the helical region 1 of PABP-1 RRM 1. The α-helix is broken by the presence of a proline residue at its C-terminus. **B.** Strip plot of a $^{15}$N edited NOESY HSQC spectrum illustrating an α-helical NOE pattern; weak/medium NOE signals were seen throughout this second helical region.

### 4.3.2 Tertiary structure analysis

The anti-parallel β-pleated sheet does not show exposed hydrophobic residues across its β-strands which are seen in other protein-protein binding RRM domains. The hydrophobic residues are buried and form an intact pocket on the adjacent side of the four stranded β-sheet. This hydrophobic pocket is further maintained by the interaction of the hydrophobic residues of the sheet with the hydrophobic residues on the two α-helices, Figure 4.12. A total of nine hydrophobic residues across the β-strands form the base of the hydrophobic pocket, $\beta_2$-strand contains buried residues I34, I41 and V43 whereas the $\beta_3$-strand hydrophobic residues are A56, V58 and F60, Figure 4.12. The N-terminal and C-terminal β-strands are home to the outlying hydrophobic residues, L14 V83 and I85 respectively, that project inward toward the

pocket. The two α-helices are packed against the β-sheet by hydrophobic interactions. Residues T24 and F32 of helix 1 project inward toward the pocket, similarly residues A69 and L70 of helix 2 are also arranged in a similar manner.

F32 appears to play a pivotal role forming hydrophobic interactions with A66, A69 and L70 of helix 2. NOE data for this residue and those hydrophobic residues within the hydrophobic pocket reveal hydrophobic interactions with the β-strands. F60 located on the $\beta_3$-strand is central making numerous hydrophobic interactions to the sheet and to the C-terminus of helix 1 and the N-terminus of helix 2. Y29 of helix 1 projects away from the helix and is angled downward toward the $\beta_2$-strand bringing into close proximity to the $\beta_2$-strand giving it a certain amount of protection. This feature is reminiscent of that seen between Y116 of helix 1 and I126 of the $\beta_2$-strand in the PABP-1 RRM 2 structure, presented in Chapter 4. NOE signals were seen between I38 and I41 on the $\beta_2$-strand and Y29 thus supporting its conformation. Further NOE signals were witnessed between I38 and I41 and V58 on the $\beta_3$-strand and F32 on helix 1; thus providing the evidence for the packing of the helices against the β-sheet.

The PABP-1 RRM 1 β-sheet contains no solvent exposed hydrophobic residues that were seen in the PABP-1 RRM 2 structure. It is also noticeable that there are not as many charged residues on the surface of the β-sheet. There are however, three tyrosine residues that are involved specifically in stacking interactions with the poly(A)-tail of RNA. NOE data indicates that Y15 is bent away from the $\beta_1$-strand toward Y55 at the same time it is rotated toward Y57 on the $\beta_3$-strand by Van der Waals interactions with Y57, Figure 4.13. Figure 4.13a shows a twist in the C-terminus of the $\beta_3$-strand. The NOE data suggests that L39 may form a hydrogen bond with N59 this would cause the β-strand to twist toward the $\beta_2$-strand. NOE signals between Q61 and L39 suggest a further hydrogen bond may form between Q61, located at the apex of loop4 between the $\beta_3$-strand and helix2, and P37 at the apex of loop 2 linking helix1 with the $\beta_2$-strand. This interaction would cause the C-terminus of the $\beta_3$-strand to twist more dramatically which is seen in Figure 4.13a. The bend in the N-terminus of the $\beta_1$-strand maybe caused by hydrophobic interactions between L14 and both A66 and L70 on helix 2. The NOE data indicates that L14 is anchored by possible hydrogen bonds with V58 and Van der waals interactions with both V58 and F60 drawing L14 inward. In opposition to this is the way that Y15 is bent away from the β-strand creating a prominent bend in the β-strand with L14 at its apex.

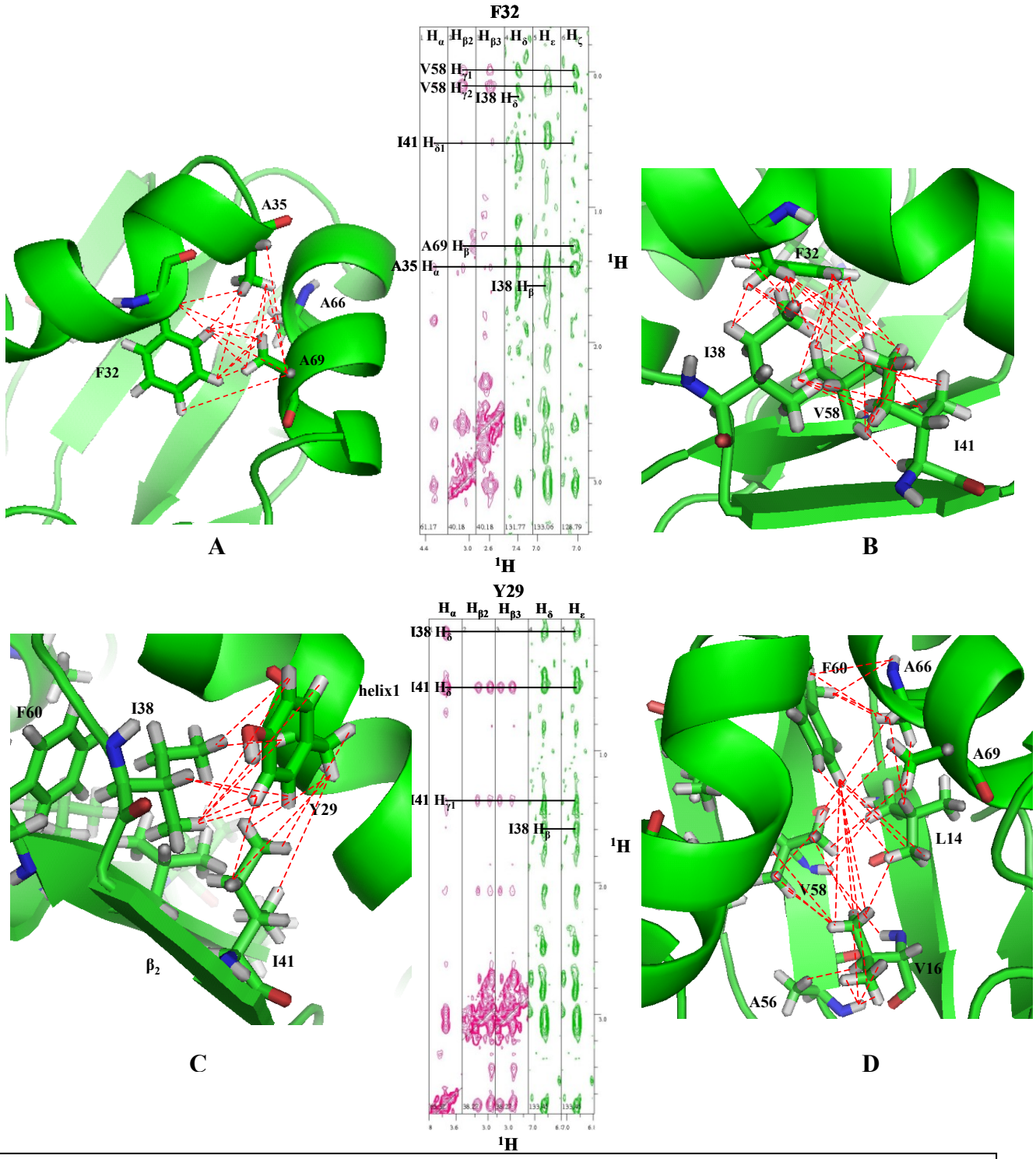



**Figure 4.12 NOE connectivites of the hydrophobic pocket supported by the associated strip plots of $^{13}$C-NOESY spectra. A.** NOE signals detected between the hydrophobic residues of the α-helices. F32 of helix 1 extends across the gap between the two helices and plays a pivotal role in maintaining the integrity of the hydrophobic pocket. The extended conformation of the C-terminus of helix 1 can be seen and its close proximity to helix 2 is confirmed by NOE data. **B.** The packing of the helices against the β-sheet is supported by the presence of hydrophobic interactions between the hydrophobic core and the hydrophobic residues within the helices. **C.** Connectivity between outlying hydrophobic residues on the $\beta_2$-strand and helix 1. Y29 on helix 1 projects away from the helix at a downward angle toward the $\beta_2$-strand. The selected $^{13}$C NOSEY spectrum strips support the position and angle of the aromatic ring, reciprocal NOE signals further supported this arrangement. **D.** Hydrophobic core located between the two helices and the β-sheet supported by $^{15}$N and $^{13}$C NOE data.

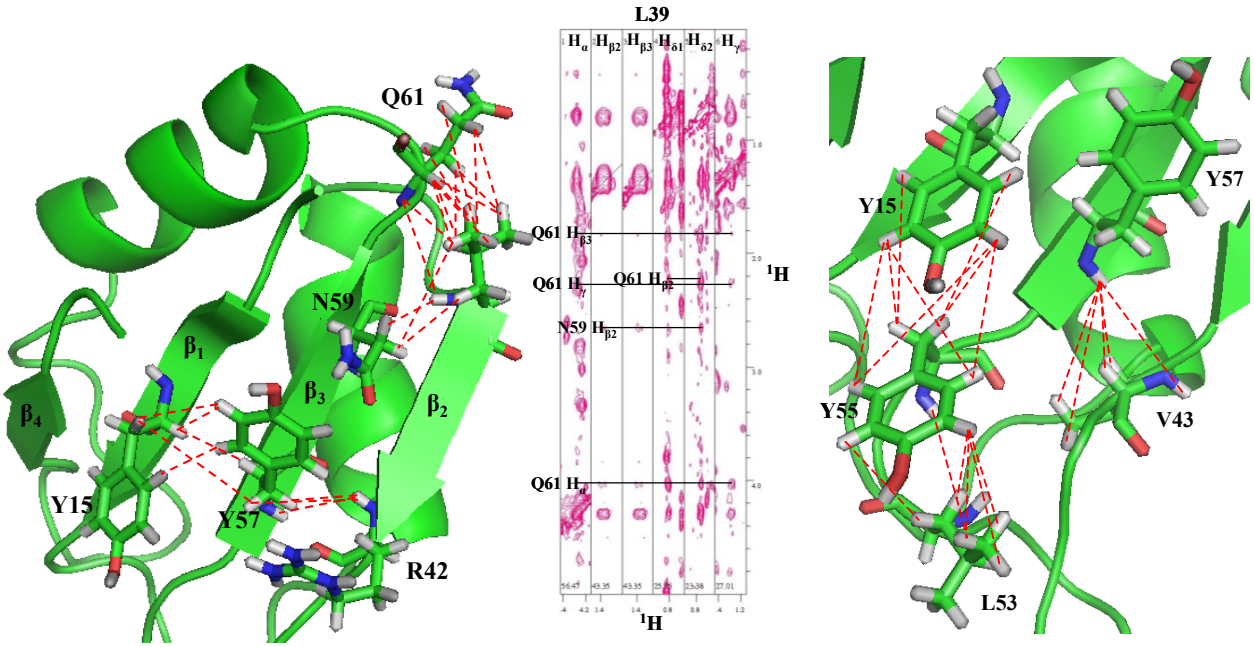



**Figure 4.13 NOE signals detected across the solvent exposed face of the β-sheet. A.** The $\beta_3$-strand indicates a twist at its C-terminus supported by NOE constraint data and hydrogen bond calculations. Y15 is bent away from the $\beta_1$-strand toward Y55. **B.** NOE signals suggest possible Van der Waals interactions between Y15 and Y57 causing Y15 to span the $\beta_1$-$\beta_3$ gap. In return Y55 is rotated away from Y15 toward the flexible loop3 region.

Loop 3, Figure 4.14a and b, is involved in the base stacking interactions of poly (A) mRNA binding to PABP-1 RRM1/2. Crystallographic studies of PABP/RNA complex revealed that Ade7 of the poly (A) mRNA is sandwiched between two copies of L53 and D46 located on loop3 of PABP-1 RRM 1. The structure of the loop appears to be dynamic showing a high degree of motility thus allowing it to make contact with RNA. This loop region differs from that seen in PABP-1 RRM 2 in that it is positively charged by three arginine residues. The NOE data indicates that R50 lies across the loop so that it is orientated parallel to R45, Figure 4.14a. From the NOE data it was predicted that a hydrogen bond may form between R45 and R50 and also between R50 and D46 suggesting the sides of the loop to pinch together. In contrast to R50, R51 projects outward away from the loop. The NOE data suggests that R51 is in close proximity to D22 of loop1 and maybe able to form a salt bridge. The loop itself has two hydrophobic residues that project outward at the apex of the turn. Potential hydrogen bonds may form between D46 and I48 and T49, salt bridges may also form between D46 and the HN groups of I48 and T49.

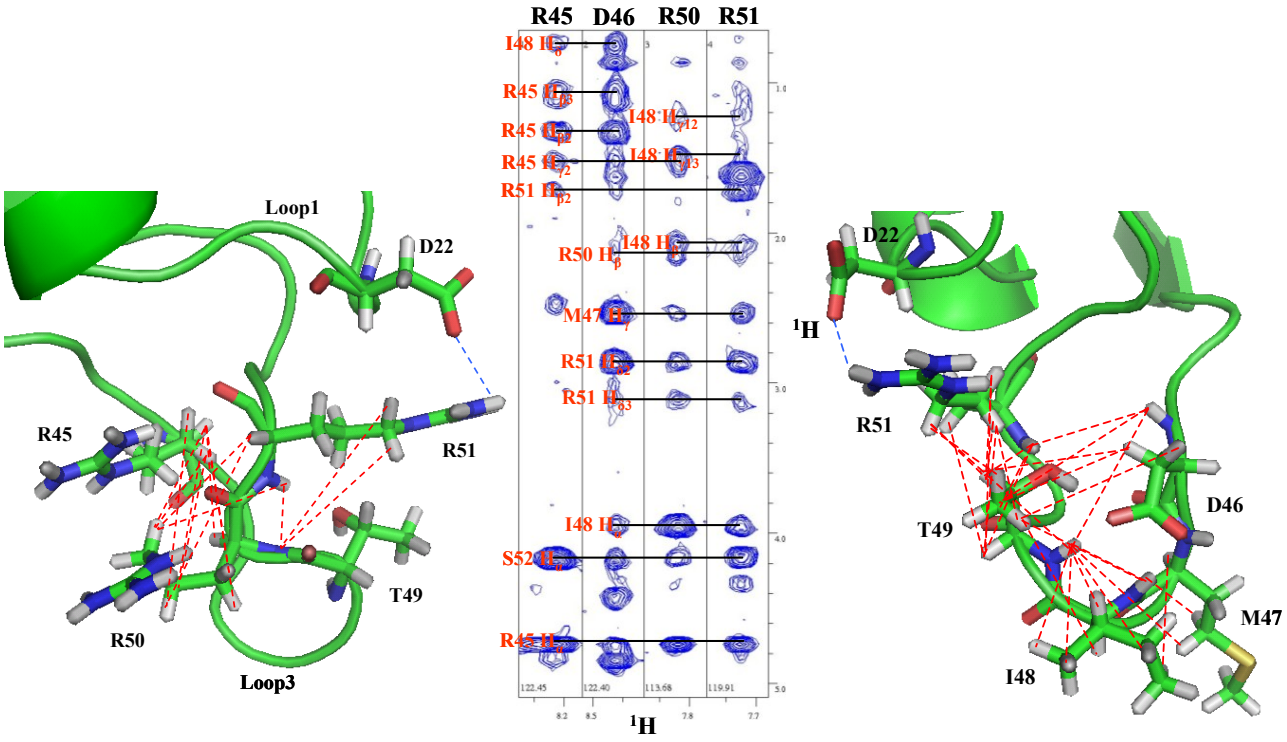



**Figure 4.14 Structure of the flexible loop3 region of PABP-1 RRM 1 with associated strip plots of $^{15}$N NOSEY spectrum.** Loop3 of PABP-1 RRM 1 displays a much different architecture than that seen for loop3 of PABP-1 RRM 2. The Figures show that this region is very positively charged with two hydrophobic residues that project outward at the apex of the turn. NOE data indicates that R45 and R50, located on opposite sides of the loop, extend parallel to one another. In contrast R51 projects away from the loop itself where the NOE data suggests it makes a salt bridge (indicated by a blue dashed line) with D22 of loop1.

### 4.3.3 Structural comparisons

The electrostatic surface representation of PABP-1 RRM 1 shows a balance of charges when compared to PABP-1 RRM 2. Figure 4.15a displays the electrostatic surface representation of the β-sheet. The surface of PABP-1 RRM 2 β-sheet gave a patch of positively charged residues and solvent exposed phenylalanines which complimented the negatively charged hydrophobic paxillin LD1 peptide. In contrast to this PABP-1 RRM 1 displays a more neutral surface with no exposed hydrophobic residues. A unique feature of PABP-1 RRM 1 compared with PABP-1 RRM 2 is the positive charge that is present on the flexible loop3 region created by R42, R45, R50 and R51. This may provide additional support for electronegative ligands such as mRNA and paxillin LD domains.

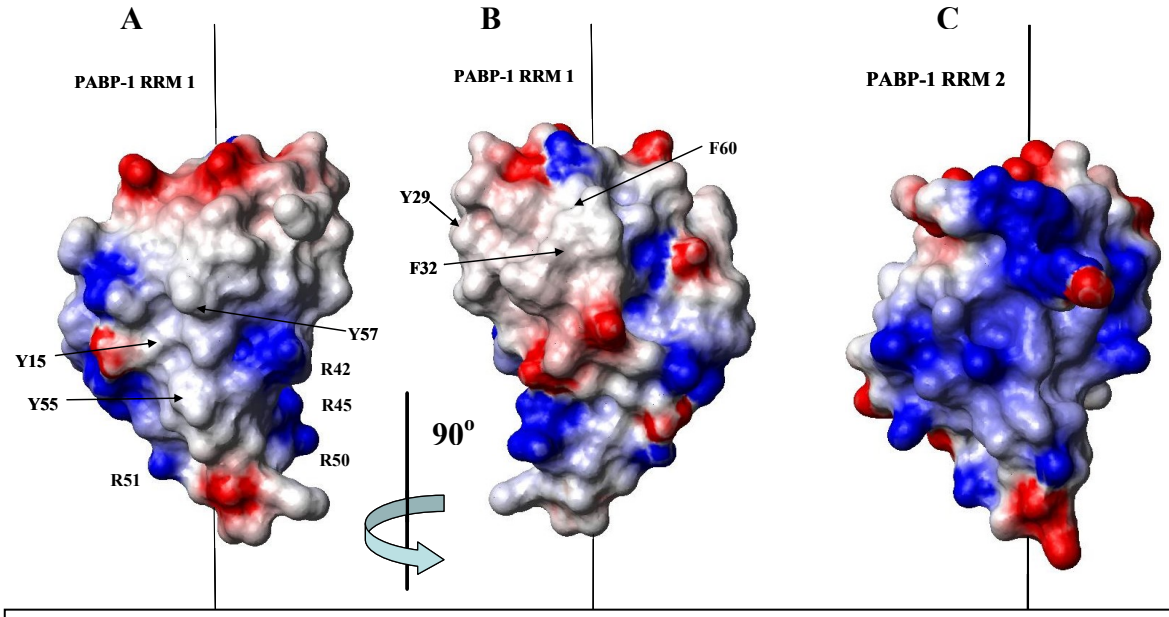



**Figure 4.15 Comparative electrostatic surface representations of PABP-1 RRM 1 and RRM 2. A.** The electrostatic solvent exposed surface of the β-sheet, positive charges are blue, negative are red and neutral are white. The aromatic rings (white) of Y15, Y55 and Y57 can be clearly seen in the centre of the β-sheet. The positive charge of R42 located at the start of loop3 can be seen to the right of the structure. **B.** The electrostatic surface charge of the helical side of PABP-1 RRM 1. The positive charges of R45 and R50 can be seen at the bottom left of the Figure. **C.** The electrostatic surface charge of the β-sheet of PABP-1 RRM 2. In contrast to PABP-1 RRM 1 the β-sheet displays a positively charged surface and also presents solvent exposed phenylalanines.

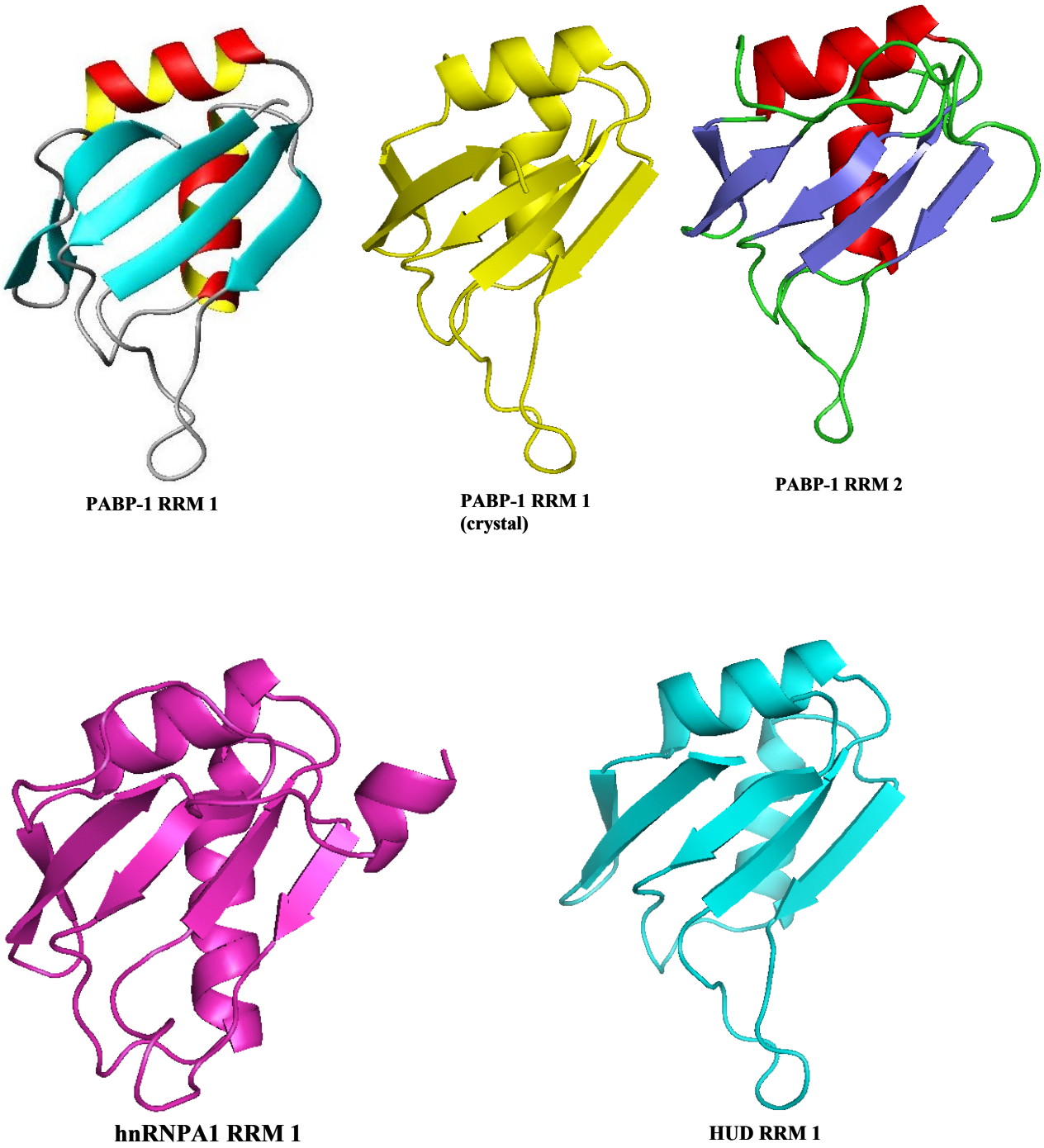



**Figure 4.16 Comparason of PABP-1 RRM 1 (top left), PABP-1 RRM 1 crystal structure (top centre), PABP-1 RRM 2 (top right), hnRNPA1 RRM 1 (bottom left) and HUD RRM 1(bottom right).** The NMR derived structure of PABP-1 RRM 1 displays very few differences when compared to the crystal structure. The NMR derived PABP-1 RRM 1 structure was determined in its unbound state and shows more flexibility in its β-strands than the crystal structure which was determined in complex with poly(A) RNA. The two NMR derived structures presented in this project display secondary structures that show a high level of homology. Indeed all of the secondary structures presented here display similar secondary structure characteristics. The presence of the flexible loop3 region is a feature that is present in a majority of RRM domain containing proteins. This might indicate that it is important to the stable interaction with RNA.

The NMR derived structure of PABP-1 RRM 1 presented in this chapter was elucidated in its unbound state. A comparison of secondary structure characteristics between the NMR derived and crystal structures of PABP-1 RRM 1 reveal very few differences. As mentioned in previous chapters the crystal structure of PABP-1 RRM1/2 (6) was elucidated in complex with poly(A) RNA and may explain the differences across the β-sheet. The orientations of Y55 and Y57 are a result of RNA stacking interactions which are evident in the crystal structure. The NMR derived structure generates NOE constraints which suggest that in their unbound states the tyrosine residues adopt a different conformation. Indeed, the degree of secondary structure homology between RRM domains of other proteins is worth noting. Both HUD (166) and hnRNPA1 show the now familiar $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ fold. The hnRNPA1 (167) protein loop3 region is less prominent and packs against the rest of the protein compared to the other RRM domains. The secondary structures of PABP-1 RRM 1 and RRM 2 presented in this project show a high degree of homology. The β-sheet of PABP-1 RRM 1 appears to be a lot less ordered than that of PABP-1 RRM 2. Indeed, the face of the β-sheet of PABP-1 RRM 2 compliments that of the paxillin LD1 domain thus facilitating its interaction. The solvent exposed face of PABP-1 RRM 1 β-sheet does not show the same degree of positive charge or exposed hydrophobic residues.

### 4.3.4 Paxillin LD1 titration

The biochemical assays reported in chapter 3 indicated a positive result for the binding interaction of PABP-1 RRM 1 with paxillin LD1. The ELISA results showed that, compared to PABP-1 RRM 2, the binding was of a reduced affinity. Based on these data this interaction was investigated using the same techniques as those described in chapter 3.

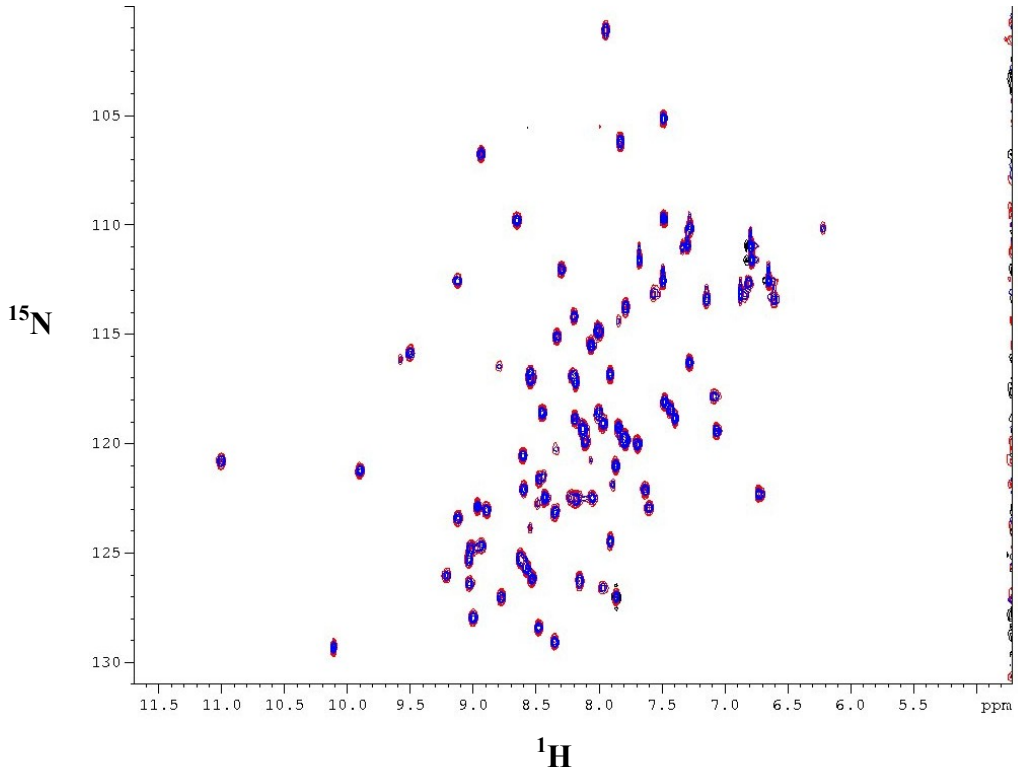
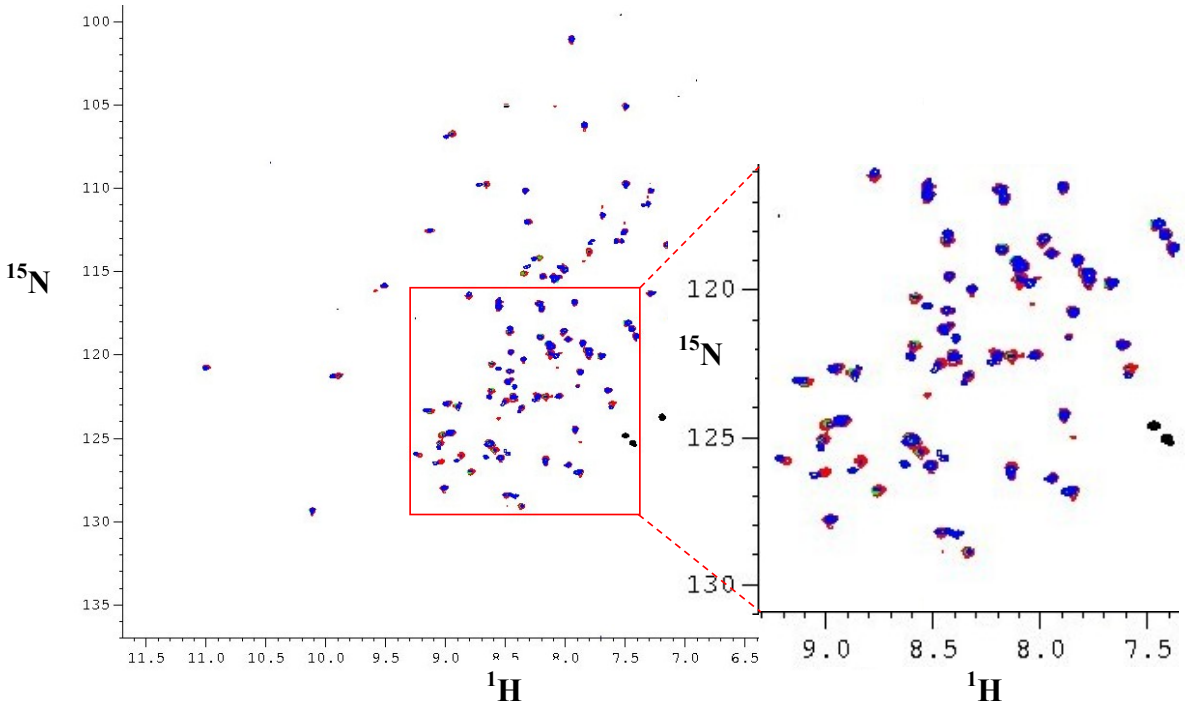



**Figure 4.17a (above) $^{1}H$ $^{15}N$ HSQC spectrum for the titration of PABP-1 RRM 1with paxillin LD1 at a 6 molar excess of the peptide.** The titration revealed no chemical shift perturbations, there were no shifts or reduction in signal intensities witnessed. **Figure 4.17b (below) $^{1}H$ $^{15}N$ HSQC spectrum for the titration of a GST labelled paxillin LD1 peptide at a ratio of 1:1.** The spectrum illustrates chemical shift perturbations indicating that the binding seen in the biochemical assays was induced by the linker region between paxillin LD1 and the GST-tag.



The $^{1}H$ $^{15}N$ HSQC spectrum in Figure 4.17a represents the paxillin LD 1 titration performed on the RRM 1 domain of PABP-1. The spectrum clearly demonstrates that no binding interaction was present between the synthetic peptide paxillin LD1 and RRM 1. In light of the ELISA data in chapter 3, this result was somewhat unexpected given the presence of a PBS motif within PABP-1 RRM1. Figure 4.18 illustrates the location of the PBS site within PABP-1 RRM1. The region extends to include helix 1 and the loop region from the $\beta_1$-strand and helix 1. NOE data indicates that the 5 hydrophobic residues within the region, L19, P21, V23, A26 and L28, are buried forming interaction between themselves and the hydrophobic core. The initial titration was performed at a 100mM NaCl concentration, in an effort to induce the binding witnessed in the ELISA assays the titration was repeated at 20mM NaCl concentration. The results were the same as those seen in Figure 4.17a, (not shown). It was curious as to why the ELISA results indicated a binding interaction where none was identified by NMR. The paxillin LD 1 synthetic peptide was 6 residues shorter than that used in the ELISA assays. In order to investigate whether a longer construct would bind it was decided to use a paxillin LD1/2 double domain construct.

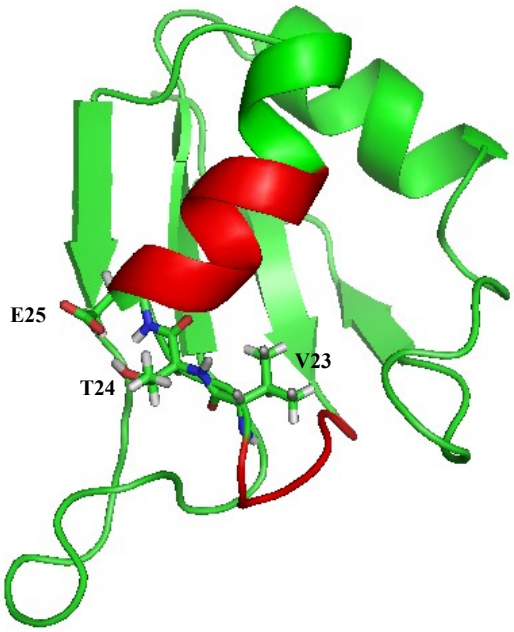



**Figure 4.18 PABP-1 RRM1 PBS domain.** Paxillin binding sub-domain highlighted in red. Residues V23, T24 and E25 show sequence homology to the PBS domains identified in actopaxin and vinculin. Hydrophobic residues within this region appear buried and inaccessible to the solvent. NOE data indicates hydrophobic interactions between V23 and other hydrophobic residues within helix 1.

Initially a GST fusion construct was used, this however proved to be labile and it was difficult to remove the GST-tag and maintain a pure sample. It was found that once the GST-tag was removed and the sample purified by gel filtration the resultant protein yield was minimal. The paxillin LD1/2 construct was moved into a pETM-30 vector (supplied by Dr Alison Woods) with an N-terminal His-tag, thus facilitating a more efficient purification method.

Despite the increase in protein yield there were still stability issues, therefore a fresh batch of paxillin LD1/2 protein was required. A titration series was performed using the double LD domain. The subsequent HSQC signals for PABP-1 RRM 1 did not indicate any chemical shift changes upon increasing additions of paxillin LD1/2 (results not shown).

A titration was performed on PABP-1 RRM 1 with the His-tag removed but using the GST fusion protein that was employed for the ELISA assays. Figure 4.17b clearly shows that an interaction is present when GST fusion paxillin LD 1 is titrated in. As part of the ELISA assay a GST negative control was used, the control protein was generated from a pGEX2TK vector. The GST-tag was removed using thrombin protease and purified by ion exchange column. The linker region, LVPRGSRRASV, appears to make weak binding interactions across the β-β loop regions. Only three PABP-1 RRM1 residues displayed noteworthy chemical shift perturbations V43, L53 and K81, none of which are located within the PBS motif of PABP-1 RRM1. These data may account for the positive result observed during the course of the ELISA assay.

### 4.4 PABP-1 RRM1/2 tandem domain

The crystal structure showed that the linker between the two domains forms a helix when bound to the poly (A) tail and supports this interaction. The biochemical results presented in chapter 3 indicated a greater binding affinity for PABP-1 RRM1/2 in complex with wild type paxillin. It would be interesting to see whether a paxillin LD1 titration of PABP-1 RRM1/2 would induce further chemical shift perturbations and whether the affected residues reside within the linker domain. It would also be interesting to determine whether these two RRM domains interact with one another in their unbound states.

#### 4.4.1 RRM domain comparison

The NMR solution structures of PABP-1 RRM 1 and 2 in their native states show the same fold as that seen for other RRM domain proteins $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$. The sequence specific alignments in Figure 4.19 show a consensus sequence of SKGFG $^{F}_{Y}{}^{I}_{V}$ showing a sequence similarity >50% across the RRM domains.

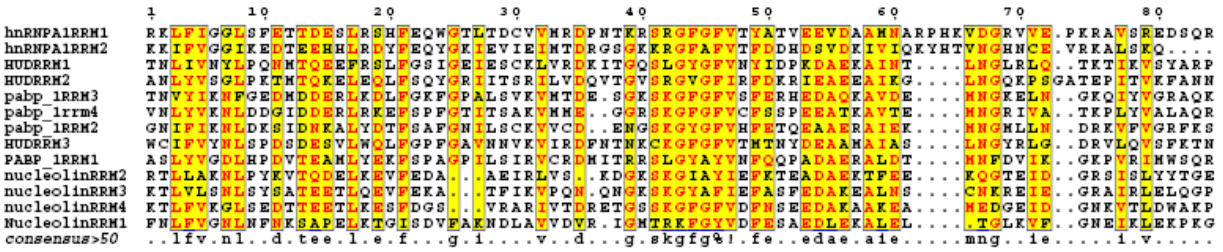



**Figure 4.19 Amino acid sequence homology and conservation between RRM domains of RNA interacting proteins.** The multiple sequence alignment of RRM domains displayed is from human form proteins of PABP-1, nucleolin, HUD and hnRNPA1. Residues with conserved sequence similarity in five or more domains are highlighted in yellow with >60% sequence similarity. The main consensus sequence SKGFG $^{F}_{Y}{}^{I}_{V}$, the '%' symbol denotes either an F or a Y residue and the '!' symbol denotes either an I or a V residue. The alignment was prepared using ClustalW and ESPript.cgi (http://npsa-pbil.ibcp.fr/cgi-bin/align clustalw.pl).

### 4.4.2 PABP-1 RRM1/2 NMR model

The $^{15}$N resolved HSQC spectrum of PABP-1 RRM1/2 was overlaid with the $^{15}$N resolved HSQC spectra of both PABP-1 RRM 1 and 2, Figure 4.20. The HSQC spectra overlaid with minimal differences in signal positions however there was spectral overlap of the PABP-1 RRM1/2 HSQC. The absence of chemical shift perturbations in the PABP-1 RRM1/2 HSQC spectrum illustrates that the two RRM domains do not interact with one another in their free states. The chemical shift maps for the two single RRM domains were used for the sequential assignments of PABP-1 RRM1/2. Unlike the NMR elucidated structures presented in this thesis the HNCACB and CBCACONH spectra of PABP-1 RRM1/2 were used to confirm the sequential assignments of the two single RRM domains. The sequential assignments of the two single domains were comparable to the HNCACB and CBCACONH spectra of PABP-1 RRM1/2, Figure 4.21. The signals appeared weaker in the inter-domain linker region which would indicate that it is largely unstructured. Due to spectral overlap and the strength of the amide backbone experiments there were only 2 residues within the inter-domain linker region that were left unassigned.

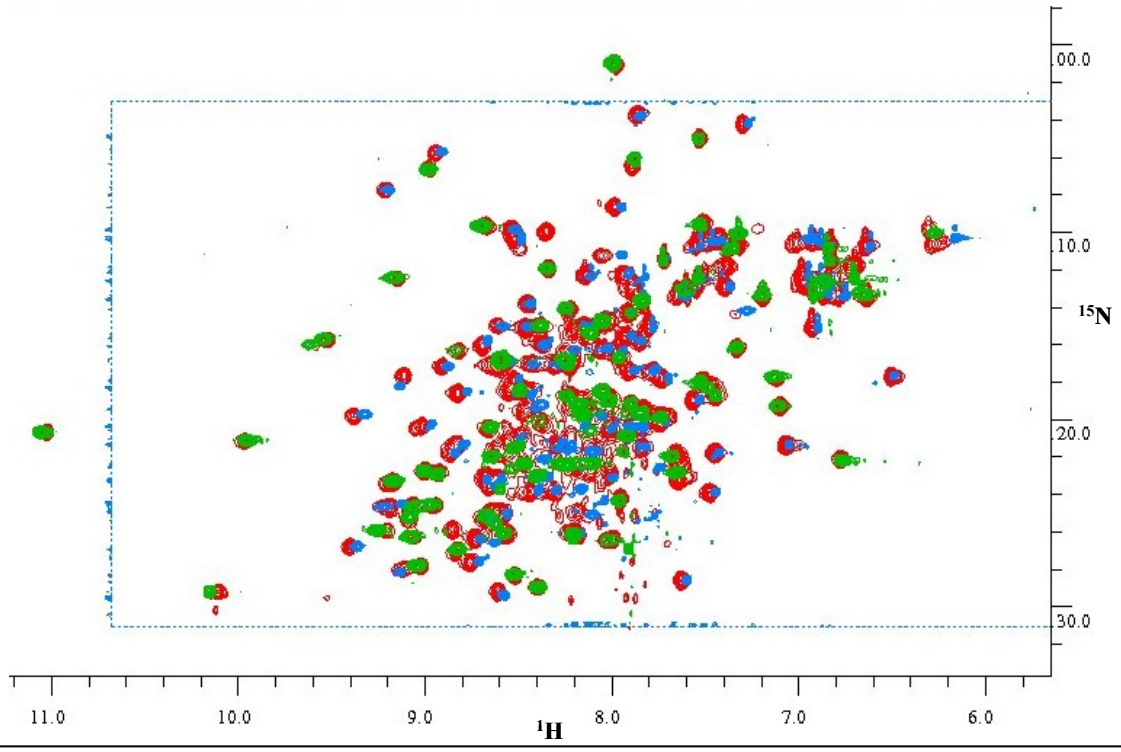



**Figure 4.20 illustrates an overlay of three $^{1}H$ $^{15}N$ HSQC spectra.** The Figure illustrates an $^{15}N$ resolved HSQC spectrum for PABP-1 RRM1/2, shown as red HSQC signals. This spectrum is overlaid with a $^{15}N$ resolved HSQC spectrum for PABP-1 RRM 2 (blue signals) and a $^{15}N$ resolved HSQC spectrum for PABP-1 RRM 1 (green signals). The two single domain RRM spectra overlay with PABP-1 RRM1/2 with minimal chemical shift difference in the signals. This would indicate that there is very little interaction between the two domains.

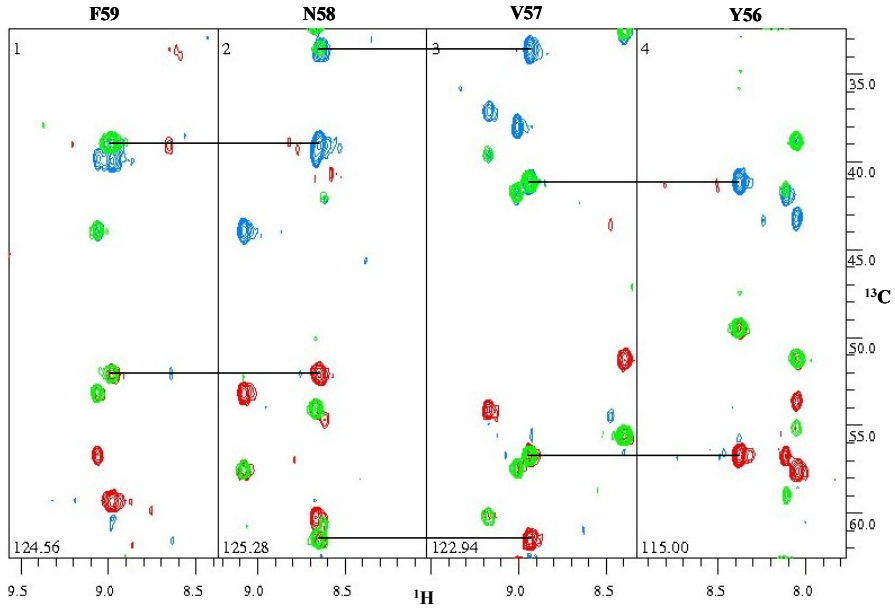



**Figure 4.21 illustrates a HNCACB and CBCACONH spectra for PABP-1 RRM1/2.** The above spectra displays inter and intra-residue signals for the RRM1 domain of PABP-1 RRM1/2. The inter-residue connections are indicated. The RRM 2 domain displayed a lot more spectral overlap than was witnessed for this particular part of the spectrum.

The strength of the amide backbone experiments made the confirmation of sequential assignments difficult from the C-terminus of the RRM 1 domain onwards. Despite this over 87% of the assignments were confirmed which was ideal to gain a model of the tandem RRM domains. In order to elucidate the structure of PABP-1 RRM1/2 domain more sensitive experiments would be needed designed for larger proteins.

### 4.4.3 NMR based secondary structure model

In order to incorporate all of the structural data that has been established for the two single domains a $^{15}N$ resolved NOESY HSQC and a $^{13}C$ resolved HSQC NOESY spectra were collected. In order to improve the sensitivity of the $^{13}C$ HSQC NOESY the protein was placed in 100% $D_2O$. Once again the spectra were used to confirm the NOE assignments of the two single RRM domains. Both NOESY spectra were analysed for NOE signals across the flexible inter-domain linker region.

The CANDID/CYANA protocol was used to create a model of the PABP-1 RRM1/2 domain based on the chemical shift data obtained from the structures presented in chapters 3 and earlier in this chapter. Intra-residue $H^N$ NOE signals, taken from the two NMR elucidated RRM structures, were rigorously checked against $^1H$ $^{15}N$ edited NOESY HSQC spectrum. Very few differences in the chemical shift data were witnessed between the single RRM domains and the double domain. Sequential NOE signals were used to generate the secondary structure model of the two RRM domains.

The model of the two domains, Figure 4.22, illustrates the typical $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ for both domains connected by the inter-domain linker region, residues 88-98. Due to the size of the double RRM domain

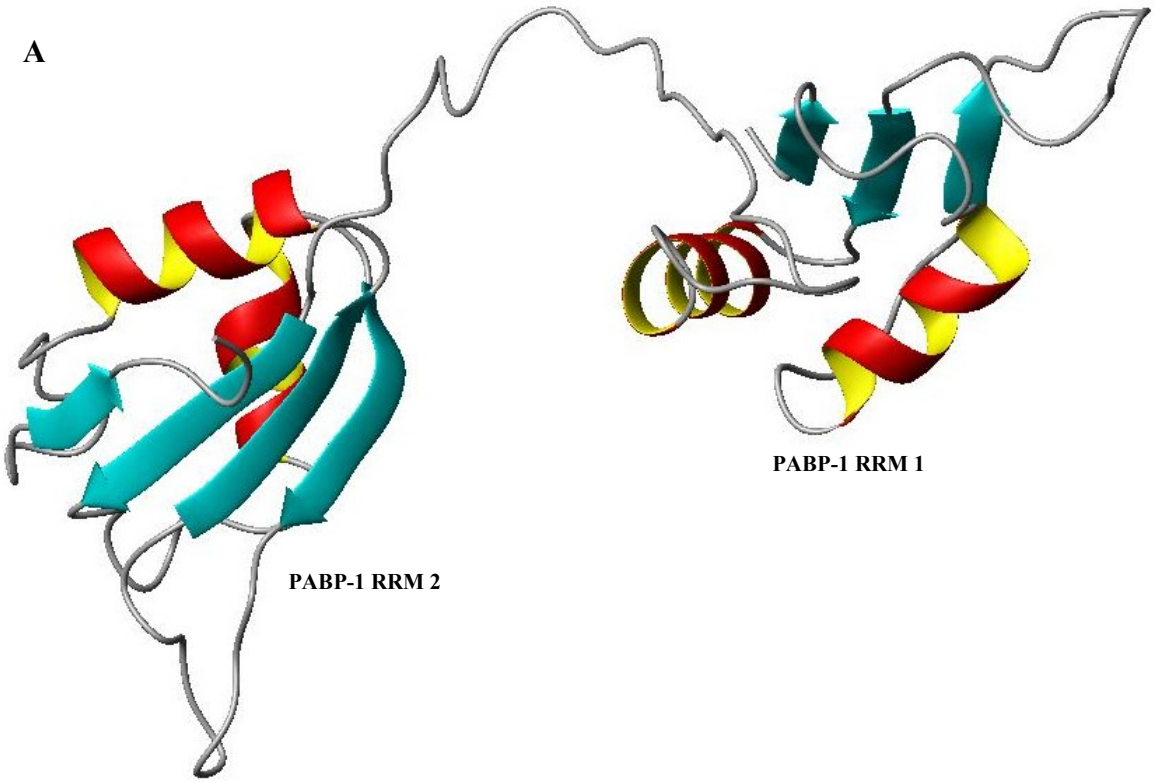
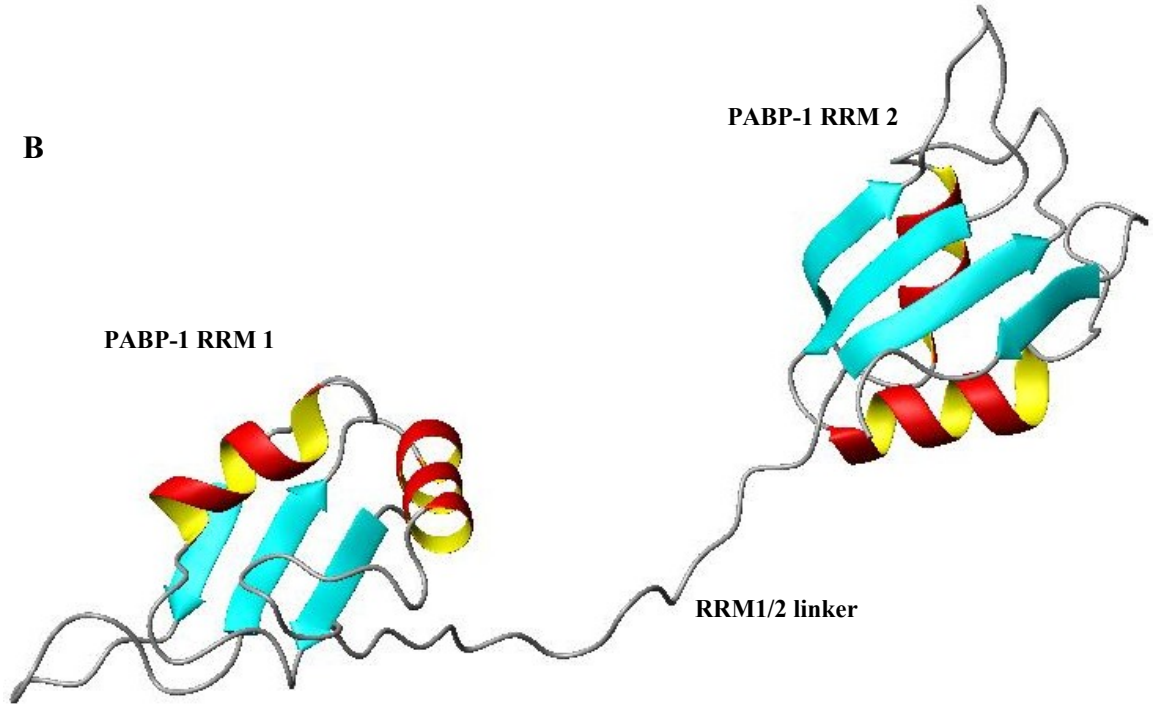



**Figure 4.22 Model of the PABP-1 RRM1/2 domain based upon derived chemical shift data of the single RRM domains. A)** The model of the tandem RRM domain was a result of the CANDID/CYANA structural calculation protocol. This structural representation was based on raw chemical shift data from the refined PABP-1 RRM 1 and 2 domains. **B)** The inter-domain linker region, residues 90-98, appears to be largely unstructured. Refinement of chemical shift data was not performed following initial attainment of the PABP-1 RRM1/2 model.



### 4.4.4 PABP-1 RRM1/2 paxillin LD1/2 interactions

As already mentioned in chapter 1 Deo *et al* established that the two domains work together to support the binding interaction of the poly (A)-tail of mRNA. They found that the inter-domain linker region between PABP-1 RRM 1 and 2 forms a helical conformation when bound to the RNA. During the course of this project it has been established that this inter-domain linker region is largely random coil in its free state. This finding was confirmed by the NOE data as there were no long range NOE signals seen between the two domains or to the inter-domain linker region. More over the NOE signals for the inter-domain linker region were mainly intra-residue and sequential inter-residue indicative of a random conformation. This inter-domain linker region may form interactions with an extended paxillin construct that includes both LD 1 and LD 2 domains. This construct had failed to produce any chemical shift changes in PABP-1 RRM1 however, PABP-1 RRM2 showed chemical shift perturbations within the N-terminal region that included several residues of the linker region.

Evidence has already been presented in chapter 3 supporting the PABP-1 RRM 2 paxillin LD1 complex. Chapter 3 also presented evidence for a binding site for paxillin LD1 within the PABP-1 RRM 1 domain. This result was found to be a false positive due to an interaction between the GST linker region of the paxillin LD1 construct and PABP-1 RRM 1. However Deo *et al* did show that the PABP-1 RRM1/2 domain increased the binding affinity of the poly (A)-tail of mRNA. Using the same premise it would be interesting to determine whether an increase in the number of affected residues could be brought about by the binding of paxillin LD1/2 to PABP-1 RRM1/2.

A titration was performed using fresh paxillin LD1/2 at a protein to ligand ratio of 1:1. It was found that upon further additions of paxillin a large amount of precipitation occurred within the sample. Chemical shift perturbations were witnessed and localised to the PABP-1 RRM2 domain, there were no additional chemical shift perturbations to the linker region or PABP-1 RRM1 domain, Figure 4.23. This finding is consistent with what has been reported so far. The linker region plays a vital role in mRNA binding by forming a helix indicating that paxillins interaction may start within the PABP-1 RRM2 domain and not interfere with PABP-1 RRM1/2 domain mRNA complex formation. The weak binding of paxillin LD1/2 domain would also indicate that the PABP-1/mRNA complex formation maybe the primary interaction.

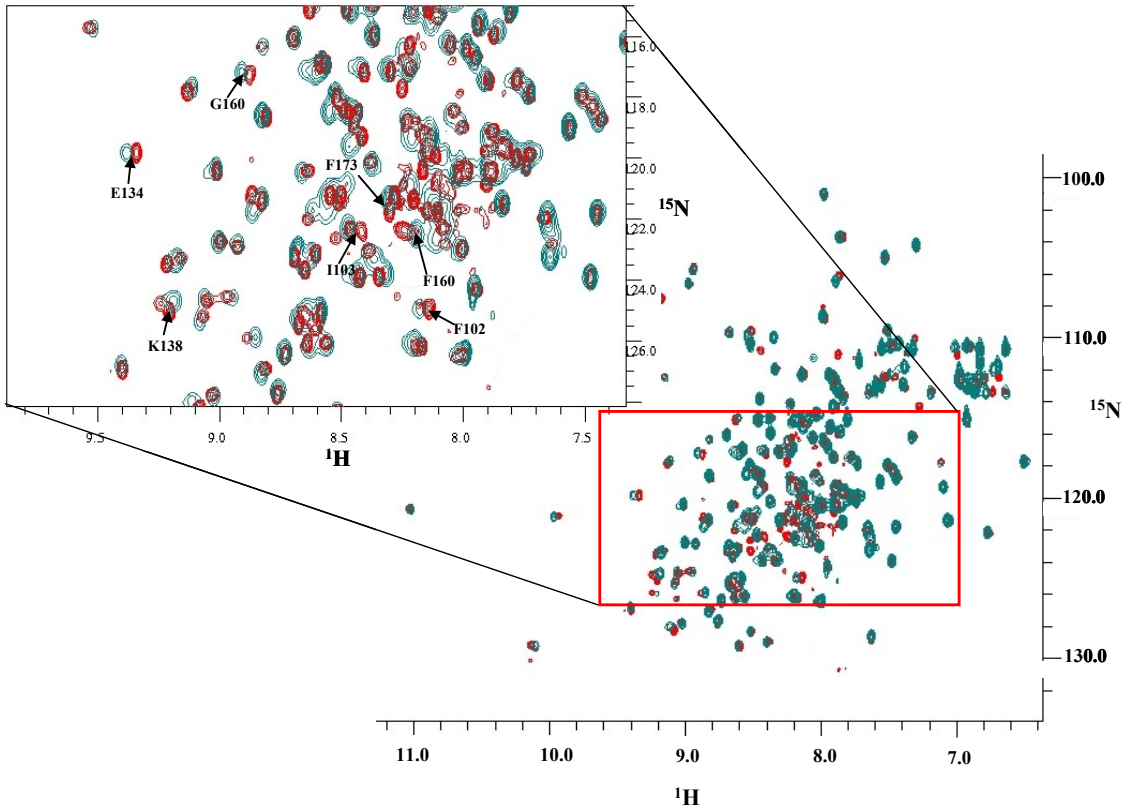



**Figure 4.23 $^{1}H^{15}N$ edited HSQC spectrum PABP-1 RRM1/2 paxillin LD1/2 titration.** The titration was performed at a 1:1 ratio, further additions suffered heavy precipitation. Analysis of the chemical shift perturbations revealed them to be localised to the RRM2 domain only. **Top left panel** illustrates a blown up region with RRM2 signals identified, slight differences were seen upon addition of the paxillin construct which were attributable to a slight increase in pH

## 4.5 Conclusion

NMR experiments were used to elucidate the structure of PABP-1 RRM 1. The structure revealed a typical RRM domain fold seen in other RRM domains from proteins such as hnRNPA1, Nucleolin and HUD. The structure produced a $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ fold with the helices stacked against the rear of the β-sheet. A hydrophobic core is present between the β-sheet and α-helices which is maintained by a number of hydrophobic, Van der Waals and hydrogen bonds across the β-sheet and between the helices. The secondary structure revealed a high degree of homology to that of PABP-1 RRM 2. However, the solvent exposed face of the β-sheet of PABP-1 RRM 1 is not as well ordered as that seen for PABP-1 RRM 2. Loop3 of PABP-1 RRM 1 is positively charged by the presence of four arginine residues this is in contrast to the other RRM domain containing proteins. Indeed, the loop3 region of PABP-1

RRM 2 is more negatively charged by comparison. Secondary structure comparisons show that PABP-1 RRM 1 is comparable to those RRM domains presented by HUD and hnRNPA1.

Titration studies with paxillin LD1 induced no chemical shift perturbations despite the positive binding results seen in chapter 4. A paxillin LD1/2 construct was employed and the titration studies repeated to see if binding could be induced in PABP-1 RRM 1. After repeated attempts with the NaCl concentration reduced to try and induce binding yielded in no chemical shift changes in $^{1}H$ $^{15}N$ HSQC spectral signals. It can only be concluded that paxillin LD1 binding does not take place in PABP-1 RRM 1 domain. This does not however, explain the reported binding seen in chapter 4. To try and repeat the conditions used in the ELISA biochemical assay a GST-fusion construct of paxillin LD1 was employed in the titration experiments with PABP-1 RRM 1. Chemical shift perturbation of $^{1}H$ $^{15}N$ HSQC spectral signals was witnessed. It can be concluded from these data that there exists a binding site for PABP-1 RRM 1/paxillin LD1 within the linker between the GST fusion tag and the protein.

In its unbound form the PABP-1 RRM1/2 inter-domain linker region is unstructured. Triple resonance NMR data gathered on the double domain enabled the assignments of NOE signals within the linker region. However, there were no NOE signals between the two domains or across the linker region that would indicate the presence of structure. The absence of inter-domain NOE signals was not surprising given that both PABP-1 RRM1 and RRM2 HSQC spectra overlay the RRM1/2 spectrum, Figure 4.20, with minimal differences indicating the absence of domain-domain interactions. The addition of paxillin LD1/2 to PABP-1 RRM1/2 failed to induce any further chemical shift perturbations than have already been reported. All changes in chemical shifts were found to be localised to the RRM2 domain. Biochemical data would suggest an interaction exists between paxillin and PABP-1 however, NMR data clearly shows that there are no interactions between the paxillin LD1/2 and PABP-1 RRM1. It maybe suggested that the interactions witnessed in the biochemical assays were transient or the fact that the RRM1 domain on its own was able to establish interactions with the LIM domains. Further investigations into the PABP-1/paxillin interaction are needed if the exact nature of this complex is to be understood.

## Chapter 5 Conclusion

It was originally thought that the PAPB-1-paxillin interaction was limited to the RRM-LD domains. Sequence alignment studies of PABP-1 against other paxillin binding partners revealed the presence of two PBS sequences (100) both located within the N-terminal RRM domains, the first within RRM1 and the second within RRM4. Mutagenesis studies revealed that disruption of these sites reduced cell migration events and that both sites operated independently of one another, (127). The biochemical data obtained during the course of this investigation have shown this to be a much more complicated interaction. The evidence indicated that an interaction site existed for paxillin within PABP-1 RRM2 domain but also within the PAPB-1 RRM1. The absence of a conserved PBS motif within PABP-1 RRM2 is surprising and further complicates the characterisation of paxillins binding partners with the knowledge that paxillin is able to bind to sites that are devoid of this motif. The C-terminal region of PABP-1 was able to bind both LD2 and LD5 domains of paxillin. Similarly the C-terminal region of paxillin gave a positive indication of binding to PABP-1's RRM4 domain. GST pull down assays of paxillin in migratory fibroblasts and their subsequent analyses suggested a stronger interaction than was observed for PABP-1's individual domains (7,127). This would suggest that the two proteins are intertwined and that no one site provides a strong interaction more over the strength of paxillin binding is a combination of the interaction sites.

The NMR structure of PABP-1 RRM2 yielded a structure with a $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ fold that is reminiscent of other RNA binding protein RRM domains such as HUD and nucleolin. The two α-helices form a V-shape on one side of four stranded β-pleated sheet. The front face of the β-sheet has three phenylalanine residues which have their aromatic rings exposed to the solvent, a fourth phenylalanine residue is present but has its aromatic ring buried within the aromatic core. In addition to this there are several positively charged residues dispersed across the face of the β-sheet with their side chains solvent exposed. The flexible loop region between $\beta_2$ and the $\beta_3$-strands, a common feature in RNA binding proteins, has been shown to make contact with polyadenylate tail of mRNA but not to interact with paxillin.

A synthetic peptide of paxillin LD1 domain was used to perform NMR based titration experiments against PABP-1 RRM2. The results of the titration showed that the chemical shift perturbations witnessed were specific even at 6 molar excesses of paxillin LD1 peptide. It appears that the interaction is weak binding, calculated to be 2.11mM at 20mM NaCl, and

to be in the fast exchange regime. This provides evidence to support the idea that paxillin's affinity of binding with PABP-1 is a sum of its individual interactions. Chemical shift perturbations highlighted several positively charged residues on the solvent exposed face of the β-sheet. This data would indicate that the binding interaction could be electrostatically driven. The idea that the PABP-1-paxillin interaction is more complicated than was first theorised is supported by the experimental titration data. Structural data on the complex show contacting residues that are unique to the PABP-1 RRM2/paxillin LD1 complex such as V131, F142 and K129. However the two PABP-1 ligands, paxillin and poly (A) mRNA, share some contacting residues, particularly F169, F102 and Y140. The crystal structure of PABP-1 RRM1/2 in complex with an 11-nucleotide poly(A) mRNA tail (6) show that the two domains form a trough that binds poly(A) mRNA to give an anti-parallel protein/RNA conformation. The floor of the trough is formed by two adjacent anti-parallel β-sheets with the β-strands; the sides of the RNA-binding trough formed by the S2-S3 loop of RRM 1, the inter-domain linker, S2′-S3′ loop of RRM 2 and part of the inter-domain linker between RRM 2 and RRM 3. With the RNA interaction site spanning both PABP-1 RRM1 and 2 domains and paxillin binding to one specific RRM domain then this suggests a competition for that site may exist. Data did show that mRNA is able to bind with greater affinity to PABP-1 RRM 2 and force paxillin LD1 out of the binding interface. Paxillin LD1 in turn was not able to completely remove all of the PABP-1/poly (A) mRNA complexes and resulted in a mixture of species to be present in the solution even at greater molar excess of the peptide. This suggests that paxillin may overlap with the mRNA to form a sandwich holding the mRNA in place during transport out of the nucleus. The weak binding nature of the PABP-1 RRM2/paxillin LD1 complex also suggests that it occurs after the primary interaction elsewhere between the proteins. An ideal candidate for this, and a candidate for further investigation, would be the interaction between paxillin's LIM domains and PABP-1 RRM4. Mutational studies performed by Woods *et al* on PABP-1 RRM4 suggested that mRNA translocation by paxillin was negated and cellular migration halted.

The NMR structure of PABP-1 RRM 1 in its unbound form yielded a similar fold to that seen in PABP-1 RRM 2. The four anti-parallel arranged β-strands presented a different appearance of solvent exposed residues compared to PABP-1 RRM 2. The surface of the β-sheet did not contain any solvent exposed phenylalanine residues. A single phenylalanine residue was present with its aromatic ring buried in the hydrophobic core, a feature that is shared with PABP-1 RRM 2. Three tyrosine residues, located on $\beta_1$ and $\beta_3$-strands, are solvent exposed

and have been implicated in stacking interactions with mRNA, (142). The α-helices position themselves on one side of the β-sheet in a similar manner as that seen for PABP-1 RRM 2, making hydrophobic interactions with the hydrophobic core. This suggests that the mRNA maybe bound to the PABP-1 RRM1 domain more tightly than to the RRM2 domain thereby allowing paxillin LD1 to occupy part of the mRNA binding site. To confirm a detailed titration series could be performed for PABP-1 RRM1 and RRM2 with mRNA in order to establish binding kinetics for the individual domains.

NMR titration experiments were carried out using the paxillin LD 1 peptide and the same sodium chloride concentrations that were used for PABP-1 RRM 2. No chemical shift perturbations were seen. Every effort was made to reproduce the result that was witnessed during the course of the ELISA binding assays. A larger paxillin fragment was used that encompassed both LD1 and LD 2 domains this too failed to induce any chemical shift perturbations. When this construct was titrated into PABP-1 RRM1/2 only chemical shift perturbations that were specific to the PABP-1 RRM 2 domain were witnessed. The site of the PBS sequence was located within the N-terminus of helix 1; many of the side chains within this region were buried and inaccessible. These data would suggest that the identified PBS site (127) is not involved in paxillin binding to PABP-1.

The linker region between the two RRM domains forms an α-helical formation when in contact with poly(A) mRNA(6). It was initially thought that the linker may provide further binding sites for paxillin in light of its weak interaction with PABP-1 RRM 2. However, further investigation failed to yield any additional chemical shift changes in the PABP-1 RRM1/2 domain spectra. The conformational changes experienced in PABP-1 RRM1/2 linker region upon RNA binding is not limited to just PABP-1 but is also witnessed in nucleolin and HUD RNA binding proteins. The formation of the helix in the linker region between RRM1 and 2 domains would increase PABP-1's affinity for mRNA but may also contribute to paxillin's weak interaction with the RRM2 domain. Of course the conformational change may also prevent paxillin from interacting with the PABP-1 RRM 2 domain when in the presence of full length protein:protein:RNA complexes.

It would be necessary to collect more detailed NMR data for the peptide in both its bound and unbound states. This would facilitate a more accurate structure of the PABP-1 RRM 2/paxillin LD1 complex. A more accurate determination of the binding constant would also

be necessary, this maybe obtained through techniques such as ITC or Biocore. In order to better study the importance of the PABP-1 RRM 2/paxillin LD1 complex it would be beneficial to create functional mutants that were capable of negating the paxillin LD1 interaction but were able to maintain mRNA binding. These mutants would then be used in cell migration assays to determine whether this site was a major contributor in paxillin binding or whether paxillins involvement as a scaffold protein is dependent on multiple sites of interaction. Continued structural research into the other sites of interaction would be required and necessary if an accurate mode of interaction be established.

The most interesting aspects revealed in this report are the absence of the PBS motifs in PABP-1 RRM2 and other paxillin interaction sites within PABP-1, as identified by the ELISA assays. This further complicates identification of paxillin's interaction partners. The identification of LIM domain binding to PABP-1 RRM4 is an interesting find and warrants further investigation. There is more and more evidence being produced that implicate zinc finger binding to RNA (171,172). This may provide clues to one of the more interesting questions as to how actin mRNA constructs are selected for transport to the lamellipodia. It is widely known that PABP-1 specifically binds to the polyadenylated tail of nascent mRNA and it does this arbitrarily. Therefore it maybe paxillin that selects for actin mRNA via its zinc finger LIM domains. The exact mode of action would likely be that PABP-1 binds to the poly(A) tail, providing the mRNA protection from enzymatic digestion, paxillin then identifies the mRNA transcript and binds the PABP-1 to form the export complex. Release of the mRNA at the lamellipodia is likely facilitated through phosphorylation of paxillin. Confirmation of this model would require extensive structural, molecular and cellular investigations but would yield a clear picture into the model. Once the model has been confirmed studies into metastatic cancer models can be performed.

**Appendix I**

## M9 Media

Solution A

| | |
|---|---|
| $Na_2HPO_4$ anh | 14.6g (27.6g of $7H_2O$) |
| $KH_2PO_4$ anh | 5.4g |
| $NH_4Cl$ | 1.0g (for $^{15}N$ and $^{13}C$ labelled samples the ammonium phosphate was replaced with 1g/l $(^{15}NH_4)_2SO_4$ 99% Cambridge Isotope Laboratories) |
| pH 7.2 | |

Solution A was autoclaved and allowed to cool before solution B was added aseptically

Solution B should be filter sterilised and not autoclaved

| | |
|---|---|
| Glucose | 4g (for $^{13}C$ labelled samples glucose was replaced with 2g/l U-$^{13}C_6$ 99% + Cambridge Isotope Laboratories) |
| $MgSO_4$ | 240mg |
| Thiamine Hydrochloride | 20mg |
| 5g/l $MNSO_4$ | 0.5ml |
| 37.5g/l$CaCl_2$ | 0.5ml |
| 1g/l$FeCl_2$ | 0.5ml |
| | |
| Amplicillin | 200mg |
| Kanamycin | 40mg/l |

## LB

| | | |
|---|---|---|
| Tryptone | 10g | |
| Yeast | 5g | |
| NaCl | 10g | pH 7.0 |

Note: for agar plates add 7.5g agar

**Nickel Column Buffers**

Phosphate Stock buffer 1

| | | |
|---|---|---|
| 20mM $Na_2HPO_4$ | 2.84g/l | |
| 500mM NaCl | 29.2g/l | |
| 20mM Imidazole | 1.36g/l | pH 7.5 |

Elution Buffer 2

| | | |
|---|---|---|
| 20mM $Na_2HPO_4$ | 2.84g/l | |
| 500mM NaCl | 29.2g/l | |
| 500mM Imidazole | 34.0g/l | pH 7.5 |

**AcTEV Protease Buffer**

Stocks

| | |
|---|---|
| 1M Tris | pH 8.0 |
| 10mM EDTA | |

To make the working buffer 2ml of each was added to 36ml of $dH_2O$ to give 50mM Tris, 0.5mM EDTA.

**Phosphate buffer**

| | | |
|---|---|---|
| 20mM $Na_2HPO_4$ | 2.84g/l | |
| 100mM NaCl | 5.84g/l | pH 6.5 |
| Made up to 1l with $dH_2O$ | | |

**10X phosphate buffered saline (PBS)**

| | | |
|---|---|---|
| $Na_2HPO_4$ | 14.4g/l | |
| $KH_2PO_4$ | 2.40g/l | |
| NaCl | 80.0g/l | |
| KCl | 2.00g/l | pH 7.0 |

**Citrate phosphate buffer**

Phosphate buffer:

0.2M $Na_2HPO_4$ 14.2g/500ml

Citrate buffer:

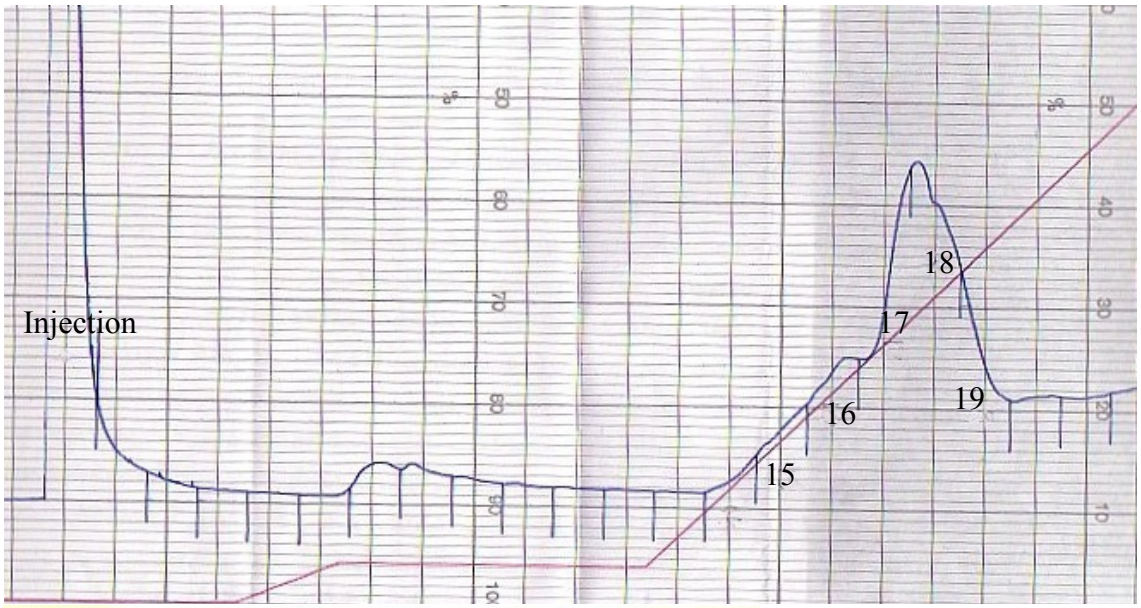
100mM Citric acid 10.6g/500ml

Citrate phosphate buffer:

49.7ml phosphate buffer

48.3ml citrate buffer.

## Appendix II



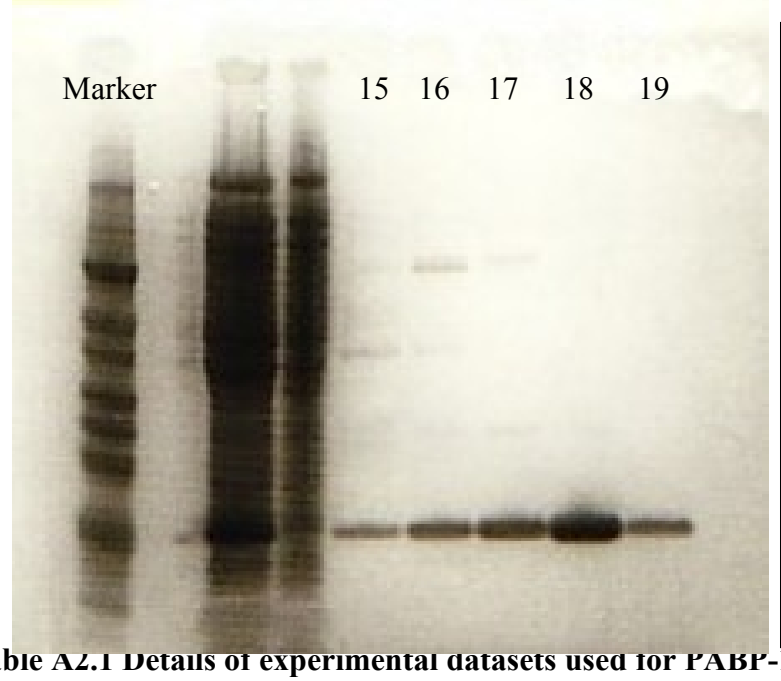




**Figure A2.1A (top) FPLC trace for PABP-1 RRM 2.** Protein yield was excellent and, as a result, was eluted from the column over a number of fractions. **A2.1B SDS-PAGE gel for PABP-1 RRM 2.** This gel shows the resultant FPLC fractions. Although the protein was eluted over a number of fractions the purity was excellent.

**Table A2.1 Details of experimental datasets used for PABP-1 RRM2.**

| Experiment Type | Acquisition Parameters (ms) | | | Field/mixing time |
|---|---|---|---|---|
| | F3 | F2 | F1 | |
| Backbone assignments | | | | |
| $^{15}$N resolved HSQC | | $^{1}$H, $\tau$aq =106.5 | $^{15}$N, $\tau$aq =79.3 | 600MHz |
| HNCACB and CBCA(CO)NH | $^{1}$H, $\tau$aq=106.5 | $^{15}$N, $\tau$aq =19.8 | $^{13}$C, $\tau$aq =4.9 | 600MHz |

| | | | | |
|---|---|---|---|---|
| HNCO | $^{1}$H, τaq=106.5 | $^{15}$N, τaq =22.0 | $^{13}$C, τaq =11.4 | 600MHz |
| Side-chain assignments | | | | |
| HBHA(CO)NH | $^{1}$H, τaq=106.5 | $^{15}$N, τaq =22.0 | $^{1}$H, τaq =15.1 | 600MHz |
| $^{13}$C resolved HSQC | | $^{1}$H, τaq =97.5 | $^{13}$C, τaq =13.3 | 800MHz |
| [H]CCH TOCSY | $^{1}$H, τaq =97.5 | $^{13}$C, τaq =9.3 | $^{13}$C, τaq =9.3 | 800MHz TOCSY mix time=18ms |
| H[C]CH TOCSY | $^{1}$H, τaq =97.5 | $^{13}$C, τaq =9.3 | $^{1}$H, τaq =26.1 | 800MHz TOCSY mix time=18ms |
| NOE assignments | | | | |
| $^{13}$C resolved NOESY-HSQC | $^{1}$H, τaq =90.8 | $^{13}$C, τaq =8.0 | $^{1}$H, τaq =27.3 | 800MHz NOESY mix time=120ms |
| $^{15}$N resolved NOESY-HSQC | $^{1}$H, τaq=102.3 | $^{15}$N, τaq =19.3 | $^{1}$H, τaq =19.8 | 800MHz NOESY mix time=150ms |
| Aromatic ring assignments | | | | |
| $^{13}$C resolved HSQC (aromatic) | | $^{1}$H, τaq =102.3 | $^{13}$C, τaq =12.8 | 800MHz |
| $^{13}$C resolved NOESY-HSQC (aromatic) | $^{1}$H, τaq=102.3 | $^{13}$C, τaq =6.7 | $^{1}$H, τaq =21.5 | 800MHz NOESY mix time=120ms |

**τaq** = approximate acquisition time in ms for that particular dimension.

Chemical Shift Ambiguity Code Definitions

| Codes | Definition |
|---|---|
| 1 | Unique |
| 2 | Ambiguity of geminal atoms or geminal methyl proton groups |
| 3 | Aromatic atoms on opposite sides of the ring (e.g. Tyr HE1 and HE2 protons) |
| 4 | Intraresidue ambiguities (e.g. Lys HG and HD protons) |
| 5 | Interresidue ambiguities (Lys 12 vs. Lys 27) |
| 9 | Ambiguous, specific ambiguity not defined |

| Atom shift assign | Residue Seq code | Residue Label | Atom Name | Atom Type | Shift/ ppm | Error/ ppm | Ambiguity Code |
|---|---|---|---|---|---|---|---|
| 1 | 3 | ASP | CA | C | 52.317 | 0.036 | 1 |
| 2 | 3 | ASP | CB | C | 41.497 | 0.033 | 1 |
| 3 | 3 | ASP | H | H | 8.175 | 0.003 | 1 |
| 4 | 3 | ASP | HA | H | 4.779 | 0.002 | 1 |
| 5 | 3 | ASP | HB2 | H | 2.579 | 0.003 | 1 |
| 6 | 3 | ASP | HB3 | H | 2.704 | 0.005 | 1 |
| 7 | 3 | ASP | N | N | 123.844 | 0.000 | 1 |
| 8 | 5 | SER | CA | C | 59.754 | 0.012 | 1 |
| 9 | 5 | SER | CB | C | 63.544 | 0.102 | 1 |
| 10 | 5 | SER | H | H | 8.447 | 0.001 | 1 |
| 11 | 5 | SER | HA | H | 4.244 | 0.002 | 1 |
| 12 | 5 | SER | HB2 | H | 3.834 | 0.003 | 1 |
| 13 | 5 | SER | HB3 | H | 3.834 | 0.003 | 1 |
| 14 | 5 | SER | N | N | 115.115 | 0.000 | 1 |
| 15 | 6 | LEU | CA | C | 55.446 | 0.017 | 1 |
| 16 | 6 | LEU | CB | C | 42.107 | 0.049 | 1 |
| 17 | 6 | LEU | CD1 | C | 25.135 | 0.068 | 2 |
| 18 | 6 | LEU | CD2 | C | 23.258 | 0.063 | 2 |
| 19 | 6 | LEU | CG | C | 27.000 | 0.007 | 1 |
| 20 | 6 | LEU | H | H | 7.758 | 0.005 | 1 |
| 21 | 6 | LEU | HA | H | 4.242 | 0.002 | 1 |
| 22 | 6 | LEU | HB2 | H | 1.537 | 0.005 | 1 |
| 23 | 6 | LEU | HB3 | H | 1.626 | 0.001 | 1 |
| 24 | 6 | LEU | HD11 | H | 0.876 | 0.004 | 2 |
| 25 | 6 | LEU | HD12 | H | 0.876 | 0.004 | 2 |
| 26 | 6 | LEU | HD13 | H | 0.876 | 0.004 | 2 |
| 27 | 6 | LEU | HD21 | H | 0.779 | 0.003 | 2 |
| 28 | 6 | LEU | HD22 | H | 0.779 | 0.003 | 2 |
| 29 | 6 | LEU | HD23 | H | 0.779 | 0.003 | 2 |
| 30 | 6 | LEU | HG | H | 1.535 | 0.001 | 1 |
| 31 | 6 | LEU | N | N | 122.673 | 0.000 | 1 |
| 32 | 7 | ARG | CA | C | 56.430 | 0.018 | 1 |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 33 | 7 | ARG | CB | C | 30.661 | 0.020 | 1 | | 73 | 11 | VAL | CA | C | 62.964 | 0.008 | 1 |
| 34 | 7 | ARG | CD | C | 43.341 | 0.007 | 1 | | 74 | 11 | VAL | CB | C | 32.623 | 0.005 | 1 |
| 35 | 7 | ARG | CG | C | 27.115 | 0.011 | 1 | | 75 | 11 | VAL | CG1 | C | 20.063 | 0.008 | 1 |
| 36 | 7 | ARG | H | H | 7.830 | 0.014 | 1 | | 76 | 11 | VAL | CG2 | C | 21.445 | 0.017 | 1 |
| 37 | 7 | ARG | HA | H | 4.185 | 0.003 | 1 | | 77 | 11 | VAL | H | H | 7.784 | 0.001 | 1 |
| 38 | 7 | ARG | HB2 | H | 1.697 | 0.003 | 1 | | 78 | 11 | VAL | HA | H | 4.147 | 0.001 | 1 |
| 39 | 7 | ARG | HB3 | H | 1.785 | 0.002 | 1 | | 79 | 11 | VAL | HB | H | 2.153 | 0.001 | 1 |
| 40 | 7 | ARG | HD2 | H | 3.103 | 0.006 | 1 | | 80 | 11 | VAL | HG11 | H | 0.934 | 0.010 | 1 |
| 41 | 7 | ARG | HD3 | H | 3.103 | 0.006 | 1 | | 81 | 11 | VAL | HG12 | H | 0.934 | 0.010 | 1 |
| 42 | 7 | ARG | HG2 | H | 1.527 | 0.005 | 1 | | 82 | 11 | VAL | HG13 | H | 0.934 | 0.010 | 1 |
| 43 | 7 | ARG | HG3 | H | 1.527 | 0.005 | 1 | | 83 | 11 | VAL | HG21 | H | 0.961 | 0.007 | 1 |
| 44 | 7 | ARG | N | N | 120.499 | 0.000 | 1 | | 84 | 11 | VAL | HG22 | H | 0.961 | 0.007 | 1 |
| 45 | 8 | LYS | CA | C | 56.398 | 0.057 | 1 | | 85 | 11 | VAL | HG23 | H | 0.961 | 0.007 | 1 |
| 46 | 8 | LYS | CB | C | 33.001 | 0.056 | 1 | | 86 | 11 | VAL | N | N | 117.404 | 0.000 | 1 |
| 47 | 8 | LYS | CD | C | 28.949 | 0.031 | 1 | | 87 | 12 | GLY | CA | C | 45.608 | 0.000 | 1 |
| 48 | 8 | LYS | CE | C | 42.176 | 0.050 | 1 | | 88 | 12 | GLY | H | H | 8.544 | 0.000 | 1 |
| 49 | 8 | LYS | CG | C | 24.803 | 0.032 | 1 | | 89 | 12 | GLY | HA2 | H | 3.955 | 0.000 | 1 |
| 50 | 8 | LYS | H | H | 8.104 | 0.000 | 1 | | 90 | 12 | GLY | HA3 | H | 3.955 | 0.000 | 1 |
| 51 | 8 | LYS | HA | H | 4.277 | 0.002 | 1 | | 91 | 12 | GLY | N | N | 109.827 | 0.000 | 1 |
| 52 | 8 | LYS | HB2 | H | 1.682 | 0.007 | 1 | | 92 | 13 | ASN | CA | C | 53.513 | 0.007 | 1 |
| 53 | 8 | LYS | HB3 | H | 1.796 | 0.007 | 1 | | 93 | 13 | ASN | CB | C | 39.601 | 0.025 | 1 |
| 54 | 8 | LYS | HD2 | H | 1.607 | 0.005 | 1 | | 94 | 13 | ASN | H | H | 7.918 | 0.001 | 1 |
| 55 | 8 | LYS | HD3 | H | 1.607 | 0.005 | 1 | | 95 | 13 | ASN | HA | H | 5.381 | 0.001 | 1 |
| 56 | 8 | LYS | HE2 | H | 2.900 | 0.007 | 1 | | 96 | 13 | ASN | HB2 | H | 2.453 | 0.002 | 1 |
| 57 | 8 | LYS | HE3 | H | 2.900 | 0.007 | 1 | | 97 | 13 | ASN | HB3 | H | 2.602 | 0.001 | 1 |
| 58 | 8 | LYS | HG2 | H | 1.339 | 0.011 | 1 | | 98 | 13 | ASN | HD21 | H | 6.229 | 0.000 | 1 |
| 59 | 8 | LYS | HG3 | H | 1.390 | 0.005 | 1 | | 99 | 13 | ASN | HD22 | H | 6.978 | 0.000 | 1 |
| 60 | 8 | LYS | N | N | 121.728 | 0.000 | 1 | | 100 | 13 | ASN | N | N | 120.469 | 0.000 | 1 |
| 61 | 9 | SER | CA | C | 58.586 | 0.017 | 1 | | 101 | 13 | ASN | ND2 | N | 110.522 | 0.000 | 1 |
| 62 | 9 | SER | CB | C | 63.967 | 0.026 | 1 | | 102 | 14 | ILE | CA | C | 59.544 | 0.010 | 1 |
| 63 | 9 | SER | H | H | 8.218 | 0.000 | 1 | | 103 | 14 | ILE | CB | C | 41.356 | 0.016 | 1 |
| 64 | 9 | SER | HA | H | 4.408 | 0.003 | 1 | | 104 | 14 | ILE | CD1 | C | 15.008 | 0.006 | 1 |
| 65 | 9 | SER | HB2 | H | 3.838 | 0.009 | 1 | | 105 | 14 | ILE | CG1 | C | 26.515 | 0.023 | 1 |
| 66 | 9 | SER | HB3 | H | 3.838 | 0.009 | 1 | | 106 | 14 | ILE | CG2 | C | 19.045 | 0.012 | 1 |
| 67 | 9 | SER | N | N | 116.345 | 0.000 | 1 | | 107 | 14 | ILE | H | H | 9.148 | 0.002 | 1 |
| 68 | 10 | GLY | CA | C | 45.602 | 0.000 | 1 | | 108 | 14 | ILE | HA | H | 5.199 | 0.004 | 1 |
| 69 | 10 | GLY | H | H | 8.437 | 0.000 | 1 | | 109 | 14 | ILE | HB | H | 1.853 | 0.001 | 1 |
| 70 | 10 | GLY | HA2 | H | 4.016 | 0.000 | 1 | | 110 | 14 | ILE | HD11 | H | 0.457 | 0.002 | 1 |
| 71 | 10 | GLY | HA3 | H | 4.016 | 0.000 | 1 | | 111 | 14 | ILE | HD12 | H | 0.457 | 0.002 | 1 |
| 72 | 10 | GLY | N | N | 110.802 | 0.000 | 1 | | 112 | 14 | ILE | HD13 | H | 0.457 | 0.002 | 1 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 113 | 14 | ILE | HG12 | H | 1.180 | 0.003 | 1 |
| 114 | 14 | ILE | HG13 | H | 1.229 | 0.001 | 1 |
| 115 | 14 | ILE | HG21 | H | 0.866 | 0.001 | 1 |
| 116 | 14 | ILE | HG22 | H | 0.866 | 0.001 | 1 |
| 117 | 14 | ILE | HG23 | H | 0.866 | 0.001 | 1 |
| 118 | 14 | ILE | N | N | 118.218 | 0.000 | 1 |
| 119 | 15 | PHE | CA | C | 54.951 | 0.036 | 1 |
| 120 | 15 | PHE | CB | C | 42.927 | 0.014 | 1 |
| 121 | 15 | PHE | H | H | 8.855 | 0.002 | 1 |
| 122 | 15 | PHE | HA | H | 5.332 | 0.001 | 1 |
| 123 | 15 | PHE | HB2 | H | 2.740 | 0.001 | 1 |
| 124 | 15 | PHE | HB3 | H | 2.937 | 0.004 | 1 |
| 125 | 15 | PHE | N | N | 121.864 | 0.000 | 1 |
| 126 | 16 | ILE | CA | C | 59.328 | 0.021 | 1 |
| 127 | 16 | ILE | CB | C | 41.678 | 0.020 | 1 |
| 128 | 16 | ILE | CD1 | C | 16.697 | 0.004 | 1 |
| 129 | 16 | ILE | CG1 | C | 29.306 | 0.008 | 1 |
| 130 | 16 | ILE | CG2 | C | 19.619 | 0.007 | 1 |
| 131 | 16 | ILE | H | H | 8.359 | 0.001 | 1 |
| 132 | 16 | ILE | HA | H | 5.152 | 0.002 | 1 |
| 133 | 16 | ILE | HB | H | 1.222 | 0.005 | 1 |
| 134 | 16 | ILE | HD11 | H | 0.300 | 0.002 | 1 |
| 135 | 16 | ILE | HD12 | H | 0.300 | 0.002 | 1 |
| 136 | 16 | ILE | HD13 | H | 0.300 | 0.002 | 1 |
| 137 | 16 | ILE | HG12 | H | 0.771 | 0.001 | 1 |
| 138 | 16 | ILE | HG13 | H | 0.771 | 0.001 | 1 |
| 139 | 16 | ILE | HG21 | H | 0.758 | 0.002 | 1 |
| 140 | 16 | ILE | HG22 | H | 0.758 | 0.002 | 1 |
| 141 | 16 | ILE | HG23 | H | 0.758 | 0.002 | 1 |
| 142 | 16 | ILE | N | N | 123.095 | 0.000 | 1 |
| 143 | 17 | LYS | CA | C | 54.876 | 0.012 | 1 |
| 144 | 17 | LYS | CB | C | 36.521 | 0.015 | 1 |
| 145 | 17 | LYS | CD | C | 29.630 | 0.021 | 1 |
| 146 | 17 | LYS | CE | C | 41.725 | 0.018 | 1 |
| 147 | 17 | LYS | CG | C | 25.524 | 0.012 | 1 |
| 148 | 17 | LYS | H | H | 9.179 | 0.002 | 1 |
| 149 | 17 | LYS | HA | H | 4.922 | 0.002 | 1 |
| 150 | 17 | LYS | HB2 | H | 1.757 | 0.002 | 1 |
| 151 | 17 | LYS | HB3 | H | 1.888 | 0.002 | 1 |
| 152 | 17 | LYS | HD2 | H | 1.521 | 0.001 | 1 |
| 153 | 17 | LYS | HD3 | H | 1.521 | 0.001 | 1 |
| 154 | 17 | LYS | HE2 | H | 2.545 | 0.003 | 1 |
| 155 | 17 | LYS | HE3 | H | 2.545 | 0.003 | 1 |
| 156 | 17 | LYS | HG2 | H | 1.296 | 0.008 | 1 |
| 157 | 17 | LYS | HG3 | H | 1.296 | 0.008 | 1 |
| 158 | 17 | LYS | N | N | 123.507 | 0.000 | 1 |
| 159 | 18 | ASN | CA | C | 54.005 | 0.012 | 1 |
| 160 | 18 | ASN | CB | C | 37.932 | 0.012 | 1 |
| 161 | 18 | ASN | H | H | 8.436 | 0.004 | 1 |
| 162 | 18 | ASN | HA | H | 4.435 | 0.001 | 1 |
| 163 | 18 | ASN | HB2 | H | 2.777 | 0.002 | 1 |
| 164 | 18 | ASN | HB3 | H | 3.728 | 0.001 | 1 |
| 165 | 18 | ASN | HD21 | H | 6.959 | 0.002 | 1 |
| 166 | 18 | ASN | HD22 | H | 8.122 | 0.001 | 1 |
| 167 | 18 | ASN | N | N | 117.091 | 0.000 | 1 |
| 168 | 18 | ASN | ND2 | N | 112.345 | 0.000 | 1 |
| 169 | 19 | LEU | CA | C | 53.747 | 0.016 | 1 |
| 170 | 19 | LEU | CB | C | 43.735 | 0.015 | 1 |
| 171 | 19 | LEU | CD1 | C | 25.624 | 0.022 | 1 |
| 172 | 19 | LEU | CD2 | C | 24.170 | 0.030 | 1 |
| 173 | 19 | LEU | CG | C | 26.574 | 0.040 | 1 |
| 174 | 19 | LEU | H | H | 7.857 | 0.001 | 1 |
| 175 | 19 | LEU | HA | H | 4.064 | 0.003 | 1 |
| 176 | 19 | LEU | HB2 | H | 1.096 | 0.001 | 1 |
| 177 | 19 | LEU | HB3 | H | 1.252 | 0.003 | 1 |
| 178 | 19 | LEU | HD11 | H | 0.549 | 0.002 | 1 |
| 179 | 19 | LEU | HD12 | H | 0.549 | 0.002 | 1 |
| 180 | 19 | LEU | HD13 | H | 0.549 | 0.002 | 1 |
| 181 | 19 | LEU | HD21 | H | 0.680 | 0.004 | 1 |
| 182 | 19 | LEU | HD22 | H | 0.680 | 0.004 | 1 |
| 183 | 19 | LEU | HD23 | H | 0.680 | 0.004 | 1 |
| 184 | 19 | LEU | HG | H | 1.347 | 0.003 | 1 |
| 185 | 19 | LEU | N | N | 115.878 | 0.000 | 1 |
| 186 | 20 | ASP | CA | C | 54.879 | 0.014 | 1 |
| 187 | 20 | ASP | CB | C | 43.679 | 0.010 | 1 |
| 188 | 20 | ASP | H | H | 8.641 | 0.005 | 1 |
| 189 | 20 | ASP | HA | H | 4.322 | 0.001 | 1 |
| 190 | 20 | ASP | HB2 | H | 2.768 | 0.001 | 1 |
| 191 | 20 | ASP | HB3 | H | 2.388 | 0.001 | 1 |
| 192 | 20 | ASP | N | N | 123.570 | 0.000 | 1 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 193 | 21 | LYS | CA | C | 59.167 | 0.020 | 1 | | 233 | 24 | ASP | CA | C | 51.446 | 0.021 | 1 | | | |
| 194 | 21 | LYS | CB | C | 32.144 | 0.015 | 1 | | 234 | 24 | ASP | CB | C | 42.257 | 0.018 | 1 | | | |
| 195 | 21 | LYS | CD | C | 29.281 | 0.019 | 1 | | 235 | 24 | ASP | H | H | 7.613 | 0.001 | 1 | | | |
| 196 | 21 | LYS | CE | C | 42.248 | 0.018 | 1 | | 236 | 24 | ASP | HA | H | 4.748 | 0.001 | 1 | | | |
| 197 | 21 | LYS | CG | C | 24.634 | 0.000 | 1 | | 237 | 24 | ASP | HB2 | H | 2.705 | 0.002 | 1 | | | |
| 198 | 21 | LYS | H | H | 8.665 | 0.003 | 1 | | 238 | 24 | ASP | HB3 | H | 3.144 | 0.005 | 1 | | | |
| 199 | 21 | LYS | HA | H | 3.927 | 0.002 | 1 | | 239 | 24 | ASP | N | N | 128.723 | 0.000 | 1 | | | |
| 200 | 21 | LYS | HB2 | H | 1.899 | 0.004 | 1 | | 240 | 25 | ASN | CA | C | 57.079 | 0.008 | 1 | | | |
| 201 | 21 | LYS | HB3 | H | 1.899 | 0.004 | 1 | | 241 | 25 | ASN | CB | C | 37.788 | 0.013 | 1 | | | |
| 202 | 21 | LYS | HD2 | H | 1.680 | 0.004 | 1 | | 242 | 25 | ASN | H | H | 8.288 | 0.002 | 1 | | | |
| 203 | 21 | LYS | HD3 | H | 1.680 | 0.004 | 1 | | 243 | 25 | ASN | HA | H | 4.140 | 0.003 | 1 | | | |
| 204 | 21 | LYS | HE2 | H | 2.985 | 0.002 | 1 | | 244 | 25 | ASN | HB2 | H | 2.788 | 0.002 | 1 | | | |
| 205 | 21 | LYS | HE3 | H | 2.985 | 0.002 | 1 | | 245 | 25 | ASN | HB3 | H | 2.788 | 0.002 | 1 | | | |
| 206 | 21 | LYS | HG2 | H | 1.394 | 0.004 | 1 | | 246 | 25 | ASN | HD21 | H | 6.902 | 0.000 | 1 | | | |
| 207 | 21 | LYS | HG3 | H | 1.394 | 0.004 | 1 | | 247 | 25 | ASN | HD22 | H | 7.473 | 0.002 | 1 | | | |
| 208 | 21 | LYS | N | N | 125.426 | 0.000 | 1 | | 248 | 25 | ASN | N | N | 117.121 | 0.000 | 1 | | | |
| 209 | 22 | SER | CA | C | 59.685 | 0.019 | 1 | | 249 | 25 | ASN | ND2 | N | 110.517 | 0.000 | 1 | | | |
| 210 | 22 | SER | CB | C | 63.710 | 0.040 | 1 | | 250 | 26 | LYS | CA | C | 59.132 | 0.027 | 1 | | | |
| 211 | 22 | SER | H | H | 8.445 | 0.000 | 1 | | 251 | 26 | LYS | CB | C | 31.939 | 0.046 | 1 | | | |
| 212 | 22 | SER | HA | H | 4.243 | 0.002 | 1 | | 252 | 26 | LYS | CD | C | 28.671 | 0.074 | 1 | | | |
| 213 | 22 | SER | HB2 | H | 3.831 | 0.001 | 1 | | 253 | 26 | LYS | CE | C | 42.263 | 0.000 | 1 | | | |
| 214 | 22 | SER | HB3 | H | 3.898 | 0.005 | 1 | | 254 | 26 | LYS | CG | C | 24.769 | 0.017 | 1 | | | |
| 215 | 22 | SER | N | N | 113.823 | 0.000 | 1 | | 255 | 26 | LYS | H | H | 7.724 | 0.004 | 1 | | | |
| 216 | 23 | ILE | CA | C | 61.672 | 0.004 | 1 | | 256 | 26 | LYS | HA | H | 4.017 | 0.002 | 1 | | | |
| 217 | 23 | ILE | CB | C | 35.821 | 0.005 | 1 | | 257 | 26 | LYS | HB2 | H | 1.814 | 0.004 | 1 | | | |
| 218 | 23 | ILE | CD1 | C | 11.802 | 0.005 | 1 | | 258 | 26 | LYS | HB3 | H | 1.899 | 0.004 | 1 | | | |
| 219 | 23 | ILE | CG1 | C | 27.604 | 0.054 | 1 | | 259 | 26 | LYS | HD2 | H | 1.587 | 0.007 | 1 | | | |
| 220 | 23 | ILE | CG2 | C | 17.217 | 0.006 | 1 | | 260 | 26 | LYS | HD3 | H | 1.587 | 0.007 | 1 | | | |
| 221 | 23 | ILE | H | H | 7.460 | 0.001 | 1 | | 261 | 26 | LYS | HE2 | H | 2.993 | 0.007 | 1 | | | |
| 222 | 23 | ILE | HA | H | 3.784 | 0.001 | 1 | | 262 | 26 | LYS | HE3 | H | 2.993 | 0.007 | 1 | | | |
| 223 | 23 | ILE | HB | H | 2.033 | 0.001 | 1 | | 263 | 26 | LYS | HG2 | H | 1.400 | 0.005 | 1 | | | |
| 224 | 23 | ILE | HD11 | H | 0.707 | 0.004 | 1 | | 264 | 26 | LYS | HG3 | H | 1.400 | 0.005 | 1 | | | |
| 225 | 23 | ILE | HD12 | H | 0.707 | 0.004 | 1 | | 265 | 26 | LYS | N | N | 120.252 | 0.000 | 1 | | | |
| 226 | 23 | ILE | HD13 | H | 0.707 | 0.004 | 1 | | 266 | 27 | ALA | CA | C | 54.903 | 0.008 | 1 | | | |
| 227 | 23 | ILE | HG12 | H | 1.246 | 0.003 | 1 | | 267 | 27 | ALA | CB | C | 18.902 | 0.006 | 1 | | | |
| 228 | 23 | ILE | HG13 | H | 1.429 | 0.004 | 1 | | 268 | 27 | ALA | H | H | 8.086 | 0.000 | 1 | | | |
| 229 | 23 | ILE | HG21 | H | 0.836 | 0.004 | 1 | | 269 | 27 | ALA | HA | H | 4.297 | 0.001 | 1 | | | |
| 230 | 23 | ILE | HG22 | H | 0.836 | 0.004 | 1 | | 270 | 27 | ALA | HB1 | H | 1.599 | 0.001 | 1 | | | |
| 231 | 23 | ILE | HG23 | H | 0.836 | 0.004 | 1 | | 271 | 27 | ALA | HB2 | H | 1.599 | 0.001 | 1 | | | |
| 232 | 23 | ILE | N | N | 123.919 | 0.000 | 1 | | 272 | 27 | ALA | HB3 | H | 1.599 | 0.001 | 1 | | | |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 273 | 27 | ALA | N | N | 121.751 | 0.000 | 1 | 313 | 31 | THR | CG2 | C | 21.303 | 0.008 | 1 |
| 274 | 28 | LEU | CA | C | 58.476 | 0.046 | 1 | 314 | 31 | THR | H | H | 7.907 | 0.001 | 1 |
| 275 | 28 | LEU | CB | C | 42.088 | 0.010 | 1 | 315 | 31 | THR | HA | H | 3.843 | 0.001 | 1 |
| 276 | 28 | LEU | CD1 | C | 25.946 | 0.042 | 1 | 316 | 31 | THR | HB | H | 4.239 | 0.001 | 1 |
| 277 | 28 | LEU | CD2 | C | 24.719 | 0.004 | 1 | 317 | 31 | THR | HG21 | H | 0.723 | 0.001 | 1 |
| 278 | 28 | LEU | H | H | 8.506 | 0.001 | 1 | 318 | 31 | THR | HG22 | H | 0.723 | 0.001 | 1 |
| 279 | 28 | LEU | HA | H | 3.960 | 0.003 | 1 | 319 | 31 | THR | HG23 | H | 0.723 | 0.001 | 1 |
| 280 | 28 | LEU | HB2 | H | 1.332 | 0.005 | 1 | 320 | 31 | THR | N | N | 115.634 | 0.000 | 1 |
| 281 | 28 | LEU | HB3 | H | 1.785 | 0.002 | 1 | 321 | 32 | PHE | CA | C | 61.453 | 0.000 | 1 |
| 282 | 28 | LEU | HD11 | H | 0.574 | 0.003 | 1 | 322 | 32 | PHE | CB | C | 39.634 | 0.008 | 1 |
| 283 | 28 | LEU | HD12 | H | 0.574 | 0.003 | 1 | 323 | 32 | PHE | CD1 | C | 132.161 | 0.000 | 3 |
| 284 | 28 | LEU | HD13 | H | 0.574 | 0.003 | 1 | 324 | 32 | PHE | CD2 | C | 132.161 | 0.000 | 3 |
| 285 | 28 | LEU | HD21 | H | 0.719 | 0.005 | 1 | 325 | 32 | PHE | CE1 | C | 131.067 | 0.000 | 3 |
| 286 | 28 | LEU | HD22 | H | 0.719 | 0.005 | 1 | 326 | 32 | PHE | CE2 | C | 131.067 | 0.000 | 3 |
| 287 | 28 | LEU | HD23 | H | 0.719 | 0.005 | 1 | 327 | 32 | PHE | CZ | C | 128.575 | 0.000 | 1 |
| 288 | 28 | LEU | N | N | 121.054 | 0.000 | 1 | 328 | 32 | PHE | H | H | 8.040 | 0.003 | 1 |
| 289 | 29 | TYR | CA | C | 62.213 | 0.037 | 1 | 329 | 32 | PHE | HA | H | 4.624 | 0.002 | 1 |
| 290 | 29 | TYR | CB | C | 38.467 | 0.014 | 1 | 330 | 32 | PHE | HB2 | H | 2.781 | 0.001 | 1 |
| 291 | 29 | TYR | CD1 | C | 133.406 | 0.000 | 3 | 331 | 32 | PHE | HB3 | H | 3.441 | 0.003 | 1 |
| 292 | 29 | TYR | CD2 | C | 133.406 | 0.000 | 3 | 332 | 32 | PHE | HD1 | H | 7.633 | 0.000 | 3 |
| 293 | 29 | TYR | CE1 | C | 118.330 | 0.000 | 3 | 333 | 32 | PHE | HD2 | H | 7.633 | 0.000 | 3 |
| 294 | 29 | TYR | CE2 | C | 118.330 | 0.000 | 3 | 334 | 32 | PHE | HE1 | H | 6.972 | 0.000 | 3 |
| 295 | 29 | TYR | H | H | 8.286 | 0.000 | 1 | 335 | 32 | PHE | HE2 | H | 6.972 | 0.000 | 3 |
| 296 | 29 | TYR | HA | H | 3.802 | 0.001 | 1 | 336 | 32 | PHE | HZ | H | 6.920 | 0.000 | 1 |
| 297 | 29 | TYR | HB2 | H | 3.093 | 0.004 | 1 | 337 | 32 | PHE | N | N | 116.272 | 0.000 | 1 |
| 298 | 29 | TYR | HB3 | H | 3.296 | 0.002 | 1 | 338 | 33 | SER | CA | C | 61.262 | 0.003 | 1 |
| 299 | 29 | TYR | HD1 | H | 7.074 | 0.000 | 3 | 339 | 33 | SER | CB | C | 62.337 | 0.016 | 1 |
| 300 | 29 | TYR | HD2 | H | 7.074 | 0.000 | 3 | 340 | 33 | SER | H | H | 8.511 | 0.001 | 1 |
| 301 | 29 | TYR | HE1 | H | 6.764 | 0.000 | 3 | 341 | 33 | SER | HA | H | 4.197 | 0.001 | 1 |
| 302 | 29 | TYR | HE2 | H | 6.784 | 0.001 | 3 | 342 | 33 | SER | HB2 | H | 3.463 | 0.001 | 1 |
| 303 | 29 | TYR | N | N | 121.424 | 0.000 | 1 | 343 | 33 | SER | HB3 | H | 3.804 | 0.001 | 1 |
| 304 | 30 | ASP | CA | C | 57.560 | 0.006 | 1 | 344 | 33 | SER | N | N | 117.968 | 0.000 | 1 |
| 305 | 30 | ASP | CB | C | 40.125 | 0.015 | 1 | 345 | 34 | ALA | CA | C | 53.581 | 0.012 | 1 |
| 306 | 30 | ASP | H | H | 8.979 | 0.000 | 1 | 346 | 34 | ALA | CB | C | 18.411 | 0.012 | 1 |
| 307 | 30 | ASP | HA | H | 4.296 | 0.001 | 1 | 347 | 34 | ALA | H | H | 7.061 | 0.000 | 1 |
| 308 | 30 | ASP | HB2 | H | 2.675 | 0.001 | 1 | 348 | 34 | ALA | HA | H | 3.884 | 0.002 | 1 |
| 309 | 30 | ASP | HB3 | H | 2.785 | 0.001 | 1 | 349 | 34 | ALA | HB1 | H | 0.701 | 0.002 | 1 |
| 310 | 30 | ASP | N | N | 120.325 | 0.000 | 1 | 350 | 34 | ALA | HB2 | H | 0.701 | 0.002 | 1 |
| 311 | 31 | THR | CA | C | 66.511 | 0.005 | 1 | 351 | 34 | ALA | HB3 | H | 0.701 | 0.002 | 1 |
| 312 | 31 | THR | CB | C | 69.027 | 0.014 | 1 | 352 | 34 | ALA | N | N | 121.434 | 0.000 | 1 |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 353 | 35 | PHE | CA | C | 59.153 | 0.003 | 1 | 393 | 38 | ILE | HD11 | H | 0.241 | 0.002 | 1 |
| 354 | 35 | PHE | CB | C | 39.213 | 0.017 | 1 | 394 | 38 | ILE | HD12 | H | 0.241 | 0.002 | 1 |
| 355 | 35 | PHE | CD1 | C | 131.372 | 0.000 | 3 | 395 | 38 | ILE | HD13 | H | 0.241 | 0.002 | 1 |
| 356 | 35 | PHE | CD2 | C | 131.372 | 0.000 | 3 | 396 | 38 | ILE | HG12 | H | 0.005 | 0.002 | 1 |
| 357 | 35 | PHE | CE1 | C | 131.438 | 0.000 | 3 | 397 | 38 | ILE | HG13 | H | 0.005 | 0.002 | 1 |
| 358 | 35 | PHE | CE2 | C | 131.438 | 0.000 | 3 | 398 | 38 | ILE | HG21 | H | 0.567 | 0.000 | 1 |
| 359 | 35 | PHE | CZ | C | 129.536 | 0.000 | 1 | 399 | 38 | ILE | HG22 | H | 0.567 | 0.000 | 1 |
| 360 | 35 | PHE | H | H | 7.394 | 0.001 | 1 | 400 | 38 | ILE | HG23 | H | 0.567 | 0.000 | 1 |
| 361 | 35 | PHE | HA | H | 4.315 | 0.001 | 1 | 401 | 38 | ILE | N | N | 127.716 | 0.000 | 1 |
| 362 | 35 | PHE | HB2 | H | 2.833 | 0.002 | 1 | 402 | 39 | LEU | CA | C | 56.215 | 0.011 | 1 |
| 363 | 35 | PHE | HB3 | H | 3.171 | 0.002 | 1 | 403 | 39 | LEU | CB | C | 42.230 | 0.011 | 1 |
| 364 | 35 | PHE | HD1 | H | 7.426 | 0.000 | 3 | 404 | 39 | LEU | CD1 | C | 22.263 | 0.004 | 1 |
| 365 | 35 | PHE | HD2 | H | 7.426 | 0.000 | 3 | 405 | 39 | LEU | CD2 | C | 25.656 | 0.002 | 1 |
| 366 | 35 | PHE | HE1 | H | 7.043 | 0.000 | 3 | 406 | 39 | LEU | CG | C | 27.170 | 0.029 | 1 |
| 367 | 35 | PHE | HE2 | H | 7.043 | 0.000 | 3 | 407 | 39 | LEU | H | H | 9.148 | 0.001 | 1 |
| 368 | 35 | PHE | HZ | H | 7.097 | 0.000 | 1 | 408 | 39 | LEU | HA | H | 4.310 | 0.001 | 1 |
| 369 | 35 | PHE | N | N | 112.981 | 0.000 | 1 | 409 | 39 | LEU | HB2 | H | 1.390 | 0.002 | 1 |
| 370 | 36 | GLY | CA | C | 44.690 | 0.026 | 1 | 410 | 39 | LEU | HB3 | H | 1.390 | 0.002 | 1 |
| 371 | 36 | GLY | H | H | 7.283 | 0.001 | 1 | 411 | 39 | LEU | HD11 | H | 0.746 | 0.003 | 1 |
| 372 | 36 | GLY | HA2 | H | 3.871 | 0.001 | 1 | 412 | 39 | LEU | HD12 | H | 0.746 | 0.003 | 1 |
| 373 | 36 | GLY | HA3 | H | 4.164 | 0.003 | 1 | 413 | 39 | LEU | HD13 | H | 0.746 | 0.003 | 1 |
| 374 | 36 | GLY | N | N | 104.274 | 0.000 | 1 | 414 | 39 | LEU | HD21 | H | 0.618 | 0.006 | 1 |
| 375 | 37 | ASN | CA | C | 53.865 | 0.003 | 1 | 415 | 39 | LEU | HD22 | H | 0.618 | 0.006 | 1 |
| 376 | 37 | ASN | CB | C | 38.591 | 0.005 | 1 | 416 | 39 | LEU | HD23 | H | 0.618 | 0.006 | 1 |
| 377 | 37 | ASN | H | H | 8.362 | 0.000 | 1 | 417 | 39 | LEU | HG | H | 1.333 | 0.005 | 1 |
| 378 | 37 | ASN | HA | H | 4.720 | 0.003 | 1 | 418 | 39 | LEU | N | N | 128.188 | 0.000 | 1 |
| 379 | 37 | ASN | HB2 | H | 2.795 | 0.001 | 1 | 419 | 40 | SER | CA | C | 57.828 | 0.006 | 1 |
| 380 | 37 | ASN | HB3 | H | 2.795 | 0.001 | 1 | 420 | 40 | SER | CB | C | 64.662 | 0.010 | 1 |
| 381 | 37 | ASN | HD21 | H | 6.858 | 0.000 | 1 | 421 | 40 | SER | H | H | 7.430 | 0.001 | 1 |
| 382 | 37 | ASN | HD22 | H | 7.846 | 0.000 | 1 | 422 | 40 | SER | HA | H | 4.555 | 0.001 | 1 |
| 383 | 37 | ASN | N | N | 115.916 | 0.000 | 1 | 423 | 40 | SER | HB2 | H | 3.748 | 0.000 | 1 |
| 384 | 37 | ASN | ND2 | N | 113.289 | 0.000 | 1 | 424 | 40 | SER | HB3 | H | 3.802 | 0.001 | 1 |
| 385 | 38 | ILE | CA | C | 62.206 | 0.003 | 1 | 425 | 40 | SER | N | N | 110.477 | 0.000 | 1 |
| 386 | 38 | ILE | CB | C | 39.614 | 0.006 | 1 | 426 | 41 | CYS | CA | C | 57.102 | 0.005 | 1 |
| 387 | 38 | ILE | CD1 | C | 14.086 | 0.006 | 1 | 427 | 41 | CYS | CB | C | 30.177 | 0.020 | 1 |
| 388 | 38 | ILE | CG1 | C | 28.842 | 0.009 | 1 | 428 | 41 | CYS | H | H | 8.414 | 0.000 | 1 |
| 389 | 38 | ILE | CG2 | C | 17.878 | 0.004 | 1 | 429 | 41 | CYS | HA | H | 4.975 | 0.002 | 1 |
| 390 | 38 | ILE | H | H | 8.728 | 0.001 | 1 | 430 | 41 | CYS | HB2 | H | 2.910 | 0.004 | 1 |
| 391 | 38 | ILE | HA | H | 3.875 | 0.002 | 1 | 431 | 41 | CYS | HB3 | H | 2.910 | 0.004 | 1 |
| 392 | 38 | ILE | HB | H | 1.622 | 0.001 | 1 | 432 | 41 | CYS | N | N | 119.322 | 0.000 | 1 |

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 433 | 42 | LYS | CA | C | 55.924 | 0.010 | 1 | | 473 | 44 | VAL | HG21 | H | 0.874 | 0.001 | 1 | | |
| 434 | 42 | LYS | CB | C | 36.416 | 0.018 | 1 | | 474 | 44 | VAL | HG22 | H | 0.874 | 0.001 | 1 | | |
| 435 | 42 | LYS | CD | C | 29.229 | 0.008 | 1 | | 475 | 44 | VAL | HG23 | H | 0.874 | 0.001 | 1 | | |
| 436 | 42 | LYS | CE | C | 42.196 | 0.031 | 1 | | 476 | 44 | VAL | N | N | 129.460 | 0.000 | 1 | | |
| 437 | 42 | LYS | CG | C | 24.379 | 0.033 | 1 | | 477 | 45 | CYS | CA | C | 57.231 | 0.011 | 1 | | |
| 438 | 42 | LYS | H | H | 8.511 | 0.001 | 1 | | 478 | 45 | CYS | CB | C | 31.632 | 0.008 | 1 | | |
| 439 | 42 | LYS | HA | H | 4.520 | 0.002 | 1 | | 479 | 45 | CYS | H | H | 8.310 | 0.000 | 1 | | |
| 440 | 42 | LYS | HB2 | H | 1.717 | 0.004 | 1 | | 480 | 45 | CYS | HA | H | 5.061 | 0.002 | 1 | | |
| 441 | 42 | LYS | HB3 | H | 1.643 | 0.002 | 1 | | 481 | 45 | CYS | HB2 | H | 2.768 | 0.003 | 1 | | |
| 442 | 42 | LYS | HD2 | H | 1.562 | 0.002 | 1 | | 482 | 45 | CYS | HB3 | H | 2.768 | 0.003 | 1 | | |
| 443 | 42 | LYS | HD3 | H | 1.562 | 0.002 | 1 | | 483 | 45 | CYS | N | N | 123.824 | 0.000 | 1 | | |
| 444 | 42 | LYS | HE2 | H | 2.895 | 0.008 | 1 | | 484 | 46 | ASP | CA | C | 52.871 | 0.008 | 1 | | |
| 445 | 42 | LYS | HE3 | H | 2.895 | 0.008 | 1 | | 485 | 46 | ASP | CB | C | 41.959 | 0.016 | 1 | | |
| 446 | 42 | LYS | HG2 | H | 1.256 | 0.001 | 1 | | 486 | 46 | ASP | H | H | 8.686 | 0.001 | 1 | | |
| 447 | 42 | LYS | HG3 | H | 1.256 | 0.001 | 1 | | 487 | 46 | ASP | HA | H | 4.758 | 0.001 | 1 | | |
| 448 | 42 | LYS | N | N | 121.482 | 0.000 | 1 | | 488 | 46 | ASP | HB2 | H | 2.609 | 0.003 | 1 | | |
| 449 | 43 | VAL | CA | C | 62.465 | 0.004 | 1 | | 489 | 46 | ASP | HB3 | H | 3.177 | 0.005 | 1 | | |
| 450 | 43 | VAL | CB | C | 33.262 | 0.009 | 1 | | 490 | 46 | ASP | N | N | 123.320 | 0.000 | 1 | | |
| 451 | 43 | VAL | CG1 | C | 21.222 | 0.010 | 1 | | 491 | 47 | GLU | CA | C | 59.063 | 0.012 | 1 | | |
| 452 | 43 | VAL | CG2 | C | 22.503 | 0.012 | 1 | | 492 | 47 | GLU | CB | C | 28.729 | 0.051 | 1 | | |
| 453 | 43 | VAL | H | H | 8.406 | 0.001 | 1 | | 493 | 47 | GLU | CG | C | 35.789 | 0.043 | 1 | | |
| 454 | 43 | VAL | HA | H | 4.014 | 0.002 | 1 | | 494 | 47 | GLU | H | H | 9.334 | 0.003 | 1 | | |
| 455 | 43 | VAL | HB | H | 1.787 | 0.001 | 1 | | 495 | 47 | GLU | HA | H | 4.107 | 0.001 | 1 | | |
| 456 | 43 | VAL | HG11 | H | 0.781 | 0.008 | 1 | | 496 | 47 | GLU | HB2 | H | 1.986 | 0.001 | 1 | | |
| 457 | 43 | VAL | HG12 | H | 0.781 | 0.008 | 1 | | 497 | 47 | GLU | HB3 | H | 2.035 | 0.009 | 1 | | |
| 458 | 43 | VAL | HG13 | H | 0.781 | 0.008 | 1 | | 498 | 47 | GLU | HG2 | H | 2.226 | 0.003 | 1 | | |
| 459 | 43 | VAL | HG21 | H | 0.689 | 0.006 | 1 | | 499 | 47 | GLU | HG3 | H | 2.269 | 0.003 | 1 | | |
| 460 | 43 | VAL | HG22 | H | 0.689 | 0.006 | 1 | | 500 | 47 | GLU | N | N | 119.827 | 0.000 | 1 | | |
| 461 | 43 | VAL | HG23 | H | 0.689 | 0.006 | 1 | | 501 | 48 | ASN | CA | C | 53.328 | 0.012 | 1 | | |
| 462 | 43 | VAL | N | N | 123.936 | 0.000 | 1 | | 502 | 48 | ASN | CB | C | 39.479 | 0.016 | 1 | | |
| 463 | 44 | VAL | CA | C | 63.984 | 0.011 | 1 | | 503 | 48 | ASN | H | H | 8.559 | 0.001 | 1 | | |
| 464 | 44 | VAL | CB | C | 31.123 | 0.017 | 1 | | 504 | 48 | ASN | HA | H | 4.851 | 0.002 | 1 | | |
| 465 | 44 | VAL | CG1 | C | 21.150 | 0.006 | 1 | | 505 | 48 | ASN | HB2 | H | 2.722 | 0.002 | 1 | | |
| 466 | 44 | VAL | CG2 | C | 22.295 | 0.032 | 1 | | 506 | 48 | ASN | HB3 | H | 2.892 | 0.002 | 1 | | |
| 467 | 44 | VAL | H | H | 8.590 | 0.001 | 1 | | 507 | 48 | ASN | HD21 | H | 6.912 | 0.001 | 1 | | |
| 468 | 44 | VAL | HA | H | 3.810 | 0.002 | 1 | | 508 | 48 | ASN | HD22 | H | 7.794 | 0.001 | 1 | | |
| 469 | 44 | VAL | HB | H | 1.194 | 0.003 | 1 | | 509 | 48 | ASN | N | N | 117.034 | 0.000 | 1 | | |
| 470 | 44 | VAL | HG11 | H | 0.828 | 0.004 | 1 | | 510 | 48 | ASN | ND2 | N | 115.164 | 0.000 | 1 | | |
| 471 | 44 | VAL | HG12 | H | 0.828 | 0.004 | 1 | | 511 | 49 | GLY | CA | C | 44.208 | 0.041 | 1 | | |
| 472 | 44 | VAL | HG13 | H | 0.828 | 0.004 | 1 | | 512 | 49 | GLY | H | H | 7.959 | 0.004 | 1 | | |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 513 | 49 | GLY | HA2 | H | 3.790 | 0.002 | 1 | 553 | 53 | TYR | HD2 | H | 6.780 | 0.000 | 3 |
| 514 | 49 | GLY | HA3 | H | 4.389 | 0.002 | 1 | 554 | 53 | TYR | N | N | 114.269 | 0.000 | 1 |
| 515 | 49 | GLY | N | N | 108.726 | 0.000 | 1 | 555 | 54 | GLY | CA | C | 45.419 | 0.018 | 1 |
| 516 | 50 | SER | CA | C | 58.542 | 0.025 | 1 | 556 | 54 | GLY | H | H | 8.924 | 0.001 | 1 |
| 517 | 50 | SER | CB | C | 63.832 | 0.022 | 1 | 557 | 54 | GLY | HA2 | H | 4.386 | 0.007 | 1 |
| 518 | 50 | SER | H | H | 8.597 | 0.012 | 1 | 558 | 54 | GLY | HA3 | H | 3.745 | 0.000 | 1 |
| 519 | 50 | SER | HA | H | 4.582 | 0.001 | 1 | 559 | 54 | GLY | N | N | 105.750 | 0.000 | 1 |
| 520 | 50 | SER | HB2 | H | 3.699 | 0.001 | 1 | 560 | 55 | PHE | CA | C | 56.130 | 0.006 | 1 |
| 521 | 50 | SER | HB3 | H | 4.003 | 0.001 | 1 | 561 | 55 | PHE | CB | C | 43.717 | 0.048 | 1 |
| 522 | 50 | SER | N | N | 115.102 | 0.000 | 1 | 562 | 55 | PHE | CD1 | C | 131.585 | 0.000 | 3 |
| 523 | 51 | LYS | CA | C | 56.098 | 0.016 | 1 | 563 | 55 | PHE | CD2 | C | 131.585 | 0.000 | 3 |
| 524 | 51 | LYS | CB | C | 32.529 | 0.023 | 1 | 564 | 55 | PHE | CE1 | C | 131.565 | 0.000 | 3 |
| 525 | 51 | LYS | CD | C | 28.846 | 0.014 | 1 | 565 | 55 | PHE | CE2 | C | 131.565 | 0.000 | 3 |
| 526 | 51 | LYS | CE | C | 41.700 | 0.023 | 1 | 566 | 55 | PHE | H | H | 8.681 | 0.001 | 1 |
| 527 | 51 | LYS | CG | C | 25.146 | 0.027 | 1 | 567 | 55 | PHE | HA | H | 5.498 | 0.001 | 1 |
| 528 | 51 | LYS | H | H | 9.225 | 0.003 | 1 | 568 | 55 | PHE | HB2 | H | 3.071 | 0.002 | 1 |
| 529 | 51 | LYS | HA | H | 4.442 | 0.002 | 1 | 569 | 55 | PHE | HB3 | H | 2.676 | 0.002 | 1 |
| 530 | 51 | LYS | HB2 | H | 1.242 | 0.004 | 1 | 570 | 55 | PHE | HD1 | H | 6.974 | 0.000 | 3 |
| 531 | 51 | LYS | HB3 | H | 2.021 | 0.002 | 1 | 571 | 55 | PHE | HD2 | H | 6.974 | 0.000 | 3 |
| 532 | 51 | LYS | HD2 | H | 1.613 | 0.005 | 1 | 572 | 55 | PHE | HE1 | H | 7.132 | 0.000 | 3 |
| 533 | 51 | LYS | HD3 | H | 1.613 | 0.005 | 1 | 573 | 55 | PHE | HE2 | H | 7.132 | 0.000 | 3 |
| 534 | 51 | LYS | HE2 | H | 3.012 | 0.002 | 1 | 574 | 55 | PHE | N | N | 115.890 | 0.000 | 1 |
| 535 | 51 | LYS | HE3 | H | 3.012 | 0.002 | 1 | 575 | 56 | VAL | CA | C | 61.691 | 0.018 | 1 |
| 536 | 51 | LYS | HG2 | H | 1.350 | 0.004 | 1 | 576 | 56 | VAL | CB | C | 34.582 | 0.012 | 1 |
| 537 | 51 | LYS | HG3 | H | 1.350 | 0.004 | 1 | 577 | 56 | VAL | CG1 | C | 21.256 | 0.010 | 1 |
| 538 | 51 | LYS | N | N | 124.811 | 0.000 | 1 | 578 | 56 | VAL | CG2 | C | 21.809 | 0.008 | 1 |
| 539 | 52 | GLY | CA | C | 45.727 | 0.064 | 1 | 579 | 56 | VAL | H | H | 8.064 | 0.003 | 1 |
| 540 | 52 | GLY | H | H | 9.211 | 0.000 | 1 | 580 | 56 | VAL | HA | H | 3.989 | 0.002 | 1 |
| 541 | 52 | GLY | HA2 | H | 3.800 | 0.003 | 1 | 581 | 56 | VAL | HB | H | 1.212 | 0.001 | 1 |
| 542 | 52 | GLY | HA3 | H | 4.145 | 0.003 | 1 | 582 | 56 | VAL | HG11 | H | -0.136 | 0.002 | 1 |
| 543 | 52 | GLY | N | N | 107.804 | 0.000 | 1 | 583 | 56 | VAL | HG12 | H | -0.136 | 0.002 | 1 |
| 544 | 53 | TYR | CA | C | 55.030 | 0.064 | 1 | 584 | 56 | VAL | HG13 | H | -0.136 | 0.002 | 1 |
| 545 | 53 | TYR | CB | C | 40.889 | 0.026 | 1 | 585 | 56 | VAL | HG21 | H | 0.207 | 0.003 | 1 |
| 546 | 53 | TYR | CD1 | C | 133.693 | 0.000 | 3 | 586 | 56 | VAL | HG22 | H | 0.207 | 0.003 | 1 |
| 547 | 53 | TYR | CD2 | C | 133.693 | 0.000 | 3 | 587 | 56 | VAL | HG23 | H | 0.207 | 0.003 | 1 |
| 548 | 53 | TYR | H | H | 7.294 | 0.001 | 1 | 588 | 56 | VAL | N | N | 119.873 | 0.000 | 1 |
| 549 | 53 | TYR | HA | H | 5.307 | 0.002 | 1 | 589 | 57 | HIS | CA | C | 53.102 | 0.018 | 1 |
| 550 | 53 | TYR | HB2 | H | 3.057 | 0.002 | 1 | 590 | 57 | HIS | CB | C | 31.344 | 0.020 | 1 |
| 551 | 53 | TYR | HB3 | H | 3.153 | 0.001 | 1 | 591 | 57 | HIS | CD2 | C | 137.752 | 0.000 | 1 |
| 552 | 53 | TYR | HD1 | H | 6.806 | 0.000 | 3 | 592 | 57 | HIS | CE1 | C | 120.032 | 0.000 | 1 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 593 | 57 | HIS | H | H | 8.639 | 0.000 | 1 | | | 633 | 60 | THR | HG21 | H | 1.262 | 0.002 | 1 | | |
| 594 | 57 | HIS | HA | H | 5.166 | 0.002 | 1 | | | 634 | 60 | THR | HG22 | H | 1.262 | 0.002 | 1 | | |
| 595 | 57 | HIS | HB2 | H | 2.918 | 0.002 | 1 | | | 635 | 60 | THR | HG23 | H | 1.262 | 0.002 | 1 | | |
| 596 | 57 | HIS | HB3 | H | 3.111 | 0.002 | 1 | | | 636 | 60 | THR | N | N | 103.817 | 0.000 | 1 | | |
| 597 | 57 | HIS | HD1 | H | 7.659 | 0.000 | 1 | | | 637 | 61 | GLN | CA | C | 58.080 | 0.005 | 1 | | |
| 598 | 57 | HIS | HE1 | H | 6.803 | 0.000 | 1 | | | 638 | 61 | GLN | CB | C | 28.510 | 0.058 | 1 | | |
| 599 | 57 | HIS | N | N | 126.479 | 0.000 | 1 | | | 639 | 61 | GLN | CG | C | 33.267 | 0.027 | 1 | | |
| 600 | 58 | PHE | CA | C | 58.628 | 0.015 | 1 | | | 640 | 61 | GLN | H | H | 8.799 | 0.002 | 1 | | |
| 601 | 58 | PHE | CB | C | 41.591 | 0.020 | 1 | | | 641 | 61 | GLN | HA | H | 3.957 | 0.002 | 1 | | |
| 602 | 58 | PHE | CD1 | C | 132.111 | 0.000 | 3 | | | 642 | 61 | GLN | HB2 | H | 2.080 | 0.005 | 1 | | |
| 603 | 58 | PHE | CD2 | C | 132.111 | 0.000 | 3 | | | 643 | 61 | GLN | HB3 | H | 1.938 | 0.002 | 1 | | |
| 604 | 58 | PHE | CE1 | C | 131.427 | 0.000 | 3 | | | 644 | 61 | GLN | HE21 | H | 6.930 | 0.000 | 1 | | |
| 605 | 58 | PHE | CE2 | C | 131.427 | 0.000 | 3 | | | 645 | 61 | GLN | HE22 | H | 7.874 | 0.000 | 1 | | |
| 606 | 58 | PHE | CZ | C | 129.422 | 0.000 | 1 | | | 646 | 61 | GLN | HG2 | H | 2.362 | 0.001 | 1 | | |
| 607 | 58 | PHE | H | H | 9.139 | 0.003 | 1 | | | 647 | 61 | GLN | HG3 | H | 2.362 | 0.001 | 1 | | |
| 608 | 58 | PHE | HA | H | 5.005 | 0.002 | 1 | | | 648 | 61 | GLN | N | N | 121.388 | 0.000 | 1 | | |
| 609 | 58 | PHE | HB2 | H | 3.248 | 0.001 | 1 | | | 649 | 61 | GLN | NE2 | N | 112.554 | 0.000 | 1 | | |
| 610 | 58 | PHE | HB3 | H | 3.402 | 0.002 | 1 | | | 650 | 62 | GLU | CA | C | 60.134 | 0.009 | 1 | | |
| 611 | 58 | PHE | HD1 | H | 7.043 | 0.001 | 3 | | | 651 | 62 | GLU | CB | C | 28.842 | 0.034 | 1 | | |
| 612 | 58 | PHE | HD2 | H | 7.043 | 0.001 | 3 | | | 652 | 62 | GLU | CG | C | 36.696 | 0.020 | 1 | | |
| 613 | 58 | PHE | HE1 | H | 7.195 | 0.000 | 3 | | | 653 | 62 | GLU | H | H | 8.793 | 0.000 | 1 | | |
| 614 | 58 | PHE | HE2 | H | 7.195 | 0.000 | 3 | | | 654 | 62 | GLU | HA | H | 3.891 | 0.001 | 1 | | |
| 615 | 58 | PHE | HZ | H | 7.531 | 0.003 | 1 | | | 655 | 62 | GLU | HB2 | H | 1.880 | 0.001 | 1 | | |
| 616 | 58 | PHE | N | N | 124.688 | 0.000 | 1 | | | 656 | 62 | GLU | HB3 | H | 2.047 | 0.007 | 1 | | |
| 617 | 59 | GLU | CA | C | 58.822 | 0.000 | 1 | | | 657 | 62 | GLU | HG2 | H | 2.255 | 0.005 | 1 | | |
| 618 | 59 | GLU | CB | C | 31.185 | 0.020 | 1 | | | 658 | 62 | GLU | HG3 | H | 2.255 | 0.005 | 1 | | |
| 619 | 59 | GLU | CG | C | 36.647 | 0.050 | 1 | | | 659 | 62 | GLU | N | N | 118.601 | 0.000 | 1 | | |
| 620 | 59 | GLU | H | H | 8.443 | 0.001 | 1 | | | 660 | 63 | ALA | CA | C | 54.960 | 0.014 | 1 | | |
| 621 | 59 | GLU | HA | H | 3.948 | 0.003 | 1 | | | 661 | 63 | ALA | CB | C | 20.130 | 0.010 | 1 | | |
| 622 | 59 | GLU | HB2 | H | 1.985 | 0.004 | 1 | | | 662 | 63 | ALA | H | H | 7.443 | 0.001 | 1 | | |
| 623 | 59 | GLU | HB3 | H | 2.363 | 0.008 | 1 | | | 663 | 63 | ALA | HA | H | 3.700 | 0.002 | 1 | | |
| 624 | 59 | GLU | HG2 | H | 2.322 | 0.011 | 1 | | | 664 | 63 | ALA | HB1 | H | 0.888 | 0.001 | 1 | | |
| 625 | 59 | GLU | HG3 | H | 2.322 | 0.011 | 1 | | | 665 | 63 | ALA | HB2 | H | 0.888 | 0.001 | 1 | | |
| 626 | 59 | GLU | N | N | 118.726 | 0.000 | 1 | | | 666 | 63 | ALA | HB3 | H | 0.888 | 0.001 | 1 | | |
| 627 | 60 | THR | CA | C | 58.495 | 0.010 | 1 | | | 667 | 63 | ALA | N | N | 121.885 | 0.000 | 1 | | |
| 628 | 60 | THR | CB | C | 72.132 | 0.000 | 1 | | | 668 | 64 | ALA | CA | C | 54.984 | 0.012 | 1 | | |
| 629 | 60 | THR | CG2 | C | 21.826 | 0.013 | 1 | | | 669 | 64 | ALA | CB | C | 18.193 | 0.012 | 1 | | |
| 630 | 60 | THR | H | H | 7.859 | 0.000 | 1 | | | 670 | 64 | ALA | H | H | 6.508 | 0.001 | 1 | | |
| 631 | 60 | THR | HA | H | 4.830 | 0.003 | 1 | | | 671 | 64 | ALA | HA | H | 3.572 | 0.001 | 1 | | |
| 632 | 60 | THR | HB | H | 4.662 | 0.002 | 1 | | | 672 | 64 | ALA | HB1 | H | 1.429 | 0.002 | 1 | | |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 673 | 64 | ALA | HB2 | H | 1.429 | 0.002 | 1 | | | 713 | 68 | ILE | HA | H | 3.142 | 0.003 | 1 | | |
| 674 | 64 | ALA | HB3 | H | 1.429 | 0.002 | 1 | | | 714 | 68 | ILE | HB | H | 1.655 | 0.003 | 1 | | |
| 675 | 64 | ALA | N | N | 117.725 | 0.000 | 1 | | | 715 | 68 | ILE | HD11 | H | 0.687 | 0.002 | 1 | | |
| 676 | 65 | GLU | CA | C | 59.132 | 0.019 | 1 | | | 716 | 68 | ILE | HD12 | H | 0.687 | 0.002 | 1 | | |
| 677 | 65 | GLU | CB | C | 29.022 | 0.004 | 1 | | | 717 | 68 | ILE | HD13 | H | 0.687 | 0.002 | 1 | | |
| 678 | 65 | GLU | CG | C | 36.529 | 0.008 | 1 | | | 718 | 68 | ILE | HG12 | H | 0.461 | 0.002 | 1 | | |
| 679 | 65 | GLU | H | H | 8.139 | 0.000 | 1 | | | 719 | 68 | ILE | HG13 | H | 0.461 | 0.002 | 1 | | |
| 680 | 65 | GLU | HA | H | 3.787 | 0.001 | 1 | | | 720 | 68 | ILE | HG21 | H | 0.818 | 0.008 | 1 | | |
| 681 | 65 | GLU | HB2 | H | 1.990 | 0.004 | 1 | | | 721 | 68 | ILE | HG22 | H | 0.818 | 0.008 | 1 | | |
| 682 | 65 | GLU | HB3 | H | 1.990 | 0.004 | 1 | | | 722 | 68 | ILE | HG23 | H | 0.818 | 0.008 | 1 | | |
| 683 | 65 | GLU | HG2 | H | 2.392 | 0.002 | 1 | | | 723 | 68 | ILE | N | N | 117.448 | 0.000 | 1 | | |
| 684 | 65 | GLU | HG3 | H | 2.392 | 0.002 | 1 | | | 724 | 69 | GLU | CA | C | 59.232 | 0.006 | 1 | | |
| 685 | 65 | GLU | N | N | 115.078 | 0.000 | 1 | | | 725 | 69 | GLU | CB | C | 29.871 | 0.034 | 1 | | |
| 686 | 66 | ARG | CA | C | 59.053 | 0.018 | 1 | | | 726 | 69 | GLU | CG | C | 36.115 | 0.098 | 1 | | |
| 687 | 66 | ARG | CB | C | 30.661 | 0.034 | 1 | | | 727 | 69 | GLU | H | H | 7.710 | 0.000 | 1 | | |
| 688 | 66 | ARG | CD | C | 43.740 | 0.011 | 1 | | | 728 | 69 | GLU | HA | H | 3.862 | 0.002 | 1 | | |
| 689 | 66 | ARG | CG | C | 28.015 | 0.020 | 1 | | | 729 | 69 | GLU | HB2 | H | 2.006 | 0.012 | 1 | | |
| 690 | 66 | ARG | H | H | 7.889 | 0.004 | 1 | | | 730 | 69 | GLU | HB3 | H | 2.066 | 0.005 | 1 | | |
| 691 | 66 | ARG | HA | H | 3.976 | 0.001 | 1 | | | 731 | 69 | GLU | HG2 | H | 2.293 | 0.004 | 1 | | |
| 692 | 66 | ARG | HB2 | H | 1.831 | 0.004 | 1 | | | 732 | 69 | GLU | HG3 | H | 2.293 | 0.004 | 1 | | |
| 693 | 66 | ARG | HB3 | H | 1.983 | 0.005 | 1 | | | 733 | 69 | GLU | N | N | 117.850 | 0.000 | 1 | | |
| 694 | 66 | ARG | HD2 | H | 3.320 | 0.005 | 1 | | | 734 | 70 | LYS | CA | C | 56.555 | 0.081 | 1 | | |
| 695 | 66 | ARG | HD3 | H | 3.320 | 0.005 | 1 | | | 735 | 70 | LYS | CB | C | 33.178 | 0.022 | 1 | | |
| 696 | 66 | ARG | HG2 | H | 1.676 | 0.004 | 1 | | | 736 | 70 | LYS | CD | C | 28.919 | 0.000 | 1 | | |
| 697 | 66 | ARG | HG3 | H | 1.676 | 0.004 | 1 | | | 737 | 70 | LYS | CE | C | 41.938 | 0.000 | 1 | | |
| 698 | 66 | ARG | N | N | 120.433 | 0.000 | 1 | | | 738 | 70 | LYS | CG | C | 24.676 | 0.000 | 1 | | |
| 699 | 67 | ALA | CA | C | 54.554 | 0.011 | 1 | | | 739 | 70 | LYS | H | H | 7.910 | 0.000 | 1 | | |
| 700 | 67 | ALA | CB | C | 19.337 | 0.008 | 1 | | | 740 | 70 | LYS | HA | H | 4.224 | 0.001 | 1 | | |
| 701 | 67 | ALA | H | H | 7.836 | 0.001 | 1 | | | 741 | 70 | LYS | HB2 | H | 1.710 | 0.001 | 1 | | |
| 702 | 67 | ALA | HA | H | 2.299 | 0.004 | 1 | | | 742 | 70 | LYS | HB3 | H | 1.710 | 0.001 | 1 | | |
| 703 | 67 | ALA | HB1 | H | 1.230 | 0.000 | 1 | | | 743 | 70 | LYS | HD2 | H | 1.601 | 0.000 | 1 | | |
| 704 | 67 | ALA | HB2 | H | 1.230 | 0.000 | 1 | | | 744 | 70 | LYS | HD3 | H | 1.601 | 0.000 | 1 | | |
| 705 | 67 | ALA | HB3 | H | 1.230 | 0.000 | 1 | | | 745 | 70 | LYS | HE2 | H | 2.961 | 0.000 | 1 | | |
| 706 | 67 | ALA | N | N | 121.564 | 0.000 | 1 | | | 746 | 70 | LYS | HE3 | H | 2.975 | 0.007 | 1 | | |
| 707 | 68 | ILE | CA | C | 65.915 | 0.007 | 1 | | | 747 | 70 | LYS | HG2 | H | 1.371 | 0.000 | 1 | | |
| 708 | 68 | ILE | CB | C | 38.873 | 0.013 | 1 | | | 748 | 70 | LYS | HG3 | H | 1.371 | 0.000 | 1 | | |
| 709 | 68 | ILE | CD1 | C | 14.511 | 0.008 | 1 | | | 749 | 70 | LYS | N | N | 114.338 | 0.000 | 1 | | |
| 710 | 68 | ILE | CG1 | C | 29.961 | 0.009 | 1 | | | 750 | 71 | MET | CA | C | 53.784 | 0.021 | 1 | | |
| 711 | 68 | ILE | CG2 | C | 17.210 | 0.049 | 1 | | | 751 | 71 | MET | CB | C | 32.454 | 0.080 | 1 | | |
| 712 | 68 | ILE | H | H | 7.894 | 0.004 | 1 | | | 752 | 71 | MET | CE | C | 16.032 | 0.000 | 1 | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 753 | 71 | MET | CG | C | 32.892 | 0.029 | 1 | |
| 754 | 71 | MET | H | H | 7.926 | 0.003 | 1 | |
| 755 | 71 | MET | HA | H | 4.792 | 0.002 | 1 | |
| 756 | 71 | MET | HB2 | H | 2.121 | 0.000 | 1 | |
| 757 | 71 | MET | HB3 | H | 2.121 | 0.000 | 1 | |
| 758 | 71 | MET | HG2 | H | 2.219 | 0.004 | 1 | |
| 759 | 71 | MET | HG3 | H | 2.219 | 0.004 | 1 | |
| 760 | 71 | MET | N | N | 112.284 | 0.000 | 1 | |
| 761 | 72 | ASN | CA | C | 56.748 | 0.023 | 1 | |
| 762 | 72 | ASN | CB | C | 38.410 | 0.011 | 1 | |
| 763 | 72 | ASN | H | H | 7.953 | 0.001 | 1 | |
| 764 | 72 | ASN | HA | H | 4.325 | 0.002 | 1 | |
| 765 | 72 | ASN | HB2 | H | 2.926 | 0.002 | 1 | |
| 766 | 72 | ASN | HB3 | H | 3.075 | 0.002 | 1 | |
| 767 | 72 | ASN | HD21 | H | 6.780 | 0.000 | 1 | |
| 768 | 72 | ASN | HD22 | H | 7.646 | 0.000 | 1 | |
| 769 | 72 | ASN | N | N | 116.361 | 0.000 | 1 | |
| 770 | 72 | ASN | ND2 | N | 113.456 | 0.000 | 1 | |
| 771 | 73 | GLY | CA | C | 45.922 | 0.021 | 1 | |
| 772 | 73 | GLY | H | H | 8.891 | 0.000 | 1 | |
| 773 | 73 | GLY | HA2 | H | 3.533 | 0.003 | 1 | |
| 774 | 73 | GLY | HA3 | H | 4.152 | 0.005 | 1 | |
| 775 | 73 | GLY | N | N | 117.240 | 0.000 | 1 | |
| 776 | 74 | MET | CA | C | 54.169 | 0.030 | 1 | |
| 777 | 74 | MET | CB | C | 33.029 | 0.023 | 1 | |
| 778 | 74 | MET | CG | C | 33.015 | 0.068 | 1 | |
| 779 | 74 | MET | H | H | 7.778 | 0.004 | 1 | |
| 780 | 74 | MET | HA | H | 4.490 | 0.000 | 1 | |
| 781 | 74 | MET | HB2 | H | 1.951 | 0.003 | 1 | |
| 782 | 74 | MET | HB3 | H | 2.128 | 0.003 | 1 | |
| 783 | 74 | MET | HG2 | H | 2.346 | 0.002 | 1 | |
| 784 | 74 | MET | HG3 | H | 2.819 | 0.001 | 1 | |
| 785 | 74 | MET | N | N | 120.003 | 0.000 | 1 | |
| 786 | 75 | LEU | CA | C | 54.639 | 0.031 | 1 | |
| 787 | 75 | LEU | CB | C | 42.724 | 0.015 | 1 | |
| 788 | 75 | LEU | CD1 | C | 23.952 | 0.032 | 1 | |
| 789 | 75 | LEU | CD2 | C | 24.886 | 0.050 | 1 | |
| 790 | 75 | LEU | CG | C | 27.231 | 0.053 | 1 | |
| 791 | 75 | LEU | H | H | 8.011 | 0.003 | 1 | |
| 792 | 75 | LEU | HA | H | 4.507 | 0.006 | 1 | |
| 793 | 75 | LEU | HB2 | H | 1.618 | 0.002 | 1 | |
| 794 | 75 | LEU | HB3 | H | 1.253 | 0.002 | 1 | |
| 795 | 75 | LEU | HD11 | H | 0.789 | 0.003 | 1 | |
| 796 | 75 | LEU | HD12 | H | 0.789 | 0.003 | 1 | |
| 797 | 75 | LEU | HD13 | H | 0.789 | 0.003 | 1 | |
| 798 | 75 | LEU | HD21 | H | 0.661 | 0.004 | 1 | |
| 799 | 75 | LEU | HD22 | H | 0.661 | 0.004 | 1 | |
| 800 | 75 | LEU | HD23 | H | 0.661 | 0.004 | 1 | |
| 801 | 75 | LEU | HG | H | 1.494 | 0.002 | 1 | |
| 802 | 75 | LEU | N | N | 120.585 | 0.000 | 1 | |
| 803 | 76 | LEU | CA | C | 54.558 | 0.000 | 1 | |
| 804 | 76 | LEU | CB | C | 43.850 | 0.015 | 1 | |
| 805 | 76 | LEU | CD1 | C | 24.601 | 0.000 | 1 | |
| 806 | 76 | LEU | CD2 | C | 24.162 | 0.007 | 1 | |
| 807 | 76 | LEU | CG | C | 27.361 | 0.024 | 1 | |
| 808 | 76 | LEU | H | H | 8.713 | 0.001 | 1 | |
| 809 | 76 | LEU | HA | H | 4.537 | 0.002 | 1 | |
| 810 | 76 | LEU | HB2 | H | 1.287 | 0.002 | 1 | |
| 811 | 76 | LEU | HB3 | H | 1.692 | 0.002 | 1 | |
| 812 | 76 | LEU | HD11 | H | 0.799 | 0.008 | 1 | |
| 813 | 76 | LEU | HD12 | H | 0.799 | 0.008 | 1 | |
| 814 | 76 | LEU | HD13 | H | 0.799 | 0.008 | 1 | |
| 815 | 76 | LEU | HD21 | H | 0.832 | 0.001 | 1 | |
| 816 | 76 | LEU | HD22 | H | 0.832 | 0.001 | 1 | |
| 817 | 76 | LEU | HD23 | H | 0.832 | 0.001 | 1 | |
| 818 | 76 | LEU | HG | H | 1.439 | 0.002 | 1 | |
| 819 | 76 | LEU | N | N | 126.468 | 0.000 | 1 | |
| 820 | 77 | ASN | CA | C | 54.987 | 0.014 | 1 | |
| 821 | 77 | ASN | CB | C | 36.330 | 0.021 | 1 | |
| 822 | 77 | ASN | H | H | 9.378 | 0.003 | 1 | |
| 823 | 77 | ASN | HA | H | 4.124 | 0.001 | 1 | |
| 824 | 77 | ASN | HB2 | H | 2.645 | 0.001 | 1 | |
| 825 | 77 | ASN | HB3 | H | 2.949 | 0.004 | 1 | |
| 826 | 77 | ASN | HD21 | H | 6.644 | 0.000 | 1 | |
| 827 | 77 | ASN | HD22 | H | 7.557 | 0.000 | 1 | |
| 828 | 77 | ASN | N | N | 126.839 | 0.000 | 1 | |
| 829 | 77 | ASN | ND2 | N | 110.740 | 0.000 | 1 | |
| 830 | 78 | ASP | CA | C | 56.174 | 0.005 | 1 | |
| 831 | 78 | ASP | CB | C | 40.308 | 0.030 | 1 | |
| 832 | 78 | ASP | H | H | 8.497 | 0.000 | 1 | |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 833 | 78 | ASP | HA | H | 4.191 | 0.002 | 1 | | | 873 | 81 | VAL | HG11 | H | 0.943 | 0.004 | 1 | | |
| 834 | 78 | ASP | HB2 | H | 2.757 | 0.002 | 1 | | | 874 | 81 | VAL | HG12 | H | 0.943 | 0.004 | 1 | | |
| 835 | 78 | ASP | HB3 | H | 2.867 | 0.002 | 1 | | | 875 | 81 | VAL | HG13 | H | 0.943 | 0.004 | 1 | | |
| 836 | 78 | ASP | N | N | 110.440 | 0.000 | 1 | | | 876 | 81 | VAL | HG21 | H | 0.965 | 0.001 | 1 | | |
| 837 | 79 | ARG | CA | C | 54.775 | 0.016 | 1 | | | 877 | 81 | VAL | HG22 | H | 0.965 | 0.001 | 1 | | |
| 838 | 79 | ARG | CB | C | 33.111 | 0.034 | 1 | | | 878 | 81 | VAL | HG23 | H | 0.965 | 0.001 | 1 | | |
| 839 | 79 | ARG | CD | C | 43.386 | 0.029 | 1 | | | 879 | 81 | VAL | N | N | 123.334 | 0.000 | 1 | | |
| 840 | 79 | ARG | CG | C | 27.228 | 0.014 | 1 | | | 880 | 82 | PHE | CA | C | 55.746 | 0.017 | 1 | | |
| 841 | 79 | ARG | H | H | 7.545 | 0.003 | 1 | | | 881 | 82 | PHE | CB | C | 41.183 | 0.018 | 1 | | |
| 842 | 79 | ARG | HA | H | 4.563 | 0.001 | 1 | | | 882 | 82 | PHE | CD1 | C | 132.089 | 0.000 | 3 | | |
| 843 | 79 | ARG | HB2 | H | 1.758 | 0.003 | 1 | | | 883 | 82 | PHE | CD2 | C | 132.089 | 0.000 | 3 | | |
| 844 | 79 | ARG | HB3 | H | 1.867 | 0.002 | 1 | | | 884 | 82 | PHE | CE1 | C | 131.327 | 0.000 | 3 | | |
| 845 | 79 | ARG | HD2 | H | 3.066 | 0.006 | 1 | | | 885 | 82 | PHE | CE2 | C | 131.327 | 0.000 | 3 | | |
| 846 | 79 | ARG | HD3 | H | 3.066 | 0.006 | 1 | | | 886 | 82 | PHE | H | H | 8.005 | 0.001 | 1 | | |
| 847 | 79 | ARG | HG2 | H | 1.620 | 0.002 | 1 | | | 887 | 82 | PHE | HA | H | 5.194 | 0.003 | 1 | | |
| 848 | 79 | ARG | HG3 | H | 1.620 | 0.002 | 1 | | | 888 | 82 | PHE | HB2 | H | 2.986 | 0.006 | 1 | | |
| 849 | 79 | ARG | N | N | 119.023 | 0.000 | 1 | | | 889 | 82 | PHE | HB3 | H | 2.888 | 0.002 | 1 | | |
| 850 | 80 | LYS | CA | C | 56.511 | 0.008 | 1 | | | 890 | 82 | PHE | HD1 | H | 7.196 | 0.000 | 3 | | |
| 851 | 80 | LYS | CB | C | 32.636 | 0.010 | 1 | | | 891 | 82 | PHE | HD2 | H | 7.196 | 0.000 | 3 | | |
| 852 | 80 | LYS | CD | C | 29.581 | 0.051 | 1 | | | 892 | 82 | PHE | HE1 | H | 7.313 | 0.000 | 3 | | |
| 853 | 80 | LYS | CE | C | 41.918 | 0.049 | 1 | | | 893 | 82 | PHE | HE2 | H | 7.313 | 0.000 | 3 | | |
| 854 | 80 | LYS | CG | C | 24.927 | 0.009 | 1 | | | 894 | 82 | PHE | N | N | 123.168 | 0.000 | 1 | | |
| 855 | 80 | LYS | H | H | 8.583 | 0.007 | 1 | | | 895 | 83 | VAL | CA | C | 60.226 | 0.030 | 1 | | |
| 856 | 80 | LYS | HA | H | 4.409 | 0.001 | 1 | | | 896 | 83 | VAL | CB | C | 34.616 | 0.013 | 1 | | |
| 857 | 80 | LYS | HB2 | H | 1.546 | 0.003 | 1 | | | 897 | 83 | VAL | CG1 | C | 23.418 | 0.105 | 1 | | |
| 858 | 80 | LYS | HB3 | H | 1.546 | 0.003 | 1 | | | 898 | 83 | VAL | CG2 | C | 22.124 | 0.079 | 1 | | |
| 859 | 80 | LYS | HD2 | H | 1.638 | 0.006 | 1 | | | 899 | 83 | VAL | H | H | 8.187 | 0.001 | 1 | | |
| 860 | 80 | LYS | HD3 | H | 1.638 | 0.006 | 1 | | | 900 | 83 | VAL | HA | H | 4.723 | 0.003 | 1 | | |
| 861 | 80 | LYS | HE2 | H | 2.963 | 0.004 | 1 | | | 901 | 83 | VAL | HB | H | 1.564 | 0.007 | 1 | | |
| 862 | 80 | LYS | HE3 | H | 2.963 | 0.004 | 1 | | | 902 | 83 | VAL | HG11 | H | 0.868 | 0.002 | 1 | | |
| 863 | 80 | LYS | HG2 | H | 1.060 | 0.002 | 1 | | | 903 | 83 | VAL | HG12 | H | 0.868 | 0.002 | 1 | | |
| 864 | 80 | LYS | HG3 | H | 1.060 | 0.002 | 1 | | | 904 | 83 | VAL | HG13 | H | 0.868 | 0.002 | 1 | | |
| 865 | 80 | LYS | N | N | 124.992 | 0.000 | 1 | | | 905 | 83 | VAL | HG21 | H | 0.770 | 0.005 | 1 | | |
| 866 | 81 | VAL | CA | C | 60.793 | 0.028 | 1 | | | 906 | 83 | VAL | HG22 | H | 0.770 | 0.005 | 1 | | |
| 867 | 81 | VAL | CB | C | 33.714 | 0.003 | 1 | | | 907 | 83 | VAL | HG23 | H | 0.770 | 0.005 | 1 | | |
| 868 | 81 | VAL | CG1 | C | 22.992 | 0.010 | 1 | | | 908 | 83 | VAL | N | N | 126.188 | 0.000 | 1 | | |
| 869 | 81 | VAL | CG2 | C | 19.756 | 0.006 | 1 | | | 909 | 84 | GLY | CA | C | 44.884 | 0.043 | 1 | | |
| 870 | 81 | VAL | H | H | 8.606 | 0.002 | 1 | | | 910 | 84 | GLY | H | H | 7.954 | 0.001 | 1 | | |
| 871 | 81 | VAL | HA | H | 4.685 | 0.004 | 1 | | | 911 | 84 | GLY | HA2 | H | 3.550 | 0.005 | 1 | | |
| 872 | 81 | VAL | HB | H | 2.117 | 0.002 | 1 | | | 912 | 84 | GLY | HA3 | H | 4.161 | 0.005 | 1 | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 913 | 84 | GLY | N | N | 111.229 | 0.000 | 1 |
| 914 | 85 | ARG | CA | C | 55.800 | 0.015 | 1 |
| 915 | 85 | ARG | CB | C | 31.426 | 0.029 | 1 |
| 916 | 85 | ARG | CD | C | 43.287 | 0.025 | 1 |
| 917 | 85 | ARG | CG | C | 27.802 | 0.039 | 1 |
| 918 | 85 | ARG | H | H | 8.520 | 0.003 | 1 |
| 919 | 85 | ARG | HA | H | 4.577 | 0.002 | 1 |
| 920 | 85 | ARG | HB2 | H | 1.741 | 0.002 | 1 |
| 921 | 85 | ARG | HB3 | H | 1.868 | 0.005 | 1 |
| 922 | 85 | ARG | HD2 | H | 3.171 | 0.007 | 1 |
| 923 | 85 | ARG | HD3 | H | 3.171 | 0.007 | 1 |
| 924 | 85 | ARG | HG2 | H | 1.611 | 0.008 | 1 |
| 925 | 85 | ARG | HG3 | H | 1.611 | 0.008 | 1 |
| 926 | 85 | ARG | N | N | 118.635 | 0.000 | 1 |
| 927 | 86 | PHE | CA | C | 57.855 | 0.000 | 1 |
| 928 | 86 | PHE | CB | C | 40.391 | 0.011 | 1 |
| 929 | 86 | PHE | CD1 | C | 132.119 | 0.000 | 3 |
| 930 | 86 | PHE | CD2 | C | 132.119 | 0.000 | 3 |
| 931 | 86 | PHE | CE1 | C | 130.563 | 0.000 | 3 |
| 932 | 86 | PHE | CE2 | C | 130.563 | 0.000 | 3 |
| 933 | 86 | PHE | CZ | C | 130.062 | 0.000 | 1 |
| 934 | 86 | PHE | H | H | 8.369 | 0.000 | 1 |
| 935 | 86 | PHE | HA | H | 4.602 | 0.000 | 1 |
| 936 | 86 | PHE | HB2 | H | 2.928 | 0.003 | 1 |
| 937 | 86 | PHE | HB3 | H | 2.928 | 0.003 | 1 |
| 938 | 86 | PHE | HD1 | H | 7.016 | 0.000 | 3 |
| 939 | 86 | PHE | HD2 | H | 7.016 | 0.000 | 3 |
| 940 | 86 | PHE | HE1 | H | 7.190 | 0.000 | 3 |
| 941 | 86 | PHE | HE2 | H | 7.190 | 0.000 | 3 |
| 942 | 86 | PHE | HZ | H | 7.104 | 0.000 | 1 |
| 943 | 86 | PHE | N | N | 122.153 | 0.000 | 1 |
| 944 | 87 | LYS | CA | C | 56.023 | 0.046 | 1 |
| 945 | 87 | LYS | CB | C | 33.518 | 0.014 | 1 |
| 946 | 87 | LYS | CD | C | 29.336 | 0.000 | 1 |
| 947 | 87 | LYS | CE | C | 42.058 | 0.033 | 1 |
| 948 | 87 | LYS | CG | C | 24.691 | 0.054 | 1 |
| 949 | 87 | LYS | H | H | 8.157 | 0.000 | 1 |
| 950 | 87 | LYS | HA | H | 4.174 | 0.006 | 1 |
| 951 | 87 | LYS | HB2 | H | 1.520 | 0.007 | 1 |
| 952 | 87 | LYS | HB3 | H | 1.635 | 0.004 | 1 |
| 953 | 87 | LYS | HD2 | H | 1.248 | 0.003 | 1 |
| 954 | 87 | LYS | HD3 | H | 1.248 | 0.003 | 1 |
| 955 | 87 | LYS | HE2 | H | 2.889 | 0.006 | 1 |
| 956 | 87 | LYS | HE3 | H | 2.889 | 0.006 | 1 |
| 957 | 87 | LYS | HG2 | H | 1.246 | 0.005 | 1 |
| 958 | 87 | LYS | HG3 | H | 1.246 | 0.005 | 1 |
| 959 | 87 | LYS | N | N | 125.164 | 0.000 | 1 |
| 960 | 88 | SER | CA | C | 58.093 | 0.018 | 1 |
| 961 | 88 | SER | CB | C | 64.124 | 0.019 | 1 |
| 962 | 88 | SER | H | H | 8.134 | 0.000 | 1 |
| 963 | 88 | SER | HA | H | 4.254 | 0.001 | 1 |
| 964 | 88 | SER | HB2 | H | 3.750 | 0.002 | 1 |
| 965 | 88 | SER | HB3 | H | 3.750 | 0.002 | 1 |
| 966 | 88 | SER | N | N | 117.424 | 0.000 | 1 |
| 967 | 89 | ARG | CA | C | 56.683 | 0.039 | 1 |
| 968 | 89 | ARG | CB | C | 30.760 | 0.000 | 1 |
| 969 | 89 | ARG | CD | C | 43.289 | 0.000 | 1 |
| 970 | 89 | ARG | CG | C | 27.068 | 0.000 | 1 |
| 971 | 89 | ARG | H | H | 8.395 | 0.000 | 1 |
| 972 | 89 | ARG | HA | H | 4.169 | 0.002 | 1 |
| 973 | 89 | ARG | HB2 | H | 1.708 | 0.000 | 1 |
| 974 | 89 | ARG | HB3 | H | 1.781 | 0.000 | 1 |
| 975 | 89 | ARG | HD2 | H | 3.112 | 0.000 | 1 |
| 976 | 89 | ARG | HD3 | H | 3.112 | 0.000 | 1 |
| 977 | 89 | ARG | HG2 | H | 1.544 | 0.000 | 1 |
| 978 | 89 | ARG | HG3 | H | 1.544 | 0.000 | 1 |
| 979 | 89 | ARG | N | N | 122.999 | 0.000 | 1 |
| 980 | 90 | LYS | CA | C | 56.705 | 0.042 | 1 |
| 981 | 90 | LYS | CB | C | 32.849 | 0.003 | 1 |
| 982 | 90 | LYS | CD | C | 28.867 | 0.000 | 1 |
| 983 | 90 | LYS | CE | C | 42.147 | 0.015 | 1 |
| 984 | 90 | LYS | CG | C | 24.649 | 0.000 | 1 |
| 985 | 90 | LYS | H | H | 8.252 | 0.004 | 1 |
| 986 | 90 | LYS | HA | H | 4.171 | 0.003 | 1 |
| 987 | 90 | LYS | HB2 | H | 1.667 | 0.007 | 1 |
| 988 | 90 | LYS | HB3 | H | 1.775 | 0.000 | 1 |
| 989 | 90 | LYS | HD2 | H | 1.610 | 0.006 | 1 |
| 990 | 90 | LYS | HD3 | H | 1.610 | 0.006 | 1 |
| 991 | 90 | LYS | HE2 | H | 2.914 | 0.004 | 1 |
| 992 | 90 | LYS | HE3 | H | 2.914 | 0.004 | 1 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 993 | 90 | LYS | HG2 | H | 1.342 | 0.005 | 1 |
| 994 | 90 | LYS | HG3 | H | 1.342 | 0.005 | 1 |
| 995 | 90 | LYS | N | N | 121.564 | 0.000 | 1 |
| 996 | 91 | GLU | CA | C | 57.103 | 0.035 | 1 |
| 997 | 91 | GLU | CB | C | 30.191 | 0.023 | 1 |
| 998 | 91 | GLU | CG | C | 36.403 | 0.020 | 1 |
| 999 | 91 | GLU | H | H | 8.240 | 0.011 | 1 |
| 1000 | 91 | GLU | HA | H | 4.142 | 0.007 | 1 |
| 1001 | 91 | GLU | HB2 | H | 1.884 | 0.002 | 1 |
| 1002 | 91 | GLU | HB3 | H | 1.984 | 0.012 | 1 |
| 1003 | 91 | GLU | HG2 | H | 2.191 | 0.004 | 1 |
| 1004 | 91 | GLU | HG3 | H | 2.191 | 0.004 | 1 |
| 1005 | 91 | GLU | N | N | 121.357 | 0.000 | 1 |
| 1006 | 92 | ARG | CA | C | 56.417 | 0.025 | 1 |
| 1007 | 92 | ARG | CB | C | 30.638 | 0.061 | 1 |
| 1008 | 92 | ARG | CG | C | 43.298 | 0.000 | 1 |
| 1009 | 92 | ARG | H | H | 8.150 | 0.004 | 1 |
| 1010 | 92 | ARG | HA | H | 4.190 | 0.002 | 1 |
| 1011 | 92 | ARG | HB2 | H | 1.700 | 0.000 | 1 |
| 1012 | 92 | ARG | HB3 | H | 1.787 | 0.002 | 1 |
| 1013 | 92 | ARG | HG2 | H | 3.114 | 0.000 | 1 |
| 1014 | 92 | ARG | N | N | 120.047 | 0.000 | 1 |
| 1015 | 93 | GLU | CA | C | 56.638 | 0.068 | 1 |
| 1016 | 93 | GLU | CB | C | 30.144 | 0.044 | 1 |
| 1017 | 93 | GLU | CG | C | 36.460 | 0.000 | 1 |
| 1018 | 93 | GLU | H | H | 8.280 | 0.000 | 1 |
| 1019 | 93 | GLU | HA | H | 4.175 | 0.003 | 1 |
| 1020 | 93 | GLU | HB2 | H | 1.885 | 0.004 | 1 |
| 1021 | 93 | GLU | HB3 | H | 2.009 | 0.000 | 1 |
| 1022 | 93 | GLU | HG2 | H | 2.196 | 0.000 | 1 |
| 1023 | 93 | GLU | N | N | 121.195 | 0.000 | 1 |
| 1024 | 94 | ALA | CA | C | 52.391 | 0.020 | 1 |
| 1025 | 94 | ALA | CB | C | 19.474 | 0.018 | 1 |
| 1026 | 94 | ALA | H | H | 8.052 | 0.001 | 1 |
| 1027 | 94 | ALA | HA | H | 4.261 | 0.002 | 1 |
| 1028 | 94 | ALA | HB1 | H | 1.328 | 0.007 | 1 |
| 1029 | 94 | ALA | HB2 | H | 1.328 | 0.007 | 1 |
| 1030 | 94 | ALA | HB3 | H | 1.328 | 0.007 | 1 |
| 1031 | 94 | ALA | N | N | 124.234 | 0.000 | 1 |
| 1032 | 95 | GLU | CA | C | 58.085 | 0.032 | 1 |
| 1033 | 95 | GLU | CB | C | 31.176 | 0.053 | 1 |
| 1034 | 95 | GLU | CG | C | 36.780 | 0.007 | 1 |
| 1035 | 95 | GLU | H | H | 7.808 | 0.030 | 1 |
| 1036 | 95 | GLU | HA | H | 4.021 | 0.007 | 1 |
| 1037 | 95 | GLU | HB2 | H | 1.832 | 0.000 | 1 |
| 1038 | 95 | GLU | HB3 | H | 1.978 | 0.004 | 1 |
| 1039 | 95 | GLU | HG2 | H | 2.151 | 0.005 | 1 |
| 1040 | 95 | GLU | HG3 | H | 2.151 | 0.005 | 1 |
| 1041 | 95 | GLU | N | N | 125.172 | 0.000 | 1 |

## Appendix III

### A3.1 PABP-1 RRM 1 chromatogram

Figure A3.1 illustrates the chromatogram results of the forward and reverse sequencing of PABP-1 RRM 1.

5′
ATGAACCCCAGTGCCCCCAGCTACCCAATGGCCTCGCTCTACGTGGGGGACCTCC
ACCCCGACGTGACCGAGGCGATGCTCTACGAGAAGTTCAGCCCGGCCGGGCCCA
TCCTCTCCATCCGGGTCTGCAGGGACATGATCACCCGCCGCTCCTTGGGCTACGC
GTATGTGAACTTCCAGCAGCCGGCGGACGCGGAGCGTGCTTTGGACACCATGAAT
TTTGATGTTATAAAGGGCAAGCCAGTACGCATCATGTGGTCTCAGCGT 3′
3′
ACGCTGAGACCACATGATGCGTACTGGCTTGCCCTTTATAACATCAAAATTCATG
GTGTCCAAAGCACGCTCCGCGTCCGCCGGCTGCTGGAAGTTCACATACGCGTAGC
CCAAGGAGCGGCGGGTGATCATGTCCCTGCAGACCCGGATGGAGAGGATGGGCC
CGGCCGGGCTGAACTTCTCGTAGAGCATCGCCTCGGTCACGTCGGGGTGGAGGTC
CCCCACGTAGAGCGAGGCCATTGGGTAGCTGGGGGCACTGGGGTTCAT 5′

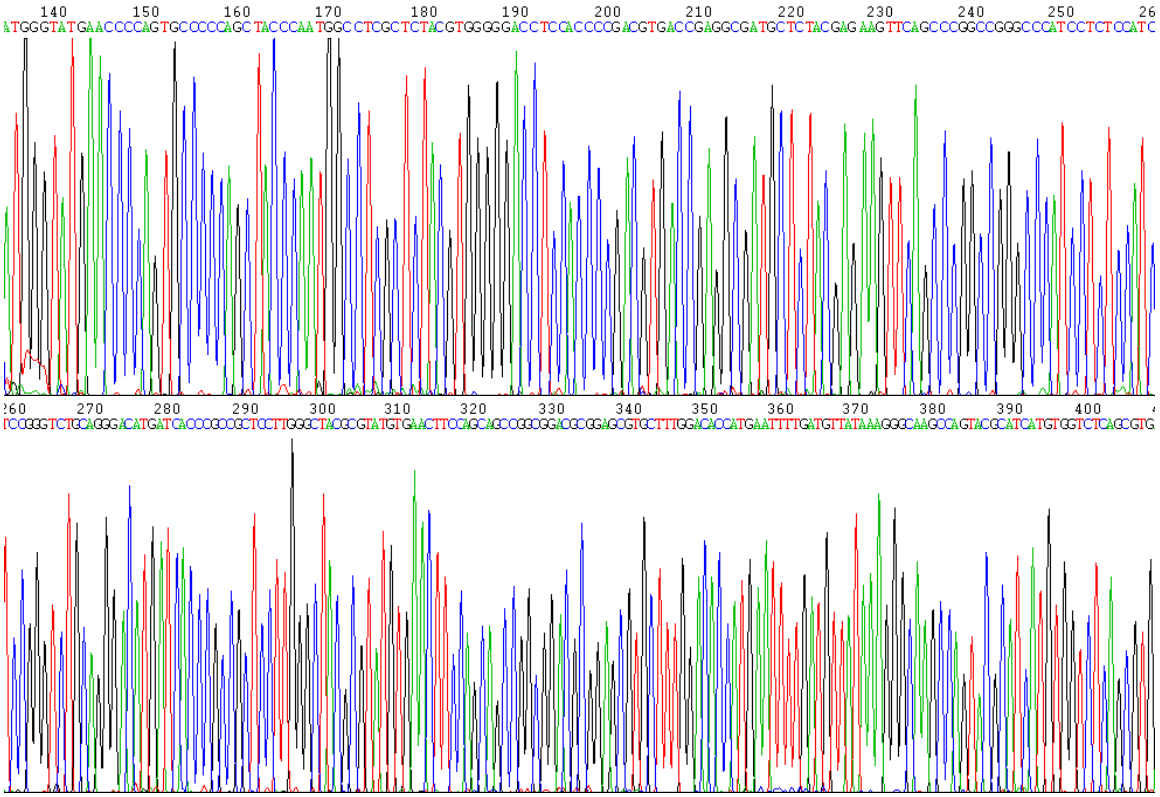


**Figure A3.1 illustrates the sequencing results for PABP-1 RRM 1.** The above Figures show the forward and reverse sequencing results with 5′ and 3′vector sequencing removed. The bottom two Figures illustrate the chromatograms of the sequencing results. These chromatograms show that there was no ambiguity in the sequencing results or single point polymorphisms.

NMR spectral data was collected from RRM 1 samples at a protein concentration of ~1.0mM in a volume of 350µl with a $H_2O/D_2O$ concentration of 90%/10% respectively. As mentioned above all spectra were collected at 298K using an 800MHz Bruker Avance spectrometer with cryo-platform for the NOESY experiments. All other experiments were performed on a 600MHz Bruker Avance spectrometer or a 600MHz Bruker Avance ultra shielded spectrometer with cryo-platform. Table A3.1 displays the datasets collected for the amino acid assignment of PABP-1 RRM 1 and includes acquisition times and mixing times.

**Table A3.1 Table of experimental datasets used for PABP-1 RRM1.**

| Experiment Type | Acquisition Parameters (ms) | | | Comments |
|---|---|---|---|---|
| | F3 | F2 | F1 | |
| Backbone assignments | | | | |
| $^{15}$N resolved HSQC (aliphatic) | | $^{1}$H, τaq=107.4 | $^{15}$N, τaq=52.6 | Field 600MHz |
| HNCACB | $^{1}$H, τaq=122.2 | $^{15}$N, τaq =23.9 | $^{13}$C, τaq =8.9 | Field 600MHz |
| CBCACONH | $^{1}$H, τaq=122.2 | $^{15}$N, τaq =23.9 | $^{13}$C, τaq =8.9 | Field 600MHz |
| HNCO | $^{1}$H, τaq=122.2 | $^{15}$N, τaq =26.5 | $^{13}$C, τaq =15.1 | Field 600MHz |
| HBHACONH | $^{1}$H, τaq=122.2 | $^{15}$N, τaq =24.3 | $^{1}$H, τaq =23.8 | Field 600MHz |
| Side-chain assignments | | | | |
| $^{13}$C resolved HSQC (aliphatic) $^{13}$C resolved HSQC (aromatic) | | $^{1}$H, τaq =111.5 | $^{13}$C, τaq =5.7 | Field 600MHz |
| [H]CCH TOCSY | $^{1}$H, τaq =97.6 | $^{13}$C, τaq =8.3 | $^{13}$C, τaq =10.1 | Field 600MHz cryo TOCSY mix time=38ms |
| H[C]CH TOCSY | $^{1}$H, τaq =97.6 | $^{13}$C, τaq =8.3 | $^{1}$H, τaq =30.5 | Field 600MHz cryo TOCSY mix time=38ms |
| NOE assignments | | | | |
| $^{13}$C resolved NOESY | $^{1}$H, τaq =76.2 | $^{13}$C, τaq =5.6 | $^{1}$H, τaq =14.3 | Field 800MHz NOESY mix time=120ms |
| $^{13}$C resolved NOESY aromatics | $^{1}$H, τaq = | $^{13}$C, τaq = | $^{1}$H, τaq = | Field 800MHz NOESY mix time=120ms |
| $^{15}$N resolved NOESY | $^{1}$H, τaq = | $^{15}$N, τaq = | $^{1}$H, τaq = | Field 800MHz NOESY mix time=120ms |

**τaq** = approximate acquisition time in ms for that particular dimension.

**A3.2**

Chemical Shift Ambiguity Code Definitions

| Codes | Definition |
|---|---|
| 1 | Unique |
| 2 | Ambiguity of geminal atoms or geminal methyl proton groups |
| 3 | Aromatic atoms on opposite sides of the ring (e.g. Tyr HE1 and HE2 protons) |
| 4 | Intraresidue ambiguities (e.g. Lys HG and HD protons) |
| 5 | Interresidue ambiguities (Lys 12 vs. Lys 27) |
| 9 | Ambiguous, specific ambiguity not defined |

| Atom shift assign | Seq code | Residue Label | Atom Name | Atom Type | Shift/ ppm | Error/ ppm | Ambiguity Code |
|---|---|---|---|---|---|---|---|
| 1 | 6 | ALA | H | H | 8.010 | 0.000 | 1 |
| 2 | 6 | ALA | N | N | 126.478 | 0.000 | 1 |
| 3 | 11 | MET | CA | C | 55.626 | 0.065 | 1 |
| 4 | 11 | MET | CB | C | 32.522 | 0.000 | 1 |
| 5 | 11 | MET | HA | H | 4.322 | 0.006 | 1 |
| 6 | 11 | MET | HB2 | H | 1.787 | 0.001 | 1 |
| 7 | 11 | MET | HB3 | H | 1.927 | 0.006 | 1 |
| 8 | 12 | ALA | CA | C | 51.261 | 0.028 | 1 |
| 9 | 12 | ALA | CB | C | 20.271 | 0.018 | 1 |
| 10 | 12 | ALA | H | H | 8.361 | 0.001 | 1 |
| 11 | 12 | ALA | HA | H | 4.718 | 0.011 | 1 |
| 12 | 12 | ALA | HB1 | H | 1.196 | 0.005 | 1 |
| 13 | 12 | ALA | HB2 | H | 1.196 | 0.005 | 1 |
| 14 | 12 | ALA | HB3 | H | 1.196 | 0.005 | 1 |
| 15 | 12 | ALA | N | N | 123.053 | 0.000 | 1 |
| 16 | 13 | SER | CA | C | 57.673 | 0.048 | 1 |
| 17 | 13 | SER | CB | C | 64.911 | 0.000 | 1 |
| 18 | 13 | SER | H | H | 8.010 | 0.004 | 1 |
| 19 | 13 | SER | HA | H | 5.368 | 0.002 | 1 |
| 20 | 13 | SER | HB2 | H | 3.732 | 0.005 | 1 |
| 21 | 13 | SER | HB3 | H | 3.843 | 0.007 | 1 |
| 22 | 13 | SER | N | N | 114.595 | 0.000 | 1 |
| 23 | 14 | LEU | CA | C | 53.306 | 0.040 | 1 |
| 24 | 14 | LEU | CB | C | 44.013 | 0.045 | 1 |
| 25 | 14 | LEU | CD1 | C | 22.749 | 0.019 | 1 |
| 26 | 14 | LEU | CD2 | C | 22.849 | 0.061 | 1 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 27 | 14 | LEU | CG | C | 25.916 | 0.031 | 1 | | | 67 | 16 | VAL | HG23 | H | 0.705 | 0.005 | 1 | | |
| 28 | 14 | LEU | H | H | 9.033 | 0.005 | 1 | | | 68 | 16 | VAL | N | N | 128.323 | 0.000 | 1 | | |
| 29 | 14 | LEU | HA | H | 4.836 | 0.007 | 1 | | | 69 | 17 | GLY | CA | C | 43.359 | 0.035 | 1 | | |
| 30 | 14 | LEU | HB2 | H | 1.046 | 0.006 | 1 | | | 70 | 17 | GLY | H | H | 9.130 | 0.004 | 1 | | |
| 31 | 14 | LEU | HB3 | H | 1.467 | 0.008 | 1 | | | 71 | 17 | GLY | HA2 | H | 4.614 | 0.293 | 1 | | |
| 32 | 14 | LEU | HD11 | H | -0.005 | 0.009 | 1 | | | 72 | 17 | GLY | HA3 | H | 3.695 | 0.439 | 1 | | |
| 33 | 14 | LEU | HD12 | H | -0.005 | 0.009 | 1 | | | 73 | 17 | GLY | N | N | 112.413 | 0.000 | 1 | | |
| 34 | 14 | LEU | HD13 | H | -0.005 | 0.009 | 1 | | | 74 | 18 | ASP | CA | C | 55.241 | 0.053 | 1 | | |
| 35 | 14 | LEU | HD21 | H | 0.012 | 0.015 | 1 | | | 75 | 18 | ASP | CB | C | 38.989 | 0.022 | 1 | | |
| 36 | 14 | LEU | HD22 | H | 0.012 | 0.015 | 1 | | | 76 | 18 | ASP | H | H | 8.203 | 0.007 | 1 | | |
| 37 | 14 | LEU | HD23 | H | 0.012 | 0.015 | 1 | | | 77 | 18 | ASP | HA | H | 4.184 | 0.005 | 1 | | |
| 38 | 14 | LEU | HG | H | 0.619 | 0.007 | 1 | | | 78 | 18 | ASP | HB2 | H | 2.961 | 0.008 | 1 | | |
| 39 | 14 | LEU | N | N | 125.159 | 0.000 | 1 | | | 79 | 18 | ASP | HB3 | H | 3.133 | 0.010 | 1 | | |
| 40 | 15 | TYR | CA | C | 56.694 | 0.011 | 1 | | | 80 | 18 | ASP | N | N | 114.045 | 0.000 | 1 | | |
| 41 | 15 | TYR | CB | C | 39.833 | 0.000 | 1 | | | 81 | 19 | LEU | CA | C | 53.840 | 0.100 | 1 | | |
| 42 | 15 | TYR | CD1 | C | 133.065 | 0.016 | 3 | | | 82 | 19 | LEU | CB | C | 43.369 | 0.043 | 1 | | |
| 43 | 15 | TYR | CD2 | C | 133.065 | 0.016 | 3 | | | 83 | 19 | LEU | CD1 | C | 26.524 | 0.020 | 1 | | |
| 44 | 15 | TYR | CE1 | C | 117.804 | 0.074 | 3 | | | 84 | 19 | LEU | CD2 | C | 23.929 | 0.037 | 1 | | |
| 45 | 15 | TYR | CE2 | C | 117.804 | 0.074 | 3 | | | 85 | 19 | LEU | CG | C | 26.507 | 0.059 | 1 | | |
| 46 | 15 | TYR | H | H | 9.027 | 0.003 | 1 | | | 86 | 19 | LEU | H | H | 8.012 | 0.002 | 1 | | |
| 47 | 15 | TYR | HA | H | 4.567 | 0.008 | 1 | | | 87 | 19 | LEU | HA | H | 4.074 | 0.010 | 1 | | |
| 48 | 15 | TYR | HB2 | H | 2.645 | 0.009 | 1 | | | 88 | 19 | LEU | HB2 | H | 1.002 | 0.006 | 1 | | |
| 49 | 15 | TYR | HB3 | H | 2.645 | 0.009 | 1 | | | 89 | 19 | LEU | HB3 | H | 1.072 | 0.012 | 1 | | |
| 50 | 15 | TYR | HD1 | H | 6.702 | 0.002 | 3 | | | 90 | 19 | LEU | HD11 | H | 0.456 | 0.005 | 1 | | |
| 51 | 15 | TYR | HD2 | H | 6.704 | 0.004 | 3 | | | 91 | 19 | LEU | HD12 | H | 0.456 | 0.005 | 1 | | |
| 52 | 15 | TYR | HE1 | H | 6.153 | 0.004 | 3 | | | 92 | 19 | LEU | HD13 | H | 0.456 | 0.005 | 1 | | |
| 53 | 15 | TYR | HE2 | H | 6.154 | 0.002 | 3 | | | 93 | 19 | LEU | HD21 | H | 0.630 | 0.007 | 1 | | |
| 54 | 15 | TYR | N | N | 124.676 | 0.000 | 1 | | | 94 | 19 | LEU | HD22 | H | 0.630 | 0.007 | 1 | | |
| 55 | 16 | VAL | CA | C | 60.819 | 0.024 | 1 | | | 95 | 19 | LEU | HD23 | H | 0.630 | 0.007 | 1 | | |
| 56 | 16 | VAL | CB | C | 32.746 | 0.041 | 1 | | | 96 | 19 | LEU | HG | H | 1.307 | 0.006 | 1 | | |
| 57 | 16 | VAL | CG1 | C | 21.101 | 0.049 | 1 | | | 97 | 19 | LEU | N | N | 114.660 | 0.000 | 1 | | |
| 58 | 16 | VAL | CG2 | C | 22.148 | 0.036 | 1 | | | 98 | 20 | HIS | CA | C | 56.424 | 0.038 | 1 | | |
| 59 | 16 | VAL | H | H | 8.487 | 0.006 | 1 | | | 99 | 20 | HIS | CB | C | 32.125 | 0.023 | 1 | | |
| 60 | 16 | VAL | HA | H | 4.359 | 0.007 | 1 | | | 100 | 20 | HIS | CE1 | C | 118.154 | 0.007 | 1 | | |
| 61 | 16 | VAL | HB | H | 1.551 | 0.006 | 1 | | | 101 | 20 | HIS | H | H | 9.215 | 0.004 | 1 | | |
| 62 | 16 | VAL | HG11 | H | 0.440 | 0.007 | 1 | | | 102 | 20 | HIS | HA | H | 4.238 | 0.009 | 1 | | |
| 63 | 16 | VAL | HG12 | H | 0.440 | 0.007 | 1 | | | 103 | 20 | HIS | HB2 | H | 2.809 | 0.013 | 1 | | |
| 64 | 16 | VAL | HG13 | H | 0.440 | 0.007 | 1 | | | 104 | 20 | HIS | HB3 | H | 3.008 | 0.009 | 1 | | |
| 65 | 16 | VAL | HG21 | H | 0.705 | 0.005 | 1 | | | 105 | 20 | HIS | HE1 | H | 7.135 | 0.002 | 1 | | |
| 66 | 16 | VAL | HG22 | H | 0.705 | 0.005 | 1 | | | 106 | 20 | HIS | N | N | 125.922 | 0.000 | 1 | | |

107 21 PRO CA C 65.150 0.044 1
108 21 PRO CB C 32.096 0.032 1
109 21 PRO CD C 50.745 0.038 1
110 21 PRO CG C 27.419 0.026 1
111 21 PRO HA H 4.120 0.009 1
112 21 PRO HB2 H 2.199 0.013 1
113 21 PRO HB3 H 1.822 0.011 1
114 21 PRO HD2 H 2.056 0.006 1
115 21 PRO HD3 H 3.331 0.005 1
116 21 PRO HG2 H 1.721 0.006 1
117 21 PRO HG3 H 1.721 0.006 1
118 22 ASP CA C 54.572 0.084 1
119 22 ASP CB C 40.333 0.051 1
120 22 ASP H H 11.007 0.003 1
121 22 ASP HA H 4.723 0.012 1
122 22 ASP HB2 H 2.637 0.014 1
123 22 ASP HB3 H 2.819 0.009 1
124 22 ASP N N 120.694 0.000 1
125 23 VAL CA C 64.601 0.058 1
126 23 VAL CB C 30.844 0.058 1
127 23 VAL CG1 C 22.191 0.072 1
128 23 VAL CG2 C 24.149 0.039 1
129 23 VAL H H 7.794 0.002 1
130 23 VAL HA H 4.139 0.005 1
131 23 VAL HB H 2.063 0.009 1
132 23 VAL HG11 H 0.824 0.009 1
133 23 VAL HG12 H 0.824 0.009 1
134 23 VAL HG13 H 0.824 0.009 1
135 23 VAL HG21 H 1.049 0.005 1
136 23 VAL HG22 H 1.049 0.005 1
137 23 VAL HG23 H 1.049 0.005 1
138 23 VAL N N 119.824 0.000 1
139 24 THR CA C 59.616 0.022 1
140 24 THR CB C 72.652 0.000 1
141 24 THR CG2 C 22.122 0.046 1
142 24 THR H H 7.450 0.011 1
143 24 THR HA H 4.383 0.006 1
144 24 THR HB H 4.711 0.006 1
145 24 THR HG21 H 1.284 0.007
1
146 24 THR HG22 H 1.284 0.007
1
147 24 THR HG23 H 1.284 0.007
1
148 24 THR N N 118.320 0.000 1
149 25 GLU CA C 62.228 0.039 1
150 25 GLU CB C 28.382 0.033 1
151 25 GLU CG C 38.950 0.025 1
152 25 GLU H H 9.906 0.003 1
153 25 GLU HA H 3.548 0.007 1
154 25 GLU HB2 H 2.165 0.006 1
155 25 GLU HB3 H 1.772 0.009 1
156 25 GLU HG2 H 2.747 0.007 1
157 25 GLU HG3 H 2.079 0.007 1
158 25 GLU N N 121.137 0.000 1
159 26 ALA CA C 55.292 0.083 1
160 26 ALA CB C 18.388 0.058 1
161 26 ALA HA H 4.128 0.008 1
162 26 ALA HB1 H 1.359 0.008 1
163 26 ALA HB2 H 1.359 0.008 1
164 26 ALA HB3 H 1.359 0.008 1
165 27 MET CA C 58.963 0.027 1
166 27 MET CB C 33.658 0.123 1
167 27 MET CG C 32.511 0.121 1
168 27 MET H H 7.471 0.006 1
169 27 MET HA H 4.083 0.009 1
170 27 MET HB2 H 2.176 0.006 1
171 27 MET HB3 H 2.033 0.012 1
172 27 MET HG2 H 2.630 0.008 1
173 27 MET HG3 H 2.630 0.008 1
174 27 MET N N 117.950 0.000 1
175 28 LEU CA C 57.670 0.022 1
176 28 LEU CB C 42.071 0.035 1
177 28 LEU CD1 C 26.531 0.030 1
178 28 LEU CD2 C 22.768 0.013 1
179 28 LEU CG C 26.446 0.101 1
180 28 LEU H H 7.799 0.005 1
181 28 LEU HA H 4.036 0.006 1
182 28 LEU HB2 H 1.101 0.009 1
183 28 LEU HB3 H 2.048 0.012 1

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 184 | 28 | LEU | HD11 | H | 0.354 | 0.006 | 1 |
| 185 | 28 | LEU | HD12 | H | 0.354 | 0.006 | 1 |
| 186 | 28 | LEU | HD13 | H | 0.354 | 0.006 | 1 |
| 187 | 28 | LEU | HD21 | H | 0.672 | 0.010 | 1 |
| 188 | 28 | LEU | HD22 | H | 0.672 | 0.010 | 1 |
| 189 | 28 | LEU | HD23 | H | 0.672 | 0.010 | 1 |
| 190 | 28 | LEU | HG | H | 1.562 | 0.011 | 1 |
| 191 | 28 | LEU | N | N | 119.585 | 0.000 | 1 |
| 192 | 29 | TYR | CA | C | 63.329 | 0.018 | 1 |
| 193 | 29 | TYR | CB | C | 38.282 | 0.041 | 1 |
| 194 | 29 | TYR | CD1 | C | 133.457 | 0.018 | 3 |
| 195 | 29 | TYR | CD2 | C | 133.448 | 0.000 | 3 |
| 196 | 29 | TYR | CE1 | C | 118.311 | 0.007 | 3 |
| 197 | 29 | TYR | CE2 | C | 118.311 | 0.007 | 3 |
| 198 | 29 | TYR | H | H | 8.972 | 0.017 | 1 |
| 199 | 29 | TYR | HA | H | 3.677 | 0.007 | 1 |
| 200 | 29 | TYR | HB2 | H | 2.989 | 0.024 | 1 |
| 201 | 29 | TYR | HB3 | H | 3.071 | 0.024 | 1 |
| 202 | 29 | TYR | HD1 | H | 6.919 | 0.009 | 3 |
| 203 | 29 | TYR | HD2 | H | 6.924 | 0.003 | 3 |
| 204 | 29 | TYR | HE1 | H | 6.703 | 0.003 | 3 |
| 205 | 29 | TYR | HE2 | H | 6.703 | 0.003 | 3 |
| 206 | 29 | TYR | N | N | 122.761 | 0.000 | 1 |
| 207 | 30 | GLU | CA | C | 59.536 | 0.024 | 1 |
| 208 | 30 | GLU | CB | C | 29.653 | 0.026 | 1 |
| 209 | 30 | GLU | CG | C | 36.414 | 0.041 | 1 |
| 210 | 30 | GLU | H | H | 7.814 | 0.002 | 1 |
| 211 | 30 | GLU | HA | H | 3.867 | 0.005 | 1 |
| 212 | 30 | GLU | HB2 | H | 2.055 | 0.019 | 1 |
| 213 | 30 | GLU | HB3 | H | 2.055 | 0.019 | 1 |
| 214 | 30 | GLU | HG2 | H | 2.169 | 0.006 | 1 |
| 215 | 30 | GLU | HG3 | H | 2.342 | 0.005 | 1 |
| 216 | 30 | GLU | N | N | 119.624 | 0.000 | 1 |
| 217 | 31 | LYS | CA | C | 57.626 | 0.024 | 1 |
| 218 | 31 | LYS | CB | C | 32.168 | 0.061 | 1 |
| 219 | 31 | LYS | CD | C | 28.222 | 0.076 | 1 |
| 220 | 31 | LYS | CE | C | 41.573 | 0.095 | 1 |
| 221 | 31 | LYS | CG | C | 24.005 | 0.055 | 1 |
| 222 | 31 | LYS | H | H | 7.287 | 0.005 | 1 |
| 223 | 31 | LYS | HA | H | 3.957 | 0.005 | 1 |
| 224 | 31 | LYS | HB2 | H | 1.597 | 0.011 | 1 |
| 225 | 31 | LYS | HB3 | H | 1.672 | 0.032 | 1 |
| 226 | 31 | LYS | HD2 | H | 1.287 | 0.008 | 1 |
| 227 | 31 | LYS | HD3 | H | 1.419 | 0.008 | 1 |
| 228 | 31 | LYS | HE2 | H | 2.473 | 0.008 | 1 |
| 229 | 31 | LYS | HE3 | H | 2.571 | 0.011 | 1 |
| 230 | 31 | LYS | HG2 | H | 0.533 | 0.015 | 1 |
| 231 | 31 | LYS | HG3 | H | 0.533 | 0.015 | 1 |
| 232 | 31 | LYS | N | N | 116.166 | 0.000 | 1 |
| 233 | 32 | PHE | CA | C | 61.180 | 0.160 | 1 |
| 234 | 32 | PHE | CB | C | 40.190 | 0.037 | 1 |
| 235 | 32 | PHE | CD1 | C | 131.772 | 0.030 | 3 |
| 236 | 32 | PHE | CD2 | C | 131.772 | 0.030 | 3 |
| 237 | 32 | PHE | CE1 | C | 133.034 | 0.039 | 3 |
| 238 | 32 | PHE | CE2 | C | 133.034 | 0.039 | 3 |
| 239 | 32 | PHE | CZ | C | 128.793 | 0.028 | 1 |
| 240 | 32 | PHE | H | H | 8.306 | 0.013 | 1 |
| 241 | 32 | PHE | HA | H | 4.341 | 0.010 | 1 |
| 242 | 32 | PHE | HB2 | H | 2.596 | 0.008 | 1 |
| 243 | 32 | PHE | HB3 | H | 3.055 | 0.008 | 1 |
| 244 | 32 | PHE | HD1 | H | 7.392 | 0.007 | 3 |
| 245 | 32 | PHE | HD2 | H | 7.392 | 0.007 | 3 |
| 246 | 32 | PHE | HE1 | H | 6.933 | 0.008 | 3 |
| 247 | 32 | PHE | HE2 | H | 6.931 | 0.007 | 3 |
| 248 | 32 | PHE | HZ | H | 7.023 | 0.008 | 1 |
| 249 | 32 | PHE | N | N | 111.895 | 0.000 | 1 |
| 250 | 33 | SER | CA | C | 63.109 | 0.000 | 1 |
| 251 | 33 | SER | H | H | 8.200 | 0.000 | 1 |
| 252 | 33 | SER | HA | H | 4.343 | 0.000 | 1 |
| 253 | 33 | SER | HB2 | H | 3.703 | 0.000 | 1 |
| 254 | 33 | SER | HB3 | H | 3.703 | 0.000 | 1 |
| 255 | 33 | SER | N | N | 117.105 | 0.000 | 1 |
| 256 | 34 | PRO | CA | C | 65.765 | 0.017 | 1 |
| 257 | 34 | PRO | CB | C | 31.573 | 0.020 | 1 |
| 258 | 34 | PRO | CD | C | 51.617 | 0.095 | 1 |
| 259 | 34 | PRO | HA | H | 4.238 | 0.009 | 1 |
| 260 | 34 | PRO | HB2 | H | 1.447 | 0.008 | 1 |
| 261 | 34 | PRO | HB3 | H | 2.294 | 0.005 | 1 |
| 262 | 34 | PRO | HD2 | H | 3.693 | 0.007 | 1 |
| 263 | 34 | PRO | HD3 | H | 3.165 | 0.007 | 1 |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 264 | 35 | ALA | CA | C | 53.591 | 0.091 | 1 | 304 | 38 | ILE | N | N | 126.088 | 0.000 | 1 |
| 265 | 35 | ALA | CB | C | 18.396 | 0.024 | 1 | 305 | 39 | LEU | CA | C | 56.465 | 0.061 | 1 |
| 266 | 35 | ALA | H | H | 7.915 | 0.001 | 1 | 306 | 39 | LEU | CB | C | 43.351 | 0.028 | 1 |
| 267 | 35 | ALA | HA | H | 4.103 | 0.007 | 1 | 307 | 39 | LEU | CD1 | C | 25.758 | 0.040 | 1 |
| 268 | 35 | ALA | HB1 | H | 1.445 | 0.008 | 1 | 308 | 39 | LEU | CD2 | C | 23.385 | 0.036 | 1 |
| 269 | 35 | ALA | HB2 | H | 1.445 | 0.008 | 1 | 309 | 39 | LEU | CG | C | 27.009 | 0.128 | 1 |
| 270 | 35 | ALA | HB3 | H | 1.445 | 0.008 | 1 | 310 | 39 | LEU | H | H | 9.002 | 0.010 | 1 |
| 271 | 35 | ALA | N | N | 116.700 | 0.000 | 1 | 311 | 39 | LEU | HA | H | 4.288 | 0.006 | 1 |
| 272 | 36 | GLY | CA | C | 44.623 | 0.016 | 1 | 312 | 39 | LEU | HB2 | H | 1.394 | 0.008 | 1 |
| 273 | 36 | GLY | H | H | 7.492 | 0.004 | 1 | 313 | 39 | LEU | HB3 | H | 1.346 | 0.008 | 1 |
| 274 | 36 | GLY | HA2 | H | 4.685 | 0.005 | 1 | 314 | 39 | LEU | HD11 | H | 0.802 | 0.007 | 1 |
| 275 | 36 | GLY | HA3 | H | 3.746 | 0.009 | 1 | 315 | 39 | LEU | HD12 | H | 0.802 | 0.007 | 1 |
| 276 | 36 | GLY | N | N | 104.974 | 0.000 | 1 | 316 | 39 | LEU | HD13 | H | 0.802 | 0.007 | 1 |
| 277 | 37 | PRO | CA | C | 63.401 | 0.043 | 1 | 317 | 39 | LEU | HD21 | H | 0.771 | 0.013 | 1 |
| 278 | 37 | PRO | CB | C | 32.119 | 0.057 | 1 | 318 | 39 | LEU | HD22 | H | 0.771 | 0.013 | 1 |
| 279 | 37 | PRO | CD | C | 49.588 | 0.016 | 1 | 319 | 39 | LEU | HD23 | H | 0.771 | 0.013 | 1 |
| 280 | 37 | PRO | CG | C | 27.793 | 0.038 | 1 | 320 | 39 | LEU | HG | H | 1.290 | 0.009 | 1 |
| 281 | 37 | PRO | HA | H | 4.334 | 0.009 | 1 | 321 | 39 | LEU | N | N | 127.823 | 0.000 | 1 |
| 282 | 37 | PRO | HB2 | H | 1.829 | 0.011 | 1 | 322 | 40 | SER | CA | C | 57.854 | 0.046 | 1 |
| 283 | 37 | PRO | HB3 | H | 2.236 | 0.010 | 1 | 323 | 40 | SER | CB | C | 64.630 | 0.050 | 1 |
| 284 | 37 | PRO | HD2 | H | 3.503 | 0.008 | 1 | 324 | 40 | SER | H | H | 7.487 | 0.006 | 1 |
| 285 | 37 | PRO | HD3 | H | 3.813 | 0.012 | 1 | 325 | 40 | SER | HA | H | 4.488 | 0.006 | 1 |
| 286 | 37 | PRO | HG2 | H | 1.970 | 0.009 | 1 | 326 | 40 | SER | HB2 | H | 3.681 | 0.007 | 1 |
| 287 | 37 | PRO | HG3 | H | 2.147 | 0.009 | 1 | 327 | 40 | SER | HB3 | H | 3.952 | 0.009 | 1 |
| 288 | 38 | ILE | CA | C | 60.696 | 0.042 | 1 | 328 | 40 | SER | N | N | 109.547 | 0.000 | 1 |
| 289 | 38 | ILE | CB | C | 40.845 | 0.023 | 1 | 329 | 41 | ILE | CA | C | 60.769 | 0.035 | 1 |
| 290 | 38 | ILE | CD1 | C | 13.351 | 0.036 | 1 | 330 | 41 | ILE | CB | C | 42.154 | 0.035 | 1 |
| 291 | 38 | ILE | CG1 | C | 28.068 | 0.018 | 1 | 331 | 41 | ILE | CD1 | C | 13.933 | 0.049 | 1 |
| 292 | 38 | ILE | CG2 | C | 17.835 | 0.033 | 1 | 332 | 41 | ILE | CG1 | C | 28.447 | 0.060 | 1 |
| 293 | 38 | ILE | H | H | 8.535 | 0.002 | 1 | 333 | 41 | ILE | CG2 | C | 17.788 | 0.052 | 1 |
| 294 | 38 | ILE | HA | H | 3.827 | 0.007 | 1 | 334 | 41 | ILE | H | H | 8.121 | 0.012 | 1 |
| 295 | 38 | ILE | HB | H | 1.576 | 0.007 | 1 | 335 | 41 | ILE | HA | H | 4.706 | 0.008 | 1 |
| 296 | 38 | ILE | HD11 | H | 0.098 | 0.006 | 1 | 336 | 41 | ILE | HB | H | 1.378 | 0.012 | 1 |
| 297 | 38 | ILE | HD12 | H | 0.098 | 0.006 | 1 | 337 | 41 | ILE | HD11 | H | 0.517 | 0.007 | 1 |
| 298 | 38 | ILE | HD13 | H | 0.098 | 0.006 | 1 | 338 | 41 | ILE | HD12 | H | 0.517 | 0.007 | 1 |
| 299 | 38 | ILE | HG12 | H | -0.047 | 0.010 | 1 | 339 | 41 | ILE | HD13 | H | 0.517 | 0.007 | 1 |
| 300 | 38 | ILE | HG13 | H | -0.047 | 0.010 | 1 | 340 | 41 | ILE | HG12 | H | 1.401 | 0.009 | 1 |
| 301 | 38 | ILE | HG21 | H | 0.416 | 0.006 | 1 | 341 | 41 | ILE | HG13 | H | 0.780 | 0.010 | 1 |
| 302 | 38 | ILE | HG22 | H | 0.416 | 0.006 | 1 | 342 | 41 | ILE | HG21 | H | 0.704 | 0.009 | 1 |
| 303 | 38 | ILE | HG23 | H | 0.416 | 0.006 | 1 | 343 | 41 | ILE | HG22 | H | 0.704 | 0.009 | 1 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 344 | 41 | ILE | HG23 | H | 0.704 | 0.009 | 1 |
| 345 | 41 | ILE | N | N | 119.782 | 0.000 | 1 |
| 346 | 42 | ARG | CA | C | 54.750 | 0.082 | 1 |
| 347 | 42 | ARG | CB | C | 33.936 | 0.077 | 1 |
| 348 | 42 | ARG | CD | C | 43.416 | 0.067 | 1 |
| 349 | 42 | ARG | CG | C | 27.169 | 0.042 | 1 |
| 350 | 42 | ARG | H | H | 8.588 | 0.009 | 1 |
| 351 | 42 | ARG | HA | H | 4.777 | 0.007 | 1 |
| 352 | 42 | ARG | HB2 | H | 1.675 | 0.007 | 1 |
| 353 | 42 | ARG | HB3 | H | 1.763 | 0.006 | 1 |
| 354 | 42 | ARG | HD2 | H | 3.105 | 0.015 | 1 |
| 355 | 42 | ARG | HD3 | H | 3.105 | 0.015 | 1 |
| 356 | 42 | ARG | HG2 | H | 1.421 | 0.008 | 1 |
| 357 | 42 | ARG | HG3 | H | 1.540 | 0.011 | 1 |
| 358 | 42 | ARG | N | N | 125.620 | 0.000 | 1 |
| 359 | 43 | VAL | CA | C | 62.654 | 0.027 | 1 |
| 360 | 43 | VAL | CB | C | 32.114 | 0.024 | 1 |
| 361 | 43 | VAL | CG1 | C | 22.160 | 0.018 | 1 |
| 362 | 43 | VAL | CG2 | C | 20.367 | 0.078 | 1 |
| 363 | 43 | VAL | H | H | 8.784 | 0.002 | 1 |
| 364 | 43 | VAL | HA | H | 3.784 | 0.005 | 1 |
| 365 | 43 | VAL | HB | H | 1.839 | 0.007 | 1 |
| 366 | 43 | VAL | HG11 | H | 0.564 | 0.008 | 1 |
| 367 | 43 | VAL | HG12 | H | 0.564 | 0.008 | 1 |
| 368 | 43 | VAL | HG13 | H | 0.564 | 0.008 | 1 |
| 369 | 43 | VAL | HG21 | H | 0.661 | 0.012 | 1 |
| 370 | 43 | VAL | HG22 | H | 0.661 | 0.012 | 1 |
| 371 | 43 | VAL | HG23 | H | 0.661 | 0.012 | 1 |
| 372 | 43 | VAL | N | N | 126.946 | 0.000 | 1 |
| 373 | 44 | CYS | CA | C | 60.765 | 0.013 | 1 |
| 374 | 44 | CYS | CB | C | 27.263 | 0.030 | 1 |
| 375 | 44 | CYS | H | H | 8.352 | 0.002 | 1 |
| 376 | 44 | CYS | HA | H | 4.139 | 0.007 | 1 |
| 377 | 44 | CYS | HB2 | H | 1.666 | 0.007 | 1 |
| 378 | 44 | CYS | HB3 | H | 2.431 | 0.006 | 1 |
| 379 | 44 | CYS | N | N | 128.839 | 0.000 | 1 |
| 380 | 45 | ARG | CA | C | 54.587 | 0.053 | 1 |
| 381 | 45 | ARG | CB | C | 34.103 | 0.054 | 1 |
| 382 | 45 | ARG | CD | C | 42.712 | 0.036 | 1 |
| 383 | 45 | ARG | CG | C | 28.436 | 0.058 | 1 |
| 384 | 45 | ARG | H | H | 8.190 | 0.006 | 1 |
| 385 | 45 | ARG | HA | H | 4.804 | 0.007 | 1 |
| 386 | 45 | ARG | HB2 | H | 1.073 | 0.010 | 1 |
| 387 | 45 | ARG | HB3 | H | 1.273 | 0.011 | 1 |
| 388 | 45 | ARG | HD2 | H | 2.980 | 0.006 | 1 |
| 389 | 45 | ARG | HD3 | H | 3.060 | 0.014 | 1 |
| 390 | 45 | ARG | HG2 | H | 1.472 | 0.011 | 1 |
| 391 | 45 | ARG | HG3 | H | 1.472 | 0.011 | 1 |
| 392 | 45 | ARG | N | N | 122.461 | 0.000 | 1 |
| 393 | 46 | ASP | CA | C | 54.612 | 0.005 | 1 |
| 394 | 46 | ASP | CB | C | 43.538 | 0.046 | 1 |
| 395 | 46 | ASP | H | H | 8.441 | 0.007 | 1 |
| 396 | 46 | ASP | HA | H | 4.391 | 0.006 | 1 |
| 397 | 46 | ASP | HB2 | H | 2.491 | 0.011 | 1 |
| 398 | 46 | ASP | HB3 | H | 2.823 | 0.009 | 1 |
| 399 | 46 | ASP | N | N | 122.329 | 0.000 | 1 |
| 400 | 47 | MET | CA | C | 57.774 | 0.043 | 1 |
| 401 | 47 | MET | CB | C | 32.209 | 0.081 | 1 |
| 402 | 47 | MET | CG | C | 32.271 | 0.042 | 1 |
| 403 | 47 | MET | HA | H | 4.103 | 0.007 | 1 |
| 404 | 47 | MET | HB2 | H | 2.052 | 0.007 | 1 |
| 405 | 47 | MET | HB3 | H | 2.629 | 0.006 | 1 |
| 406 | 47 | MET | HG2 | H | 2.558 | 0.004 | 1 |
| 407 | 47 | MET | HG3 | H | 2.558 | 0.004 | 1 |
| 408 | 48 | ILE | CA | C | 62.739 | 0.052 | 1 |
| 409 | 48 | ILE | CB | C | 37.604 | 0.123 | 1 |
| 410 | 48 | ILE | CD1 | C | 11.505 | 0.032 | 1 |
| 411 | 48 | ILE | CG1 | C | 27.838 | 0.025 | 1 |
| 412 | 48 | ILE | CG2 | C | 17.213 | 0.083 | 1 |
| 413 | 48 | ILE | H | H | 8.453 | 0.002 | 1 |
| 414 | 48 | ILE | HA | H | 4.011 | 0.007 | 1 |
| 415 | 48 | ILE | HB | H | 2.052 | 0.007 | 1 |
| 416 | 48 | ILE | HD11 | H | 0.767 | 0.003 | 1 |
| 417 | 48 | ILE | HD12 | H | 0.767 | 0.003 | 1 |
| 418 | 48 | ILE | HD13 | H | 0.767 | 0.003 | 1 |
| 419 | 48 | ILE | HG12 | H | 1.197 | 0.005 | 1 |
| 420 | 48 | ILE | HG13 | H | 1.473 | 0.006 | 1 |
| 421 | 48 | ILE | HG21 | H | 0.805 | 0.004 | 1 |
| 422 | 48 | ILE | HG22 | H | 0.805 | 0.004 | 1 |
| 423 | 48 | ILE | HG23 | H | 0.805 | 0.004 | 1 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 424 | 48 | ILE | N | N | 118.432 | 0.000 | 1 |
| 425 | 49 | THR | CA | C | 62.712 | 0.020 | 1 |
| 426 | 49 | THR | CB | C | 70.697 | 0.140 | 1 |
| 427 | 49 | THR | CG2 | C | 21.385 | 0.051 | 1 |
| 428 | 49 | THR | H | H | 8.655 | 0.003 | 1 |
| 429 | 49 | THR | HA | H | 4.128 | 0.008 | 1 |
| 430 | 49 | THR | HB | H | 4.285 | 0.008 | 1 |
| 431 | 49 | THR | HG21 | H | 1.171 | 0.004 | 1 |
| 432 | 49 | THR | HG22 | H | 1.171 | 0.004 | 1 |
| 433 | 49 | THR | HG23 | H | 1.171 | 0.004 | 1 |
| 434 | 49 | THR | N | N | 109.638 | 0.000 | 1 |
| 435 | 50 | ARG | CA | C | 57.635 | 0.040 | 1 |
| 436 | 50 | ARG | CB | C | 26.631 | 0.038 | 1 |
| 437 | 50 | ARG | CD | C | 42.699 | 0.045 | 1 |
| 438 | 50 | ARG | CG | C | 27.153 | 0.030 | 1 |
| 439 | 50 | ARG | H | H | 7.790 | 0.003 | 1 |
| 440 | 50 | ARG | HA | H | 3.915 | 0.009 | 1 |
| 441 | 50 | ARG | HB2 | H | 2.016 | 0.011 | 1 |
| 442 | 50 | ARG | HB3 | H | 2.116 | 0.005 | 1 |
| 443 | 50 | ARG | HD2 | H | 3.087 | 0.009 | 1 |
| 444 | 50 | ARG | HD3 | H | 3.087 | 0.009 | 1 |
| 445 | 50 | ARG | HG2 | H | 1.511 | 0.006 | 1 |
| 446 | 50 | ARG | HG3 | H | 1.511 | 0.006 | 1 |
| 447 | 50 | ARG | N | N | 113.590 | 0.000 | 1 |
| 448 | 51 | ARG | CA | C | 56.027 | 0.065 | 1 |
| 449 | 51 | ARG | CB | C | 31.069 | 0.065 | 1 |
| 450 | 51 | ARG | CD | C | 43.331 | 0.057 | 1 |
| 451 | 51 | ARG | CG | C | 26.527 | 0.058 | 1 |
| 452 | 51 | ARG | H | H | 7.697 | 0.005 | 1 |
| 453 | 51 | ARG | HA | H | 4.114 | 0.010 | 1 |
| 454 | 51 | ARG | HB2 | H | 1.560 | 0.011 | 1 |
| 455 | 51 | ARG | HB3 | H | 1.691 | 0.007 | 1 |
| 456 | 51 | ARG | HD2 | H | 2.851 | 0.006 | 1 |
| 457 | 51 | ARG | HD3 | H | 3.096 | 0.015 | 1 |
| 458 | 51 | ARG | HG2 | H | 1.438 | 0.007 | 1 |
| 459 | 51 | ARG | HG3 | H | 1.663 | 0.001 | 1 |
| 460 | 51 | ARG | N | N | 119.879 | 0.000 | 1 |
| 461 | 52 | SER | CA | C | 58.421 | 0.040 | 1 |
| 462 | 52 | SER | CB | C | 64.290 | 0.134 | 1 |
| 463 | 52 | SER | HA | H | 4.142 | 0.006 | 1 |
| 464 | 52 | SER | HB2 | H | 3.756 | 0.008 | 1 |
| 465 | 52 | SER | HB3 | H | 3.907 | 0.005 | 1 |
| 466 | 53 | LEU | CA | C | 54.599 | 0.009 | 1 |
| 467 | 53 | LEU | CB | C | 42.131 | 0.011 | 1 |
| 468 | 53 | LEU | CD1 | C | 25.872 | 0.016 | 1 |
| 469 | 53 | LEU | CD2 | C | 22.195 | 0.123 | 1 |
| 470 | 53 | LEU | CG | C | 27.102 | 0.078 | 1 |
| 471 | 53 | LEU | H | H | 9.036 | 0.003 | 1 |
| 472 | 53 | LEU | HA | H | 4.288 | 0.005 | 1 |
| 473 | 53 | LEU | HB2 | H | 0.989 | 0.010 | 1 |
| 474 | 53 | LEU | HB3 | H | 1.606 | 0.008 | 1 |
| 475 | 53 | LEU | HD11 | H | 0.700 | 0.011 | 1 |
| 476 | 53 | LEU | HD12 | H | 0.700 | 0.011 | 1 |
| 477 | 53 | LEU | HD13 | H | 0.700 | 0.011 | 1 |
| 478 | 53 | LEU | HD21 | H | 0.812 | 0.005 | 1 |
| 479 | 53 | LEU | HD22 | H | 0.812 | 0.005 | 1 |
| 480 | 53 | LEU | HD23 | H | 0.812 | 0.005 | 1 |
| 481 | 53 | LEU | HG | H | 1.303 | 0.008 | 1 |
| 482 | 53 | LEU | N | N | 126.282 | 0.000 | 1 |
| 483 | 54 | GLY | CA | C | 45.679 | 0.079 | 1 |
| 484 | 54 | GLY | H | H | 8.947 | 0.005 | 1 |
| 485 | 54 | GLY | HA2 | H | 3.431 | 0.009 | 1 |
| 486 | 54 | GLY | HA3 | H | 4.132 | 0.003 | 1 |
| 487 | 54 | GLY | N | N | 106.606 | 0.000 | 1 |
| 488 | 55 | TYR | CA | C | 55.152 | 0.026 | 1 |
| 489 | 55 | TYR | CB | C | 40.271 | 0.027 | 1 |
| 490 | 55 | TYR | CD1 | C | 133.778 | 0.038 | 3 |
| 491 | 55 | TYR | CD2 | C | 133.791 | 0.048 | 3 |
| 492 | 55 | TYR | CE1 | C | 118.303 | 0.027 | 3 |
| 493 | 55 | TYR | CE2 | C | 118.303 | 0.027 | 3 |
| 494 | 55 | TYR | H | H | 6.806 | 0.010 | 1 |
| 495 | 55 | TYR | HA | H | 5.239 | 0.005 | 1 |
| 496 | 55 | TYR | HB2 | H | 2.128 | 0.008 | 1 |
| 497 | 55 | TYR | HB3 | H | 2.893 | 0.006 | 1 |
| 498 | 55 | TYR | HD1 | H | 6.481 | 0.002 | 3 |
| 499 | 55 | TYR | HD2 | H | 6.480 | 0.002 | 3 |
| 500 | 55 | TYR | HE1 | H | 6.792 | 0.008 | 3 |
| 501 | 55 | TYR | HE2 | H | 6.792 | 0.008 | 3 |
| 502 | 55 | TYR | N | N | 112.536 | 0.000 | 1 |
| 503 | 56 | ALA | CA | C | 49.596 | 0.017 | 1 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 504 | 56 | ALA | CB | C | 25.915 | 0.070 | 1 | | | 544 | 60 | PHE | CB | C | 40.192 | 0.031 | 1 | | |
| 505 | 56 | ALA | H | H | 8.608 | 0.008 | 1 | | | 545 | 60 | PHE | CD1 | C | 131.485 | 0.000 | 3 | | |
| 506 | 56 | ALA | HA | H | 4.960 | 0.004 | 1 | | | 546 | 60 | PHE | CD2 | C | 131.485 | 0.000 | 3 | | |
| 507 | 56 | ALA | HB1 | H | 0.824 | 0.008 | 1 | | | 547 | 60 | PHE | CE1 | C | 130.365 | 0.000 | 3 | | |
| 508 | 56 | ALA | HB2 | H | 0.824 | 0.008 | 1 | | | 548 | 60 | PHE | CE2 | C | 130.365 | 0.000 | 3 | | |
| 509 | 56 | ALA | HB3 | H | 0.824 | 0.008 | 1 | | | 549 | 60 | PHE | CZ | C | 128.588 | 0.000 | 1 | | |
| 510 | 56 | ALA | N | N | 120.416 | 0.000 | 1 | | | 550 | 60 | PHE | H | H | 8.962 | 0.022 | 1 | | |
| 511 | 57 | TYR | CA | C | 56.742 | 0.081 | 1 | | | 551 | 60 | PHE | HA | H | 4.472 | 0.007 | 1 | | |
| 512 | 57 | TYR | CB | C | 41.324 | 0.132 | 1 | | | 552 | 60 | PHE | HB2 | H | 2.964 | 0.004 | 1 | | |
| 513 | 57 | TYR | CE1 | C | 118.057 | 0.039 | 3 | | | 553 | 60 | PHE | HB3 | H | 3.564 | 0.006 | 1 | | |
| 514 | 57 | TYR | CE2 | C | 118.057 | 0.039 | 3 | | | 554 | 60 | PHE | HD1 | H | 6.942 | 0.000 | 3 | | |
| 515 | 57 | TYR | H | H | 8.345 | 0.005 | 1 | | | 555 | 60 | PHE | HD2 | H | 6.942 | 0.000 | 3 | | |
| 516 | 57 | TYR | HA | H | 5.430 | 0.009 | 1 | | | 556 | 60 | PHE | HE1 | H | 6.933 | 0.000 | 3 | | |
| 517 | 57 | TYR | HB2 | H | 2.473 | 0.006 | 1 | | | 557 | 60 | PHE | HE2 | H | 6.933 | 0.000 | 3 | | |
| 518 | 57 | TYR | HB3 | H | 2.810 | 0.005 | 1 | | | 558 | 60 | PHE | HZ | H | 6.823 | 0.000 | 1 | | |
| 519 | 57 | TYR | HE1 | H | 6.938 | 0.002 | 3 | | | 559 | 60 | PHE | N | N | 124.567 | 0.000 | 1 | | |
| 520 | 57 | TYR | HE2 | H | 6.933 | 0.005 | 3 | | | 560 | 61 | GLN | CA | C | 59.001 | 0.012 | 1 | | |
| 521 | 57 | TYR | N | N | 114.980 | 0.000 | 1 | | | 561 | 61 | GLN | CB | C | 29.706 | 0.037 | 1 | | |
| 522 | 58 | VAL | CA | C | 61.552 | 0.039 | 1 | | | 562 | 61 | GLN | CG | C | 34.100 | 0.127 | 1 | | |
| 523 | 58 | VAL | CB | C | 33.808 | 0.111 | 1 | | | 563 | 61 | GLN | H | H | 8.008 | 0.002 | 1 | | |
| 524 | 58 | VAL | CG1 | C | 20.956 | 0.033 | 1 | | | 564 | 61 | GLN | HA | H | 4.030 | 0.006 | 1 | | |
| 525 | 58 | VAL | CG2 | C | 20.269 | 0.021 | 1 | | | 565 | 61 | GLN | HB2 | H | 2.258 | 0.005 | 1 | | |
| 526 | 58 | VAL | H | H | 8.902 | 0.005 | 1 | | | 566 | 61 | GLN | HB3 | H | 1.848 | 0.007 | 1 | | |
| 527 | 58 | VAL | HA | H | 3.963 | 0.005 | 1 | | | 567 | 61 | GLN | HG2 | H | 2.234 | 0.015 | 1 | | |
| 528 | 58 | VAL | HB | H | 1.373 | 0.011 | 1 | | | 568 | 61 | GLN | HG3 | H | 2.234 | 0.015 | 1 | | |
| 529 | 58 | VAL | HG11 | H | -0.013 | 0.007 | 1 | | | 569 | 61 | GLN | N | N | 118.443 | 0.000 | 1 | | |
| 530 | 58 | VAL | HG12 | H | -0.013 | 0.007 | 1 | | | 570 | 62 | GLN | CA | C | 51.606 | 0.000 | 1 | | |
| 531 | 58 | VAL | HG13 | H | -0.013 | 0.007 | 1 | | | 571 | 62 | GLN | CB | C | 29.430 | 0.000 | 1 | | |
| 532 | 58 | VAL | HG21 | H | 0.150 | 0.010 | 1 | | | 572 | 62 | GLN | H | H | 9.619 | 0.000 | 1 | | |
| 533 | 58 | VAL | HG22 | H | 0.150 | 0.010 | 1 | | | 573 | 62 | GLN | N | N | 116.055 | 0.000 | 1 | | |
| 534 | 58 | VAL | HG23 | H | 0.150 | 0.010 | 1 | | | 574 | 63 | PRO | CA | C | 65.851 | 0.035 | 1 | | |
| 535 | 58 | VAL | N | N | 122.942 | 0.000 | 1 | | | 575 | 63 | PRO | CB | C | 32.241 | 0.089 | 1 | | |
| 536 | 59 | ASN | CA | C | 52.158 | 0.039 | 1 | | | 576 | 63 | PRO | CD | C | 51.444 | 0.040 | 1 | | |
| 537 | 59 | ASN | CB | C | 39.036 | 0.034 | 1 | | | 577 | 63 | PRO | CG | C | 27.674 | 0.123 | 1 | | |
| 538 | 59 | ASN | H | H | 8.606 | 0.003 | 1 | | | 578 | 63 | PRO | HA | H | 4.225 | 0.007 | 1 | | |
| 539 | 59 | ASN | HA | H | 5.312 | 0.005 | 1 | | | 579 | 63 | PRO | HB2 | H | 1.958 | 0.009 | 1 | | |
| 540 | 59 | ASN | HB2 | H | 2.644 | 0.006 | 1 | | | 580 | 63 | PRO | HB3 | H | 2.396 | 0.008 | 1 | | |
| 541 | 59 | ASN | HB3 | H | 2.871 | 0.012 | 1 | | | 581 | 63 | PRO | HD2 | H | 3.860 | 0.007 | 1 | | |
| 542 | 59 | ASN | N | N | 125.287 | 0.000 | 1 | | | 582 | 63 | PRO | HD3 | H | 3.969 | 0.009 | 1 | | |
| 543 | 60 | PHE | CA | C | 59.557 | 0.040 | 1 | | | 583 | 63 | PRO | HG2 | H | 2.142 | 0.007 | 1 | | |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 584 | 63 | PRO | HG3 | H | 2.142 | 0.007 | 1 | | | 624 | 68 | ARG | HB2 | H | 1.910 | 0.007 | 1 | | |
| 585 | 64 | ALA | CA | C | 54.972 | 0.140 | 1 | | | 625 | 68 | ARG | HB3 | H | 1.910 | 0.007 | 1 | | |
| 586 | 64 | ALA | CB | C | 18.472 | 0.101 | 1 | | | 626 | 68 | ARG | HD2 | H | 3.271 | 0.008 | 1 | | |
| 587 | 64 | ALA | H | H | 8.553 | 0.003 | 1 | | | 627 | 68 | ARG | HD3 | H | 3.198 | 0.007 | 1 | | |
| 588 | 64 | ALA | HA | H | 4.116 | 0.010 | 1 | | | 628 | 68 | ARG | HG2 | H | 1.677 | 0.004 | 1 | | |
| 589 | 64 | ALA | HB1 | H | 1.336 | 0.016 | 1 | | | 629 | 68 | ARG | HG3 | H | 1.677 | 0.004 | 1 | | |
| 590 | 64 | ALA | HB2 | H | 1.336 | 0.016 | 1 | | | 630 | 68 | ARG | N | N | 118.723 | 0.000 | 1 | | |
| 591 | 64 | ALA | HB3 | H | 1.336 | 0.016 | 1 | | | 631 | 69 | ALA | CA | C | 55.785 | 0.053 | 1 | | |
| 592 | 64 | ALA | N | N | 117.023 | 0.000 | 1 | | | 632 | 69 | ALA | CB | C | 18.347 | 0.070 | 1 | | |
| 593 | 65 | ASP | CA | C | 56.092 | 0.149 | 1 | | | 633 | 69 | ALA | H | H | 7.918 | 0.012 | 1 | | |
| 594 | 65 | ASP | CB | C | 39.659 | 0.025 | 1 | | | 634 | 69 | ALA | HA | H | 3.759 | 0.004 | 1 | | |
| 595 | 65 | ASP | H | H | 7.074 | 0.006 | 1 | | | 635 | 69 | ALA | HB1 | H | 1.289 | 0.008 | 1 | | |
| 596 | 65 | ASP | HA | H | 4.267 | 0.009 | 1 | | | 636 | 69 | ALA | HB2 | H | 1.289 | 0.008 | 1 | | |
| 597 | 65 | ASP | HB2 | H | 2.303 | 0.006 | 1 | | | 637 | 69 | ALA | HB3 | H | 1.289 | 0.008 | 1 | | |
| 598 | 65 | ASP | HB3 | H | 2.811 | 0.010 | 1 | | | 638 | 69 | ALA | N | N | 124.351 | 0.000 | 1 | | |
| 599 | 65 | ASP | N | N | 119.300 | 0.000 | 1 | | | 639 | 70 | LEU | CA | C | 59.006 | 0.021 | 1 | | |
| 600 | 66 | ALA | CA | C | 54.553 | 0.029 | 1 | | | 640 | 70 | LEU | CB | C | 41.693 | 0.075 | 1 | | |
| 601 | 66 | ALA | CB | C | 18.367 | 0.045 | 1 | | | 641 | 70 | LEU | CD1 | C | 24.637 | 0.025 | 1 | | |
| 602 | 66 | ALA | H | H | 6.734 | 0.002 | 1 | | | 642 | 70 | LEU | CD2 | C | 25.358 | 0.038 | 1 | | |
| 603 | 66 | ALA | HA | H | 2.931 | 0.006 | 1 | | | 643 | 70 | LEU | CG | C | 27.068 | 0.062 | 1 | | |
| 604 | 66 | ALA | HB1 | H | 1.289 | 0.007 | 1 | | | 644 | 70 | LEU | H | H | 7.855 | 0.012 | 1 | | |
| 605 | 66 | ALA | HB2 | H | 1.289 | 0.007 | 1 | | | 645 | 70 | LEU | HA | H | 3.535 | 0.008 | 1 | | |
| 606 | 66 | ALA | HB3 | H | 1.289 | 0.007 | 1 | | | 646 | 70 | LEU | HB2 | H | 1.034 | 0.009 | 1 | | |
| 607 | 66 | ALA | N | N | 122.189 | 0.000 | 1 | | | 647 | 70 | LEU | HB3 | H | 1.553 | 0.008 | 1 | | |
| 608 | 67 | GLU | CA | C | 59.539 | 0.028 | 1 | | | 648 | 70 | LEU | HD11 | H | 0.167 | 0.006 | 1 | | |
| 609 | 67 | GLU | CB | C | 29.753 | 0.074 | 1 | | | 649 | 70 | LEU | HD12 | H | 0.167 | 0.006 | 1 | | |
| 610 | 67 | GLU | CG | C | 36.565 | 0.071 | 1 | | | 650 | 70 | LEU | HD13 | H | 0.167 | 0.006 | 1 | | |
| 611 | 67 | GLU | H | H | 8.214 | 0.004 | 1 | | | 651 | 70 | LEU | HD21 | H | 0.413 | 0.006 | 1 | | |
| 612 | 67 | GLU | HA | H | 3.650 | 0.007 | 1 | | | 652 | 70 | LEU | HD22 | H | 0.413 | 0.006 | 1 | | |
| 613 | 67 | GLU | HB2 | H | 2.061 | 0.008 | 1 | | | 653 | 70 | LEU | HD23 | H | 0.413 | 0.006 | 1 | | |
| 614 | 67 | GLU | HB3 | H | 2.173 | 0.009 | 1 | | | 654 | 70 | LEU | HG | H | 1.188 | 0.006 | 1 | | |
| 615 | 67 | GLU | HG2 | H | 2.320 | 0.013 | 1 | | | 655 | 70 | LEU | N | N | 119.158 | 0.000 | 1 | | |
| 616 | 67 | GLU | HG3 | H | 2.416 | 0.005 | 1 | | | 656 | 71 | ASP | CA | C | 56.953 | 0.037 | 1 | | |
| 617 | 67 | GLU | N | N | 116.777 | 0.000 | 1 | | | 657 | 71 | ASP | CB | C | 42.088 | 0.022 | 1 | | |
| 618 | 68 | ARG | CA | C | 59.012 | 0.039 | 1 | | | 658 | 71 | ASP | H | H | 8.081 | 0.016 | 1 | | |
| 619 | 68 | ARG | CB | C | 29.762 | 0.065 | 1 | | | 659 | 71 | ASP | HA | H | 4.372 | 0.003 | 1 | | |
| 620 | 68 | ARG | CD | C | 43.353 | 0.026 | 1 | | | 660 | 71 | ASP | HB2 | H | 2.723 | 0.004 | 1 | | |
| 621 | 68 | ARG | CG | C | 27.845 | 0.090 | 1 | | | 661 | 71 | ASP | HB3 | H | 2.601 | 0.007 | 1 | | |
| 622 | 68 | ARG | H | H | 7.402 | 0.008 | 1 | | | 662 | 71 | ASP | N | N | 115.416 | 0.000 | 1 | | |
| 623 | 68 | ARG | HA | H | 4.075 | 0.010 | 1 | | | 663 | 72 | THR | CA | C | 63.244 | 0.085 | 1 | | |

664 72 THR CB C 71.563 0.039 1
665 72 THR CG2 C 21.544 0.019 1
666 72 THR H H 7.838 0.004 1
667 72 THR HA H 4.442 0.005 1
668 72 THR HB H 4.158 0.005 1
669 72 THR HG21 H 1.217 0.008 1
670 72 THR HG22 H 1.217 0.008 1
671 72 THR HG23 H 1.217 0.008 1
672 72 THR N N 106.071 0.000 1
673 73 MET CA C 55.779 0.020 1
674 73 MET CB C 33.327 0.038 1
675 73 MET CG C 33.245 0.106 1
676 73 MET H H 8.139 0.002 1
677 73 MET HA H 4.579 0.005 1
678 73 MET HB2 H 1.759 0.007 1
679 73 MET HB3 H 1.759 0.007 1
680 73 MET HG2 H 2.503 0.005 1
681 73 MET HG3 H 2.503 0.005 1
682 73 MET N N 119.179 0.000 1
683 74 ASN CA C 54.610 0.058 1
684 74 ASN CB C 37.460 0.139 1
685 74 ASN H H 7.084 0.003 1
686 74 ASN HA H 4.141 0.005 1
687 74 ASN HB2 H 2.545 0.010 1
688 74 ASN HB3 H 2.963 0.004 1
689 74 ASN N N 117.707 0.000 1
690 75 PHE CA C 59.553 0.024 1
691 75 PHE CB C 36.512 0.021 1
692 75 PHE CD1 C 131.579 0.080 3
693 75 PHE CD2 C 131.579 0.080 3
694 75 PHE CE1 C 132.411 0.000 3
695 75 PHE CE2 C 132.411 0.000 3
696 75 PHE H H 7.975 0.005 1
697 75 PHE HA H 4.533 0.003 1
698 75 PHE HB2 H 3.291 0.005 1
699 75 PHE HB3 H 3.291 0.005 1
700 75 PHE HD1 H 7.092 0.003 3
701 75 PHE HD2 H 7.092 0.003 3
702 75 PHE HE1 H 7.095 0.000 3
703 75 PHE HE2 H 7.095 0.000 3
704 75 PHE N N 118.892 0.000 1
705 76 ASP CA C 54.945 0.101 1
706 76 ASP CB C 40.721 0.105 1
707 76 ASP H H 7.875 0.005 1
708 76 ASP HA H 4.452 0.006 1
709 76 ASP HB2 H 2.288 0.009 1
710 76 ASP HB3 H 2.627 0.006 1
711 76 ASP N N 120.882 0.000 1
712 77 VAL CA C 62.721 0.049 1
713 77 VAL CB C 32.195 0.050 1
714 77 VAL CG1 C 21.506 0.020 1
715 77 VAL CG2 C 21.515 0.031 1
716 77 VAL H H 8.190 0.010 1
717 77 VAL HA H 3.940 0.006 1
718 77 VAL HB H 1.704 0.006 1
719 77 VAL HG11 H 0.598 0.007 1
720 77 VAL HG12 H 0.598 0.007 1
721 77 VAL HG13 H 0.598 0.007 1
722 77 VAL HG21 H 0.818 0.005 1
723 77 VAL HG22 H 0.818 0.005 1
724 77 VAL HG23 H 0.818 0.005 1
725 77 VAL N N 118.741 0.000 1
726 78 ILE CA C 61.394 0.100 1
727 78 ILE CB C 40.000 0.123 1
728 78 ILE CD1 C 13.633 0.041 1
729 78 ILE CG1 C 26.489 0.041 1
730 78 ILE CG2 C 17.726 0.049 1
731 78 ILE H H 8.170 0.011 1
732 78 ILE HA H 4.086 0.005 1
733 78 ILE HB H 1.684 0.009 1
734 78 ILE HD11 H 0.588 0.011 1
735 78 ILE HD12 H 0.588 0.011 1
736 78 ILE HD13 H 0.588 0.011 1
737 78 ILE HG12 H 0.758 0.015 1
738 78 ILE HG13 H 1.445 0.009 1

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 739 | 78 | ILE | HG21 | H | 0.719 | 0.008 | 1 |
| 740 | 78 | ILE | HG22 | H | 0.719 | 0.008 | 1 |
| 741 | 78 | ILE | HG23 | H | 0.719 | 0.008 | 1 |
| 742 | 78 | ILE | N | N | 126.183 | 0.000 | 1 |
| 743 | 79 | LYS | CA | C | 56.591 | 0.067 | 1 |
| 744 | 79 | LYS | CB | C | 29.873 | 0.119 | 1 |
| 745 | 79 | LYS | CD | C | 28.392 | 0.014 | 1 |
| 746 | 79 | LYS | CE | C | 41.839 | 0.152 | 1 |
| 747 | 79 | LYS | CG | C | 25.572 | 0.074 | 1 |
| 748 | 79 | LYS | HA | H | 3.604 | 0.006 | 1 |
| 749 | 79 | LYS | HB2 | H | 1.340 | 0.006 | 1 |
| 750 | 79 | LYS | HB3 | H | 1.709 | 0.005 | 1 |
| 751 | 79 | LYS | HD2 | H | 1.197 | 0.009 | 1 |
| 752 | 79 | LYS | HD3 | H | 1.197 | 0.009 | 1 |
| 753 | 79 | LYS | HE2 | H | 2.470 | 0.010 | 1 |
| 754 | 79 | LYS | HE3 | H | 2.591 | 0.007 | 1 |
| 755 | 79 | LYS | HG2 | H | 1.050 | 0.007 | 1 |
| 756 | 79 | LYS | HG3 | H | 1.129 | 0.009 | 1 |
| 757 | 80 | GLY | CA | C | 45.762 | 0.057 | 1 |
| 758 | 80 | GLY | H | H | 7.953 | 0.002 | 1 |
| 759 | 80 | GLY | HA2 | H | 3.391 | 0.008 | 1 |
| 760 | 80 | GLY | HA3 | H | 4.012 | 0.004 | 1 |
| 761 | 80 | GLY | N | N | 100.921 | 0.000 | 1 |
| 762 | 81 | LYS | CA | C | 52.742 | 0.057 | 1 |
| 763 | 81 | LYS | CB | C | 34.638 | 0.030 | 1 |
| 764 | 81 | LYS | CD | C | 29.005 | 0.055 | 1 |
| 765 | 81 | LYS | CE | C | 41.597 | 0.072 | 1 |
| 766 | 81 | LYS | CG | C | 24.080 | 0.042 | 1 |
| 767 | 81 | LYS | H | H | 7.632 | 0.002 | 1 |
| 768 | 81 | LYS | HA | H | 5.038 | 0.005 | 1 |
| 769 | 81 | LYS | HB2 | H | 2.030 | 0.193 | 1 |
| 770 | 81 | LYS | HB3 | H | 3.014 | 0.005 | 1 |
| 771 | 81 | LYS | HD2 | H | 1.600 | 0.009 | 1 |
| 772 | 81 | LYS | HD3 | H | 1.695 | 0.008 | 1 |
| 773 | 81 | LYS | HE2 | H | 2.859 | 0.007 | 1 |
| 774 | 81 | LYS | HE3 | H | 2.929 | 0.014 | 1 |
| 775 | 81 | LYS | HG2 | H | 1.506 | 0.006 | 1 |
| 776 | 81 | LYS | HG3 | H | 1.506 | 0.006 | 1 |
| 777 | 81 | LYS | N | N | 121.986 | 0.000 | 1 |
| 778 | 82 | PRO | CA | C | 62.536 | 0.085 | 1 |
| 779 | 82 | PRO | CB | C | 32.089 | 0.085 | 1 |
| 780 | 82 | PRO | CD | C | 51.486 | 0.053 | 1 |
| 781 | 82 | PRO | CG | C | 27.729 | 0.061 | 1 |
| 782 | 82 | PRO | HA | H | 4.459 | 0.007 | 1 |
| 783 | 82 | PRO | HB2 | H | 1.775 | 0.009 | 1 |
| 784 | 82 | PRO | HB3 | H | 2.249 | 0.010 | 1 |
| 785 | 82 | PRO | HD2 | H | 3.715 | 0.023 | 1 |
| 786 | 82 | PRO | HD3 | H | 3.838 | 0.009 | 1 |
| 787 | 82 | PRO | HG2 | H | 2.138 | 0.007 | 1 |
| 788 | 82 | PRO | HG3 | H | 2.138 | 0.007 | 1 |
| 789 | 83 | VAL | CA | C | 60.784 | 0.032 | 1 |
| 790 | 83 | VAL | CB | C | 33.111 | 0.113 | 1 |
| 791 | 83 | VAL | CG1 | C | 22.106 | 0.032 | 1 |
| 792 | 83 | VAL | CG2 | C | 20.590 | 0.000 | 1 |
| 793 | 83 | VAL | H | H | 8.608 | 0.003 | 1 |
| 794 | 83 | VAL | HA | H | 4.481 | 0.007 | 1 |
| 795 | 83 | VAL | HB | H | 2.056 | 0.007 | 1 |
| 796 | 83 | VAL | HG11 | H | 0.937 | 0.003 | 1 |
| 797 | 83 | VAL | HG12 | H | 0.937 | 0.003 | 1 |
| 798 | 83 | VAL | HG13 | H | 0.937 | 0.003 | 1 |
| 799 | 83 | VAL | HG21 | H | 0.929 | 0.009 | 1 |
| 800 | 83 | VAL | HG22 | H | 0.929 | 0.009 | 1 |
| 801 | 83 | VAL | HG23 | H | 0.929 | 0.009 | 1 |
| 802 | 83 | VAL | N | N | 121.996 | 0.000 | 1 |
| 803 | 84 | ARG | CA | C | 54.161 | 0.061 | 1 |
| 804 | 84 | ARG | CB | C | 32.260 | 0.019 | 1 |
| 805 | 84 | ARG | CD | C | 43.525 | 0.156 | 1 |
| 806 | 84 | ARG | CG | C | 27.586 | 0.129 | 1 |
| 807 | 84 | ARG | H | H | 7.603 | 0.003 | 1 |
| 808 | 84 | ARG | HA | H | 4.608 | 0.007 | 1 |
| 809 | 84 | ARG | HB2 | H | 1.529 | 0.011 | 1 |
| 810 | 84 | ARG | HB3 | H | 1.529 | 0.011 | 1 |
| 811 | 84 | ARG | HD2 | H | 3.166 | 0.006 | 1 |
| 812 | 84 | ARG | HD3 | H | 3.166 | 0.006 | 1 |
| 813 | 84 | ARG | HG2 | H | 1.335 | 0.006 | 1 |
| 814 | 84 | ARG | HG3 | H | 1.523 | 0.008 | 1 |
| 815 | 84 | ARG | N | N | 122.846 | 0.000 | 1 |
| 816 | 85 | ILE | CA | C | 60.246 | 0.033 | 1 |
| 817 | 85 | ILE | CB | C | 39.693 | 0.071 | 1 |
| 818 | 85 | ILE | CD1 | C | 14.085 | 0.038 | 1 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 819 | 85 | ILE | CG1 | C | 28.108 | 0.162 | 1 |
| 820 | 85 | ILE | CG2 | C | 20.303 | 0.032 | 1 |
| 821 | 85 | ILE | H | H | 8.627 | 0.005 | 1 |
| 822 | 85 | ILE | HA | H | 5.049 | 0.005 | 1 |
| 823 | 85 | ILE | HB | H | 1.300 | 0.008 | 1 |
| 824 | 85 | ILE | HD11 | H | 0.684 | 0.006 | 1 |
| 825 | 85 | ILE | HD12 | H | 0.684 | 0.006 | 1 |
| 826 | 85 | ILE | HD13 | H | 0.684 | 0.006 | 1 |
| 827 | 85 | ILE | HG12 | H | 0.797 | 0.009 | 1 |
| 828 | 85 | ILE | HG13 | H | 0.797 | 0.009 | 1 |
| 829 | 85 | ILE | HG21 | H | 0.685 | 0.009 | 1 |
| 830 | 85 | ILE | HG22 | H | 0.685 | 0.009 | 1 |
| 831 | 85 | ILE | HG23 | H | 0.685 | 0.009 | 1 |
| 832 | 85 | ILE | N | N | 125.072 | 0.000 | 1 |
| 833 | 86 | MET | CA | C | 54.246 | 0.118 | 1 |
| 834 | 86 | MET | CB | C | 37.184 | 0.020 | 1 |
| 835 | 86 | MET | H | H | 9.130 | 0.004 | 1 |
| 836 | 86 | MET | HA | H | 4.572 | 0.010 | 1 |
| 837 | 86 | MET | HB2 | H | 1.938 | 0.010 | 1 |
| 838 | 86 | MET | HB3 | H | 2.131 | 0.005 | 1 |
| 839 | 86 | MET | N | N | 123.310 | 0.000 | 1 |
| 840 | 87 | TRP | CA | C | 58.713 | 0.000 | 1 |
| 841 | 87 | TRP | CB | C | 30.209 | 0.058 | 1 |
| 842 | 87 | TRP | CD1 | C | 127.521 | 0.000 | 1 |
| 843 | 87 | TRP | CE3 | C | 123.784 | 0.000 | 1 |
| 844 | 87 | TRP | CZ2 | C | 118.848 | 0.000 | 1 |
| 845 | 87 | TRP | CZ3 | C | 113.831 | 0.000 | 1 |
| 846 | 87 | TRP | H | H | 8.480 | 0.000 | 1 |
| 847 | 87 | TRP | HA | H | 4.724 | 0.009 | 1 |
| 848 | 87 | TRP | HB2 | H | 2.938 | 0.006 | 1 |
| 849 | 87 | TRP | HB3 | H | 3.329 | 0.001 | 1 |
| 850 | 87 | TRP | HD1 | H | 7.210 | 0.002 | 1 |
| 851 | 87 | TRP | HE3 | H | 6.985 | 0.000 | 1 |
| 852 | 87 | TRP | HZ2 | H | 7.121 | 0.008 | 1 |
| 853 | 87 | TRP | HZ3 | H | 7.222 | 0.004 | 1 |
| 854 | 87 | TRP | N | N | 121.526 | 0.000 | 1 |
| 855 | 88 | SER | CA | C | 58.904 | 0.120 | 1 |
| 856 | 88 | SER | CB | C | 63.931 | 0.023 | 1 |
| 857 | 88 | SER | H | H | 8.553 | 0.006 | 1 |
| 858 | 88 | SER | HA | H | 4.425 | 0.007 | 1 |
| 859 | 88 | SER | HB3 | H | 3.830 | 0.016 | 1 |
| 860 | 88 | SER | N | N | 116.548 | 0.000 | 1 |

**Reference:**


1. Arnaout, M. A., Goodman, S.L., Xiong, J.P. (2007) *Curr Opin Cell Biol* **19**, 495-507
2. Zamir, E., Geiger, B. (2001) *J Cell Sci* **114**, 3577-3579
3. Larsen, M., Artym, V.V., Green, J.A., Yamada, K.M. (2006) *Curr Opin Cell Biol* **18**, 463-471
4. Gingras, A. R., Ziegler, W.H., Bobkov, A.A., Gordon Joyce, M., Fasci, D., Rothemund, S., Ritter, A., Grossman, J., Patel, B., Bate, N., Goult, B.T., Emsley, J., Barsukov, I.L., Roberts, G.C.K., Liddington, R.C., Critchley, D.R. (2009) *J Biol Chem* **284**(13), 8866-8876
5. Lietha, D., Cai, X., Ceccarelli, D.F.J., Li, Y., Schaller, M.D., Eck, M.J. (2007) *Cell* **129**, 1177-1185
6. Deo, R. C., Bonanno, J. B., Sonenberg, N., and Burley, S. K. (1999) *Cell* **98**(6), 835-845
7. Woods, A. J., Roberts, M. S., Choudhary, J., Barry, S. T., Mazaki, Y., Sabe, H., Morley, S. J., Critchley, D. R., and Norman, J. C. (2002) *J Biol Chem* **277**(8), 6428-6437
8. Langer, H. F., and Chavakis, T. (2009) *J Cell Mol Med* **13**(7), 1211-1220
9. Croft, D. R., Olson, M.F. (2008) *Cancer Cell* **14**, 349-351
10. Olson, M. F., Sahai, E. (2008) *Clin Exp Metastasis*
11. Wyckoff, J. B., Pinner, S.E., Gschmeissner, S., Condeelis, J.S., Sahai, E. (2006) *Curr Opin Cell Biol* **16**, 1515-1523
12. Sanz-Moreno, V., Gadea, G., Ahn, J., Paterson, H., Marra, P., Pinner, S., Sahai, E., and Marshall, C. J. (2008) *Cell* **135**(3), 510-523
13. Hynes, R. O. (1996) *Dev Biol* **180**, 402-412
14. Oelz, D., Schmeiser, C., Small, J.V. (2008) *cell Adh Migr.* **2**, 117-126
15. Le Clainche, C., and Carlier, M. F. (2008) *Physiol Rev* **88**(2), 489-513
16. Broussard, J., Webb, D.J., Kaverina, I. (2008) *Curr Opin Cell Biol* **20**, 85-90
17. Ezratty, E. J., Partridge, M.A., Gunderson, G.G. (2005) *Nat Cell Biol* **7**, 581-590
18. Wittmann, T., Bokoch, G.M., Waterman-Storer, C.M. (2003) *J Cell Biol* **161**, 845-851
19. Small, J. V., and Kaverina, I. (2003) *Curr Opin Cell Biol* **15**(1), 40-47
20. Golubovskaya, V. M., Kweh, F.A., Cance, W.G. (2009) *J Cell Mol Biol* **24**, 503-510
21. Huttenlocher, A. (2005) *Nat Cell Biol* **7**(4), 336-337
22. Hynes, R. O. (2002) *Cell* **69**, 11-25
23. Hanks, S. K., Ryzhova, L., Shin, N.Y., Brabek, J. (2003) *Front Biosci* **8**, d982-d996
24. Berrier, A., Yamada, K.M. (2007) *J Cell Physiol* **213**, 565-573
25. Zaidel-Bar, R., Itzkovitz, S., Maayan, A., Iyengar, R., Geiger, B. (2007) *Nat Cell Biol* **9**, 858-867
26. Chan KT, C. C., Huttenlocher A. (2009) *J Cell Biology* **185**, 357-370
27. Brunton, V. G., and Frame, M. C. (2008) *Curr Opin Pharmacol* **8**(4), 427-432
28. Cox, B. D., Natarajan, M., Stettner, M.R., Gladson, C.L. (2006) *J Cell Biochem* **99**, 36-52
29. Critchley, D. R. (2009) *Annu Rev Biophys* **38**, 235-254
30. Ziegler, W. H., Liddington, R. C., and Critchley, D. R. (2006) *Trends Cell Biol* **16**(9), 453-460
31. Humphries, J. D., Wang, P., Streuli, C., Geiger, B., Humphries, M.J., Ballestrem, C. (2007) *J Cell Biol* **179**, 1043-1057
32. Bois, P. R. J., Borgon, R.A., Vonrhein, C., Izard, T. (2005) *Mol Cell Biol* **25**, 6112-6122

33. Papagrigoriou, E., Gingras, A.R., Barsukov, I.L., Bate, N., Fillingham, I., Patel, B., Frank, R., Ziegler, W.H., Roberts, G.C.K., Critchley, D.R., Emsley, J. (2004) *Embo J* **23**, 2942-2951
34. Turner, C. E. (2000) *J Cell Sci* **113 Pt 23**, 4139-4140
35. Deakin, N. O., Turner, C.E. (2009) *J Cell Sci* **121**, 2435-2444
36. Wang, X., Fukuda, K., Velyvis, A., Wu, C., Gronenburg, A., Qin, J. (2008) *J Biol Chem* **283**, 21113-21119
37. Nikolopoulos, S. N., Turner, C.E. (2002) *J Biol Chem* **277**, 1568-1575
38. Efimov, A., Schiefermeier, N., Grigoriev, I., Brown, M. C., Turner, C. E., Small, J. V., and and Kaverina, I. (2008) *J Cell Sci* **121**, 196-204
39. Webb, D. J., Donais, K., Whitmore, L. A., Thomas, S. M., Turner, C. E., Parsons, J., and T. and Horwitz, A. F. (2004) *Nat Cell Biol* **6**, 154-161
40. Ito, A., Kataoka, T. R., Watanabe, M., Nishiyama, K., Mazaki, Y., Sabe, H., Kitamura, and Y. and Nojima, H. (2000) *Embo J* **19**, 562-571
41. Geiger, B. (2001) *Nature Review* **2**, 793
42. Van Slambrouck S, J. A., Romero AE, Steelant WF. (2009) *Int J Oncol.* **34**, 1717-1726
43. Retta, S. F., Cassara, G., Amato, M., Alessandro, R., Pellegrino, M., Degani, S., Leo, G., Silengo, L., Tarone, G. (2001) *Mol Cell Biol* **12**, 3126-3138
44. Lee, J. O., Rieu, P., Arnaout, M. A., and Liddington, R. (1995) *Cell* **80**(4), 631-638
45. Isaji, T., Sato, Y., Zhao, Y., Miyoshi, E., Wada, Y., Taniguchi, N., Gu, J. (2006) *J Biol Chem* **281**, 33258-33267
46. Xiong, J. P., Stehle, T., Diefenbach, B., Zhang, R., Dunker, R., Scott, D.L., Joachimiak, A., Goodman, S.L., Arnaout, M.A. (2002) *Science* **296**, 151-155
47. Zhu, J., Luo, B. H., Xiao, T., Zhang, C., Nishida, N., and Springer, T. A. (2008) *Mol Cell* **32**(6), 849-861
48. Garcia-Alvarez, B., Pereda, J.M., Calderwood, D.A., Ulmer, T.S., Critchley, D., Campbell, I.D., Ginsberg, M.H., Liddington, R.C. (2003) *Mol Cell Biol* **11**, 49-58
49. Anthis, N. J., Wegener, K. L., Ye, F., Kim, C., Goult, B. T., Lowe, E. D., Vakonakis, I., Bate, N., Critchley, D. R., Ginsberg, M. H., and Campbell, I. D. (2009) *Embo J* **28**, 3623-3632
50. Yan, B., Calderwood, D.A., Yaspan, B., and Ginsberg, M.H. (2001) *J Biol Chem* **276**, 28164–28170
51. Gingras, A. R., Ziegler, W. H., Frank, R., Barsukov, I. L., Roberts, G. C., Critchley, D. R., and Emsley, J. (2005) *J Biol Chem* **280**(44), 37217-37224
52. Gingras, A. R., Bate, N., Goult, B.T., Hazelwood, L., Canestrelli, I., Grossmann, J.G., Liu, H., Putz, N.S., Roberts, G.C., Volkmann, N., Hanein, D., Barsukov, I.L., Critchley, D.R. (2008) *Embo J* **27**, 458-469
53. Wegener, K. L., Partridge, A. W., Han, J., Pickford, A. R., Liddington, R. C., Ginsberg, M. H., and Campbell, I. D. (2007) *Cell* **128**(1), 171-182
54. Rodius, S., Chaloin, O., Moes, M., Schaffner-Reckinger, E., Landrieu, I., Lippens, G., Lin, M., Zhang, J., Kieffer, N. (2008) *J Biol Chem* **283**, 24212-24223
55. Fillingham, I., Gingras, A.R., Papagrigoriou, E., Patel, B., Emsley, J., Critchley, D.R., Roberts, G.C., Barsukov, I.L. (2005) *Structure* **13**, 65-74
56. Gingras, A. R., Vogel, K.P., Steinhoff, H.J., Ziegler, W.H., Patel, B., Emsley, J., Critchley, D.R., Roberts, G.C., Barsukov, I.L. (2006) *Biochemistry* **45**, 1805-1817
57. Izard, T., Evans, G., Borgon, R.A., Rush, C.L., Bricogne, G., Bois, P.R. (2004) *Nature* **427**, 171-175

58. Bakolitsa, C., Cohen, D. M., Bankston, L. A., Bobkov, A. A., Cadwell, G. W., Jennings, L., Critchley, D. R., Craig, S. W., and Liddington, R. C. (2004) *Nature* **430**(6999), 583-586
59. Goksoy, E., Ma, Y. Q., Wang, X., Kong, X., Perera, D., Plow, E. F., and Qin, J. (2008) *Mol Cell* **31**(1), 124-133
60. Goult, B. T., Bate, N., Anthis, N. J., Wegener, K. L., Gingras, A. R., Patel, B., Barsukov, I. L., Campbell, I. D., Roberts, G. C., and Critchley, D. R. (2009) *J Biol Chem* **284**(22), 15097-15106
61. del Rio, A., Perez-Jimenez, R., Liu, R., Roca-Cusachs, P., Fernandez, J.M., Sheetz, M.P. (2009) *Science* **323**, 638-641
62. Sieg, D. J., Hauck, C.R., Ilic, D., Klingbeil, C.L., Schaefer, E., Damsky, C.H., Schlaepfer, D.D. (2000) *Nat Cell Biol* **2**, 249-256
63. Ceccarelli, D. F., Song, H. K., Poy, F., Schaller, M. D., and Eck, M. J. (2006) *J Biol Chem* **281**(1), 252-259
64. Nowakowski, J., Cronin, C.N., McRee, D.E., Knuth, M.W., Nelson, C.G., Pavletich, N.P., Rodgers, J., Sang, B.-C., Scheibe, D.N., Swanson, R.V., Thompson, D.A. (2002) *Structure* **10**, 1659-1667
65. Calalb, M. B., Polte, T. R., and Hanks, S. K. (1995) *Mol Cell Biol* **15**(2), 954-963
66. Schaller, M. D., Borgman, C.A., Cobb, B.S., Vines, R.R., Reynolds, A.B, Parsons, J.T. (1992) *Proc Natl Acad Sci U S A* **89**, 5192-5196
67. Gao, G., Prutzman, K.C., King, M.L., Scheswohl, D.M., DeRose, E.F., London, R.E., Schaller, M.D., Campbell, S.L. (2004) *J Biol Chem* **279**, 8441-9451
68. Serrels, B., Serrels, A., Brunton, V. G., Holt, M., McLean, G. W., Gray, C. H., Jones, G. E., and Frame, M. C. (2007) *Nat Cell Biol* **9**(9), 1046-1056
69. Cooper, L. A., Shen, T.L., Guan, J.L. (2003) *Mol Cell Biol* **23**, 8030-8041
70. Hoellerer, M. K., Noble, M. E., Labesse, G., Campbell, I. D., Werner, J. M., and Arold, S. T. (2003) *Structure* **11**(10), 1207-1217
71. Bertulucci, C. M., Guibao, C.D., Zheng, J.J. (2007) *J Biochemistry*
72. Abbi, S., Ueda, H., Zheng, C., Cooper, L.A., Zhao, J., Christopher, R., Guan, J.L. (2002) *Mol Cell Biol* **13**, 3178-3191
73. Nagoshi, Y., Yamamoto, G., Irie, T., and Tachikawa, T. (2006) *Med Mol Morphol* **39**(3), 154-160
74. Glenney Jr, J. R., Zokas, L. (1989) *J Cell Biol* **108**, 2401-2408
75. Mazaki, Y., Hashimoto, S., Sabe, H. (1997) *J Biol Chem* **272**, 7437-7444
76. Critchley, D. R. (2000) *Curr Opin Cell Biol* **12**(1), 133-139
77. Schaller, M. D., Parsons, J.T. (1995) *Mol Cell Biol* **15**, 2635-2645
78. Deakin NO, B. M., Warwood S, Schoelermann J, Mostafavi-Pour Z, Knight D, Ballestrem C, Humphries MJ. (2009) *J Cell Sci* **122**, 1654-1664
79. Goldfinger, L. E., Han, J., Kiosses, W.B., Howe, A.K., Ginsberg, M.H. (2003) *J Cell Biol* **162**, 731-741
80. Nishiya, N., Kiosses, W. B., Han, J., and Ginsberg, M. H. (2005) *Nat Cell Biol* **7**(4), 343-352
81. Brown, M. C., Perrotta, J.A., Turner, C.E. (1996) *J Cell Biol* **135**, 1109-1123
82. Tong, X., Salgia, R., Li, J. L., Griffin, J. D., and Howley, P. M. (1997) *J Biol Chem* **272**(52), 33373-33376
83. Tumbarello, D. A., Brown, M. C., and Turner, C. E. (2002) *FEBS Lett* **513**(1), 114-118
84. Brown, M. C., Turner, C.E. (2004) *Physiol Rev* **84**, 1315-1339
85. Nikolopoulos, S. N., Turner, C.E. (2000) *J Cell Biol* **151**, 1435-1448
86. Nikolopoulos, S. N., Turner, C.E. (2001) *J Biol Chem* **276**, 23499-23505

87. Turner, C. E., Brown, M.C., Perrotta, J.A., Riedy, M.C., Nikolopoulos, S.N., McDoanld, A.R., Bagrodia, S., Thomas, S., Leventhal, P.S. (1999) *J Cell Biol* **145**, 851-863
88. Hervy, M., Hoffman, L., Beckerle, M.C. (2006) *Curr Opin Cell Biol* **18**, 524-532
89. Kadrmas, J. L., Beckerle, M.C. (2004) *Nat Rev Mol Cell Biol* **5**, 920-931
90. Nix, D. A., Beckerle, M.C. (1997) *J Cell Biol* **138**, 1139-1147
91. Fradelizi, J., Noireaux, V., Plastino, J., Menichi, B., Louvard, D., Sykes, C., Golsteyn, R.M., Friederich, E. (2001) *Nat Cell Biol* **3**, 699-707
92. Sadler, I., Crawford, A.W., Michelsen, J.W., Beckerle, M.C. (1992) *J Cell Biol* **119**, 1573-1587
93. Dawid, I. B., Breen, J.J., Toyama, R. (1998) *Trends Genetics* **14**, 156-162
94. Perez-Alvarado, G. C., Miles, C., Michelsen, J.W., Louis, H.A., Winge, D.R., Beckerle, M.C., Summers, M.F. (1994) *Nat Struct Biol* **1**, 388-398
95. Cote, J. F., Turner, C.E., Tremblay, M.L. (1999) *J Biol Chem* **274**, 20550-20560
96. Jamieson, J. S., Tumbarello, D. A., Halle, M., Brown, M. C., Tremblay, M. L., and Turner, C. E. (2005) *J Cell Sci* **118**(Pt 24), 5835-5847
97. Angers-Loustau, A. (1999) *J Cell Biol* **144**, 1019-1031
98. Herreros, L., Rodriguez-Fernandez, J.L., Brown, M.C., Alonso-Lebrero, J.L., Cabanas, C., Sanchez-Madrid, F., Longo, N., Turner, C.E., Sanchez-Mateos, P. (2000) *J Biol Chem* **275**, 26436-26440
99. Waterman-Storer, C. M., Worthylake, R.A., Liu, B.P., Burridge, K., Salmon, E.D. (1999) *Nat Cell Biol* **1**, 45-50
100. Wood, C. K., Turner, C. E., Jackson, P., and Critchley, D. R. (1994) *J Cell Sci* **107 (Pt 2)**, 709-717
101. Tachibana, K., Sato, T., D'Avirro, N., and Morimoto, C. (1995) *J Exp Med* **182**(4), 1089-1099
102. Scheswohl, D. M., Harrell, J. R., Rajfur, Z., Gao, G., Campbell, S. L., Schaller, M. D. (2008) *J Mol Signal* **3**, 1
103. Schaller, M. D. (2001) *Oncogene* **20**, 6459-6472
104. Richardson, A., Parsons, T. (1996) *Nature* **380**, 538-540
105. Panetti, T. S. (2002) *Front Biosci* **7**, d143-150
106. Schneider IC, H. C., Waterman CM. (2009) *J Mol Biol* **20**(13), 3155-3167
107. Zaidel-Bar, R., Milo, R., Kam, Z., Geiger, B. (2007) *J Cell Sci* **120**, 137-148
108. Thomas, S. M., Soriano, P., Imamoto, A. (1995) *Nature* **376**, 267-271
109. Yang, Q., Co, D., Sommercorn, J., Tonks, N.K. (1993) *J Biol Chem* **268**, 6622-6628
110. Nishiya, N., Iwabuchi, Y., Shibanuma, M., Cote, J.F., Tremblay, M.L., Nose, K. (1999) *J Biol Chem* **274**, 9847-9853
111. Bellis, S. L., Perrotta, J.A., Curtis, M.S., Turner, C.E. (1997) *J Biol Chem* **325**, 375-381
112. Ku, H., Meier, K.E. (2000) *J Biol Chem* **275**, 11333-11340
113. Webb, D. J., Schroeder, M.J., Brame, C.J., Whitmore, L., Shabanowitz, J., Hunt, D.F., Horwitz, R.A. (2005) *J Cell Sci* **118**, 4925-4929
114. Ishino, K., Kaneyama, J.R., Shibanuma, M., Nose, K. (2000) *J Cell Biochem* **76**, 411-419
115. Kim-Kaneyama, J. R., Suzuki, W., Ichikawa, K., Ohki, T., Kohno, Y., Sata, M., Nose, K., Shibanuma, M. (2005) *J Cell Sci* **118**, 937-949
116. Wu, R. F., Xu, Y.C., Ma, Z., Nwariaku, F.E., Sarosi, G.A., Terada, L.S. (2005) *J Cell Biol* **171**, 893-904
117. Lewis, J. M., Schwartz, M.A. (1998) *J Biol Chem* **273**, 14225-14230
118. Kain, K. H., Gooch, S., Klemke, R.L. (2003) *Oncogene* **22**, 6071-6080

119. Czaplinski, K., and Singer, R. H. (2006) *Trends Biochem Sci* **31**(12), 687-693
120. Kislauskis, E. H., Zhu, X., and Singer, R. H. (1997) *J Cell Biol* **136**(6), 1263-1270
121. de Hoog, C. L., Foster, L. J., and Mann, M. (2004) *Cell* **117**(5), 649-662
122. Bullock, S. L. (2007) *Semin Cell Biol* **18**, 194-201
123. Brownawell, A. M., Macara, I.G. (2002) *J Cell Biol* **156**, 53-64
124. Monecke, T., Güttler, T., Neumann, P., Dickmanns, A., Görlich, D., Ficner, R. (2009) *Science* **324**, 1087-1091
125. Gwizdek, C., Ossareh-Nazari, B., Brownawell, A.M., Evers, S., Macara, I.G., Dargemont, C. (2004) *J Biochemistry* **279**, 884-891
126. Shiohamaa, A., Sasakia, T., Nodac, S., Minoshimaa, S., Shimizub, N. (2007) *Exp Cell Res* **313**, 4196-4207
127. Woods, A. J., Kantidakis, T., Sabe, H., Critchley, D. R., and Norman, J. C. (2005) *Mol Cell Biol* **25**(9), 3763-3773
128. Lewis, M. K., Burgess, R.R. (1982) *The Enzymes* **15**, 109-153
129. Young, R. A. (1991) *Annu Rev Biochem* **60**, 689-715
130. Sachs, A. B., Wahle, E. (1993) *J Biol Chem* **268**, 22955-22958
131. Whale, E., Keller, W. (1993) *Trends Biochem Sci* **21**, 247-250
132. Blobel, G. (1973) *Proc Natl Acad Sci U S A* **70**, 924-928
133. Baer, B. W., and Kornberg, R. D. (1980) *Proc Natl Acad Sci U S A* **77**(4), 1890-1892
134. Gallie, D. R. (1991) *Genes Dev* **5**(11), 2108-2116
135. Marintchev, A., Wagner, G. (2004) *Quarterly reviews of Biophysics* **37**, 197-284
136. Hershey, P. E., McWhirter, S.M., Gross, J.D., Wagner, G., Alber, T., Sachs, A.B. (2000) *J Biol Chem* **274**, 21297-21304
137. Adam, S. A., Nakagawa, T., Swanson, M. S., Woodruff, T. K., and Dreyfuss, G. (1986) *Mol Cell Biol* **6**(8), 2932-2943
138. Burd, C. G., Matunis, E. L., and Dreyfuss, G. (1991) *Molecular Cell Biology* **11**, 3419-3424
139. Kuhn, U., and Pieler, T. (1996) *J Mol Biol* **256**(1), 20-30
140. Kessler, S. H., and Sachs, A. B. (1998) *Mol Cell Biol* **18**(1), 51-57
141. Oberstrass FC, A. S., Erat M, Hargous Y, Henning A, Wenter P, Reymond L, Amir-Ahmady B, Pitsch S, Black DL, Allain FH. (2005) *Science* **309**, 2054-2057
142. Deo, R. C., Sonenberg, N., and Burley, S. K. (2001) *Proc Natl Acad Sci U S A* **98**(8), 4414-4419
143. Mili S, M. K., Macara IG. (2008) *Nature* **453**, 115-119
144. Amrani, N., Minet, M., Le Gouar, M., Lacroute, F., and Wyers, F. (1997) *Mol Cell Biol* **17**(7), 3694-3701
145. Kozlov, G., Trempe, J. F., Khaleghpour, K., Kahvejian, A., Ekiel, I., and Gehring, K. (2001) *Proc Natl Acad Sci U S A* **98**(8), 4409-4413
146. Bodenhausen, G., Ruben, D.J. (1980) *Chem. Phys. Lett.* **69**, 185-189
147. Grzesiek, S., and Bax, A. (1993) *J Biomol NMR* **3**(2), 185-204
148. Wittekind, M., Meuller, L. (1993) *J Magn. Reson.* **101**, 201-205
149. Sklenar, V., Peterson, R. D., Rejante, M. R., and Feigon, J. (1993) *J Biomol NMR* **3**(6), 721-727
150. Grzesiek, S., Bax, A. (1992) *J Magn. Reson.* **96**, 432-440
151. Clubb, R. T., Thanabal, V., and Wagner, G. (1992) *J Biomol NMR* **2**(2), 203-210
152. Vuister, G. W., and Bax, A. (1992) *J Biomol NMR* **2**(4), 401-405
153. Braunschweiler, L., Ernst, R.R. (1983) *J Magn. Reson.* **53**, 521-528
154. Bax, A., Clore, G.M., Gronenborn, A.M. (1990) *J Magn. Reson.* **88**, 425-431
155. Zuiderweg, E. R. P., McIntosh, L.P., Dahlquist, F.W., Fesik, S.W. (1990) *J Magn. Reson.* **86**, 210-216

156. Marion, D., Driscoll, P.C., Kay, L.E., Windfield, P.T., Bax, A., Gronenborn, A.M. (1989) *J Biol Chem* **28**, 6150-6156
157. Marion, D., Kay, L.E., Sparks, S.W., Torchia, D.A., Bax, A. (1989) *J Am Chem Soc* **111**, 1515-1517
158. Piotto, M., Saudek, V., Skelenar, V. (1992) *J.Biomol. NMR* **2**, 661-665
159. Vranken, W. F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J., Laue, E.D. (2005) *Proteins* **59**, 687-696
160. Wuthrich, K. (1986) NMR for proteins and nucleic acids: Polypeptide secondary structures in proteins by NMR. In: Wuthrich, K. (ed). John Wiley and Sons Inc., New York
161. Wishart, D. S., Bigam, C.G., Holm, A., Hodges, R.S., Sykes, B.D. (1994) *J.Biomol. NMR* **5**, 67-81
162. Guntert, P., Mumenthaler, C., Wuthrich, K. (1997) *J Mol Biol* **273**, 283-298
163. Laskowski, R. (1996) PROCHECK-NMR. In., 3.5.4 Ed.
164. Koradi, R., Billeter, M., Withrich, K. (1996) *J Mol Graph.* **14**, 51-55
165. Johansson, C., Finger, L. D., Trantirek, L., Mueller, T. D., Kim, S., Laird-Offringa, I. A., and Feigon, J. (2004) *J Mol Biol* **337**(4), 799-816
166. Wang, X., and Tanaka Hall, T. M. (2001) *Nat Struct Biol* **8**(2), 141-145
167. Avis, J. M., Allain, F. H., Howe, P. W., Varani, G., Nagai, K., and Neuhaus, D. (1996) *J Mol Biol* **257**(2), 398-411
168. Urbaniak, M. D., Muskett, F.W., Finucane, M.D., Caddick, S., Woolfson, D.N. (2002) *Biochemistry* **41**, 11731-11739
169. Zwahlen, C., Legault, P., Vincent, S., Greenblatt, J., Konrat, R., Kay, L.E. (1997) *J Am Chem Soc* **119**, 6711-6721
170. Dominguez, C., Boelens, R., and Bonvin, A. M. (2003) *J Am Chem Soc* **125**(7), 1731-1737
171. Leung, S. W., Apponi, L.H., Cornejo, O.E., Kitchen, C.M., Valentini, S.R., Pavlath, G.K., Dunham, C.M., Corbett, A.H. (2009) *Gene* **439**, 71-78
172. Loughlin, F. E., Mansfield, R.E., Vaz, P.M., McGrath, A.P., Setiyaputra, S., Gamsjaeger, R., Chen, E.S., Morris, B.J., Guss, J.M., Mackay, J.P. (2009) *Proc Natl Acad Sci U S A* **14**, 5581-5586