University of Kent

MSc by Research Thesis

Characterising a Novel Interaction between Rap1b and Rhea sheds light on new Mechanisms for Focal Adhesion Assembly

Author: W M CASTLE

Degree: BIOCHEMISTRY

Word Count: 22478

Thesis Advisor: B T GOULT

2016

Contents page:

List of figures

List of tables

Abstract

For the first time, we reveal a direct interaction between Rap1b and the fly homolog of talin, Rhea. Using a combination of biochemical and biophysical techniques, the Rap1 binding site on Rhea has been successfully mapped. Additionally, we reveal that an acidic-to-basic K17E substitution, on Rhea, completely abolishes Rap1 binding. Our collaborators have shown that this mutation results in non-viable embryos and our data links the Rap1:Rhea interaction to this lethal phenotype. The implications of our findings support currently proposed mechanisms of RIAM-independent integrin activation, that would challenge our understanding of focal adhesion formation. Furthermore, we propose a double-dependent Rap1 integrin-activation pathway, involving Rap1 directly interacting with the FERM domain, alongside the known Rap1-dependent recruitment of talin.

Optimisations have allowed us to express both the wild-type and mutant Rhea F0 domain in E.coli BL21(DE3) cells. Efficient purification via Ni-NTA-based affinity chromatography results in yields of ~50-60 mg/litre being obtained. Using circular dichroism, it is shown that substitution of the K17 residue does not interfere with the structural integrity of Rhea; both proteins have identical full spectrum measurements and T_m values.

Optimal expression of the conserved G-domain of mouse Rap1b was achieved in the CK600K cell line. This region is highly conserved to that in fly (90% identical). NMR was used to show direct interaction between drosophila Rhea F0 and Rap1b; whilst additionally confirming that Rap1b was unable to induce chemical shifts in the F0-K17E mutant. Triple resonance NMR experiments revealed the location of the Rap1 binding site on the wild-type Rhea F0, with V15, K17, T18, K37 and E40 being highlighted at the centre of this interaction. Structural models of Rap1:Rhea F0 binding agree with our findings, with the 5 highlighted residues seen to make close contact with the Rap1 switch I domain.

Together this work confirms a direct interaction between Rhea and Rap1 whilst providing biochemical validation for the lethal phenotypes observed in mutant flies. It also provides further insight into new mechanisms of focal adhesion formation and integrin activation.

Declaration and copyright statement

I confirm here that I am the sole author of this thesis and that all work produced here is my own, unless clearly stated otherwise.

List of Abbreviations

1D	1-dimensional
2D	2-dimensional
3D	3-dimensional
ml	millilitre
µl	microlitre
g	grams
OD	Optical Density
mg	milligrams
M	Mole
MAPK	Mitogen-Activated Protein Kinase
mM	millimole
µM	micromole
L	Litre
nm	nanometre
rpm	rotations per minute
pI	isoelectric point
Ni	Nickel
FPLC	Fast Protein Liquid Chromatography
EGFR	Epidermal Growth Factor Receptor
SH2	Src Homology Domain 2
ERK	Extra-cellular signal regulated Kinase
SOS	Son of Sevenless
GFR	Growth Factor Receptor
LB	Luria Bertani

WT	Wild Type
X	times (as in 2x would mean 2 times)
Hr	hour
H	hydrogen
kDa	Kilodalton
Da	Dalton
IPTG	Isopropyl ß-D-1-thiogalactopyranoside
UT	untagged
PMSF	phenylmethane sulfonyl fluoride
FID	Free Induction Decay
NMR	Nuclear Magnetic Resonance
HSQC	Heteronuclear single quantum coherence spectroscopy
C	Carbon
N	Nitrogen
CA	Carbon Alpha
CB	Carbon Beta
CO	Carbon in the Carbonyl group of amino acid
NTA	nitrilotriacetic acid
CD	Circular Dichroism
Prep	Preparation
DTT	Dithiothreitol
MWCO	Molecular Weight Cut Off
K	Kelvin
Kan	Kanamycin
Amp	Ampicillin

Acknowledgements

I would like to thank all members of the Goult Lab for their support, advice and expertise over the last year. Special thanks, to my supervisor Ben Goult, without whom my work would not have been possible, and especially for all your help and expertise in NMR.

Thanks to Marie Anderson and Rosie Gough for your encouragement and training over the last year; you have both been fantastic Lab partners and I have learnt a tremendous amount from each of you. You have also both supported me during some of the more difficult aspects of the course, which I do not take for granted.

Many thanks to Guy Tanentzapf at the University of British Columbia, Vancouver, for collaborating with us on this exciting project, and to Dr Igor Barsukov at the University of Liverpool, for all of the cells and constructs you have gifted us.

I would also like to thank the members of the Toseland Lab, especially Ália Dos Santos, for all your help with CD, and for many entertaining conversations. Thanks also to Chris and Natali for your help along the way, and Alex for being a class A guy, who is always helpful and positive.

Thanks to the people in the Rossman lab for letting us use your equipment so frequently. Particular thanks to Matt Badham for being one of the most helpful and supportive guys I have met, and for frequently going out of your way to try and make my life easier.

Thanks to everyone in MADCAP who listened to my presentations and questioned me on aspects of my work over the last year. You have all been welcoming, lovely people; especially Sam, Karen and Holly. You have been incredibly kind and helpful over the year, not least for lending me so many cells, chemicals and concentrators.

I need to also thank my parents, who always support me so much, especially for their financial help over the last year. I would never have reached as far as I have without you both. Also to my sister for her support and last but definitely not least, to the huge support that my girlfriend has given me over the last 6 years. Thanks for believing in me and letting me achieve more than I thought I ever could.

To anyone and everyone else who has helped or guided me over the last year, thank you too.

Section 1: Introduction

1.1 The GTPase Molecular Switch Mechanism

In order to understand the interactions between Rap1b and Rhea, first these proteins must be put into structural and functional context. Rap1b is a monomeric GTPase. A key feature of the small monomeric GTPases is that they can alternate between active and inactive forms, thereby acting as molecular switches (Cherfils and Zeghouf, 2013). In the active form, GTPases are bound to GTP, contrasting their binding to GDP when inactive. This GTPase switch mechanism is intrinsically tied to the regulation of a large array of cellular functions (Cherfils and Zeghouf, 2011). The cycling of GTPases between their GTP/GDP bound forms is regulated by Guanine nucleotide Exchange Factors (GEFs) which activate the GTPase, and GTPase Activating Proteins (GAPs) which hydrolyse the GTP to GDP inactivating the GTPase, (Wennerberg, 2005) (see Figure 1, left).

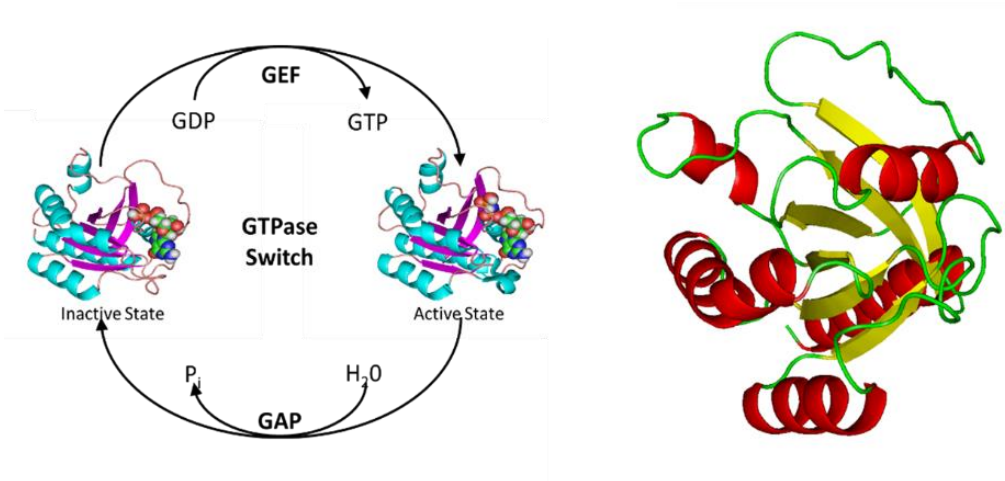


Figure 1: The GTPase Switch Cycle and GTPase structure
(left) The classic, simplified, model of the GTPase switch showing how a small GTPase can switch between inactive and active states by exchanging GDP for GTP. As these proteins have intrinsically slow GTPase activity, inactivation is sped up by GEFs. (right) the secondary structure of a typical GTPase; usually comprises of 5 helical and 6 beta structures.

The study of this molecular switch domain traces back to the Nobel Prize winning discovery of the heterotrimeric G-proteins, named by their ability to bind guanine nucleotides, by Alfred Gilman and Martin Rodbell in the 1970s (reviewed here: Birnbaumer, 2007). Small GTPases are homologous to the α-subunit of heterotrimeric GTPases, and therefore share the common G domain. Separating the monomeric GTPases from the G proteins however, is their ability to function independently of the beta and gamma subunits, allowing them to function within the cytosol. The crystal structures of Ras bound to GTP and GDP were solved between 1989 and 1990 (Pai et al., 1989 and Milburn et al., 1990). GDP dissociation, is required for protein activation. This occurs intrinsically, however is inefficient. GEFs speed up this process, facilitating GDP to GTP exchange. The intrinsic GTPase activity within the proteins is also a slow process. GAPs speed up this process, of hydrolysing GTP to GDP, in order for GTPases to be inactivated (Vetter and Wittinghofer, 2001).

The G domain shared between all members of the Ras family is typically between 20-25 kDa. It is comprised of 5 alpha helical and 6 beta structures (see Figure 1, right).

Key features, conserved between the Ras GTPases, are the P-loop (involved in phosphate binding), and the two switch domains, designated Switch I and Switch II (see Figure 2A). These switch domains are the only structural elements that change between the active and inactive forms of the protein (see Figure 2 A-D). When bound to GTP, the gamma phosphate is able to hydrogen bond to conserved T35 and G60 residues (in switch I and II respectively), which changes the conformation of the switch domains, pulling them towards the nucleotide binding site. As the gamma phosphate is cleaved, the two switch domains spring back to their original, inactive conformation. Reactivation of the GTPase requires GDP molecules to be removed and exchanged for GTP (Vetter and Wittinghofer, 2001).

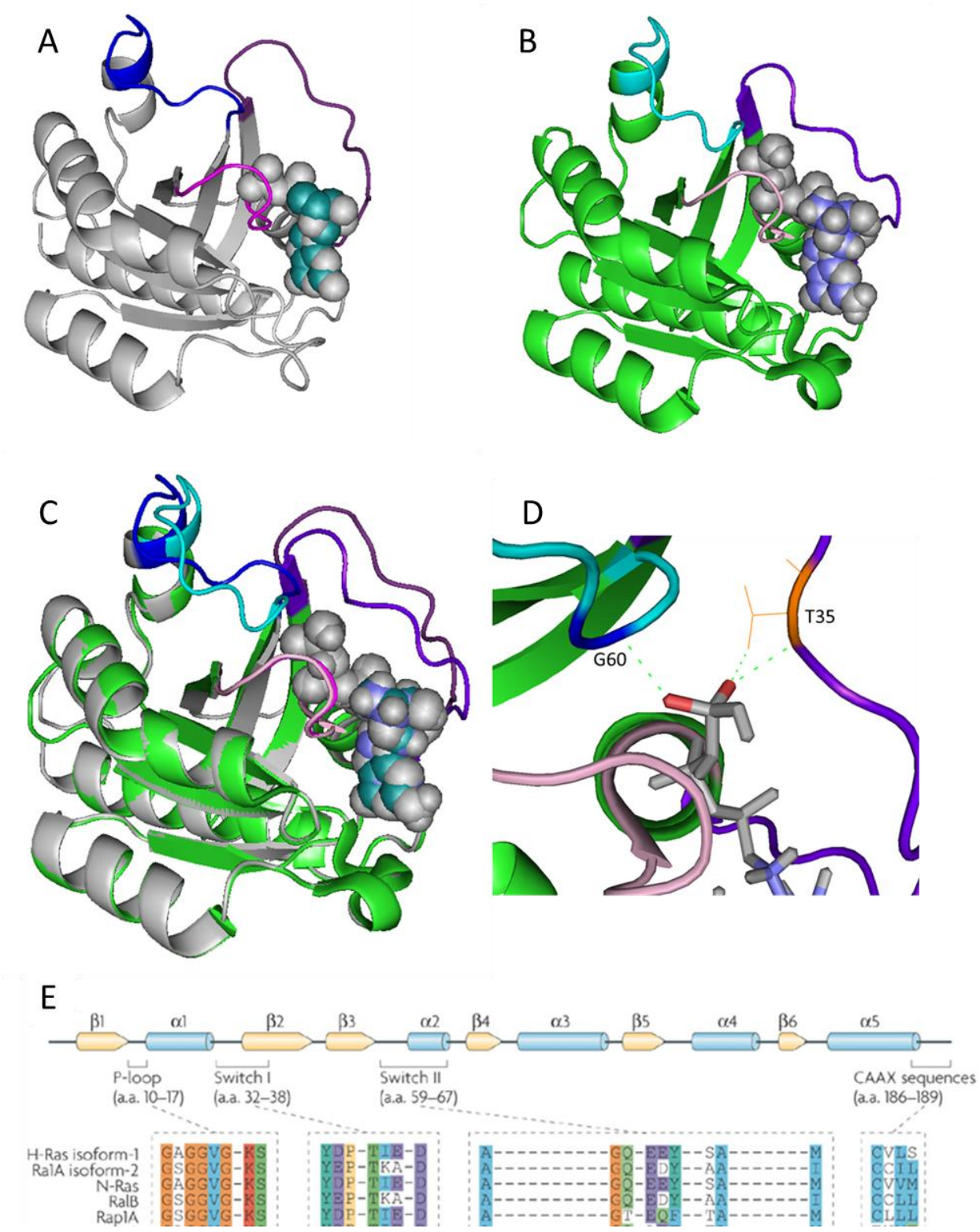


Figure 2: Structure of P-loop and Switch Regions

(A) the GDP bound form of a GTPase with the switch I coloured in dark purple, the switch II in blue and P-loop in pink. (**B)** the GTP bound, active form and (**C)** the overlay shows that the switch domains are the only two regions of the protein that change between its active and inactive forms. This is because the gamma phosphate forms key hydrogen bonds **(D)** that re-configure the switch domains. **(E)** The P-loop and switch domains are

highly conserved between H-Ras, N-Ras and Rap1 *(Karnoub and Weinberg, 2008)*.

1.2 Overview of the Ras GTPase family

The small GTPases belong to the Ras superfamily, of which there are over 150 members. This superfamily can be subdivided into five major subfamilies, each with distinct functions: the Ras, Rab, Rho, Arf and Ran families (Wennerberg, 2005) (see table 1). There are 61 Rab, 36 Ras, 27 Arf, 20 Rho, 1 Ran and 9 other GTPase human genes (Wennerberg, 2005). Due to their role in oncogenesis, the Ras subfamily has been extensively studied; their name derives from "Rat Sarcoma" inducing viruses identified in the 1960s (reviewed in Malumbres and Barbacid, 2003). The Ras proteins are major regulators of cell survival and proliferation (Zhang and Liu, 2002). The major role of the Rab proteins, closely related in function to the Arf family, is to regulate and coordinate each stage of the transport of intracellular vesicles (Zerial and McBride, 2001). The Rab family is the largest subfamily of the Ras proteins (Rojas, A and Valencia A. 2014). The Rho family has a principal involvement in regulating actin and cytoskeleton dynamics (Jaffe and Hall, 2005). Finally, the single human member of the Ran family, Ran, is responsible for formation of, and transport through, the nuclear envelope, in addition to mitotic spindle assembly (Stewart, 2007). Although other species (particularly plant) have multiple Ran proteins, there is only one human Ran protein, and it is the most abundantly expressed GTPase protein in human (Rojas, A and Valencia A. 2014). Ran proteins are distinct enough from the other GTPases to be considered their own family. Ran, and Arf proteins are unique from the Ras, Rab and Rho families, in that their structural changes are different following GTP hydrolysis. Unlike the changes observed in the switch domains of the other three families, the Ran and Arf proteins form an extra switch I domain beta-strand when in the GDP bound state (Goldberg, 1998).

The role of Ras is best understood in the classically studied Erk Mitogen-activated protein kinase (MAPK) pathway (see Figure 3) in which growth factors binding to receptors, such as Epidermal Growth Factor Receptor (EGFR), on the cell surface, inducing a cascade of intracellular kinase activity. Upon growth factor binding, EGFR dimerizes and auto-phosphorylates tyrosine residues in its cytoplasmic domain. These phosphorylated residues bind to the Src Homology (SH2) (Felder et

al., 1993) of the Shc protein. Following subsequent Shc phosphorylation, Grb2 binds to Shc via its own SH2 domain (Walker et al., 1998). Grb2 also has an N terminal SH3 domain that is able to bind proline-rich regions of the Ras-GEF SOS. SOS activates Ras, which activates Raf. Raf then phosphorylates and activates MEK, which in turn does the same to ERK. ERK is then translocated to the nucleus where it activates cyclinD, fos, CREB and other survival and proliferation factors (Kim and Choi, 2010). As expected for a key regulator of cell proliferation, this pathway is rife with potential for malignant transformation. Constitutively active GFRs, Ras and ERK are all capable of cell transformation, with Ras mutants reported to be present in 30% of all cancers. This figure is however, potentially misleading, due to screening bias (Prior, Lewis and Mattos, 2012). Never the less, in certain cancers, Ras mutations have a very high prevalence; particularly colorectal cancer (Prior, Lewis and Mattos, 2012).

Table 1: The Ras Family

Family	Subfamily	Members (PDB accession codes)
Ras	Ras	H-Ras, K-Ras, N-Ras (3con), M-Ras, R-Ras (2fn4), R-Ras2/TC21 (2ery)
Ras extended	Ral	RalA, RalB
	RheB	RheB, RheBL1 (3oes)
	Rap	Rap1A, Rap1B, Rap2A, Rap2B Ras-3 (4ku4, Ras-like protein from Cryphonectria parasitica) Di-Ras1 (2gf0), Di-Ras2 (2erx)
	RGK	Rad, Gem/Kir, Rem1 (2nzj), Rem2 (3cbq, 3q85, 4aii)
	Rerg	Rerg (2atv), RasL12 (3c5c)
Rho	Rho	RhoA, RhoB, RhoC, EhRho1 (3ref, 3reg, complex with Diaphanous-Protein: 4dvg)
	Rac	Rac1, Rac2, Rac3, Rop4, Rop5, Rop7, Rop9
	Cdc42	Cdc42, RhoUA (2q3h), TC10 (2atx)
	RhoD	RhoD (2j1l), Rnd1 (2cls, complexes with plexin: 2rex, 3q3j), RhoE/Rnd3
Rab	Rab	Rab1,2,3,4,5,6,7,8,9,11,12,14,18,21,22,23,25,26,27,28,30,31,33,35,43
	RasEF	RasEF (2p5s), IFT27 (2yc2/2yc4)
	Ypt	Ypt1,7,8,32,51
	Sec	Sec4
Arf	Arf	Arf1,2,4,5,6,8, Arl2,3,5,6,8,10,13
Sar	Sar	Sar1
Ran	Ran	Ran, RanE (4djt, Nuclear GTP-binding protein from Encephalitozoon cuniculi)

Table 1: An overview of the members of the Ras family (Vetter, 2014).

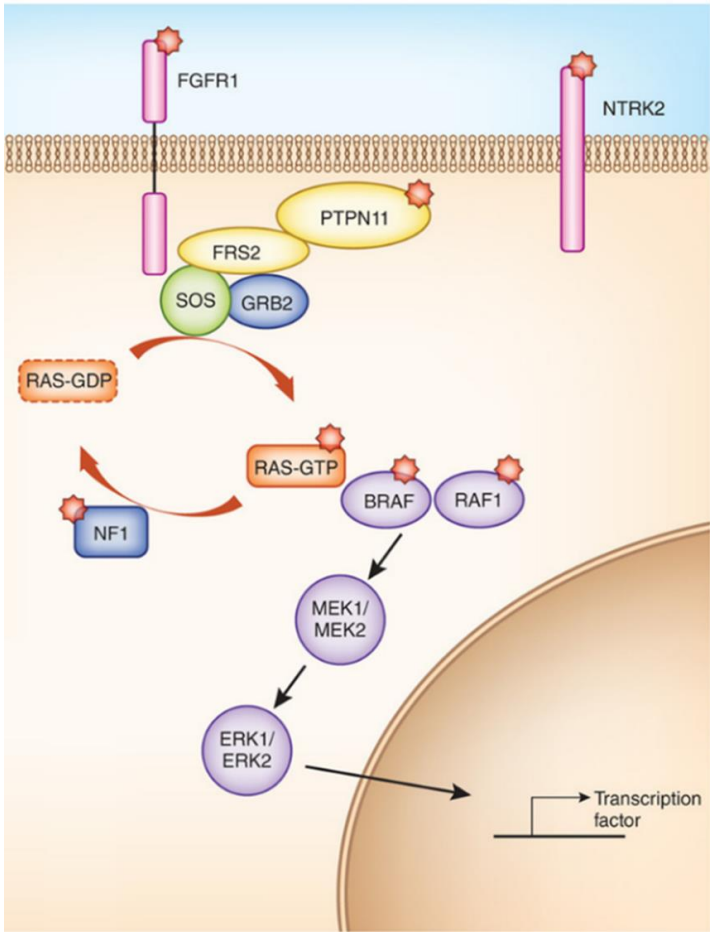


Figure 3: **The Ras MAPK Pathway (Turcan and Chan, 2013)**
Growth factors binding to receptors, such as Epidermal Growth Factor Receptor (EGFR), on the cell surface, inducing a cascade of intracellular kinase activity. Upon growth factor binding, EGFR dimerizes and auto-phosphorylates tyrosine residues in its cytoplasmic domain. These phosphorylated residues bind to the Src Homology (SH2) (Felder et al., 1993) of the Shc protein. Following subsequent Shc phosphorylation, Grb2 binds to Shc via its own SH2 domain (Walker et al., 1998). Grb2 also has an N terminal SH3 domain that is able to bind proline-rich regions of the Ras-GEF SOS. SOS activates Ras, which activates Raf. Raf then phosphorylates and activates MEK, which in turn does the same to ERK. ERK is then translocated to the nucleus where it activates cyclinD, fos, CREB and other survival and proliferation factors (Kim and Choi, 2010).

1.3 The Rap subfamily: key regulators of focal adhesion assembly

The Rap proteins were discovered in 1988 (Pizon et al 1988) due to their high sequence identity with human K-Ras. Sequence alignment of Rap1 with K-Ras gives the two proteins a sequence identity of 58%. The Rap proteins are among the closest related homologs to Ras, with the five isoforms sharing approximately 50% identity to Ras (Pannekoek and Bos 2014) (see Figure 4). The five Rap isoforms are

all highly conserved with the exception of their hypervariable C terminus; responsible for subcellular localisation of the Rap (Prior and Hancock, 2012).

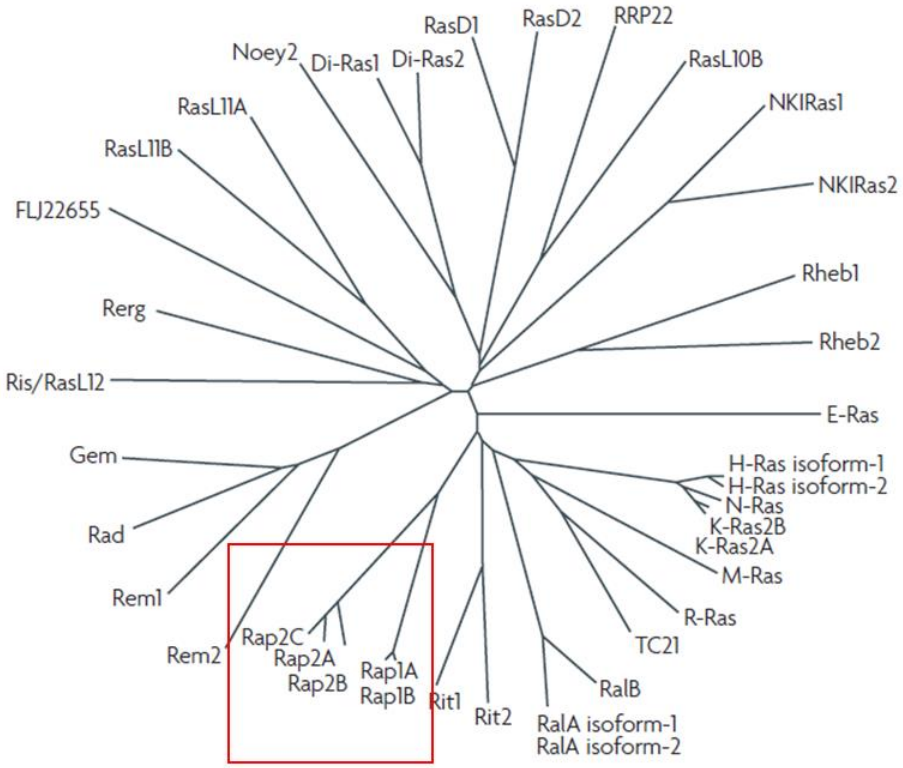


Figure 4: Relationship of Rap to members of the Ras Family
Phylogenetic tree of Ras proteins showing that there are five Rap proteins: Rap1A and Rap1B, Rap2A, Rap2B and Rap2C. The Rap isoforms have been indicated in red. The close relationship of the Rap proteins to the different Ras proteins is indicated by the proximity of the Rap and Ras branches of the phylogenetic tree (Karnoub and Weinberg, 2008).

Interestingly, Rap proteins were initially highlighted by their ability, when over expressed, to suppress cell transformation induced by Ras (Kitayama et al., 1989). It is possible that the reversal of this phenotype is due to competitive binding of Ras and Rap for Raf, as both proteins have very similar binding domains for their effectors (Chen et al., 1997), with Rap being unable to functionally activate the MAPK pathway. However, there is evidence that Rap proteins are able to influence MAPK pathways that challenges this idea. In Aplysia californica for example, the Ras and Rap1 homologs are able to both activate and suppress the MAPK pathway in a context dependent manner (Ye et al., 2008). The predominant, current theory is that Rap is able to reverse the Ras-induced transformation due to its ability to regulate integrins, enhancing cell adhesion, and thereby altering the invasive capability of transformed cells. Although, the mechanisms remain unclear it is now

known that the Rap1 pathway is intrinsically linked to the metastatic potential of tumours. If for example, the Rap pathway is upregulated by depleting Rap1GAP, human cancer cells have a tendency towards increased aggressive migration (Tsygankova, Wang and Meinkoth, 2013). Similarly, activate Rap1 increases murine metastatic incidence when introduced into prostate cancer cells, while upregulating Rap1GAP does the converse (Bailey 2009). It is now known that Rap is intrinsically linked to integrin-mediated cell adhesion (reviewed: Bos, 2005).

Altered Rap1 activity has been studied in developmental models. Here it has been shown that lower organisms have a single isoform of Rap1, and its deletion leads to lethality (Knox, 2002). This highlights the critical role of Rap1 in morphogenesis; the event of highest cell migration in an organism's life cycle. In the mouse model, it is hypothesised that Rap1a and Rap1b may have redundancy, as deletion of either of these genes results in a similar number of embryos surviving to adulthood (Chrzanowska-Wodnicka et al., 2015). Rap1 appears to have a crucial role in cell-cell junction formation (Knox, 2002). In drosophila, Rap1 has been shown to be a key regulator of cell migration, with Rap1 deletion causing defects in the migration of certain cell populations (Asha et al., 1999). Observed phenotypes include dorsal closure defects and severely disrupted mesodermal cell migration (see Figure 5).

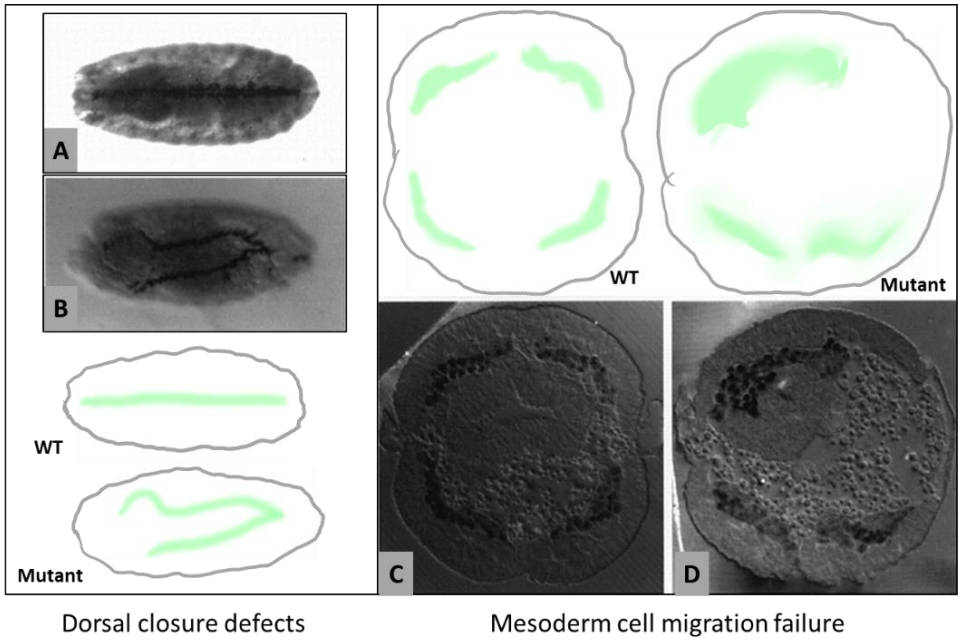


Figure 5: Dorsal closure and mesoderm cell migration failure in Rap1 deletion embryos

*Dorsal closure in **A** wild-type, and **B** Rap1 deleted drosophila, dorsal closure defects are observed following Rap1 deletion. **C** (wild-type) and **D** (mutant) show cell migration failure for mesoderm cells (Asha et al., 1999). Original illustrations of each are included for clarity.*

The inherent importance of Rap can be traced back through its evolutionary history. Rap was likely present in the last eukaryotic common ancestor, over a billion years ago, and possibly precedes this time. Supporting evidence for this comes from a Rap orthologue being present in *Dictyostelium discoideum* and *Naegleria gruberi*. Rap GAP and GEF orthologues are also present in *Naegleria gruberi* (van Dam, Bos and Snel, 2011).

1.4 Rap1, Rhea and Focal Adhesion Assembly

As mentioned previously, alongside its role in cell-cell junctions, the function of Rap1 is intrinsically tied to that of integrin activation and the coordination of focal adhesion assembly (Goult et al 2010). Without the ability to adhere to the extracellular matrix (ECM), the development of current multicellular organisms could not have arisen (Aszodi et al, 2006). Co-ordinating this process is highly complex and involves many additional proteins (Campbell and Humphries, 2011). Integrin receptors, cell surface receptors that make the direct linkage between the cell and ECM, are at the heart of this process. Two main methods are able to begin the assembly of a focal adhesion complex: where the signal is received from inside, or outside of the cell. In the first, “inside out” pathway, intracellular signaling molecules lead to integrin activation. In the second, “outside in” pathway, complex formation is initiated via integrin binding to the ECM (Shattil, Kim and Ginsberg, 2010., Goult et al 2010). A crucial adaptor protein in this process is Talin, the fly homolog of which is named Rhea. Talin has two distinct domains: a FERM head domain and a Rod domain. Like other FERM Domains, the talin head domain is comprised of F1-F3, but is atypical in that it has an additional N-terminal F0 domain containing an extra ubiquitin like fold (similar in structure to the F1 domain) (Goult et al., 2010). The rod domain contains 13 compact four-, and five- helical bundles (Goult et al., 2013) (see Figure 6).

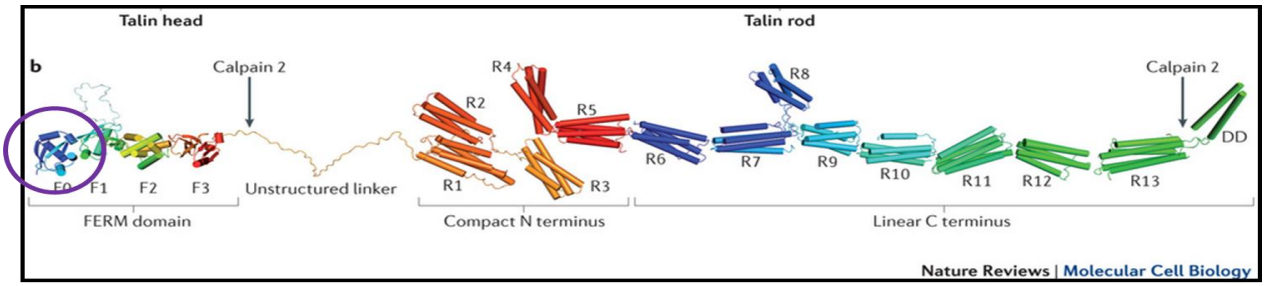


Figure 6 Structural overview of Talin, including the binding site of Rap1 on F0 (Calderwood et al 2013).
The N-terminal contains an F0 domain which is atypical of FERM domains. The FERM domain also contains the F1, F2 and F3 domains, typical of FERM domains. F3 contains a phosphotyrosine binding-like domain. This site binds to the cytoplasmic tails of ß-integrins. Proceeding the FERM domain, and an unstructured linker region, is the Talin Rod domain. This comprises of thirteen helical bundles, each containing 4 or 5 helixes. F0 contains a Rap1 binding site. RIAM binds to the talin R2/R3 subunits of the rod domain

The current model for inside-out integrin activation involves Rap1 leading to integrin activation via a Rap1 effector, RIAM (Rap1-GTP-interacting adaptor molecule) binding to membrane associated active Rap1, and recruiting talin (Goult et al., 2013). RIAM binds to the talin R2/R3 subunits of the rod domain (Shattil, Kim and Ginsberg, 2010., Goult et al., 2013). An additional hypothesis proposes that RIAM is autoinhibited, and the talin binding (TB) domain on RIAM, covers its pleckstrin homology (PH) domain (Zhang et al., 2013). Binding of the RIAM Ras association (RA) domain to Rap1 releases the TB domain, which subsequently binds to the talin rod, and allows the PH domain to tether RIAM to the membrane (Zhang et al., 2013) (see Figure 7). Wegener et al., (2007) demonstrated that integrin cytoplasmic beta tails are able to bind to talin F3 via a membrane proximal helix and an NPxY motif in the integrin tail (see Figure 8). Usually, integrin is kept in a low affinity state via a salt-bridge between the α and ß cytoplasmic tails, but Anthis et al. 2009, have shown that this salt-bridge is disrupted by talin F3, causing integrin activation.

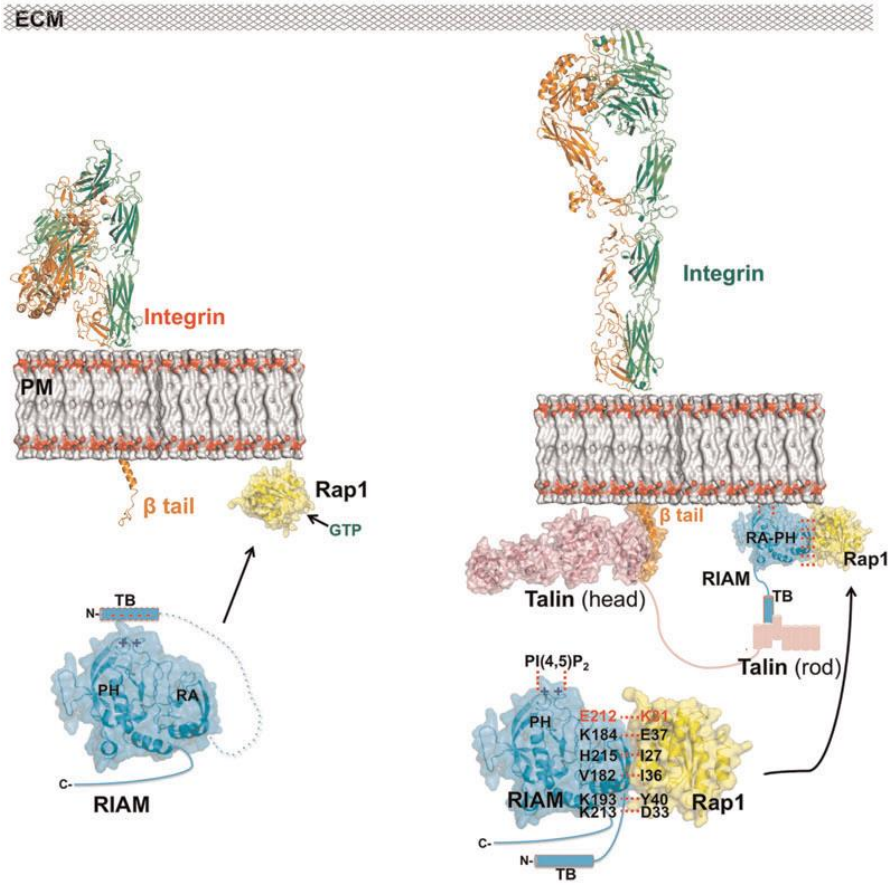


Figure 7: Model for inside-out integrin activation (Zhang et al., 2013)

The inactive state of integrin is shown on the left, with active integrin being shown on the right. (Zhang et al., 2013). On the left, active Rap1 is bound to GTP and anchored to the membrane, awaiting RIAM binding. RIAM is shown to be in a possible inhibited, inactive form, where the talin binding region (TB) is covering the Ras Association- (RA) and pleckstrin homology- (PH) domains. On the right, RIAM has bound to active Rap1 via the RA domain. This has released the TB domain, and revealed the PH domain. The PH domain allows RIAM to anchor to the membrane, alongside Rap1. Talin can then be recruited, which binds to the RIAM TB domain. The talin head domain can then bind to the β-integrin cytoplasmic tail, activating integrin.

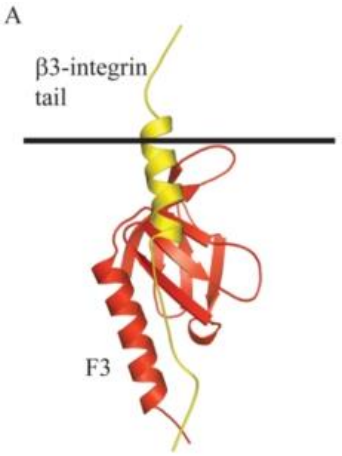


Figure 8: Binding of β3 integrin tail to the talin F3 domain
Talin is in red, integrin yellow, and the membrane in black (Anthis et al., 2009. Goult et al., 2009).

Interestingly, studies performed previously by the Goult lab have shown that Rap is able to bind directly to the mouse talin F0 domain (see Figure 9), suggesting alternative Rap1 dependent adhesion mechanisms (Goult et al., 2010).

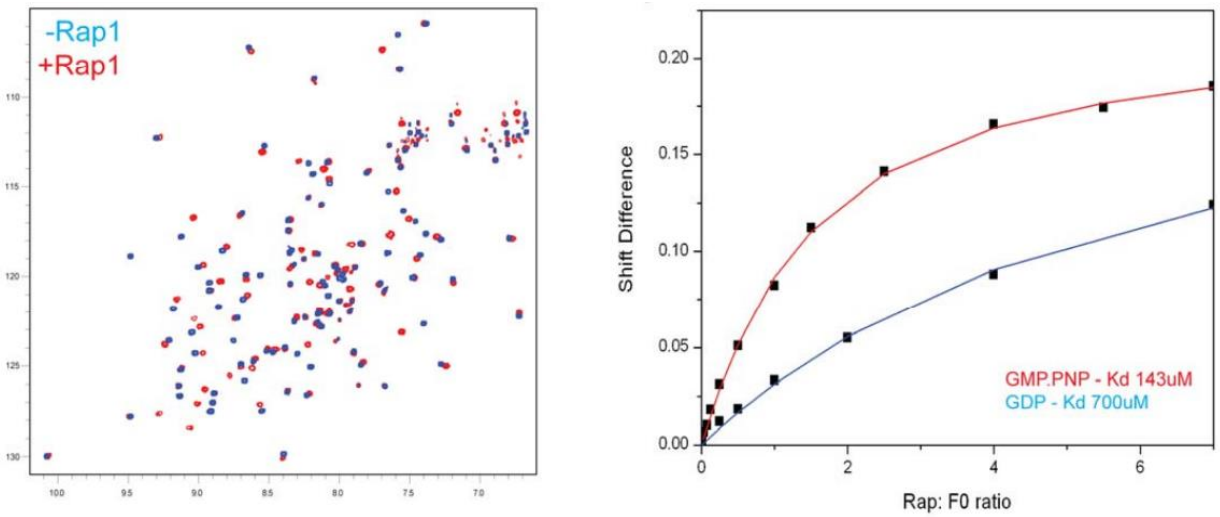


Figure 9: Rap is able to bind to the F0 domain of Talin
*(**left**) NMR shifts induced in mouse talin F0 domain by the addition of the Rap protein and (**right**) Binding of Rap1 to F0 is considerably tighter when Rap is in the GTP bound state (figure taken from Goult et al., 2010).*

1.5 Overview of Nuclear Magnetic Resonance

1.5.1 Basic Principles of NMR: NMR in 1 Dimension

The content of the following section, on the basic principles of NMR, is from a series of video tutorials produced by Magritek Ltd and delivered by Sir Paul Callaghan. All information from this section will be assumed to have come from this source, unless cited otherwise. The following description of NMR will avoid the quantum mechanics necessary for a deep understanding of NMR, in an attempt to keep the understanding at a level that is more appropriate to those with little Biophysical background.

Understanding NMR requires a basic knowledge of how atomic nuclei behave in an applied magnetic field (B_0). A key property of nuclei, that allows us to study them via NMR, is magnetism. This property results in the magnetic dipole of the nuclei aligning with a magnetic field when placed in it. The further key properties of atomic nuclei that aid this understanding, are angular momentum (spin) and the fact that, in the case of ^{1}H, ^{13}C and ^{15}N, the nuclei can exist in one of two orientations (Nerz-Stormes, 2016), or quantum states (see Figure 10A). The first is known as “up spin” in which the orientation is aligned with the magnetic field, which is the lowest energy state for the nuclei to reside. The second orientation is “down spin,” which is aligned against the magnetic field, therefore occupying a higher energy level. In order to disrupt the nuclei from their natural orientation, firstly, a “torque” (a second electromagnetic field) must be applied, as a pulse, orthogonally to the magnetic field. Secondly, the “torque” must oscillate (varying with time) at the same frequency (resonance) as the precession frequency of the nuclear spin. When nuclei are at 90° to the magnetic field, they will precess, rotating around the axis of the magnetic field, and undergo relaxation until they re-orientate back along the axis of the magnetic field (see Figure 10E). With time, as the nuclei re-orientate themselves, the precession decreases until the angular momentum is back at equilibrium. We can say that precession is dependent on the orientation of the nuclei; it is not observed when the nuclei are in either of their two natural states. Precession frequency is the time it takes for the nuclei to

precess a single time, and remains constant at all orientations aside from in the up or down spin states.

The Boltzmann factor defines the proportion of atomic nuclei in the higher and lower energy quantum states and is defined by the proportion of magnetic energy to thermal energy (Rattle, 1995). The proportion of atomic nuclei occupying the lower energy orientation will increase as either the temperature decreases, or the magnetic energy increases. At absolute zero all spins will be in the low energy orientation, however, at room temperature, thermal energy becomes a factor in the Boltzmann distribution, and there is a distribution of atomic nuclei in each state; the thermal energy will allow the atomic nuclei to occupy the down spin orientation. As expected however, there will be a slight preference for the lower energy state, so there will be slightly more atomic nuclei occupying this quantum state (see Figure 10C). This imbalance in the populations is called a "spin excess." At room temperature, the Boltzmann distribution states that we will have a near equal number of atomic nuclei in each state. The populations occupying the up and down spins will cancel each other out and are not seen. The spin excess however, is visible. This is due to the fact that the low proportion in the spin excess still equates to a very large number of actual atomic nuclei in a given sample. The Boltzmann population allows us to predict two ways of increasing the spin excess. Firstly, we could perform the NMR experiments at a lower temperature. Secondly, we can increase the strength of B_0, whilst fixing the temperature.

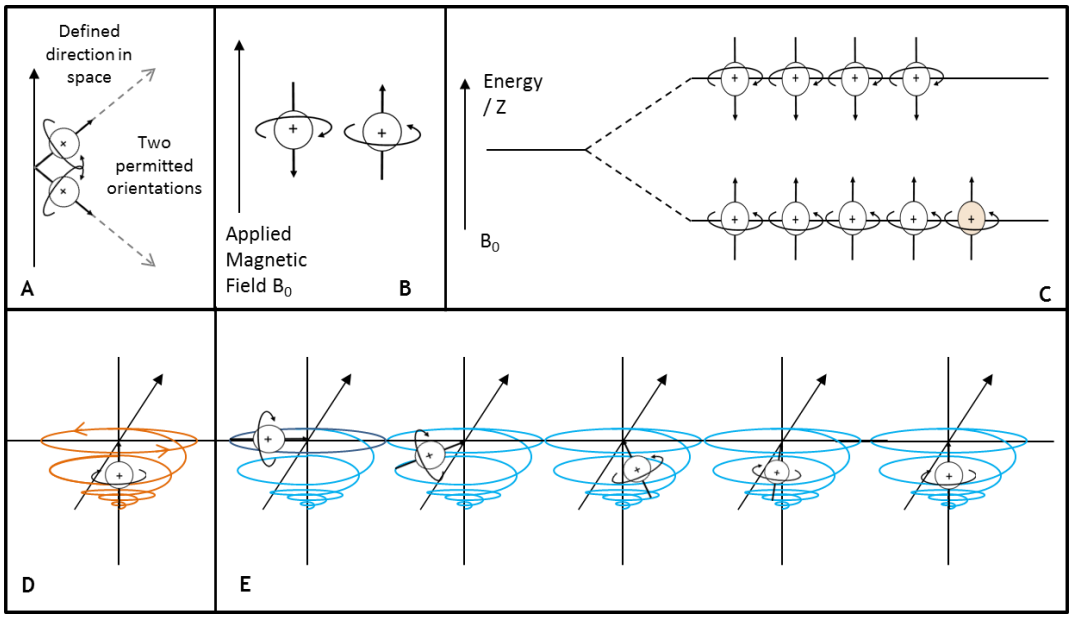


Figure 10: Schematic cartoons illustrating the basic principles of NMR

***A**: ^{1}H, ^{15}N and ^{13}C nuclei are restricted to two orientations in space. **B** when the nuclei are placed in an applied magnetic field, their orientations will align with ("up" spin), or against ("down" spin), the magnetic field. **C** aligning with the magnetic field is the lowest energy state, aligning against it requires energy. At absolute zero, the entire population is in the up spin orientation. At room temperature, thermal energy is present, allowing the nuclei to be in either orientation. There is a small tendency for the population to have slightly more nuclei in the lower energy state (orange), which is called the spin excess. The spin excess is observed in NMR, as signals from the up and down spin nuclei cancel out. **D** firing an oscillating pulse at 90° in resonance with the Larmor frequency of the nuclei will disrupt the orientation of the spin so that **E** the nuclei begin to precess. This precession induces an oscillating voltage in the NMR detector coil, which fades as the nuclei return to their original orientation.*

The precession of an atomic nucleus is known as Larmor precession; which takes place at the Larmor Frequency. This frequency is directly proportional to the magnetic field (B_0); a larger B_0 will increase the precession frequency (the Larmor Frequency is defined as $\omega = \gamma B_0$ where gamma is the gyromagnetic ratio (dependent on the nature of each atomic nucleus)). The last feature of the 1D NMR experiment is the coil. This is responsible for producing the "torque:" a transverse, oscillating magnetic field at the same frequency as the Larmor Frequency of the atomic nuclei, applied as a pulse. Not only does this coil produce the oscillating magnetic field, it also serves as a detector. As the nuclei precess, their fluctuating magnetic fields produce an oscillating voltage in the coil, which corresponds to its precession. This signal decays over time, as the precession returns to its equilibration state, and is known as Free Induction Decay (FID). The FID is a collection of the Larmor frequencies of all of the protons in a unique chemical environment within the sample. Each of these will therefore be decaying at different rates. A FID can be converted via Fourier transform, from a function of time, to a function of frequency, thereby giving us the 1D NMR spectrum (see Figure 11) (Rattle, 1995).

Due to shielding from electrons, different hydrogen environments correspond to peaks in different regions of the NMR spectrum.

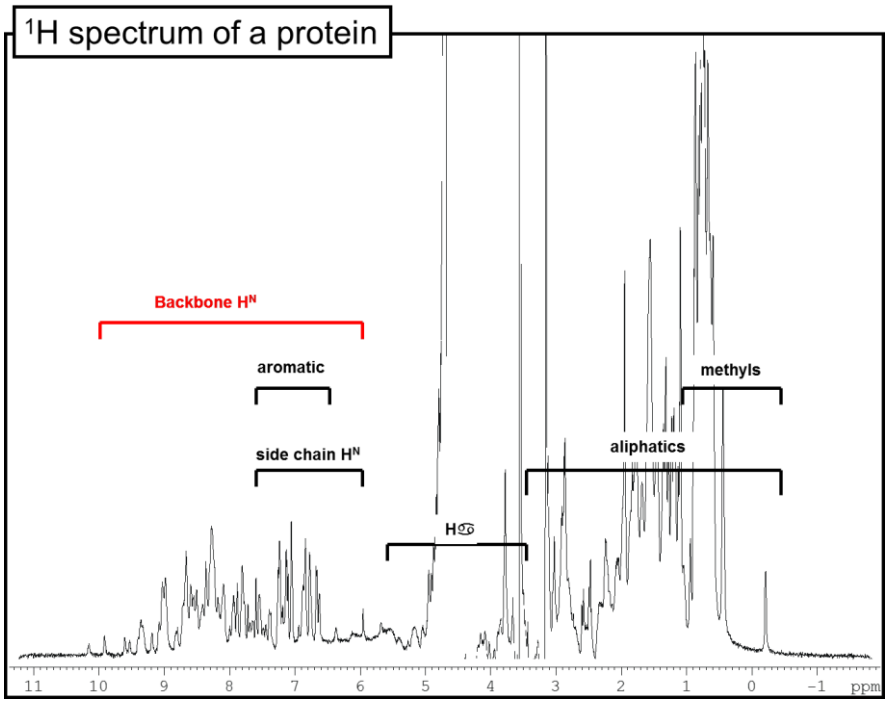


Figure 11: The different regions of a typical NMR spectrum (Albert Einstein College of Medicine, 2012).

See Figure 13A for a visual representation of proton NMR.

1.5.2 Basic Principles of NMR: A second dimension, HSQC

Having determined a background for the physics behind NMR, an important two-dimensional experiment, the HSQC (heteronuclear single quantum coherence) will be briefly explained. HSQC was first developed in 1980 (Bodenhausen and Ruben, 1980) and is an abundantly used technique for studying proteins. Simply put, the experiment involves the transfer of magnetization between the proton and a ^{15}N (or ^{13}C) atomic nucleus it is bonded to (Pascal, 2008). In a protein, the backbone of each amino acid has a proton attached to a nitrogen within its amide group (Figure 13A). The naturally occurring isotope of nitrogen, nitrogen-14 is not magnetically active, therefore, to be able to observe the nitrogen signals, the proteins are grown up in minimal media, containing the ^{15}N isotope as the sole nitrogen source. The HSQC plots the relationship of the proton spectrum against the coupled nitrogen. Each amino acid within the protein therefore gives a distinct peak (Figure 13B), with the exception of proline (Pascal, 2008). As the ^{15}N is the only source of nitrogen in the media, it will also be incorporated into the side chains of any nitrogen-containing residues. The advantage of this second

dimension, is that the peaks are dispersed, unlike the peaks in a 1D spectrum, making them more easily observed and distinguished, with less data overlap (Rattle, 1995). Like the 1D spectra, different amino acids are generally restricted to distinct, broad regions of the HSQC spectra (Figure 12).

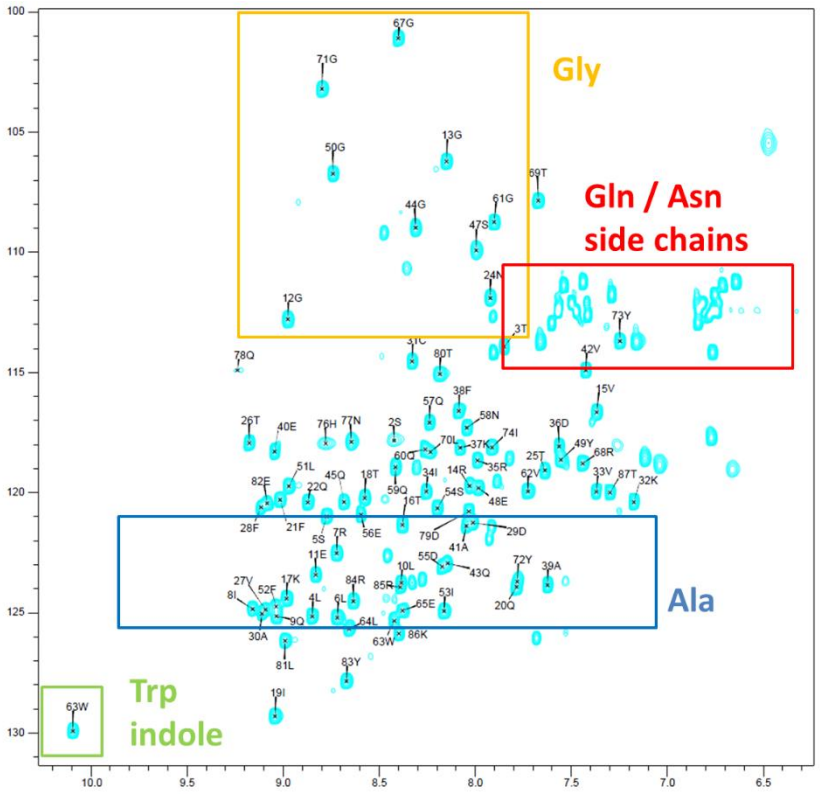


Figure 12: HSQC spectrum of Rhea F0-WT showing broad regions where certain residues typically appear.

1.5.3 Basic Principles of NMR: Triple Resonance Experiments for Backbone Assignment.

Having performed a HSQC, it is possible to go one step further and add a third dimension to the spectra and couple the amide to adjoining carbons. Triple resonance NMR allows the assignment of the spectra (identifying the exact amino acid that each peak corresponds to), thus allowing for more meaningful conclusions to be drawn from the data. Like the HSQC, triple resonance experiments involve looking at the correlation between ^{1}H and ^{15}N whilst adding the additional dimension of ^{13}C (Figure 13B-C). This allows the peaks of the HSQC to be viewed in 3D, so that more relational information can be obtained from the spectra. For example, it now allows us to view overlapping peaks. It also allows us to do the backbone assignment of the protein. In the assignments done in this

study, a series of five different triple resonance experiments were performed to complete this task. The principle of backbone assignment is that these experiments will give a series of peaks that identify different carbon resonances within the amino acid chain.

We have to consider a frame of reference for the backbone; and the common nomenclature designates the amino acid we are looking at as "i," with the preceding amino acid, in the N-terminal direction, as "i-1" and "i+1" being the amino acid in the other direction (Figure 13A). The experiments transfer magnetization, by coupling spins, between carbons in i and i-1. This allows us to detect which amino acid we are looking at, and which it is likely next to. A more detailed explanation of the different experiments, and assignment procedure, is covered in the methods section.

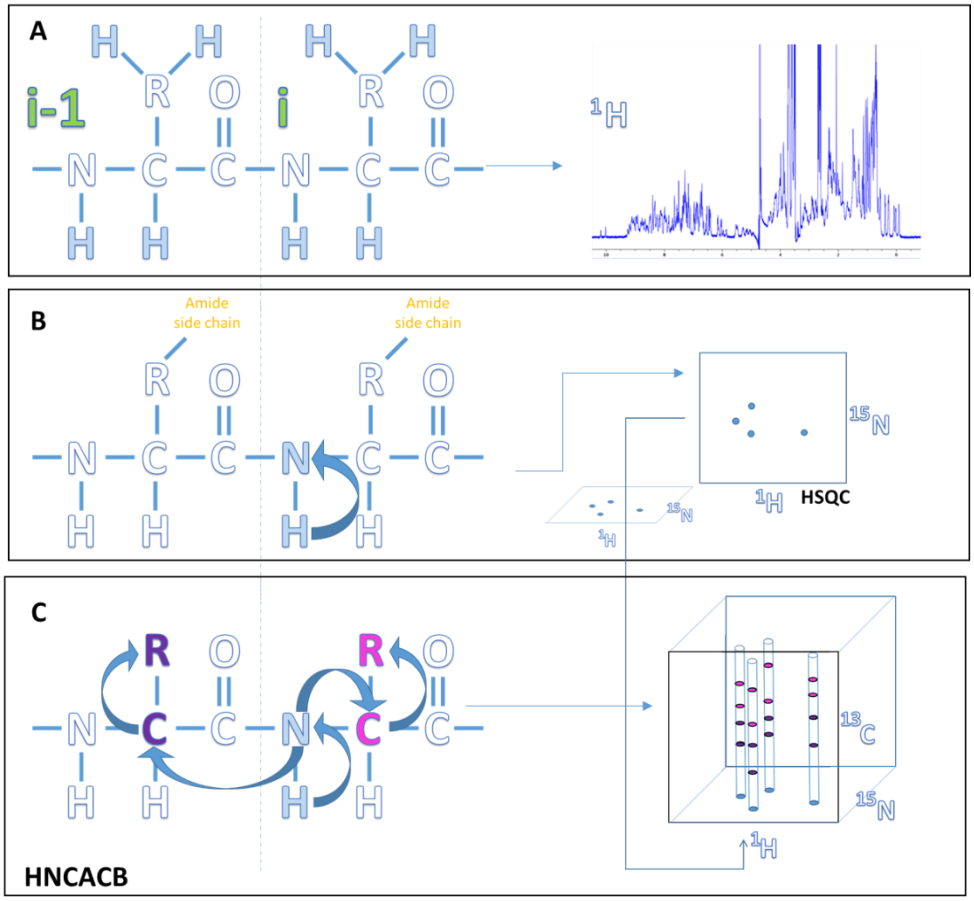


Figure 13: Schematic Representation of the difference between 1D, 2D HSQC and 3D NMR

A) *1D NMR allows us to view the chemical shifts of protons*

B) *HSQC transfers magnetisation from protons to ^{15}N and back to give us a second dimension, and observable peaks for each amino acid residue. Side chains are also*

observed. The arrows indicate the magnetisation transfer.

C) *Transferring the magnetisation to different ^{13}C species across the backbone, gives us a third dimension, allowing backbone assignment to be performed.*

1.6 Project aims:

The aim of this study is to produce work that can contribute towards a collaborative paper between Dr Goult and Dr Tanentzapf, at the University of British Columbia, Vancouver. This paper will take a multidisciplinary approach to understanding a new interaction between Rap1 and Rhea/Talin. Dr Tanentzapf uses CRISPR to insert mutations into endogenous genes and then images and studies observed phenotypes in drosophila embryos. Alongside this, the Goult lab use biochemical and biophysical techniques to understand the biochemistry that underpins these protein interactions. A successful collaboration of this nature was performed previously in their paper on talin autoinhibition (Ellis et al., 2013), where studies were performed that produced talin-null mutant embryos, labelled for integrin and F-actin. They demonstrated that the talin-null mutant could be rescued with talin-GFP but that recue failed with the autoinhibition defective talin E1777A mutant protein. Dr Goult provided the biochemistry that compared the WT talin and E1777A mutants, and characterised the autoinhibition process.

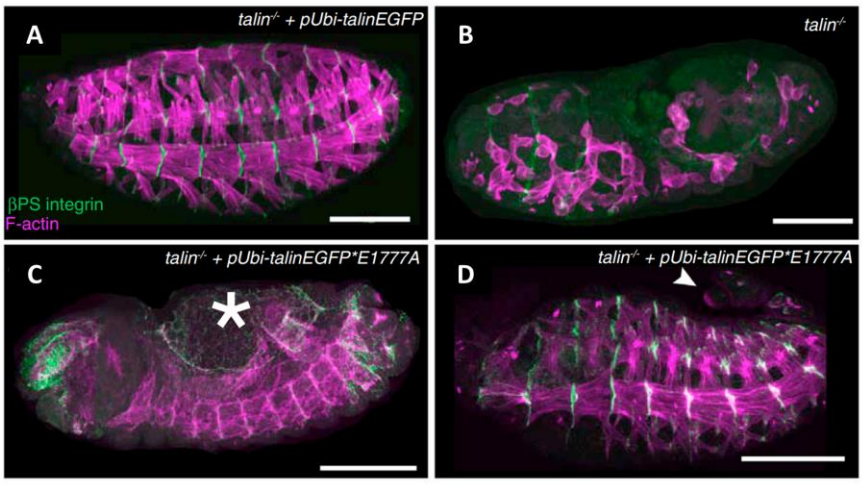


Figure 14: Talin Autoinhibition Mutants
In their previous collaboration, Dr Goult and Dr Tanentzapf showed that talin-null embryos (B) could be rescued with talin-GFP (A) but not with the autoinhibition defective talin E1777A protein (C-D). The asterix in C identifies an open dorsal hole. The arrow in D shows an un-retracted tail (Ellis et al., 2013).

The aim of this study is to produce a similar collaboration. Dr Tanentzapf has demonstrated that the Rhea F0-K17E mutation is embryonic lethal (unpublished). Dr Goult has identified a new interaction between talin F0 and Rap1 (Goult et al., 2010) whilst showing that the equivalent mouse talin F0-K15E mutation perturbs Rap1 binding (unpublished). We therefore want to demonstrate that these interactions are also occurring in drosophila, to link this interaction to the observed lethal phenotypes.

We have therefore set out to answer the following questions:

- Does the Rhea F0-WT protein bind Rap1b?
- Does this interaction still occur in the Rhea F0-K17E mutant, or is Rap1 binding perturbed?
- Can we identify additional key Rhea F0-WT residues involved in binding to Rap1?
- Can we identify the location of the Rap1 binding site on structural models of Rhea F0?
- Do our findings agree with the structural models of Rhea:Rap1 binding?
- What conclusions will the data allow to be drawn in relation to the current understanding of focal adhesion formation?

Section 2: Methods

Methods 2.1: Materials and Reagents

Table 2: Materials and Reagents

	Source	Composition
Acrylamide	BIO-RAD	40% Acrylamide/Bis Solution
Ampicillin	MELFORD Biolaboratories Ltd.	
IPTG	MELFORD Biolaboratories Ltd.	
DTT	MELFORD Biolaboratories Ltd.	
Instant Blue™ stain	expedeon	
Ni-NTA Buffer A		20 mM Tris, 500 mM NaCl, 20 mM Imidazole, /L, pH8
Ni-NTA Buffer B		20 mM Tris, 500 mM NaCl, 500 mM Imidazole, /L, pH8
Ni-NTA Buffer B (Batch)		20 mM Tris, 500 mM NaCl, 300 mM Imidazole, /L, pH8

Q buffer A		20 mM Tris, 50 mM NaCl, 2 mM DTT, /L, pH8
Q buffer B		20 mM Tris, 1000 mM NaCl, 2 mM DTT, /L, pH8
2M9 Solution A, for ^{15}N labelling (per 750ml)		9.375 g Na_2HPO_4, 5.625 g KH_2PO_4, made to 750 ml and autoclaved.
2M9 Solution B, for ^{15}N labelling (per 750ml)	$^{15}NH_4Cl$: Cambridge Isotope Laboratories, Inc. BME vitamins solution 100x: SIGMA Life Science	3.0 g Glucose, 7.5 ml Milli Q 'ultrapure water', 7.5 ml BME vitamins, 1.5 ml $MgSO_4$ (1 M), 0.075 ml $CaCl_2$ (1 M), 0.75 ml Ampicillin (1 mg/ml stock), 0.75 g $^{15}NH_4Cl$. Filter sterilised with 0.2 µl filter
2M9 Solution B, for $^{13}C^{15}N$ labelling (per 750ml)	D-Glucose-$^{13}C_6$ was sourced from SIGMA Life Science	Same as 2M9 Solution B, for ^{15}N labelling, with the substitution of D-Glucose-$^{13}C_6$
LB Broth: Low Salt Powder	MELFORD Biolaboratories Ltd.	Tryptone 10g/l Sodium Chloride 5g/l Yeast Extract 5g/l
NMR phosphate buffer (per L)		1.9 g NaH_2PO_4, 0.9 g Na_2HPO_4, 2.9 g NaCl, 0.3 g DTT, pH 6.5, made to 1000 ml*
HisPur™ Ni-NTA Superflow Agarose	Thermo Scientific	

PBS		8 g NaCl, 0.2 g KCl, 1.44 g Na_2HPO_4, 0.2 g KH_2PO_4 /L, pH 7.4
Glycerol Stock (50%)	Glycerol: Fisher Chemical	20 ml glycerol and 20 ml water, filter sterilised.
TEMED	BIO-RAD	
Ammonium Persulphate	Fisher Chemical	
Kanamycin sulfate	SIGMA Life Science	
$CaCl_2$ / glycerol		0.1 M $CaCl_2$, 10% glycerol: 150 ml stock contained 15 ml glycerol and 2.205 g $CaCl_2$ and 135 ml Milli-Q ultrapure water
Sample Buffer		
SnakeSkin Dialysis Tubing	Thermo Scientific	10K MWCO

All buffers were made with Type1 reagent-grade water from a Milli-Q Academic Ultrapure Water System

Methods 2.2: Universal Protocols

2.2.1 Transformation Protocol

Competent cells were in 100 µl aliquots. These were thawed on ice and divided into 50 µl aliquots. Immediately after thawing, 2 µl of DNA vector was added to each prior to gently mixing. The cells were returned to ice for a further 30 minutes and then heat shocked at 42°C for 45 seconds. Following this, cells were allowed to recover for a further 2 minutes on ice. The transformed cells were then added to the overnight cultures.

2.2.2 Overnight Cultures

Following the transformation, the cells were added directly to 10 ml of LB and antibiotic (Amp for Rhea constructs in BL21(DE3) cells / Amp + Kan for Rap1 in CK600K cells). The following morning, glycerol stocks were produced and scrapings of these were added to subsequent overnight preparations.

2.2.3 Glycerol Stocks

300 µl of filter sterilised glycerol (50%) was added to 750 µl of the overnight culture and stored at -80°C.

2.2.4 Competent Cells preparation.

A scraping of glycerol CK600K cell stock was cultured overnight in 5 ml of LB + Kan. The following day, 5 ml culture was added to a 500 ml flask of LB + Kan. This flask was incubated at 37 °C until the cells reached an OD_{595} of 0.6-0.8. The cells were then divided into 10x 50 ml fractions and left to cool on ice for 10 minutes. The fractions were centrifuged at 4000 rpm for 10 minutes at 4°C and the supernatant discarded. Each pellet was resuspended in 10 ml of ice cold $CaCl_2$ / glycerol and left on ice for 15 minutes, prior to a subsequent spin at 3000 rpm for 10 minutes for 4°C. The supernatant was carefully removed and discarded. Pellets were resuspended in 1 ml of ice cold $CaCl_2$ / glycerol and aliquoted into 100 µl fractions. Aliquots were flash frozen in liquid nitrogen before storing at -80°C.

Competent BL21(DE3) cells were prepared previously by the Toseland Lab.

2.2.5 LB, Ampicillin and Kanamycin Stock Concentrations

LB was prepared at 20 g per litre and autoclaved prior to use.

Ampicillin stocks were at a concentration of 50 mg/ml and ampicillin at 100 mg/ml. These were added to each prep to achieve a 1:1000 dilution.

2.2.6.1 Solubility Test Version 1

Prior to induction and harvesting, 1 ml of cells were taken from the culture, centrifuged for 2 minutes and the pellets were frozen at -20°C. Pre-induction pellets were resuspended in 500 µl of PBS and sonicated 2x on ice for 20 seconds each, with 20 seconds between sonications. Sonications were done on an MSE Soniprep 150 at 14 amplitude microns. At the different stages of the process, samples were taken for analysis by SDS-PAGE. 50 µl was taken and boiled with 50 µl of sample buffer for 5 minutes at 95°C. The post-induction pellets were prepared in the same way. The first fraction represented the whole-cell protein contents of the sample. The remaining 450µl, was centrifuged for 5 minutes, and 50 µl was taken to represent the soluble fraction of proteins. The supernatant was discarded and the insoluble pellet resuspended in 450 µl of PBS. A further 50 µl was taken, and both 50 µl fractions were boiled with 50 µl of sample buffer for 5 minutes at 95°C. When analysing with SDS-PAGE, loading volumes were adjusted according to the OD of pre-/post-induction fractions.

2.2.6.2 Solubility Test Version 2

This method was used alongside the Lysis and Centrifugation Protocol (see 2.2.8), to determine soluble protein levels present in the 30 ml resuspended cell pellets. After sonication, 30 µl of sample was taken, and boiled with 200 µl of sample buffer, to represent “whole cell” contents of the prep. Following centrifugation, 30 µl of sample was taken again, and boiled with 200 µl of sample buffer, to represent the “soluble” contents of the prep. Due to the difficult nature of resuspending an insoluble pellet of this size, the insoluble fraction was not taken for representation on a gel.

2.2.7 Harvesting Protocol

Cultures were pooled in 1 L centrifuge pots and centrifuged at 4000 rpm for 10 minutes. The pellets were resuspended, and aliquoted, in 30 ml of Ni-NTA-A buffer (if His-tagged) or Q-Buffer A (if untagged) per half litre of prep. This was done to allow effective sonication, by ensuring the cell density was not too high in each resuspended pellet. Resuspended pellets were stored at -20°C.

2.2.8 Cell Lysis and Centrifugation Protocol

30ml resuspended cell pellets were defrosted in a beaker of water at room temperature. This freeze-thaw process was used as part of the cell lysis step (Johnson and Hecht, 1994). Cells were sonicated on ice with 6 cycles of 20 seconds on, 40 seconds off. Each sonicated pellet was then centrifuged for 20 minutes at 18,000 rpm. The soluble fraction was retained, with the insoluble pellet discarded.

2.2.9 SDS-PAGE

Samples were boiled at 95°C for 5 minutes prior to SDS-PAGE analysis. Typically, 50 µl of sample was boiled with 50 µl of sample buffer (with the exception of the section 2.2.6.2 solubility test, where 30 µl of sample was boiled with 200 µl of sample buffer).

Denaturing SDS-PAGE gels were run at 200V for approximately 40 minutes. Gel lanes were loaded with 10 µl of sample, however this volume was adjusted, where appropriate, to account for different concentrations; such as in pre- and post-induction expression tests. Gels were made with a 12% separating and 4% stacking component, in pre-made gel cassettes, according to Novex™ gel casting instructions as shown below:

Table 3: The composition of the gels used for SDS-PAGE

Separating Gel 12%		**Stacking Gel 4%**	
40% acrylamide	7.5 ml	40% acrylamide	1.25 ml
Separating Gel Buffer	9.4 ml	Stacking gel buffer	4.2 ml
10% SDS	250 µl	10% SDS	125 µl
50% sucrose	4.0 ml	Water	6.05 ml
water	3.3 ml	TEMED	5.0 µl
TEMED	6.25 µl	Ammonium Persulphate	1.0 ml
Ammonium Persulphate	625 µl		

A*mmonium persulphate* was at a concentration of 50 mg/ml.

Separating gel buffer (1 M Tris-HCl, pH8.8): 30 g of Tris in 250 ml of water, adjusted to correct pH with HCl

Stacking gel buffer (0.375 M Tris HCl, pH6.8): 11.4 g of Tris in 250 ml of water, adjusted to correct pH with HCl

2.2.10 Ni-NTA Purification Protocol

In order to purify the target proteins, affinity chromatography was principally used; a technique that takes advantage of protein binding interactions. By tagging proteins with a His-tag, they have a high affinity for Ni-NTA resin, meaning they can be isolated from cell lysates (Alberts et al 2002). In this type of purification, firstly, the soluble cell lysates are passed through a Ni-NTA affinity column. The His-tagged proteins will bind to the column whilst the majority of un-tagged proteins will pass straight through. A low imidazole (20 mM) Ni-NTA buffer A is washed through the column to remove as much non-specific binding as possible.

Then a gradient of increasing imidazole concentration is run (up to 500 mM in these experiments) and the eluted fractions are collected. At a specific concentration of imidazole, the His-tagged proteins are eluted from the column and in the process are purified and concentrated. Non-specifically bound impurities are likely separated, from the target protein, by eluting at different concentrations of imidazole. FPLC was carried out at room temperature. Fractions were collected from across the A280 peak and run on a gel for identification, with the purest fractions pooled.

2.2.11.1 Buffer Exchange Protocol 1: Desalting Column

Protein was loaded onto a desalting column and exchanged into Q-Buffer A using an Akta Purifier FPLC, at 4°C.

2.2.11.2 Buffer Exchange Protocol 2: pd10

The pd10 column was washed with 25 ml of phosphate buffer, then 2.5 ml of protein was added to the column. The protein was then eluted with 3.5 ml phosphate buffer. This was repeated as required dependent on the volume of protein. Any remaining protein was made up to 2.5 ml with phosphate buffer and then run through the column as before.

2.2.11.3 Buffer Exchange Protocol 3: Dialysis

To exchange from Q-Buffer A to phosphate (NMR) buffer, proteins were dialysed against 3 L of phosphate buffer at 4°C overnight.

2.2.12 Mono-Q Chromatography Purification Protocol

In order to get the proteins above 90% purity, and suitable for biochemical assays, a subsequent purification step via ion-exchange chromatography was used after

Ni-NTA chromatography. Proteins were exchanged into Q-Buffer A and run through an anion exchange Mono-Q column. The column is positively charged and therefore, negatively charged proteins stick to the column and can be eluted with a salt gradient. When proteins are in a buffer with a pH greater than their pI, they have a net negative charge, as basic groups of the protein tend to be uncharged, while negative charges arise from carboxylic acid groups (Albert et al 2002). This process was carried out at room temperature.

2.2.13 Protocol to remove uncleaved protein and cleaved His-tags

100 µl of Ni-NTA Resin was cleaned, in Q-Buffer A (same buffer as the proteins were in), and then added to the protein. The sample was agitated gently for 15 minutes to allow binding to the resin to occur. This removed any uncleaved protein, cleaved His-tags and TEV from the solution. To remove the resin, and any precipitate, the sample was then filter sterilised with a sterile 0.2-micron filter. SDS-PAGE analysis was then used to determine to protein purity.

Methods 2.3.1: Rhea Expression and Purification

2.3.1.1 Rhea Expression Trial

For the initial expression trial, overnights of BL21(DE3) cells transformed with pet151-TOPO DNA vectors containing either F0-WT or F0-K17E were put on. These overnights were cultured in autoclaved, bevelled flasks containing 750 ml of LB + Amp. The volumes of each overnight culture added were adjusted according to their respective OD_{595}. This ensured that the cultures would be at the same starting density at time 0. The cultures were incubated and shaken at 37°C for around 4 hours, until the OD_{595} reached 0.5. The cells were induced for 3 hours with 0.25 mM IPTG. The cells were harvested prior to a solubility test being performed (see 2.2.6.1) and then discarded.

2.3.1.2 Rhea Expression Optimisation

Further cultures were grown of F0-WT and F0-K17E, as before, but upon reaching the correct OD, the flasks were then taken out of the incubator, and placed into room temperature water to cool. Once cool the flasks were induced with 200 μM IPTG and incubated with shaken overnight at 18°C. Cells were harvested (see 2.2.7) and stored at -20°C.

2.3.1.3 Rhea Purification

Cells pellets were lysed and centrifuged (see 2.2.8), and the soluble fractions were purified using Ni-NTA FPLC (see 2.2.10). The main A280 peaks for each protein corresponded to fractions A11-A15, with a second peak arising between B15-B14. These fractions were analysed with SDS-PAGE prior to pooling the A12-A15 fractions for F0-WT and the A11-A15 fractions for Rhea F0-WT and F0-K17E. The protein was desalted into Q-Buffer A (see 2.2.11.1) and cleaved overnight with TEV (a cysteine protease that cleaves between the Q\S of the ENLYFQ\S sequence that links the His-tag to the protein) at 4°C. The following day, the prep was cloudy, indicating a large amount of precipitation. A fraction was taken and then the sample was spun for two minutes, and transferred to a clean falcon tube, to separate it from the precipitate. The sample was purified via ion exchange chromatography (section 2.2.12). The F0-WT and F0-K17E proteins were present in the flowthrough; the proteins bound very weakly to the column, but were washed out with the low salt buffer. To check that the TEV cleavage had been successful in removing the His-tag, uncleaved and cleaved fractions were analysed by SDS-PAGE. The flowthroughs were aliquoted into 1 ml fractions, flash-frozen with liquid nitrogen and stored at -20°C.

2.3.1.4 Rhea ^{15}N F0-WT Expression and Purification

Isotopic labelling was performed by culturing in 2M9 minimal media. 2x 750 ml flasks of 2M9 solution A were prepared and autoclaved in bevelled flasks. The 2M9 solution B is not able to be autoclaved and so was prepared and filter

sterilised into 2x falcon tubes. Prior to growth the 2M9 solution B was added to the 2M9 solution A to produce 2M9 minimal media.

In order for the cells to become acclimatised to minimal media, overnight cultures were grown in 2M9. A scraping of the glycerol stocks of F0-WT and F0-K17E was cultured overnight with 10 ml 2M9 solution A and 270 µl of 2M9 solution B. The following morning, after successful growth, fresh glycerol stocks were made of these cells, for use with future preps in minimal media. Adjusted volumes of overnight culture were added to each flask, which were then grown up and purified as described in section 2.3.1.3. A notable difference is that additional TEV was added to ensure complete cleavage of the proteins.

2.3.1.5 Rhea ^{13}C ^{15}N F0-K17E Expression and Purification

Double labelled Rhea F0-WT was produced using the same conditions as described in 2.3.1.4, aside from the use of isotopically labelled, D-Glucose-$^{13}C_6$. Additional TEV was required to obtain complete cleavage prior to Q-Column Purification, and this was achieved by adding more TEV at room temperature for 40 minutes. Subsequently, cleaved His-tags and uncleaved protein were removed (see 2.2.13). The protein was put in a 3,000 Da concentrator and spun for 2x 20 minutes at 4000 rpm and concentrated down to 1 ml, where the concentration was measured at 5 mg/ml.

The 1 ml of concentrated protein was split; 500 µl was used immediately for NMR analysis, and the other 500 µl was flash frozen and stored at -20°C.

Methods 2.3.2: Rhea CD Analysis

2.3.2.1 Full Spectrum Measurement

Rhea F0-WT was at 0.500 mg/ml and F0-K17E at 0.496 mg/ml for CD analysis. Full spectrum measurements were done at 20°C, with standard sensitivity, data pitch of 0.5 nm, 200 nm/min scanning speed, 0.5 second response, 2.0 nm band

width, accumulation of 8 and ran from 260-190 nm. All measurements were performed on a Jasco J715 and converted to Mean Residue Ellipticity as described previously (Kelly, Jess and Price, 2005).

2.3.2.2 Variable Temperature Experiment

The Variable temperature measurement was done with the same parameters as section 2.3.2.1, at a wavelength of 208 nm, starting at 20°C and ending at 90°C, with a data pitch of 2°C, delay time of 0 seconds and temperature slope of 2°C/minute.

Methods 2.3.3: Rhea NMR Analysis

2.3.3. 1D and HSQC of Rhea

All NMR experiments were performed on a Bruker 600 MHz spectrometer equipped with CryoProbe. To prepare samples for NMR analysis, 25 µl of D_20 was added to 425 µl of ^{15}N labelled sample to yield a 450 µl sample suitable for NMR. Rhea F0-WT was at 0.90 mg/ml and F0-K17E at 1.10 mg/ml. Experiments were done at 298 K, at pH 6.5.

Methods 2.3.4: Rhea Backbone Assignment

2.3.4.1 Triple Resonance Protocol

$^{13}C^{15}N$ Rhea F0-WT was at 5.0 mg/ml. 25 µl of D_20 was added to 425 µl of double labelled sample to yield a 450 µl sample suitable for NMR. HNCO, HNCA, HN(CO)CACB, HNCACB and HN(CO)CA triple resonance experiments were performed. Experiments were done at 298.0 K, at pH 6.5. The experiments were processed using TOPSPIN.

2.3.4.2 Backbone Assignment Overview

The information from this section comes from a series of lecture notes, for a graduate NMR course, entitled “Triple resonance experiments and structure determination”, from Mark Girvin and Sean Cahill at the Albert Einstein College of Medicine, NY, 2012 who cite the following authors: Cavanagh, J. (2007). *Protein NMR spectroscopy*. Amsterdam: Academic Press. Evans, J. (1995). *Biomolecular NMR spectroscopy*. Oxford: Oxford University Press. Additional information comes from Goult et al., 2015

As explained in the introduction, 3D NMR involves coupling spins, so that magnetization transfer occurs between different atomic species within the protein backbone. For the assignment in this study, 5 experiments were carried out to allow peaks to be observed for different carbons at i and i-1.

Note that the carbonyl is typically referred to as “CO,” the alpha carbon as “CA”, and the beta carbon as “CB” which is the first carbon of the side chain (present on all amino acids aside from glycine).

The first experiment was the HNCO (Figure 15A). In this experiment, the magnetization was transferred from the amide proton to the nitrogen. From there, it was transferred to the carbon of the carbonyl of the preceding residue, i-1. From CO, the magnetization was passed back to the nitrogen, and then hydrogen, for acquisition. Having performed a HNCO we observe a single peak for the CO at i-1 for each amino acid.

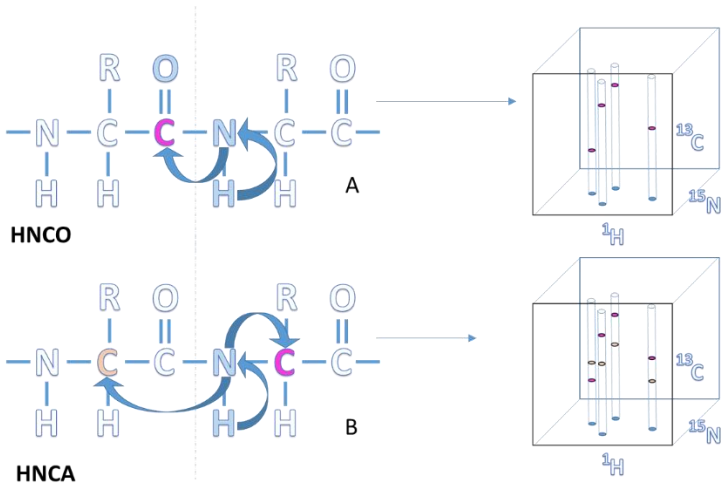


Figure 15: Transfer of Magnetisation in HNCO and HNCA

The next experiment was the HNCA (Figure 15B). The magnetization was transferred from the proton to the nitrogen again, but this time it principally transferred to the CA at i. We therefore see a peak for this CA. We also observe a peak for the CA-1, as the magnetization skips across the i-1 CO. Usually this peak is smaller than for the CA.

The HN(CO)CACB was one of the most important experiments, in that it transferred the magnetization from the proton, to the nitrogen, skipping across the CO, to the CA-1 and CB-1. We therefore observe peaks for the CA-1 and CB-1 (Figure 16C).

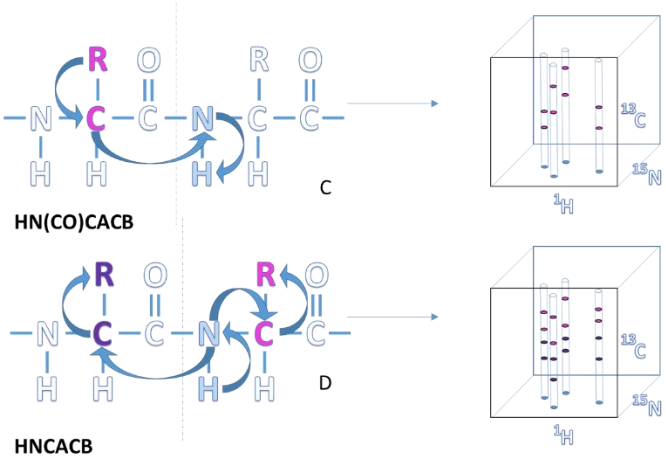


Figure 16: Transfer of magnetisation in HN(CO)CACB and HNCACB

The next triple resonance experiment performed was the HNCACB. After transferring from the hydrogen to the nitrogen (Figure 16D), the magnetization passed to the CA and CB at i. We therefore observe peaks for these. We also see peaks, for the CA-1 and CB-1, as the magnetization skips over the CO and passes to these two carbons. This experiment therefore results in four observable chemical shift peaks CA, CB, CA-1 and CB-1.

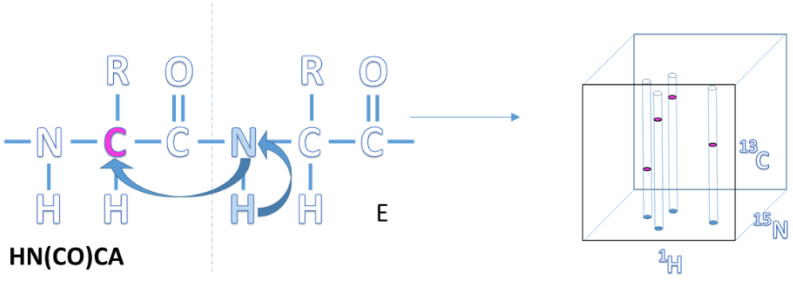


Figure 17: Transfer of magnetisation in HN(CO)CA

The final experiment used for the assignment was the HN(CO)CA, where the magnetization passed to the CA -1 (Figure 17E), which we observe as a single peak.

2.3.4.3 Using this Information for Backbone Assignment

The HNCACB tells us the chemical shifts for the CA, CB, CA-1 and CB-1. For each residue, their CA and CB values lie at defined chemical shift values (see Figure 19) This means that from the HNCACB, we can predict which amino acid we are looking at from the CA and CB at i, and which it is next to, from the CA-1 and CB-1. We simply need to work out which of the 4 peaks are which, and to do this, we use the other triple resonance experiments, which can be overlaid using the CCPNMR Analysis software. In order to determine which 2 peaks belong to i, and which belong to i-1, the HN(CO)CACB can be overlaid with the HNCACB, which will mark the CA-1 and CB-1. The HNCA can then be overlaid to determine which 2 of the 4 peaks are the alpha carbons. This means that we can differentiate the 4 peaks using just those 3 experiments.

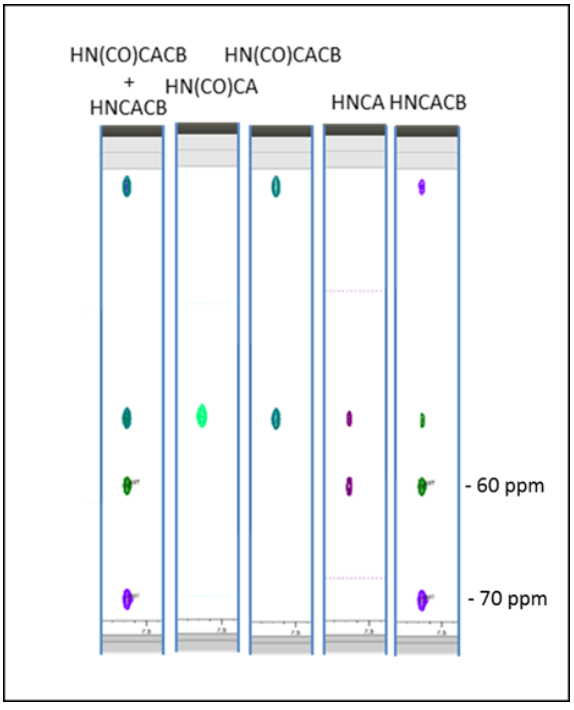


Figure 18: CCPNMR strips identifying CA, CB, CA-1 and CB-1

As seen in Figure 18, looking from right to left, we first see the four peaks in the HNCACB. Toggling on the HNCA allows us to determine that the two middle peaks

of the HNCACB are alphas. The HN(CO)CACB tells us that the top peak of the HNCACB must be the CB-1, whereas the next peak is the CA-1. This reveals the third peak down to be the CA own (at i), leaving the bottom peak to be the CB own. The HN(CO)CA can be used as a toggle to quickly determine which peak is the CA-1 when looking at the HN(CO)CACB.

When we look at the chemical shift values of each peak, the CA at i is at 60 ppm, and the CB at i is at 70 ppm. This tells us that we are most likely observing a threonine (Figure 19) which is a diagnostic residue for backbone assignment. This is due to the fact that there are only two residues with a beta chemical shift higher than an alpha chemical shift. The chemical shifts of the CA-1 and CB-1 tell us that we could be observing an arginine.

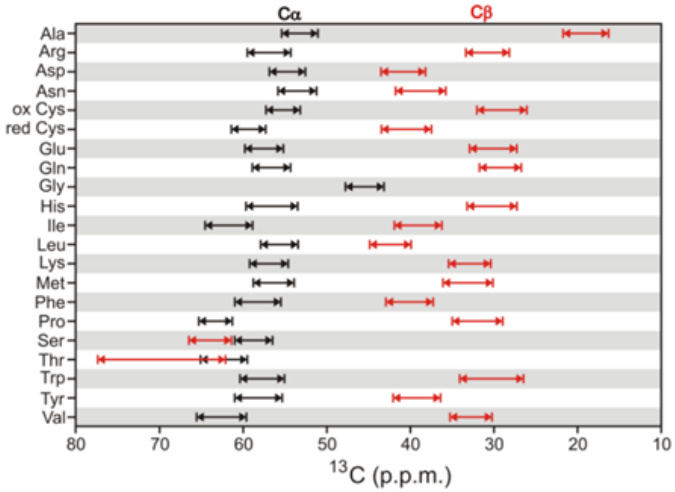


Figure 19: The range of CA and CB chemical shifts allow for residue identification (Girvin and Cahill, 2012)

We can't be sure of this initially however. What we need to do is find the spin system that has a corresponding, identical chemical shift for its own alpha and beta, as that of the alpha and beta minus one of the threonine. We see this corresponding spin system in Figure 20, in the second row from the left (the threonine at i+1, to the right). For the i-1 amino acid, it only has one peak for its own i-1 (i-2 for the threonine), which can be confirmed to be a CA-1 using the HN(CO)CA. There is only one amino acid without a beta carbon, meaning that the downstream amino acid must be a glycine. If we then look at the entire sequence of amino acids for the Rhea F0, we can try to identify G-R-T. In this example, it

turns out that this combination is unique, only occurring once in the sequence. We can therefore assign these residues as 67G, 68R and 69T.

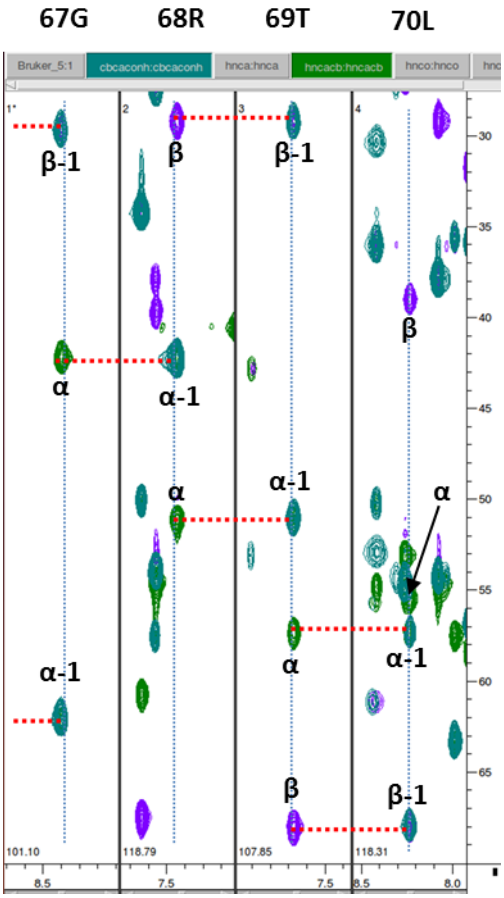


Figure 20: An example strip showing sequential, aligned residues in CCPNMR

This can be confirmed when looking for the i-3 residue of the 69T. In the sequence, this should be a proline. Proline, does not have an amide and therefore does not have its own spin system that can be observed at i, however they can always be seen at the i-1 of its upstream amino acid. Looking at the chemical shift of the i-3, we see values that are diagnostic of a proline, and these CA-1 and CB-1 have no matching beta and alpha own, when looking at all the other spin systems.

The way that this is carried out in CCPNMR, is by propagating the peaks of the CA-1 and CB-1 to the HN(CO)CACB strips, and the CA and CB owns to the HNCACB strips. Under the assignment options of the program, you can then set the HN(CO)CACB as a query, and the HNCACB as a match. This means that when you click on the strips for a given peak, it will automatically highlight the CA-1 and CB-1, and search through every other strip, to look for amino acids that have similar chemical shifts for their CA and CB owns. In this way you can “walk” your

way down the protein backbone, matching up neighbouring amino acids, until you find diagnostic regions that tell you where in the sequence you are observing. When selecting the peaks from the HSQC, it can be noted that there are peaks which are not from the protein backbone. These can easily be distinguished as side chain amino acids, as they have no proper spin systems, when navigated in HNCACB.

Having completed the assignment, the chemical shifts were used to calculate the dihedral angles based on chemical shift indexing with the program DANGLE which forms part of the CCPNMR package. This created a Ramachandran Plot output and determined the secondary structure elements of the F0 domain (Figure 36 left).

Methods 2.4: Rap Expression and Purification

2.4.1 Rap1 Expression Trial and purification in BL21(DE3)

Rap1-His and -untagged constructs were transformed into E.coli BL21(DE3) cells as described previously. Note that “Rap1” when mentioned in the context of the Drosophila form, is referring to “Ras-like protein 3”, but will be named as “Rap1” from here onwards. The constructs were already prepared previously by the Goult Lab. Overnight cultures were grown in LB + Amp, and subsequently transferred to 750 ml flasks of LB + Amp as described in section 2.3.1.1. The flasks were induced overnight at 20°C, with 0.25 mM IPTG and harvested as described previously (section 2.2.7). A solubility test was performed (sections 2.2.6.1 and 2.2.6.2) and the pellets lysed and centrifuged (section 2.2.8). Purification was then attempted with Ni-NTA chromatography (see 2.2.10), with the observed peak being analysed with SDS-PAGE.

2.4.2 Rap1 Expression Trial and purification in CK600K

CK600K cells were kindly gifted by Dr Igor Barsukov (University of Liverpool). The Rap1 constructs were transformed into these cells after being made competent

(section 2.2.5) and glycerol stocks were made, as described in 2.2.3. Further cultures were grown of Rap1, as before, with the addition of Kan, due to the CK600K cell line having Kanamycin resistance. The cultures were induced with 0.20 mM IPTG, overnight at 18°C and harvested as described in section 2.2.7. Cells were lysed and centrifuged, before being subjected to a solubility test (see 2.2.6.2). The soluble fraction was purified with Ni-NTA chromatography as before. Two potential Rap1 bands were identified via SDS-PAGE, although both were in very low yields. The higher molecular weight band was deemed unlikely to be Rap1 due to its size. The lower molecular weight band aligned with the His-Rap1 marker and was an appropriate size. Both were sent to be analysed by mass spectrometry.

Additionally, TEV was added to the fractions eluted from the Ni Column, containing the Rap1 targets, to see if an incidence of cleavage could be observed, which would indicate the presence of the His-tagged Rap1. These were left at room temperature overnight, mixing with an Eppendorf flea.

2.4.3 Mouse Rap1 Expression Trial in CK600K

The expression trial from section 2.4.2 was repeated, but this time, using a mouse Rap1b construct, pre-transformed into CK600K cells (also gifted by Dr Igor Barsukov). The cells arrived as a lawn of plated cells, so a scraping was used to inoculate 10 ml of LB + ampicillin + kanamycin. These were left overnight, incubated and shaking at 37°C. 4 flasks of LB+Amp+Kan were innoculated with 2 ml of Rap1 overnight culture. These flasks were incubated and shaken at 37°C for around 3 hours, until the OD_{595} reached 0.8. before induction with 0.5 mM IPTG overnight at 30°C. Cells were harvested as described in section 2.2.7. Expression was analysed with a solubility test (see 2.2.6.1).

2.4.4.1 Mouse Rap1 in CK600K purification

Cell pellets were lysed and centrifuged as described previously. 1 mM PMSF was added after lysis. The construct was untagged and therefore needed to be

purified via ion exchange chromatography. The construct comprised of residues 1-166 and had a theoretical pI of 4.69 according to UniProt. Therefore, the protein was run on a Mono-Q column using Q-Buffer A at pH 8, to ensure the protein was negatively charged and would bind to the column.

As the column volume was only 5 ml, and the entire soluble fraction was being loaded onto the column, it was theorised that the column may get saturated with contaminants, preventing the Rap1b from binding on the first Q-run. This phenomenon had been previously observed within the Goult Lab, during previous untagged purification studies. In such cases, the protein would be present in the flow-through, requiring multiple cycles of Q column until enough contaminants had been removed to allow target binding. With this in mind, the cytosol was loaded onto the column and the flow through was then run onto the column a further four times, and the eluted fractions were taken for each.

The target was present in the A10-11 fractions of the first run, so these were analysed with SDS-PAGE and sent for MALDI Mass Spectrometry analysis. The untagged protein needed to be isolated using size exclusion chromatography so the A10-11 fractions were pooled and left overnight at 4°C, whilst the preparative Superdex S200 gel filtration column equilibrated. The gel filtration column had a 5 ml loop, so the sample was concentrated to 5 ml, and loaded onto the column. The protein was eluted through the column with the same buffer that the protein was already in: Q-Buffer A.

Fractions from each peak were run across 2 gels, which located Rap1 in the tallest peak of the evaluation. The peak showed the protein was purest in fractions E6 to E11, so these were pooled and purified with a final Q-column run.

2.4.4.2 Mouse Rap1 in CK600K purification for NMR analysis

Having optimised the purification protocol for producing pure Rap1, the prep was scaled up to 3 litres, so as to obtain enough protein for the NMR experiments.

Note that 2 mM $MgCl_2$ was added to the buffers used for the Q-column, (Q-buffer-A and B) in this prep. The first half litre prep was run on the Q column, and the fractions analysed with SDS-PAGE.

After confirming the proteins presence, in the A10-11 fractions (as in section 2.4.4.1), it was assumed with reasonable certainty that the results would be the same for the remaining five half-litre preps. Each half litre flask of cells was run separately on the Q column for ease of isolating the Rap; by loading a higher volume onto the column, it risked the Rap failing to bind to the column.

The 5 remaining Q-columns were run, and in each case, the A10 and A11 fractions were collected and pooled with those from the first Q-column run. This resulted in around 50 ml of pooled elution. This was concentrated to 5 ml in a 10 kDa concentrator and run on the gel filtration column as before. The peak containing the proteins fractions were pooled to approximately 8.5 ml and 1 mM PMSF was added. The protein was exchanged into phosphate buffer using a pd10 column (see 2.2.11.2) prior to undergoing NMR analysis with Rhea.

Methods 2.5: Rhea/Rap1 NMR Experiments

2.5 Rhea F0 alone and Rhea F0-WT and F0-K17E with Rap1

The Rhea F0-WT and F0-K17E was at a concentration of 39.42 µM in a Shigemi NMR tube. The Rap1 was added at a five-fold excess. However, densitrometry of the SDS-PAGE bands using ImageJ revealed that it was a four-fold excess, ~160 µM Rap1. NMR experiments were performed at pH 6.5 and 298K, with a total volume of 350 µl, using the plunger in the Shigemi NMR tube.

Section 3: Results

Results 3.1: Rhea Expression and Purification

3.1.1. Rhea Expression Trial

Initially, for the first expression trial, both the Rhea F0-WT and F0-K17E were predominantly insoluble (Figure 21). In spite of this, the cultures grew well, with a large overexpression of protein being observed.

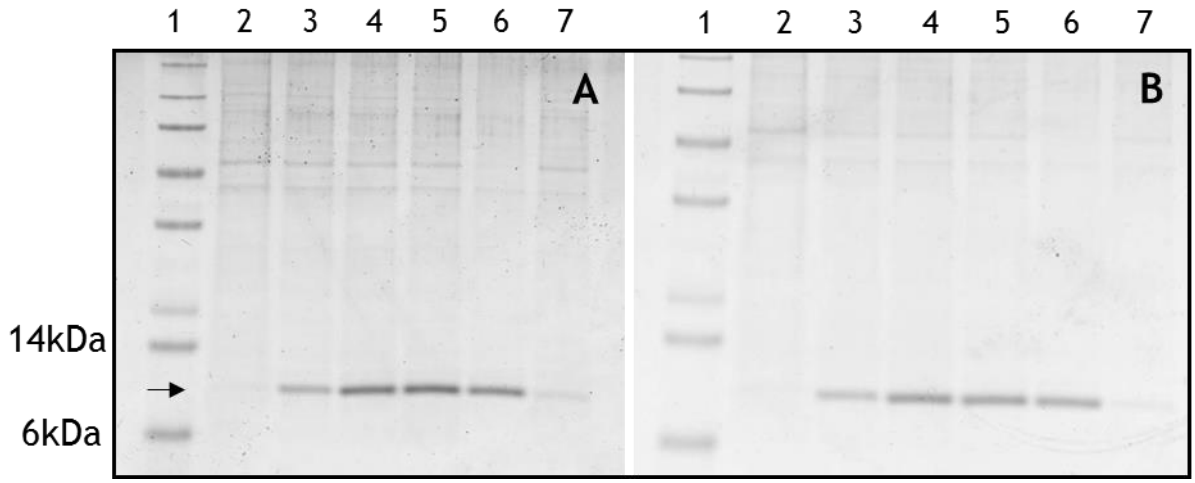


Figure 21: Trial 1 solubility test for Rhea F0-WT and K17E.
*solubility tests for Rhea F0-WT (**A**) and Rhea F0-K17E (**B**). A1/B1 are marker lanes. A2/B2 represent pre-induction fractions, A3/B3 show post induction +1hr, A4/B4 post induction +2hrs, A5/B5 post induction +3hrs/whole cell contents of harvested cells, A6/B6 insoluble fraction, A7/B7 soluble fraction. Arrow marks proteins of interest.*

Expression levels remained relatively constant from two hours onwards (Lanes 2-4), with expression appearing to double between the first and second hours. Despite good expression, the soluble fraction of each protein was largely void of protein, being almost entirely present in the insoluble pellet. It can be estimated from the gel that we are looking at upwards of 90% insolubility. It is

likely that at these induction conditions, the protein is being made too quickly, forming aggregates and ending up as inclusion bodies.

3.1.2 Rhea Expression Optimisation

In the second trial, the expression conditions were adjusted to slow the rate of expression, this was achieved by lowering the temperature to 18°C and the IPTG concentration to 200 µM. The optimisations resulted in considerably improved yields, with the protein being completely soluble (Figure 22).

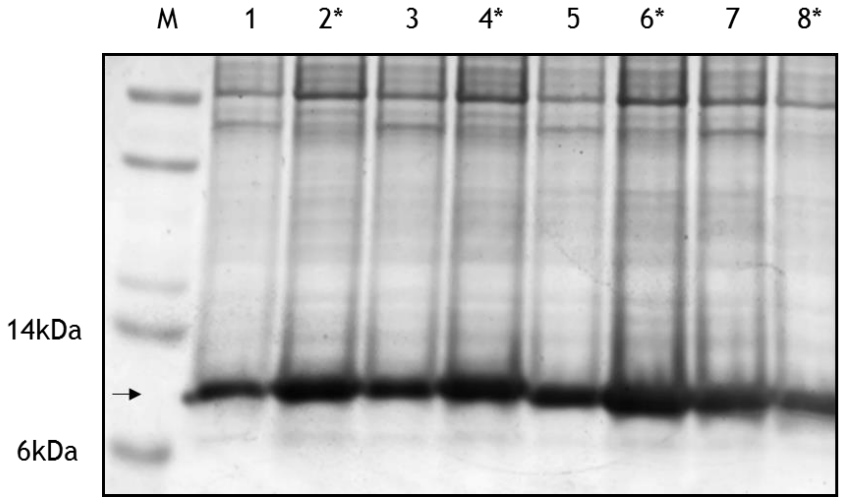


Figure 22: Trial 2 solubility test for Rhea F0-WT and K17E.
M is the marker lane. Lanes 1 and 3 show the whole cell fractions for the two sonicated pellets of F0-WT, with 2 and 4* showing their soluble fractions respectively. Lanes 5 and 7 show the whole cell fractions for the two sonicated pellets of F0-K17E, with 6* and 8* showing their soluble fractions respectively.*

3.1.3 Rhea Purification

Having determined the conditions to express soluble Rhea F0-WT and F0-K17E, they were successfully purified as evidenced by a large A280 peak observed for both proteins (Figure 23A+B). Some of the protein was lost through purification due to a leak on the Akta. Fortunately, the protein was abundant, so this lost protein was not critical.

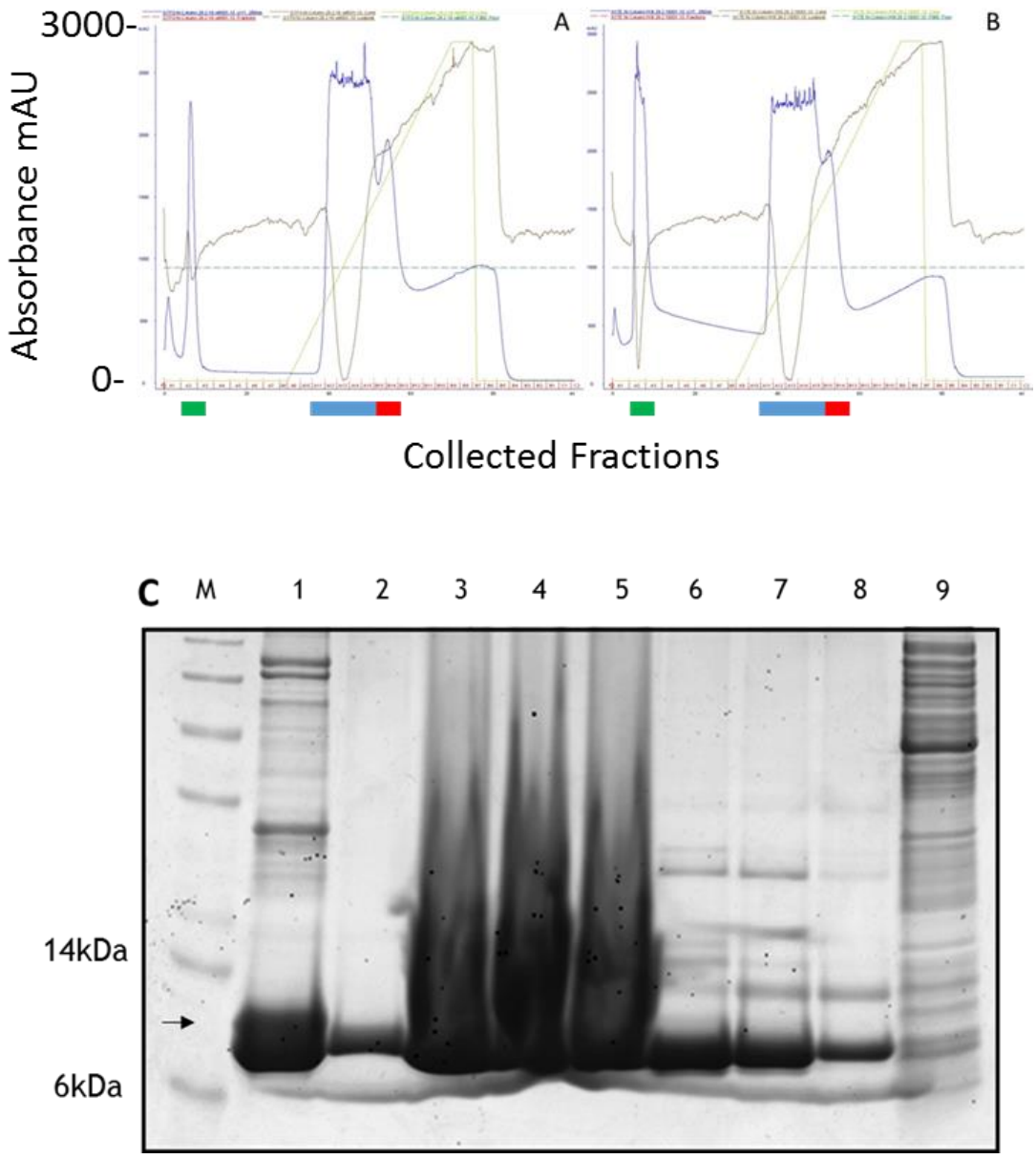


Figure 23: Ni-NTA purification of Rhea F0

A-B *the nickel column evaluations of Rhea F0-WT (**A**) and Rhea F0-K17E (**B**) with the main peak being displayed between fractions A11-A15 (blue bar). A second peak was observed between B15-B14 (red bar), but SDS-PAGE showed it be largely non-specific contaminants bound to the column. The peak at A2 (green bar) was due to a leak of high imidazole buffer, a technical issue of the Akta system used.* ***C*** *the fractions eluted from the Ni-column for F0-K17E (F0-WT data not shown). M displays the marker. 1. shows the A2 fraction, 2. A11, 3. A12, 4. A13, 5. A14, 6. A15. 7. B15, 8. B14, 9. The flow-through.*

The proteins were cleaved to remove their His-tags; the results of which can be seen in Figure 24.

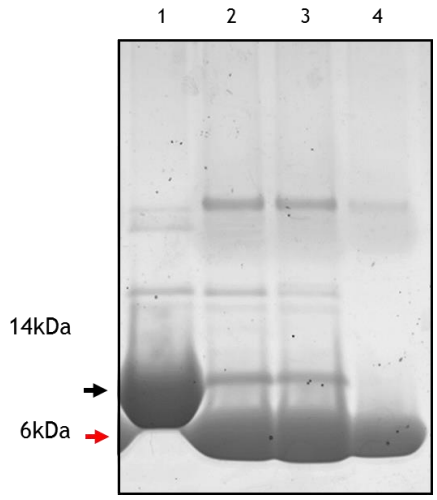


Figure 24: TEV cleavage of Rhea F0

1. F0-WT desalted, before TEV addition, 2. F0-WT + TEV before spin, 3. F0-WT + TEV after spin, 4. F0-WT flow-through of Q-column, Arrow (black) shows protein prior to cleavage, arrow (red) shows protein after cleavage.

The subsequent round of Q-purification resulted in highly pure Rhea F0-WT, having separated it from contaminants (Figure 25).

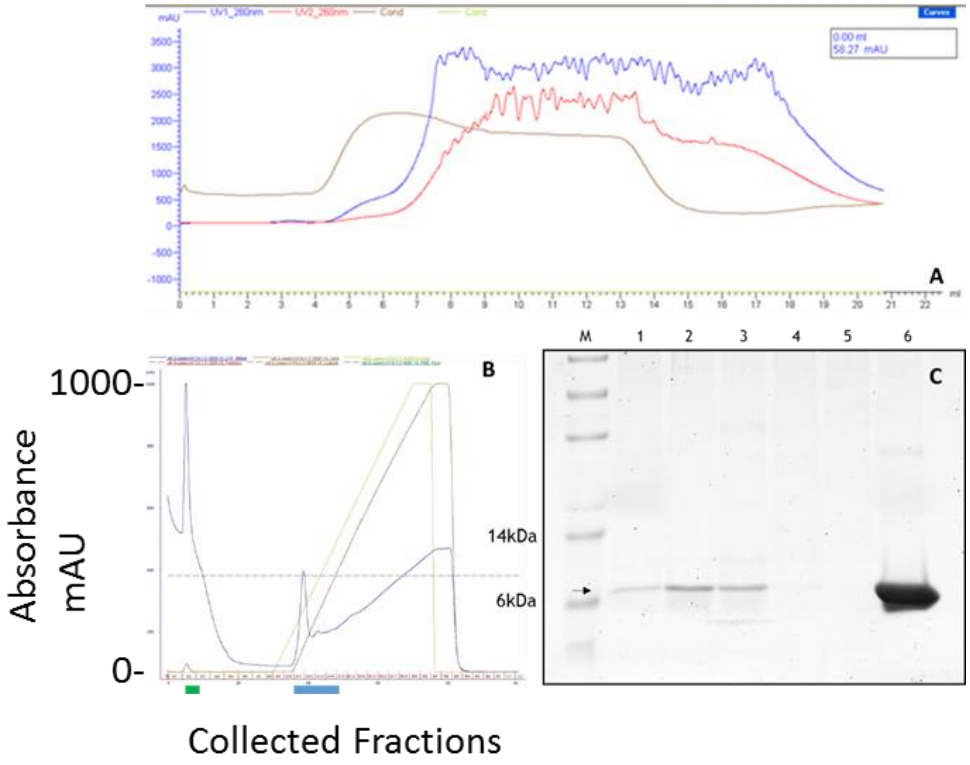


Figure 25: Ion Exchange Purification of Rhea F0

A) A large volume of protein did not bind to the Q-column and was detected in the flowthrough. The large peak in A280 absorbance (blue trace) is predominantly F0-WT protein. B) shows the evaluation from the Akta for F0-K17E (F0-WT data not shown). A2 is marked with the green bar and A11-14 with the blue bar. As expected from the loading, only a very small peak was observed with most F0 present in the flowthrough. B) shows the corresponding fractions from B, where M. Marker, 1. A2, 2. A11, 3. A12, 4. A13, 5. A14, 6. The flow-through as shown in A.

The presence of small bands eluted from the Q-column (Figure 25) suggest that this purification process was effective as a "reverse Q"; the contaminants bound to the column, therefore being separated from the protein.

3.1.4 Rhea ^{15}N F0-WT Expression and Purification

^{15}N labelling is relatively costly, so we wanted to be sure that Rhea could be successfully expressed and purified before undergoing this process. Having achieved this, the Rhea F0-WT and F0-K17E were labelled with ^{15}N. This was achieved using 2M9 minimal media where the sole nitrogen source was ^{15}N-labelled ammonium sulphate. The solubility tests showed that the optimised conditions from section 2.3.1.2 were highly repeatable in this minimal media prep, with comparably high expression and solubility of the proteins (Figure 26).

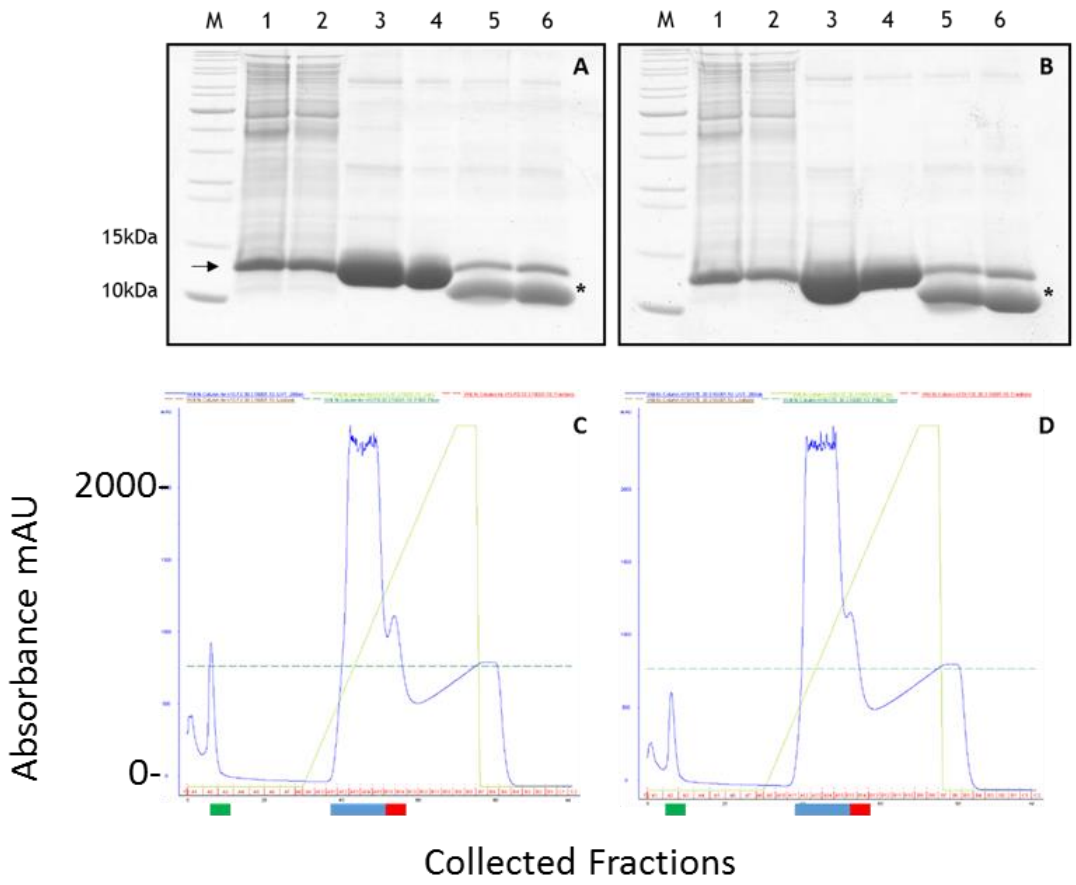


Figure 26: Expression and Purification of Labelled Rhea F0

A) *^{15}N F0-WT /* ***B)*** *^{15}N F0-K17E: M. marker, 1. Whole cell post-induction fraction after sonication, 2. Soluble protein fraction after centrifugation, 3. Pooled A11-A15 of Ni Column, 4. Desalted protein, 5. Protein after TEV cleavage, 6. Protein after TEV cleavage after centrifugation to remove precipitation. Arrow indicates protein location prior to cleavage. * marks protein after cleavage* ***C)*** *Ni Evaluation of ^{15}NF0-WT.* ***D)*** *Ni Evaluation of ^{15}NF0-K17E. A2 fractions on* ***C) + D)*** *are marked (green bar), as are the A11-15 fractions (blue bar) and the contaminant peak (red bar).*

When purifying the proteins on the Ni-NTA and Q- columns, they behaved identically to when unlabelled. The Reverse-Q process was more dramatic in this case, with the contaminants, that were removed by the column, being more visible on the gel (Figure 27A+B). The gel also reveals that the additional TEV was successful to fully cleave the protein, with a gel being run to clearly demonstrate the size difference between the His-tagged and fully cleaved protein (Figure 27C).

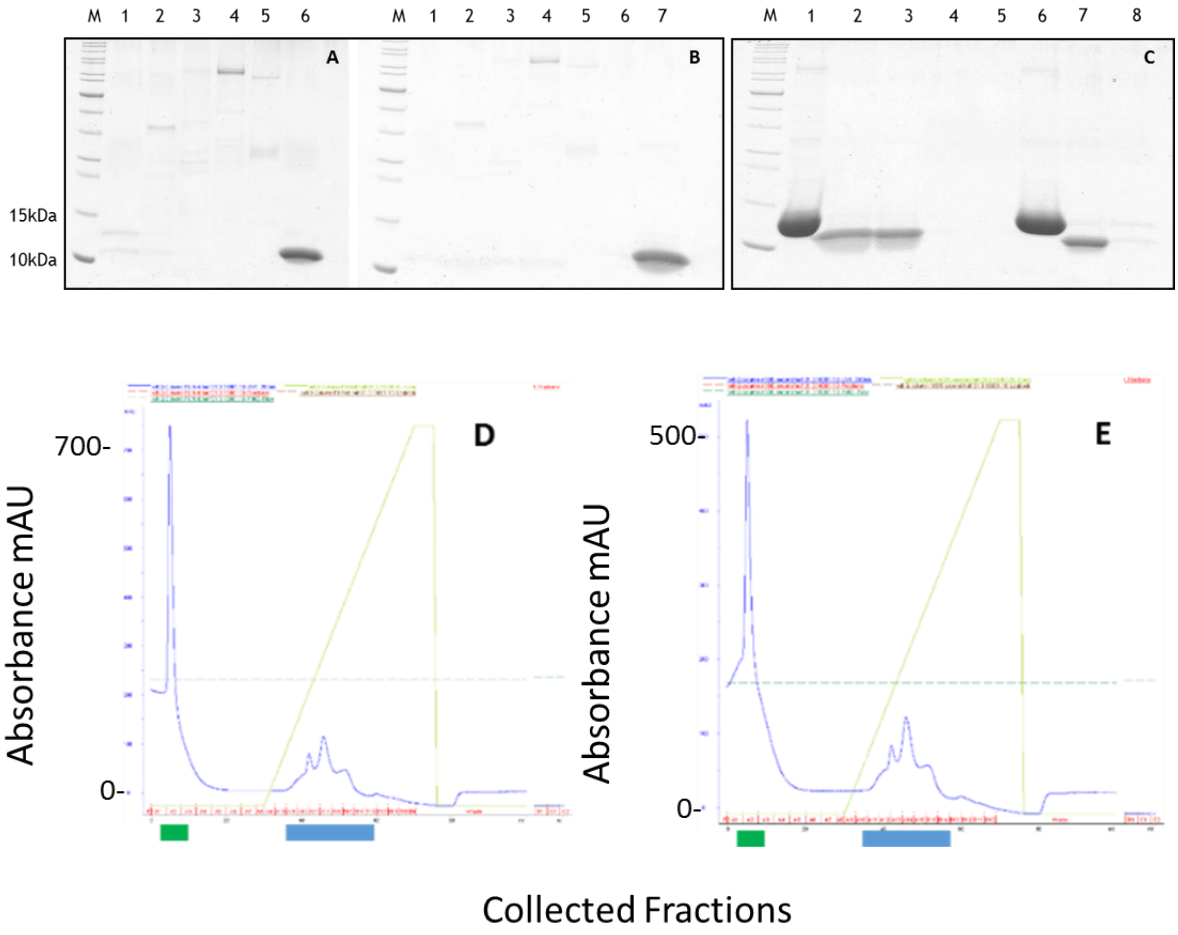


Figure 27: Ion Exchange Purification and Cleavage of Labelled Rhea F0

A) *fractions from Mono-Q for ^{15}NF0-K17E, M. marker, 1. A2, 2. A12, 3. A13, 4. A14, 5. B15, 6. Flow-through.* ***B)*** *fractions from Mono-Q for ^{15}NF0-WT, M. marker, 1. A2, 2. A12, 3. A13, 4. A14, 5. B15, 6. Flow-through (small peak) 7. flow-through (main peak).* ***C)*** *M. marker, 1. ^{15}NF0-WT desalted before TEV cleavage, 2-3. ^{15}NF0-WT flow-through 4-5. ^{15}NF0-WT A2 6. ^{15}NF0-K17E desalted before TEV cleavage, 7. ^{15}NF0-K17E flow-through, 8. ^{15}NF0-K17E A2. All fractions taken were 50 µl, boiled with 50 µl of sample buffer for 5 minutes at 95°C.* ***D)*** *Q-Evaluation of Rhea F0-K17E and* ***E)*** *Q-Evaluation of Rhea F0-WT, with A2 (green bar) fraction, and A12-B15 (blue bar) indicated.*

3.1.5 Rhea ^{13}C ^{15}N F0-K17E Expression and Purification

Having optimised the conditions to give correctly folded, soluble protein suitable for NMR analysis, a doubly labelled ^{13}C ^{15}N protein was successfully produced and purified using the same methods (see Figure 28). Uncleaved protein and cleaved His-tags were successfully removed from the prep.

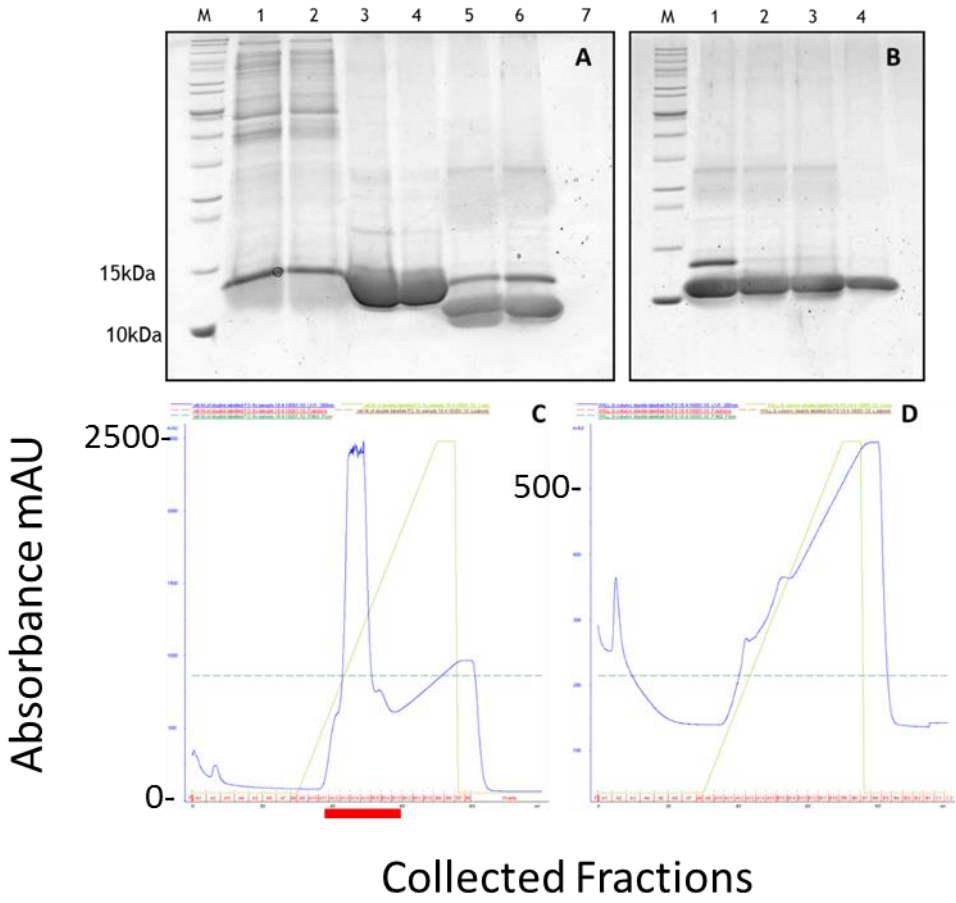


Figure 28: Expression and Purification of Double Labelled Rhea F0

A) *M. marker, 1. Sonicated $^{13}C^{15}$NF0-WT pellet, 2. Centrifuged $^{13}C^{15}$NF0-WT, 4. A13-A15 fractions pooled from Ni-Column, 5. Pooled fractions desalted into Q Buffer A, 6. $^{13}C^{15}$NF0-WT after cleavage with TEV before spin, 7. $^{13}C^{15}$NF0-WT after cleavage with TEV after spin.* ***B)*** *M. marker, 1. $^{13}C^{15}$NF0-WT after cleavage with TEV after spin, 2. $^{13}C^{15}$NF0-WT after Ni Resin and additional TEV, 3. $^{13}C^{15}$NF0-WT after Ni Resin and additional TEV then subsequent spin and filtration, 4. Flow-through of Q column showing purified $^{13}C^{15}$NF0-WT.* ***C)*** *Evaluation from Ni-Column showing peak between A13-A15 fractions (red bar).* ***D)*** *Evaluation from Mono-Q column run, indicating low levels of weak binding of contaminants (gel of fractions not shown).*

This protein was used to perform triple resonance NMR experiments for the Rhea F0 backbone assignments as described in the Method section 2.3.4.

Results 3.2: Unlabeled Rhea CD Analysis

CD analyses allowed us to check whether the folding and stability of the wild type and mutant Rhea F0 remained the same, to ensure that gross structural changes had not occurred from the K to E substitution.

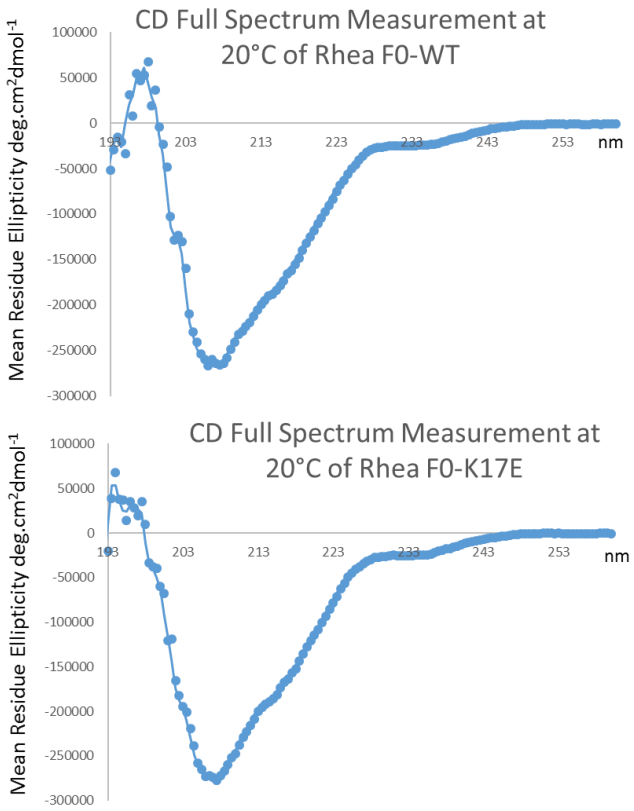


Figure 29: The Full Spectrum Measurements of Rhea F0-WT and F0-K17E
The top figure shows the circular dichroism full spectrum measurement of Rhea F0-WT at 20˚C. It confirms that the protein is stable and folded, containing a mixture of alpha and beta secondary structure. The main peak is at ~208 nm. The bottom figure shows the equivalent full spectrum measurement for the Rhea F0-K17E mutant protein. Both measurements are identical, revealing that the mutant protein is similarly folded. These spectra are overlaid in figure 30.

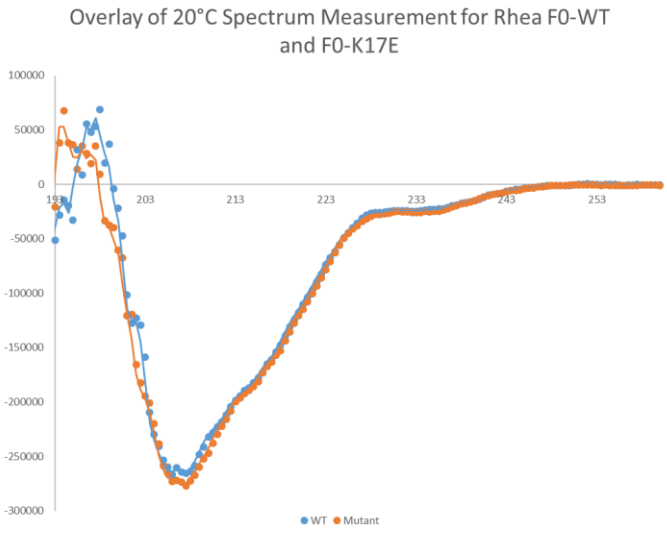


Figure 30: Overlay of full spectrum measurements of Rhea F0-WT and F0-K17E

Overlaying the full spectrum measurements for Rhea F0-WT and F0-K17E shows that the two proteins have identical folding. The ratio of alpha to beta structure is identical between both proteins. This allows us to confirm that the K17E substitution has not resulted in gross structural changes within the Rhea F0 domain. This is an important control for the NMR experiments, as it allows us to conclude that disrupted binding can be directly linked to the amino acid substitution, and not therefore as a result of the mutant protein being unrecognizable compared to its wild-type counterpart; if the changes to the protein are too severe, we would logically not expect binding to occur and subsequently could not conclude that K17 is a key residue involved in Rap1 binding.

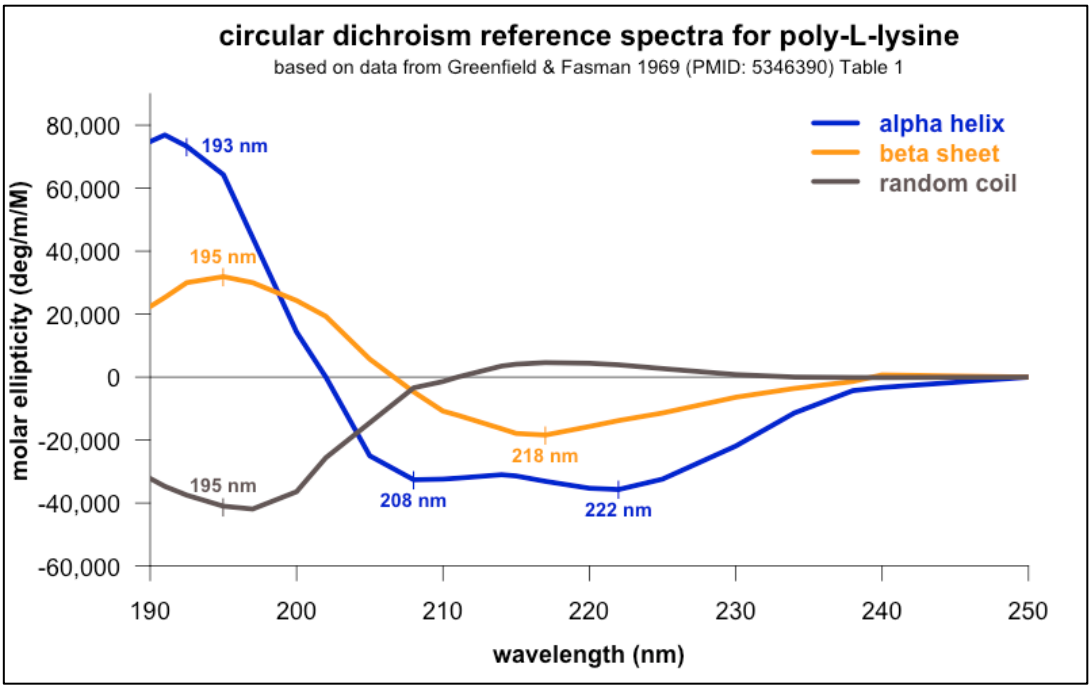


Figure 31: Typical Reference Spectra for structural determination using Circular Dichroism (Minikel, 2015)
The above figure shows the standard reference spectra for alpha helix (blue), beta sheet (orange) and random coil (grey). This data can be compared to the peaks obtained in a given protein's CD spectrum, to predict the secondary structure of the protein. A protein that is mostly alpha helical for example, will display characteristic peaks at 208- and 222 nm. The main peak observed in a protein mostly comprised of beta sheet is typically around 218 nm, whereas random coil typically peaks at 195 nm. These values are idealistic however, and proteins comprised of all three elements will reveal peaks somewhere between these values. The full spectrum measurements of Rhea F0-WT and F0-K17E for example, indicate the presence of alpha helix (with a negative peak around 208 nm) and beta sheet (with a second peak occurring at around 215 nm). When denaturing the F0 proteins, the measurements at 208 nm clearly demonstrated a shift to random coil following denaturation (data not shown).

The CD analysis revealed that the protein contained a combination of alpha and beta secondary elements, determined with reference spectra (see Figure 31), in good agreement with the structure of mouse F0. An overlay of full spectrum measurements of the wild type and mutant Rhea F0 proteins confirm that their

folding remains the same via CD (Figure 29, Figure 30), and their thermal stabilities are similarly almost identical. The melting curves at 208 nm for each protein have the same T_m (midpoint temperature of transition, where 50% of the protein molecules are unfolded (Allen and Cooper, 1999)) of 74˚C (Figure 32). This shows that the mutation of the surface exposed K17 residue has not perturbed the structural integrity. This is an important result as it allows us to conclude, with confidence, that the biological phenotype is not due to a secondary effect such as destabilisation of the Rhea.

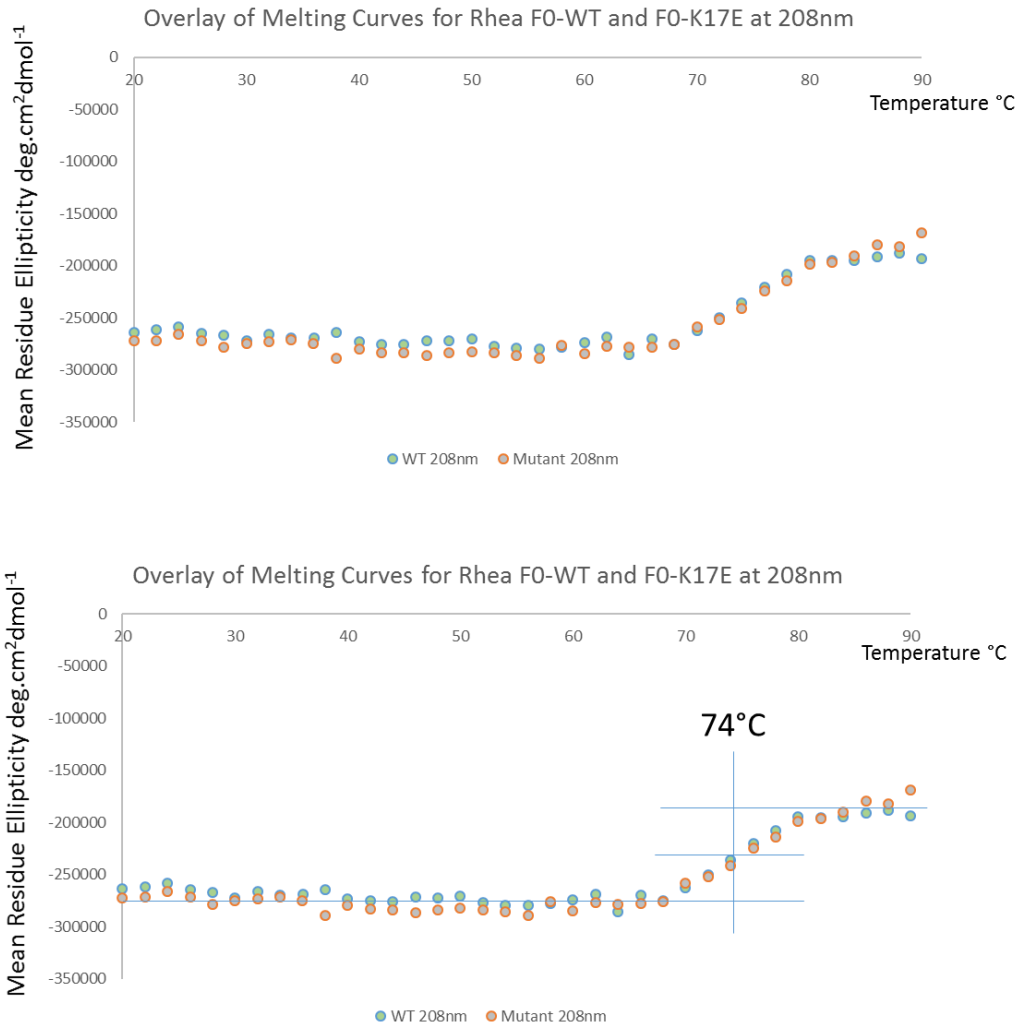


Figure 32: Overlay of melting curves of Rhea F0-WT and F0-K17E and determination of Tm for each protein

The top figure shows the denaturation (melting curve) of Rhea F0-WT (green) and F0-K17E (orange) from 20˚C to 90˚C, at 208 nm. The melting curve for both proteins is identical and reveal a Tm (midpoint temperature of transition, where 50% of the protein molecules are unfolded (Allen and Cooper, 1999) of 74˚C (calculation of which is shown in the bottom figure). This shows that the mutation of the surface exposed K17 residue has not perturbed the structural integrity.

Results 3.3: ^{15}N Rhea NMR Analysis

Having successfully purified the ^{15}N labelled protein, 1D and HSQC NMR spectra were obtained for each. Both spectra show that the proteins were correctly folded, with the HSQC displaying evenly dispersed peaks (Figure 33), and the 1D spectra displaying shifted methyls, indicative of a folded monomeric protein. Overlaying the spectra confirmed that we indeed had isolated the wild type and mutant versions of the same proteins, with a heavy degree of overlap of the peaks of the two spectra but small scattered differences between them.

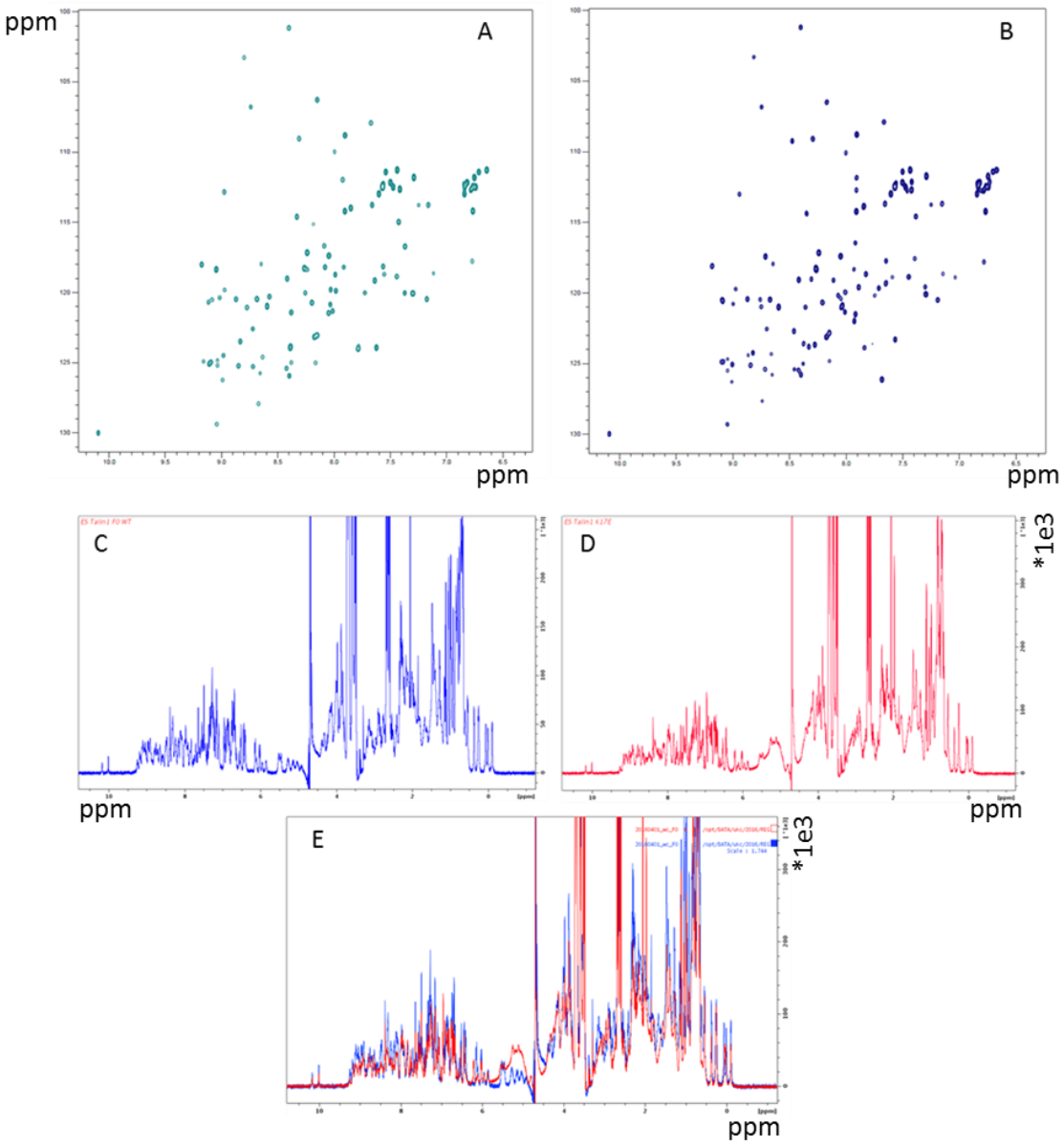


Figure 33: HSQC and 1D spectra of Rhea F0-WT and F0-K17E
Wild-Type (A) and Mutant (B) Rhea F0 HSQC spectra. Wild-Type (C) and Mutant (D) Rhea F0 1D spectra and overlay (E). The spectra of Rhea F0-K17E and F0-WT show that the proteins were correctly folded, with the HSQC displaying evenly dispersed peaks. The 1D spectra displayed shifted methyls, indicative of a folded monomeric protein.

Overlaying the spectra confirmed that we indeed had isolated the wild type and mutant versions of the same proteins, with a heavy degree of overlap of the peaks of the two spectra but small scattered differences between them.

Results 3.4: Rhea Backbone Assignment

The assignment of Rhea F0 was successfully completed (Figure 34) with all amide peaks being assigned as well as the CA, CB and CO resonances for each residue, the only missing assignments were the CO immediately before proline residues. As explained in the methods, this is due to proline residues not having an amide and therefore not having their own spin system. The Rhea F0 domain contains three prolines, which were therefore not assigned. For a detailed breakdown of how the assignment was carried out, see the methods section 2.3.4.

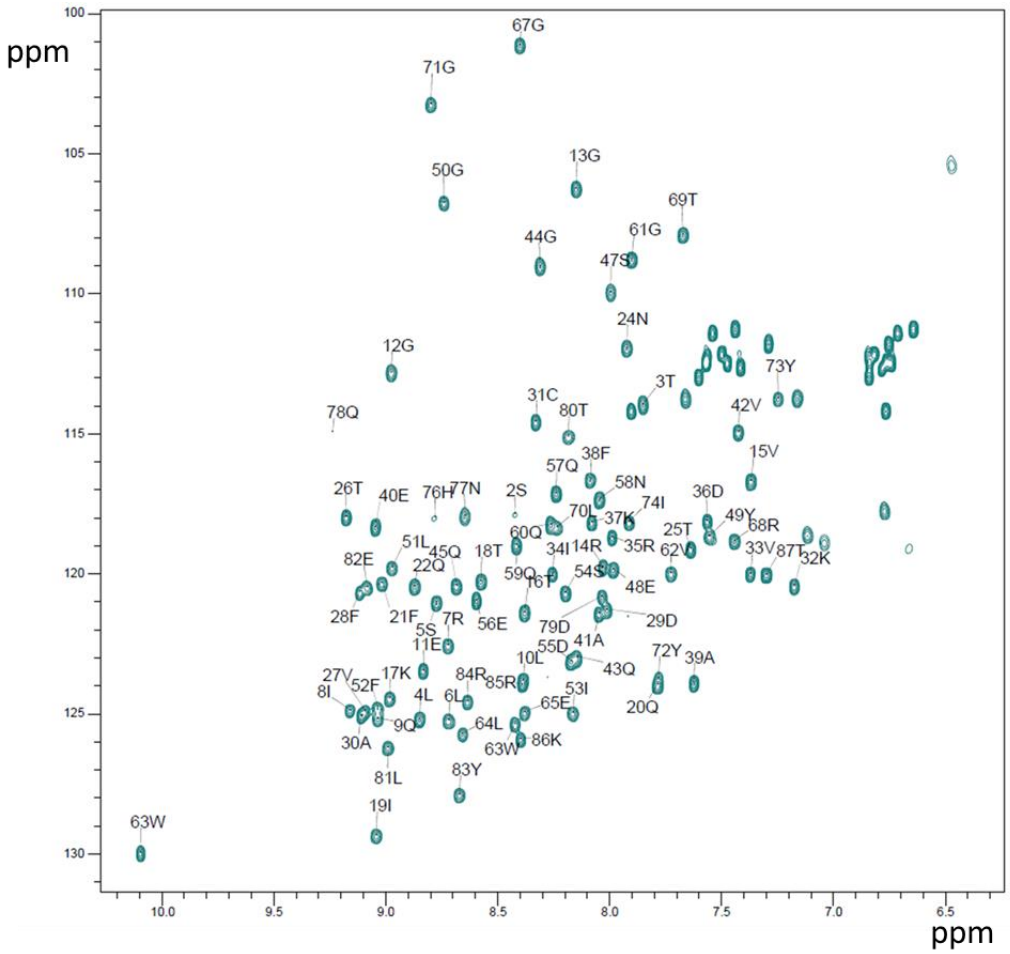


Figure 34: The completed assignment of Rhea F0-WT.

A completed section of the assignment is shown below (Figure 35).

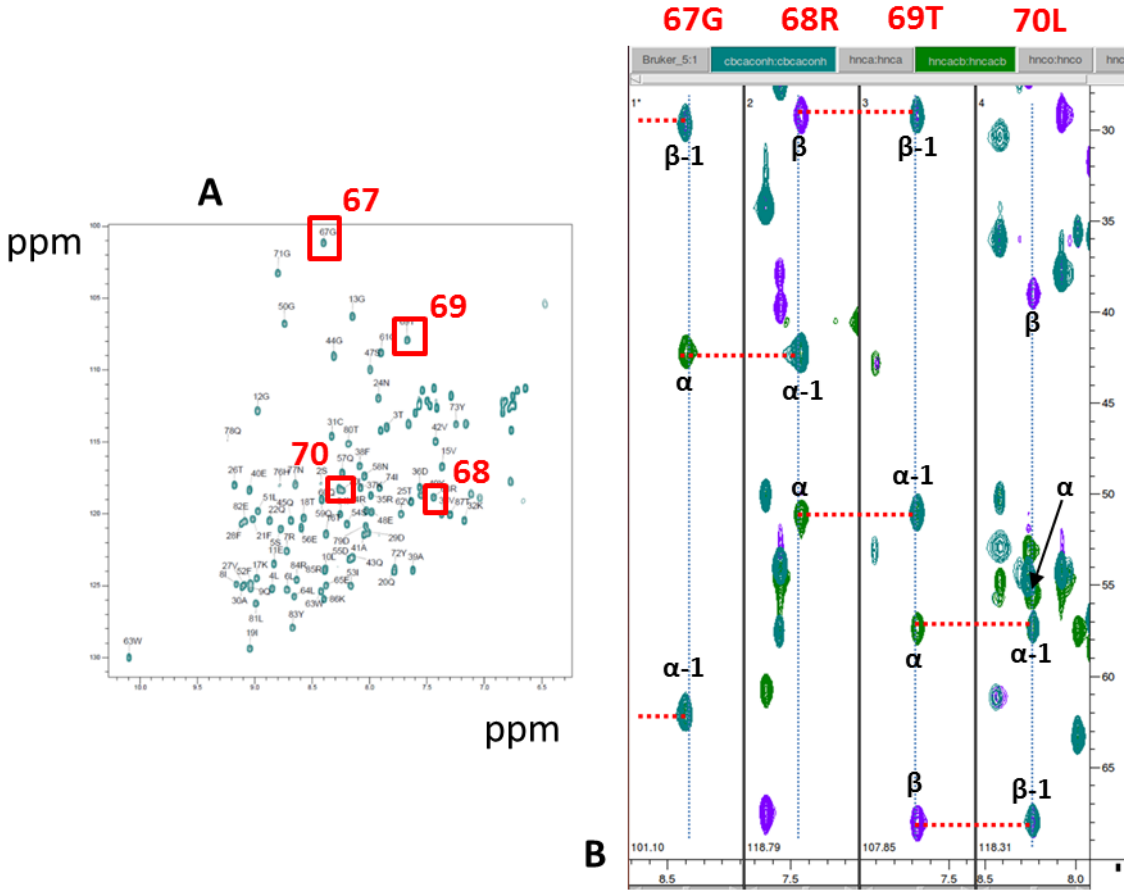


Figure 35: Completed Assignment Strips from CCPNMR for Rhea F0

The location of 4 assigned sequential amino acids on the HSQC spectra (A) and their corresponding strips (B), showing the HNCACB and HN(CO)CACB. The HNCACB shows the alpha and beta carbons of each amino acid, in addition to the alpha and beta carbons of the residue at i-1 in the N terminal direction. The alpha and beta carbon chemical shifts of 69 are diagnostic of a threonine. The arginine next to it was harder to assign with certainty, however, it is next to a glycine; clear from not having a beta carbon and a characteristic alpha position. The sequence G-R-T only appears once in the amino acid sequence, meaning that we can assign these residues with certainty. Further certainty comes from there being a proline at the next residue downstream of 67G. Proline do not have an amide and so do not produce a spin system, and as expected, there is no matching alpha and beta for the alpha- and beta-1 chemical shifts observed in the 67G HNCACB. This method allows us to assign the entire spectrum.

Having completed the assignment, the chemical shifts of all the CA, CO and CB residues were assigned to their corresponding residues, completing the backbone

assignment. These assignments were then exported in NMRSTAR format and deposited into the Biomagnetic Resonance Bank (BMRB), a public repository for NMR data.

As well as providing sequential information, the chemical shift assignments were used to calculate the dihedral angles based on chemical shift indexing with the program DANGLE which forms part of the CCPNMR package. This created a Ramachandran Plot output and determined the secondary structure elements of the F0 domain (Figure 36 left). By comparing these to a predicted structural model created using Phyre, which used mouse talin F0 as a template, we could confirm that the assignments were correct and also confirm that Rhea F0 has the same secondary structure elements as mouse talin F0 (Figure 36 right). These results reveal that the assignment was successful.

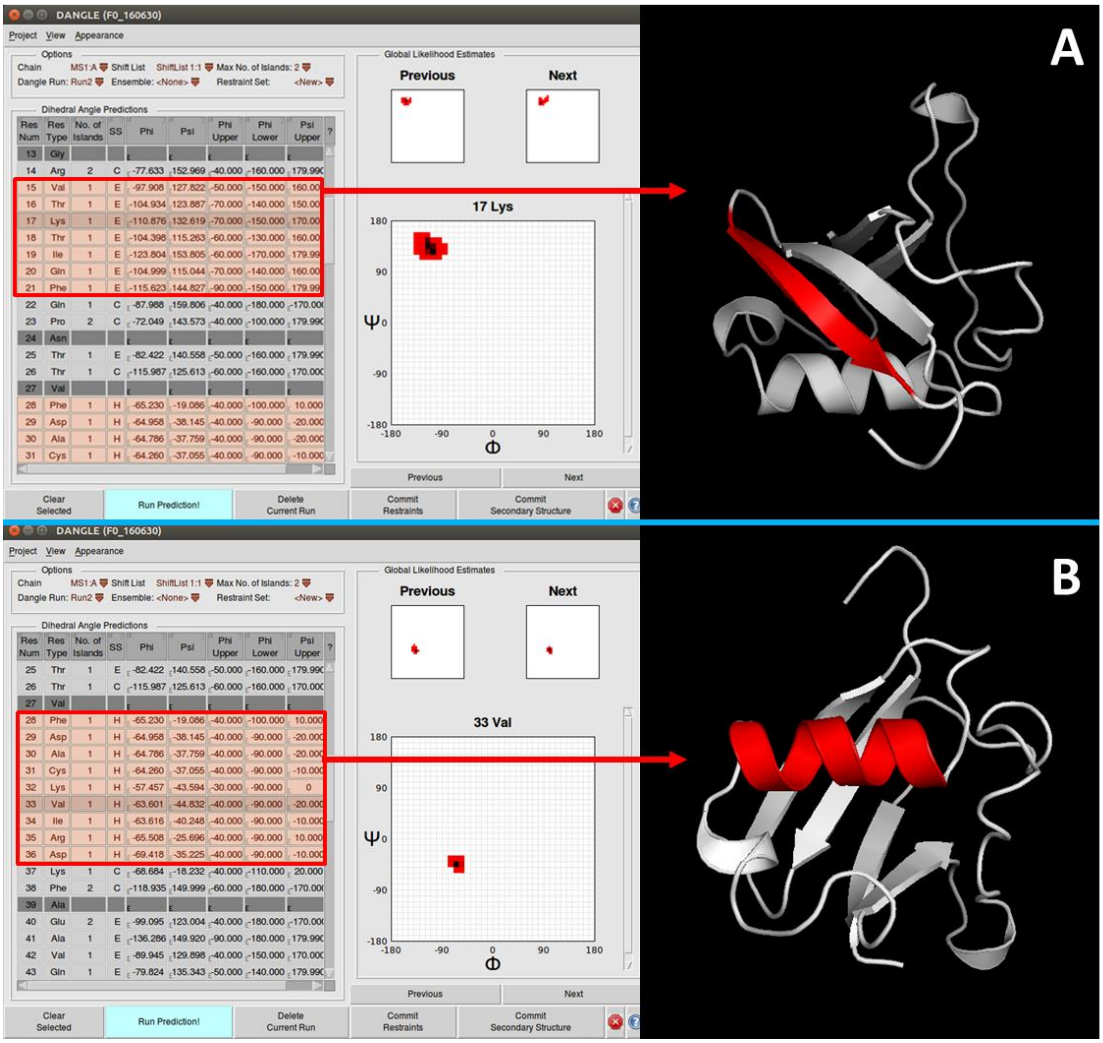


Figure 36: Ramachandran Plot of F0 structure based on the chemical shift assignments

*The Ramachandran Plot predicted sheet structure between residues 16-21 (A) and a helix between residues 28-36 (B) which maps accurately to corresponding regions of the Phyre model (right) (*Kelley LA et al. 2015).

Results 3.5: Rap Expression and Purification

3.5.1 Rap1 Expression Trial and purification in BL21(DE3)

Full length Drosophila Ras-like Protein 3 was found to be insoluble in BL21(DE3) E.coli cells. The cells were very slow to grow, suggesting that the constructs were inducing toxicity within the cells. Additionally, the thawed cell pellets were very stringy, suggesting the presence of a large amount of DNA. It appears that the tagged construct expressed better, however both constructs were insoluble (Figure 37A). Sonicating the cells in a larger volume was attempted, to test whether ineffective sonication had resulted in the protein being in the insoluble debris pellet (Figure 37B). The protein appeared to remain insoluble. Due to their overexpression, the proteins were easily identifiable on the gel, with their bands appearing at the expected size of around 20kDa. The untagged construct was also slightly smaller on the gel, further confirming each proteins identity on the gel.

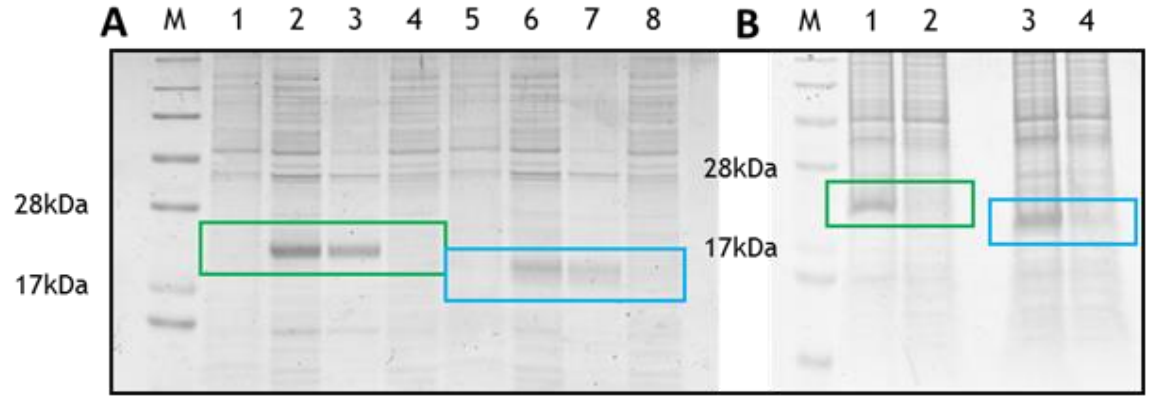


Figure 37: Solubility test of Rap1 in BL21(DE3) cells
A) *M. marker, 1. Pre-induction His-Rap1, 2. Whole cell fraction for His-Rap1 (also shown in* ***B1)****, 3. Insoluble fraction of His-Rap1, 4. Soluble fraction of His-Rap1 (also shown in* ***B2)****, 5. Pre-induction UT-Rap1, 6. Whole cell fraction for UT-Rap1 (also shown in* ***B3)****,*

*7. Insoluble fraction of UT-Rap1, 8. Soluble fraction of UT-Rap1 (also shown in **B4**). His-Rap1 is indicated with the green box, UT-Rap1 with the blue box.*

Attempting to purify the soluble fraction of the His-tagged construct was unsuccessful, despite an initially promising A280 peak being observed in the Ni-NTA purification process (Figure 38).

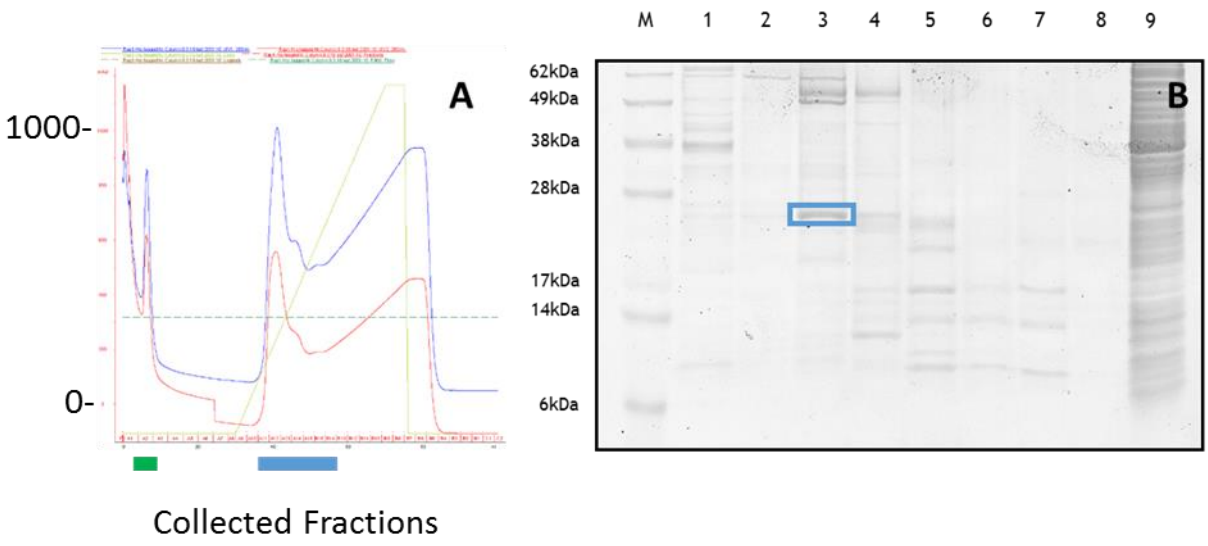


Figure 38: Attempting to purify Rap1 from BL21(DE3) cells
***A)** evaluation for the His-Rap1 soluble fraction. A280 is shown in green and A260 in red. The green bar indicates the location of A2, the Blue bar indicates A11-B14 **B) Fractions from Q evaluation**. M. marker, 1. A2, 2. A11, 3. A12, 4. A13, 5. A14, 6. A15, 7. B15, 8. B14, 9. Flow-through. The only possible target is highlighted with the blue box, however, it was determined to be slightly too large and at too low a yield for meaningful assay.*

The peak at A280nm was re-examined and shown to also have high absorbance at A260nm, suggesting that the peak could be the result of a large amount of DNA. This idea is supported by the consistency of the pellets upon thawing, and the slow cell growth rate observed. There was a potential target identified from the gel as being the Rap1 protein (blue box, Figure 38B), however, it was in too low a yield for use in the NMR experiments. Additionally, the identified target appears slightly too large to be the Rap1.

3.5.2 Rap1 Expression Trial and purification in CK600K

We were gifted CK600K cells by Dr Igor Barsukov (University of Liverpool), who has expertise in working with small GTPases. These cells were more tolerable of the Rap1 vectors than BL21(DE3) cells, due to a faster growth rate. The results of this expression test was similarly disappointing however, with there being no visible overexpression on the gel (Figure 39).

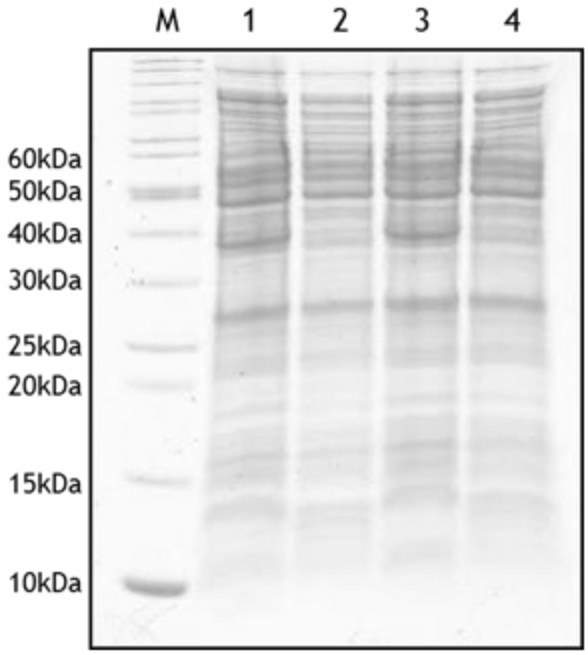


Figure 39: Solubility Test of Rap1 in CK600K
solubility test of drosophila Rap1-His (1 + 2) and Rap1-UT (3 + 4). M. marker, 1/3. whole cell fractions, 2/4. soluble fractions

The gel did reveal several proteins at approximately the right size of the Rap1 tagged construct, however, these proteins were the same size across both the tagged and untagged construct lanes, suggesting that they were not the Rap protein. In order to look for the presence of any Rap1, even if at low expression, the soluble fraction of the His-tagged construct prep was purified with Ni-NTA affinity chromatography (Figure 40).

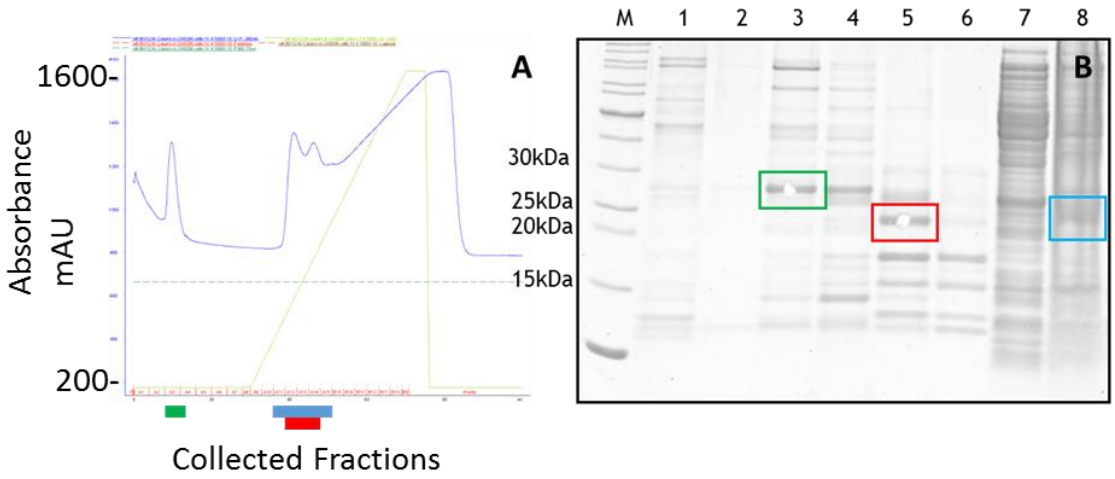


Figure 40: Ni-NTA chromatography identifies two Rap1 targets
A) *Ni Evaluation of His-Rap1 in CK600K cells displaying a small double peak (400 mAU) between the A12 and A14 eluted fractions. The green bar indicates A3, the blue indicates A11-A15 and the red indicates A12-14* ***B)*** *fractions from Ni column elution. M. marker, 1. A3, 2. A11, 3. A12, 4. A13, 5. A14, 6. A15, 7. flow-through, 8. Sonicated His-Rap1 from 3.1.5 prep. Green box indicates a potential (but unlikely) target for Rap1, the red box indicates a more likely candidate for Rap1 (despite low yield), the blue box indicates the presence of His-Rap1 (used as a marker). Green and red targets were analysed with mass spec.*

The results were again disappointing, with only a small peak being observed on the evaluation. Two possible candidates were identified that could have been Rap1, one of which was a correct size match using a Rap1 marker (Figure 40B).

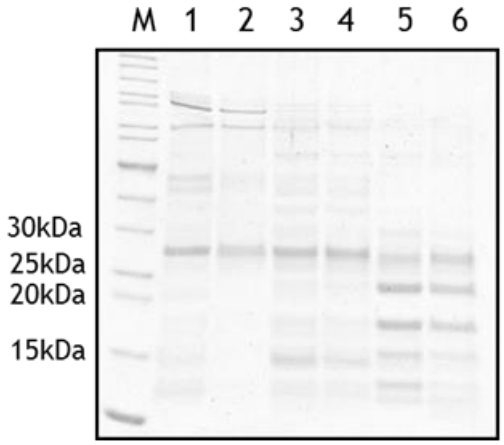


Figure 41: Attempting to cleave two Rap1 targets
TEV addition to the A12-14 elutions from the Ni-Column; no cleavage was observed. M. marker, 1. A12 before TEV addition, 2. A12 after TEV addition, 3. A13 before TEV addition, 4. A13 after TEV addition, 5. A14 before TEV addition, 6. A14 after TEV addition.

No cleavage was achieved with these proteins (Figure 41), making it likely that neither targets were Rap1. To confirm this, stabs of the gel (shown in Figure 40B) were taken and identified as E. coli 30S ribosomal protein S2 and cAMP-activated global transcriptional regulator CRP (Figure 42), via Mass Spectroscopy, confirming that they were not Rap1 but E.coli proteins.

1. RS2_ECO27 **Mass:** 26784 **Score:** 144 **Expect:** 2.2e-009 **Matches:** 20
30S ribosomal protein S2 OS=Escherichia coli O127:H6 (strain E2348/69 / EPEC) GN=rpsB PE=3 SV=1
RS2_ECO45 **Mass:** 26784 **Score:** 144 **Expect:** 2.2e-009 **Matches:** 20
30S ribosomal protein S2 OS=Escherichia coli O45:K1 (strain S88 / ExPEC) GN=rpsB PE=3 SV=1
RS2_ECO57 **Mass:** 26784 **Score:** 144 **Expect:** 2.2e-009 **Matches:** 20
30S ribosomal protein S2 OS=Escherichia coli O157:H7 GN=rpsB PE=3 SV=2
RS2_ECO5E **Mass:** 26784 **Score:** 144 **Expect:** 2.2e-009 **Matches:** 20
30S ribosomal protein S2 OS=Escherichia coli O157:H7 (strain EC4115 / EHEC) GN=rpsB PE=3 SV=1
RS2_ECO7I **Mass:** 26784 **Score:** 144 **Expect:** 2.2e-009 **Matches:** 20
30S ribosomal protein S2 OS=Escherichia coli O7:K1 (strain IAI39 / ExPEC) GN=rpsB PE=3 SV=1
RS2_ECO81 **Mass:** 26784 **Score:** 144 **Expect:** 2.2e-009 **Matches:** 20
30S ribosomal protein S2 OS=Escherichia coli O81 (strain ED1a) GN=rpsB PE=3 SV=1
RS2_ECOBW **Mass:** 26784 **Score:** 144 **Expect:** 2.2e-009 **Matches:** 20
30S ribosomal protein S2 OS=Escherichia coli (strain K12 / MC4100 / BW2952) GN=rpsB PE=3 SV=1
RS2_ECODH **Mass:** 26784 **Score:** 144 **Expect:** 2.2e-009 **Matches:** 20
30S ribosomal protein S2 OS=Escherichia coli (strain K12 / DH10B) GN=rpsB PE=3 SV=1
RS2_ECOK1 **Mass:** 26784 **Score:** 144 **Expect:** 2.2e-009 **Matches:** 20
30S ribosomal protein S2 OS=Escherichia coli O1:K1 / APEC GN=rpsB PE=3 SV=1
RS2_ECOL5 **Mass:** 26784 **Score:** 144 **Expect:** 2.2e-009 **Matches:** 20
30S ribosomal protein S2 OS=Escherichia coli O6:K15:H31 (strain 536 / UPEC) GN=rpsB PE=3 SV=1
RS2_ECOL6 **Mass:** 26784 **Score:** 144 **Expect:** 2.2e-009 **Matches:** 20
30S ribosomal protein S2 OS=Escherichia coli O6:H1 (strain CFT073 / ATCC 700928 / UPEC) GN=rpsB PE=3 SV=2
RS2_ECOLC **Mass:** 26784 **Score:** 144 **Expect:** 2.2e-009 **Matches:** 20
30S ribosomal protein S2 OS=Escherichia coli (strain ATCC 8739 / DSM 1576 / Crooks) GN=rpsB PE=3 SV=1

1. CRP_ECO57 **Mass:** 23796 **Score:** 179 **Expect:** 6.9e-013 **Matches:** 15
cAMP-activated global transcriptional regulator CRP OS=Escherichia coli O157:H7 GN=crp PE=3 SV=1
CRP_ECOL6 **Mass:** 23796 **Score:** 179 **Expect:** 6.9e-013 **Matches:** 15
cAMP-activated global transcriptional regulator CRP OS=Escherichia coli O6:H1 (strain CFT073 / ATCC 700928 / UPEC) GN=crp PE=3 SV=1
CRP_ECOLI **Mass:** 23796 **Score:** 179 **Expect:** 6.9e-013 **Matches:** 15
cAMP-activated global transcriptional regulator CRP OS=Escherichia coli (strain K12) GN=crp PE=1 SV=1
CRP_ENTAE **Mass:** 23812 **Score:** 179 **Expect:** 6.9e-013 **Matches:** 15
cAMP-activated global transcriptional regulator CRP OS=Enterobacter aerogenes GN=crp PE=1 SV=1
CRP_SALTY **Mass:** 23812 **Score:** 179 **Expect:** 6.9e-013 **Matches:** 15
cAMP-activated global transcriptional regulator CRP OS=Salmonella typhimurium (strain LT2 / SGSC1412 / ATCC 700720) GN=crp PE=1 SV=1
CRP_SHIFL **Mass:** 23796 **Score:** 179 **Expect:** 6.9e-013 **Matches:** 15
cAMP-activated global transcriptional regulator CRP OS=Shigella flexneri GN=crp PE=3 SV=1
Y3711_CLOAB **Mass:** 10409 **Score:** 32 **Expect:** 3.7e+002 **Matches:** 3
Uncharacterized protein CA_C3711 OS=Clostridium acetobutylicum (strain ATCC 824 / DSM 792 / JCM 1419 / LMG 5710 / VKM B-1787) GN=CA_C3711 PE=4 SV=2

Figure 42: Mass Spec of the two possible Rap1 target bands
Upper *the mass spec of the higher molecular weight target was identified as E. coli 30S ribosomal protein S2 and* ***lower*** *the mass spec of the lower molecular weight target was identified as E. coli cAMP-activated global transcriptional regulator CRP, with a high degree of certainty.*

3.5.3 Mouse Rap1 Expression Trial in CK600K

It was observed in the literature that most structures of Rap1b had previously been based off a truncated protein (residues 1-166, spanning the conserved G-Domain), in CK600K cells, using a ptac expression system ((Edreira et al., 2009),

(Kupzig et al., 2006), (Tucker et al 1986)). Dr Igor Barsukov (University of Liverpool) gifted us this construct, which was originally supplied to Dr David Critchley, by Dr Alfred Wittinghofer at the Max Planck Institute of Physiology; a leading expert in the structural analysis of small GTPases.

The resulting expression and solubility test from this construct was promising with a prominent, overexpressed band in the soluble fraction (Figure 43 lane 4).

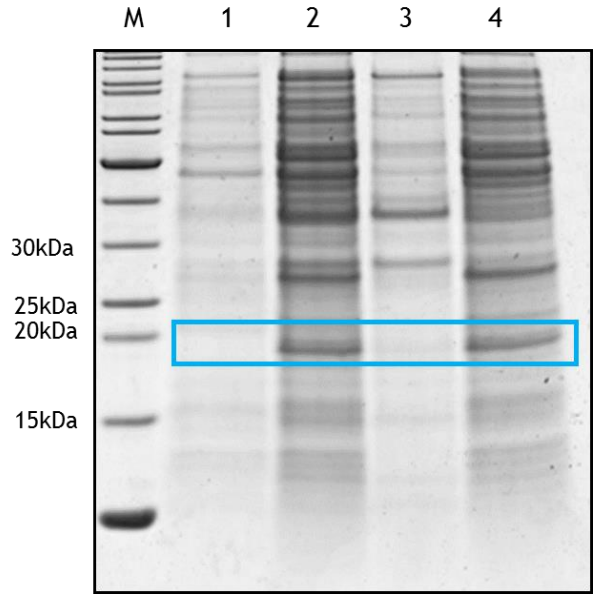


Figure 43: Solubility test showing Soluble Rap1
M. marker, 1. Pre-induction, 2. Post-induction "whole cell" 3. Post-induction insoluble, 4. Post-induction soluble. Fractions were loaded to adjust for pre- and post-induction ODs (10.4µl or pre-induction and 4.5µl of post induction fractions). The blue box indicates the mouse Rap1b location.

The truncated Rap1b is around 2kDa smaller than the full length protein, as determined by UniProt, with a molecular weight of 18 kDa. At this size, there was a promising target band on the gel that was overexpressed and 100% soluble.

3.5.4.1 Mouse Rap1 in CK600K purification

As the construct was untagged, purification was successfully achieved using Q-column ion exchange chromatography (Figure 44).

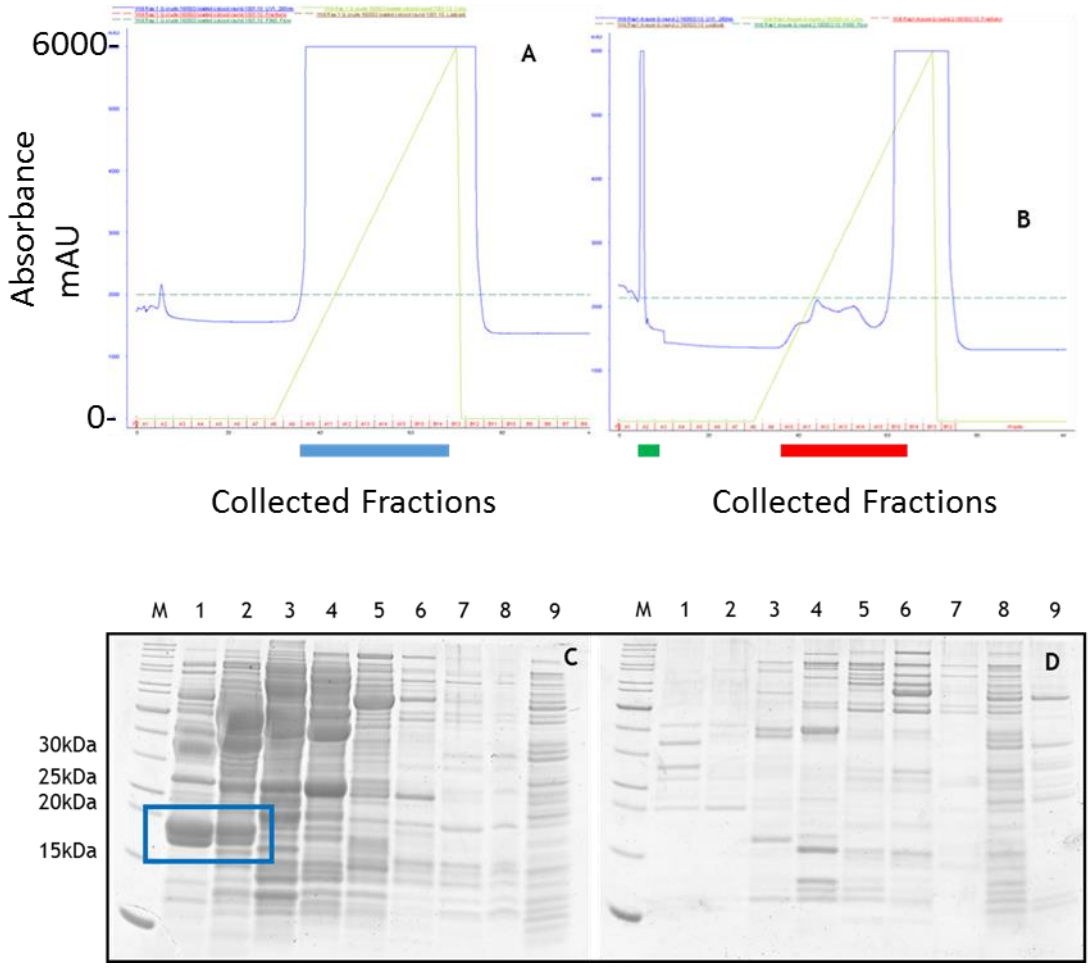


Figure 44: Ion Exchange purification of Rap1 (1)
A) *Q-Evaluation of first cycle of Q indicating presence of large volumes of protein. A10-B14 is indicated by the blue bar* ***B)*** *Flow through of first round was loaded onto Q-column again. The evaluation is shown. The green bar indicates A2 and the red bar indicates A10-B15. Significantly less protein eluted from the column at the lower part of the salt gradient. A large amount of protein remained bound to the column until the high part of the salt gradient washed through the column. Data for successive runs are not shown.* ***C)*** *gel fractions from across peak of first Q-cycle. M. marker, 1. A10, 2. A11, 3. A12, 4. A13, 5. A14, 6. A15, 7. B15, 8. B14, 9. Flow-through.* ***D)*** *M. marker, 1. A2, 2. A10, 3. A11, 4. A12, 5. A13, 6. A14, 7. B15, 8. Flow through from first round, 9. Flow through from second round. The blue box indicates the location of Rap1b.*

It was found that the target protein bound to the column in the first round of Q-Purification (Figure 45A+C). This protein was concentrated down to 5 ml and loaded onto a preparative Superdex S200 gel filtration column.

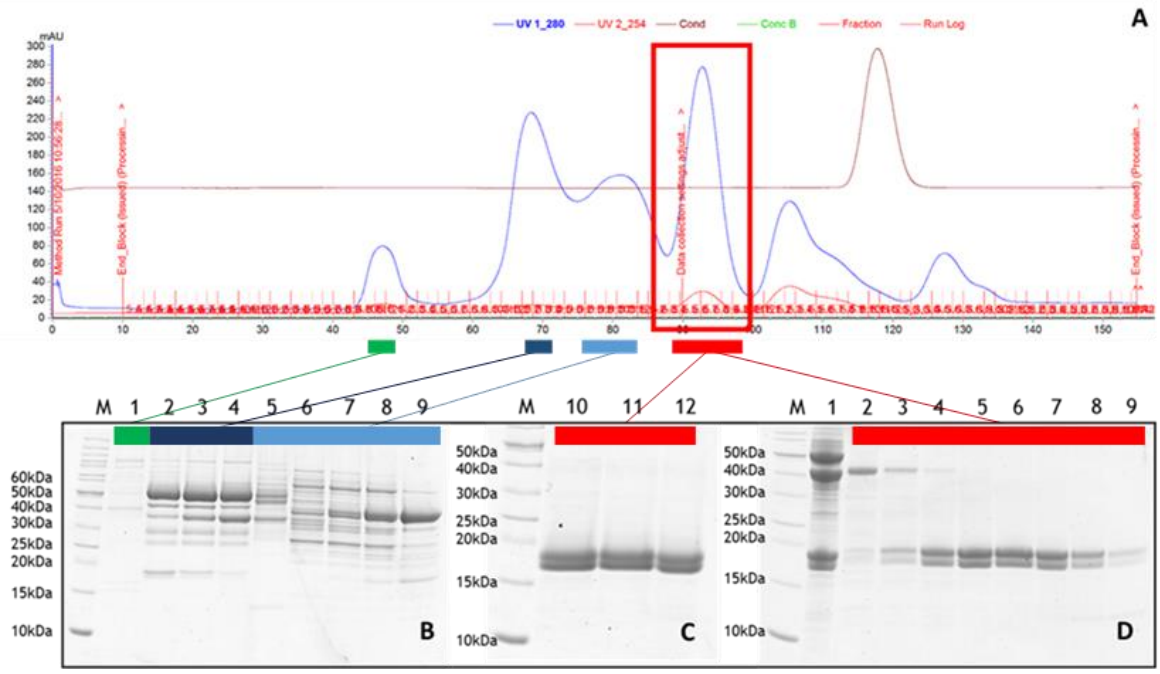


Figure 45: Gel filtration purification of Rap1: Part 1
A) *gel filtration evaluation of the pooled A10 and A11 fractions from the Q-column. C1 was taken from the region indicated by the green bar, the dark blue bar indicates D2-D4, the light blue D7-D12 and the red E4-E11. The* ***red*** *box indicates the Rap1 peak.* ***B)*** *and* ***C)*** *fractions taken from peaks across the evaluation to identify Rap. M. marker, 1. C1, 2. D2, 3. D3, 4. D4, 5. D7, 6. D10, 7. D11, 8. D12, 9. E2, 10. E7, 11. E8, 12, E9.* ***D)*** *running the Rap1 peak 1. A11 fraction from Q column 2. E4, 3. E5, 4. E6, 5. E7, 6. E8, 7. E9, 8. E10, 9. E11.*

Gel filtration was successful in isolating pure Rap1b (Figure 45). An interesting observation is that the isolated protein was present in two bands on the gel.

Both of these bands were sent for Mass Spec analysis and the results confirmed that they were both Rap1b (Figure 46). It is possible that the two bands on the gel represent the nucleotide-bound state of the protein, with one band representing the GDP bound form, and the other the GTP bound form. Alternatively, one band could represent both GTP and GDP forms, with the other band being an apo form.

RAP1B_BOVIN **Mass:** 21040 **Score:** 342 **Matches:** 9(5) **Sequences:** 9(5) **emPAI:** 3.13
Ras-related protein Rap-1b OS=Bos taurus GN=RAP1B PE=2 SV=1
Check to include this hit in error tolerant search or archive report

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
14	1258.5592	1257.5519	1257.5513	0.48	0	42	0.016	1	U	K.YDPTIEDSYR.K
18	1298.7190	1297.7117	1297.7102	1.14	1	38	0.082	1	U	R.VVGKEQGQNLAR.Q
20	1319.6321	1318.6248	1318.6187	4.62	1	21	4	1	U	K.CDLEDERVVGK.E
22	1386.6516	1385.6444	1385.6463	-1.39	1	29	0.47	1	U	K.YDPTIEDSYRK.Q
30	1494.7788	1493.7715	1493.7878	-10.92	0	46	0.016	1	U	K.INVNEIFYDLVR.Q
38	1554.6921	1553.6848	1553.6933	-5.42	0	48	0.0038	1	U	R.QWNNCAFLESSAK.S
47	1643.8663	1642.8590	1642.8600	-0.61	1	17	11	5	U	R.VKDTDDVPMILVGNK.C
49	1665.9041	1664.8968	1664.9138	-10.18	0	54	0.0017	1	U	K.SALTVQFVQGIFVEK.Y
59	1709.9044	1708.8972	1708.9148	-10.32	1	47	0.011	1	U	K.SKINVNEIFYDLVR.Q

Proteins matching the same set of peptides:
RAP1B_CHICK **Mass:** 21040 **Score:** 342 **Matches:** 9(5) **Sequences:** 9(5)
Ras-related protein Rap-1b OS=Gallus gallus GN=RAP1B PE=2 SV=1
RAP1B_HUMAN **Mass:** 21040 **Score:** 342 **Matches:** 9(5) **Sequences:** 9(5)
Ras-related protein Rap-1b OS=Homo sapiens GN=RAP1B PE=1 SV=1
RAP1B_MACFA **Mass:** 21040 **Score:** 342 **Matches:** 9(5) **Sequences:** 9(5)
Ras-related protein Rap-1b OS=Macaca fascicularis GN=RAP1B PE=2 SV=1
RAP1B_MOUSE **Mass:** 21040 **Score:** 342 **Matches:** 9(5) **Sequences:** 9(5)
Ras-related protein Rap-1b OS=Mus musculus GN=Rap1b PE=1 SV=2
RAP1B_PANTR **Mass:** 21040 **Score:** 342 **Matches:** 9(5) **Sequences:** 9(5)
Ras-related protein Rap-1b OS=Pan troglodytes GN=RAP1B PE=2 SV=1
RAP1B_PONAB **Mass:** 21040 **Score:** 342 **Matches:** 9(5) **Sequences:** 9(5)
Ras-related protein Rap-1b OS=Pongo abelii GN=RAP1B PE=2 SV=1
RAP1B_XENLA **Mass:** 21049 **Score:** 342 **Matches:** 9(5) **Sequences:** 9(5)
Ras-related protein Rap-1b OS=Xenopus laevis GN=rap1b PE=2 SV=1
RAP1B_XENTR **Mass:** 21049 **Score:** 342 **Matches:** 9(5) **Sequences:** 9(5)
Ras-related protein Rap-1b OS=Xenopus tropicalis GN=rap1b PE=2 SV=1

RAP1B_BOVIN **Mass:** 21040 **Score:** 351 **Matches:** 9(4) **Sequences:** 9(4) **emPAI:** 3.13
Ras-related protein Rap-1b OS=Bos taurus GN=RAP1B PE=2 SV=1
Check to include this hit in error tolerant search or archive report

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
15	1258.5533	1257.5461	1257.5513	-4.18	0	34	0.077	1	U	K.YDPTIEDSYR.K
18	1298.7079	1297.7006	1297.7102	-7.42	1	37	0.11	1	U	R.VVGKEQGQNLAR.Q
20	1319.6149	1318.6076	1318.6187	-8.43	1	29	0.58	1	U	K.CDLEDERVVGK.E
21	1386.6417	1385.6344	1385.6463	-8.59	1	35	0.11	1	U	K.YDPTIEDSYRK.Q
30	1494.7810	1493.7737	1493.7878	-9.41	0	45	0.02	1	U	K.INVNEIFYDLVR.Q
39	1554.6857	1553.6784	1553.6933	-9.56	0	49	0.0029	1	U	R.QWNNCAFLESSAK.S
49	1643.8474	1642.8401	1642.8600	-12.12	1	22	3.4	1	U	R.VKDTDDVPMILVGNK.C
51	1665.9019	1664.8946	1664.9138	-11.53	0	59	0.00055	1	U	K.SALTVQFVQGIFVEK.Y
60	1709.9029	1708.8957	1708.9148	-11.19	1	41	0.039	1	U	K.SKINVNEIFYDLVR.Q

Proteins matching the same set of peptides:
RAP1B_CHICK **Mass:** 21040 **Score:** 351 **Matches:** 9(4) **Sequences:** 9(4)
Ras-related protein Rap-1b OS=Gallus gallus GN=RAP1B PE=2 SV=1
RAP1B_HUMAN **Mass:** 21040 **Score:** 351 **Matches:** 9(4) **Sequences:** 9(4)
Ras-related protein Rap-1b OS=Homo sapiens GN=RAP1B PE=1 SV=1
RAP1B_MACFA **Mass:** 21040 **Score:** 351 **Matches:** 9(4) **Sequences:** 9(4)
Ras-related protein Rap-1b OS=Macaca fascicularis GN=RAP1B PE=2 SV=1
RAP1B_MOUSE **Mass:** 21040 **Score:** 351 **Matches:** 9(4) **Sequences:** 9(4)
Ras-related protein Rap-1b OS=Mus musculus GN=Rap1b PE=1 SV=2
RAP1B_PANTR **Mass:** 21040 **Score:** 351 **Matches:** 9(4) **Sequences:** 9(4)
Ras-related protein Rap-1b OS=Pan troglodytes GN=RAP1B PE=2 SV=1
RAP1B_PONAB **Mass:** 21040 **Score:** 351 **Matches:** 9(4) **Sequences:** 9(4)
Ras-related protein Rap-1b OS=Pongo abelii GN=RAP1B PE=2 SV=1
RAP1B_XENLA **Mass:** 21049 **Score:** 351 **Matches:** 9(4) **Sequences:** 9(4)
Ras-related protein Rap-1b OS=Xenopus laevis GN=rap1b PE=2 SV=1
RAP1B_XENTR **Mass:** 21049 **Score:** 351 **Matches:** 9(4) **Sequences:** 9(4)
Ras-related protein Rap-1b OS=Xenopus tropicalis GN=rap1b PE=2 SV=1

Figure 46: Mass Spec Identification of Rap1b
*Mass Spec analysis of the top band of the gel (*Figure 44C*) shown (top) and of the lower band of the gel (*Figure 44C*) shown (bottom). Both evaluations indicate that the Rap1b protein had been isolated and was present in both bands on the gel.*

3.5.4.2 Mouse Rap1 in CK600K purification for NMR analysis

Having optimised the purification protocol, the procedure was scaled up to 3 litres, to produce sufficient Rap1b for NMR analysis. The results of the scaled up experiment reveal that the Rap1 Q-Purification was comparable to in section 3.5.4.1, as shown by the A10-11 fractions eluted from the first half litre prep (see Figure 47 lanes 1 and 2).

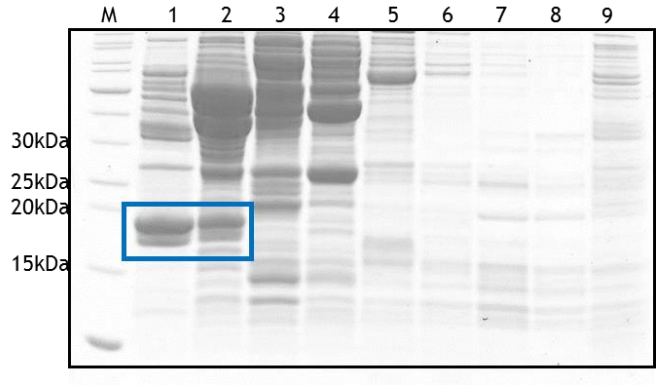


Figure 47: Fractions from the Ion Exchange purification of Rap1 (2)
Q-fractions from across peak of first ½ L prep. M. marker, 1. A10, 2. A11, 3. A12, 4. A13, 5. A14, 6. A15, 7. B15, 8. B14, 9. Flow-through. Blue box indicated the location of Rap1.

It should be expected that the intensity of Rap1 seen on a gel should be approximately the same, in the pooled A10-11 fractions from all 6 half litre preps, as for the A10-11 of the first 500 ml prep. As seen below (Figure 48 lane 1), the Rap1 intensity appears to have the same intensity as in the first 500 ml prep (Figure 47 lanes 1 and 2).

After pooling the six A10-11 fractions and concentrating to 5 ml, a large excess of Rap1b was observed (Figure 48 right, lane 2). In spite of this, the Rap1 that eluted from the gel filtration column was at a lower yield than expected; the peak indicated a yield of protein approximately the same, if not slightly less than the amount observed during the optimization process (Figure 45). Nevertheless, the peak was run on a gel and shown to be pure (Figure 48 right, lanes 3-8). The exact reasons for this loss on the gel filtration are not clear but one possible explanation was that some precipitation of the Rap1 had occurred at the high concentration.

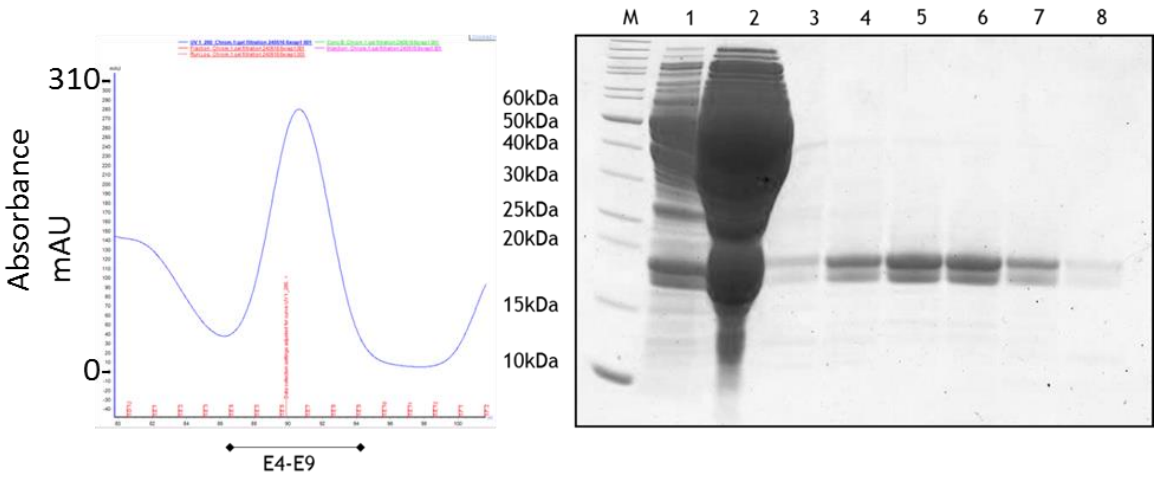


Figure 48: Gel filtration purification of Rap1: Part 2
***left**: gel filtration evaluation zoomed on the Rap1 peak between fractions E4-E9. **Right**: gel filtration fractions M. marker, 1. the pooled A10-A11 fractions from the 6Q rounds, 2. Fraction 1 after concentrating to 5 ml, 3. E4, 4. E5, 5. E6, 6. E7, 7. E8, 8. E9.*

Results 3.6: Rhea/Rap1 NMR Experiments

Having purified the Rap1 protein, Rhea F0-WT and F0-K17E, NMR was used to look at the binding of Rap1 to labelled Rhea. The NMR experiments showed that Rhea F0-WT was able to bind to the Rap1 protein (Figure 49, Figure 50, Figure 53, Figure 54), with significant shifts being observed. The F0-K17E protein however, was unable to bind the Rap1 protein, despite Rap1 being in high excess (Figure 51, Figure 52, Figure 53, Figure 54).

Having assigned the peaks in the HSQC spectra (section 3.4) it was possible to identify the residues that the peaks that shifted were. To do this chemical shift mapping was used. This is the process of colouring chemical shifts onto the surface of a protein structure, in order to determine where a binding site lies. By colour coding the mapping according to the size of the shift difference, the key residues involved in binding are highlighted on the structure. 5 residues had particularly large chemical shifts, which mapped to the end of the F0-WT domain (Figure 54, Figure 55, Figure 56), revealing the Rap1 binding surface on Rhea F0. These shifts were localized around the K17 location, with V15 and T18 being in close proximity in the amino acid sequence. Two of the other shifts, K37 and E40, are close to the K17 in tertiary structure. The final shift at Q9 is likely to be a

secondary consequence of the shift caused in T18, which is in close proximity to Q9 (Figure 55).

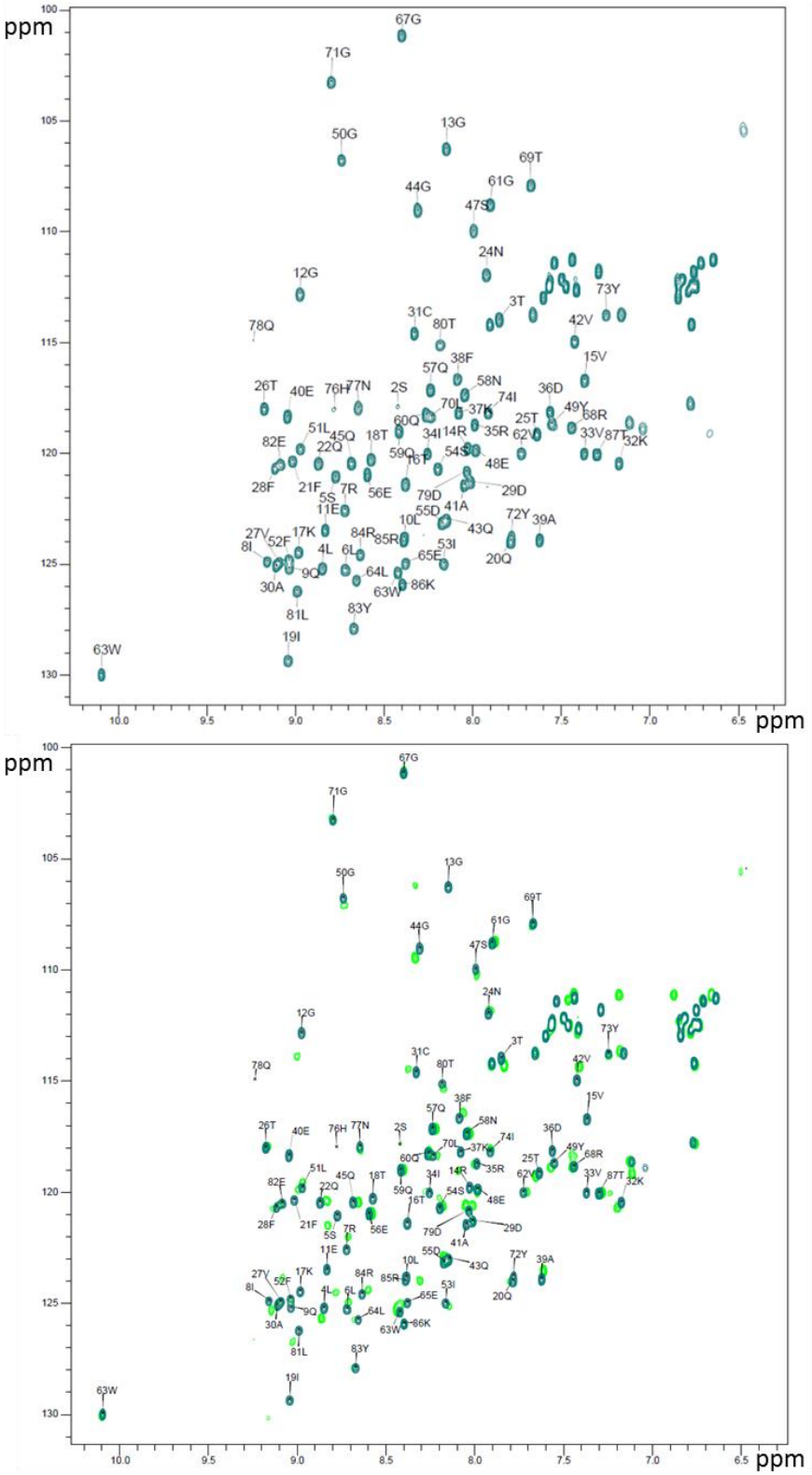


Figure 49: Rhea F0-WT chemical shifts induced by Rap1
Top, the spectra of wild-type Rhea FO, Bottom, the addition of Rap1 induces chemical shifts (light green overlay). Significant shifts are observed

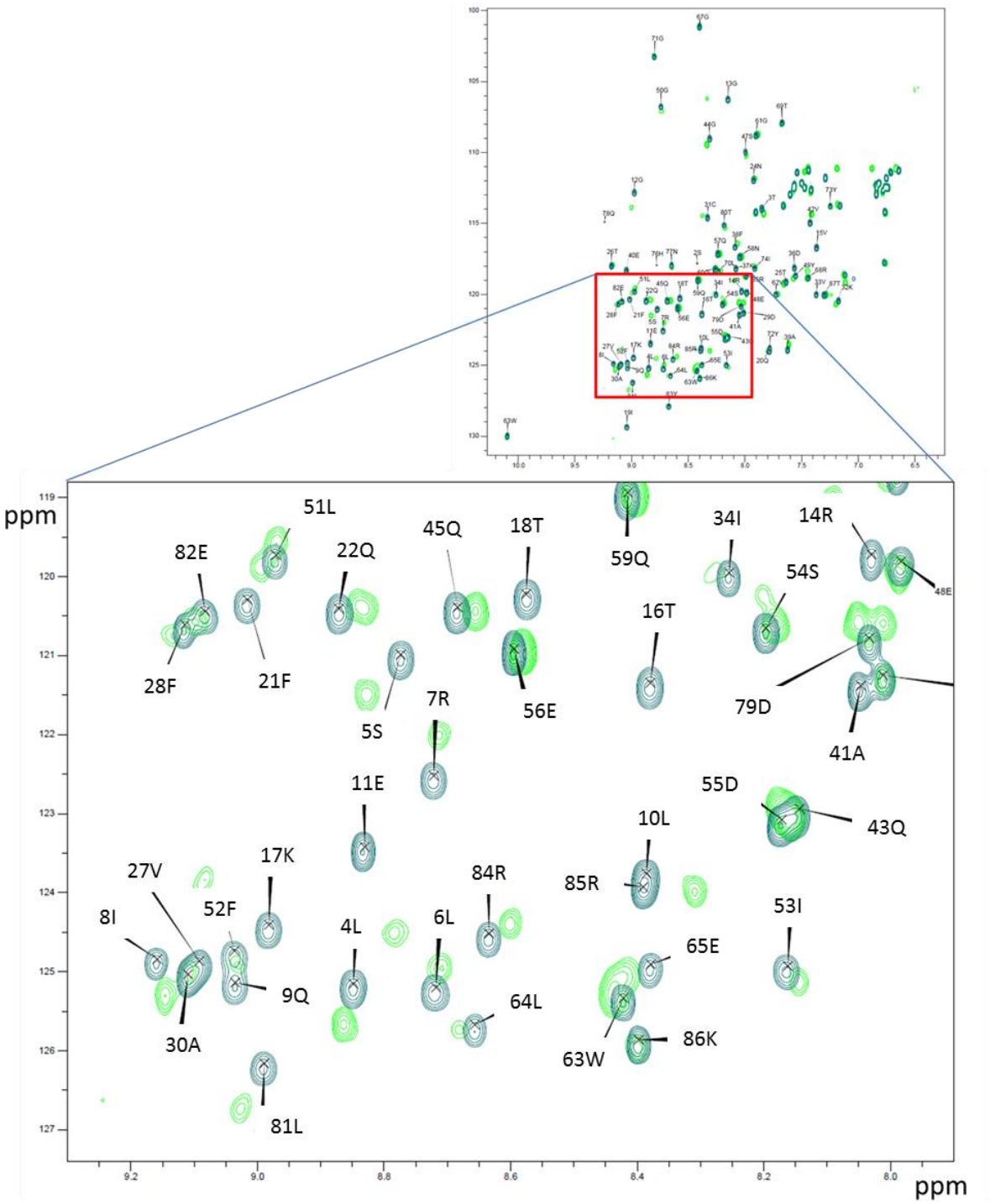


Figure 50: Rhea F0-WT chemical shifts induced by Rap1: Zoomed Section
A zoomed section of the Rhea F0-WT plus Rap1 spectra is shown to make the shifts more readily observed

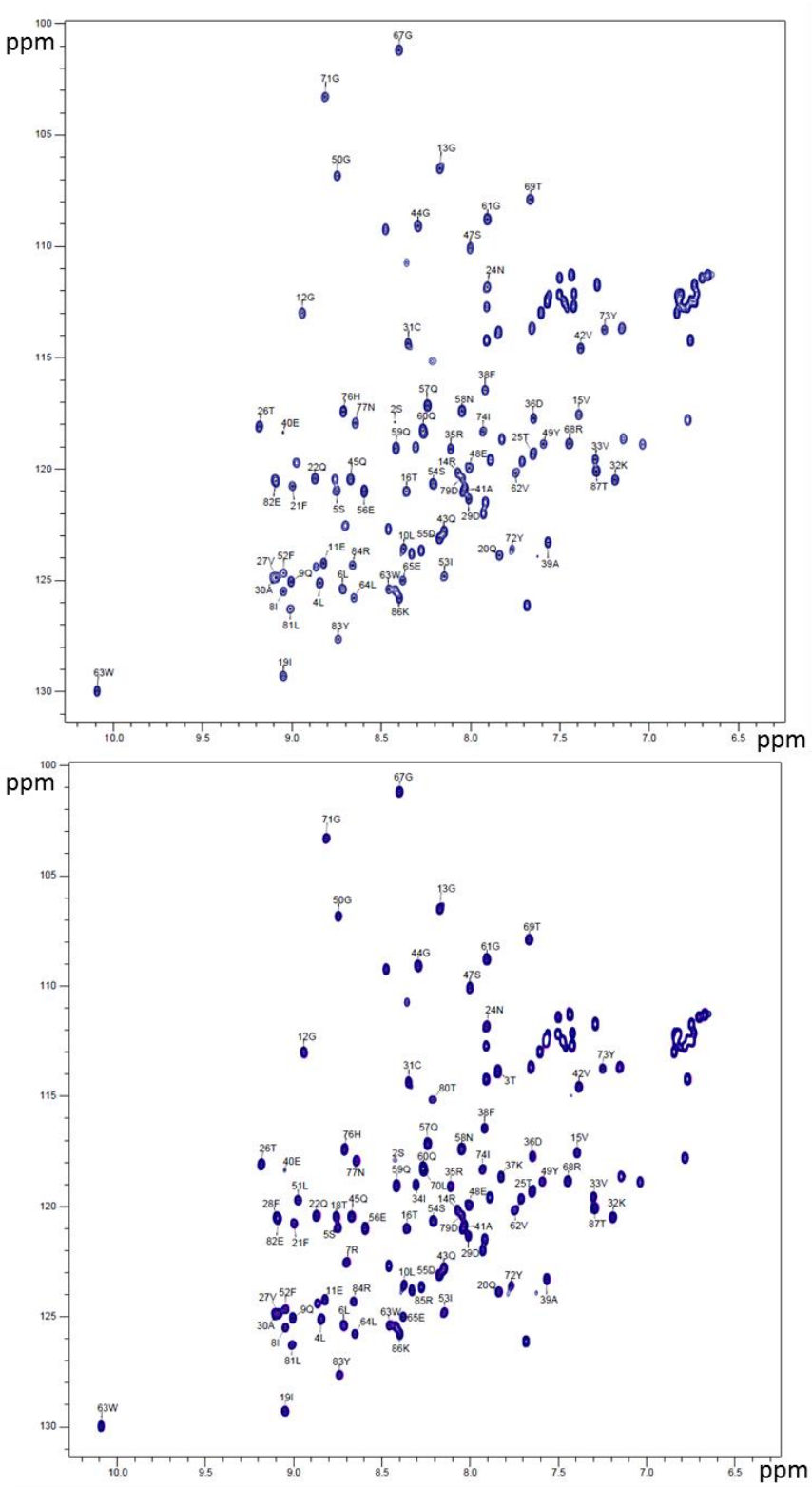


Figure 51: Rhea F0-K17E chemical shifts induced by Rap1.
Top, the K17E spectra and Bottom, the K17E spectra on its own (blue) and in the presence of a 4-fold excess of Rap1 (pink) The Rap1 does not induce significant shifts

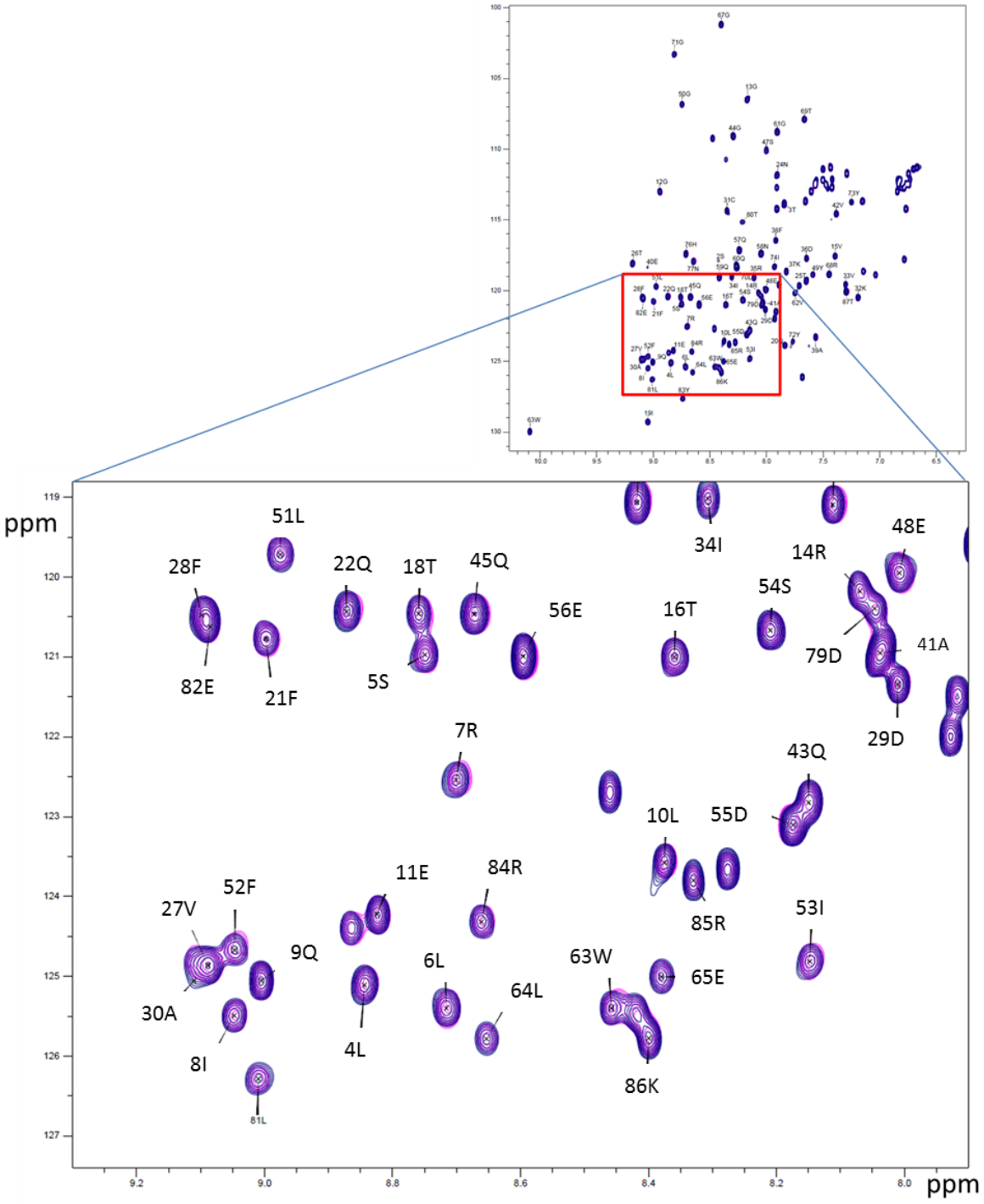


Figure 52: Rhea F0-K17E chemical shifts induced by Rap1: Zoomed Section
A zoomed section of the Rhea F0-K17E plus Rap1 spectra is shown to make the lack of chemical shift changes upon addition of Rap1 more readily observed

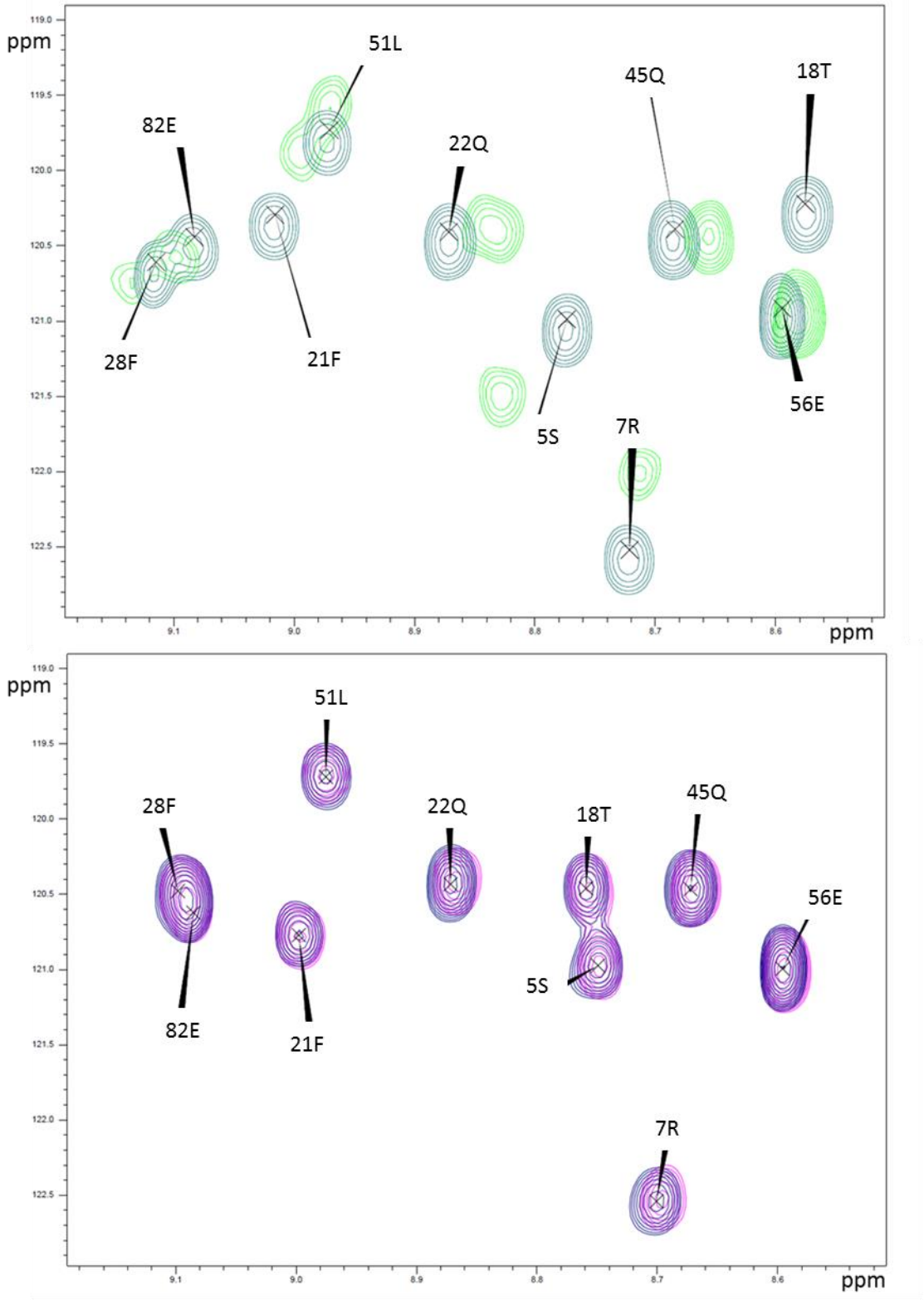


Figure 53: Comparison of the Rhea F0-WT and F0-K17E shifts induced by Rap1
Further zoomed sections of the Rhea F0-WT and F0-K17E spectra, with Rap1, to compare the shifts. Shift partners were determined using the "nearest neighbour" rule. The assignment of the K17E spectra was achieved using a nearest neighbour approach, for the majority of peaks assignments could be made with certainty, however, those which had shifted as a result of the mutation it was not possible to transfer the assignments with absolute certainty; we cannot be entirely sure which peaks correspond to their wild type equivalents. This doesn't matter in this case, as no shifts are observed.

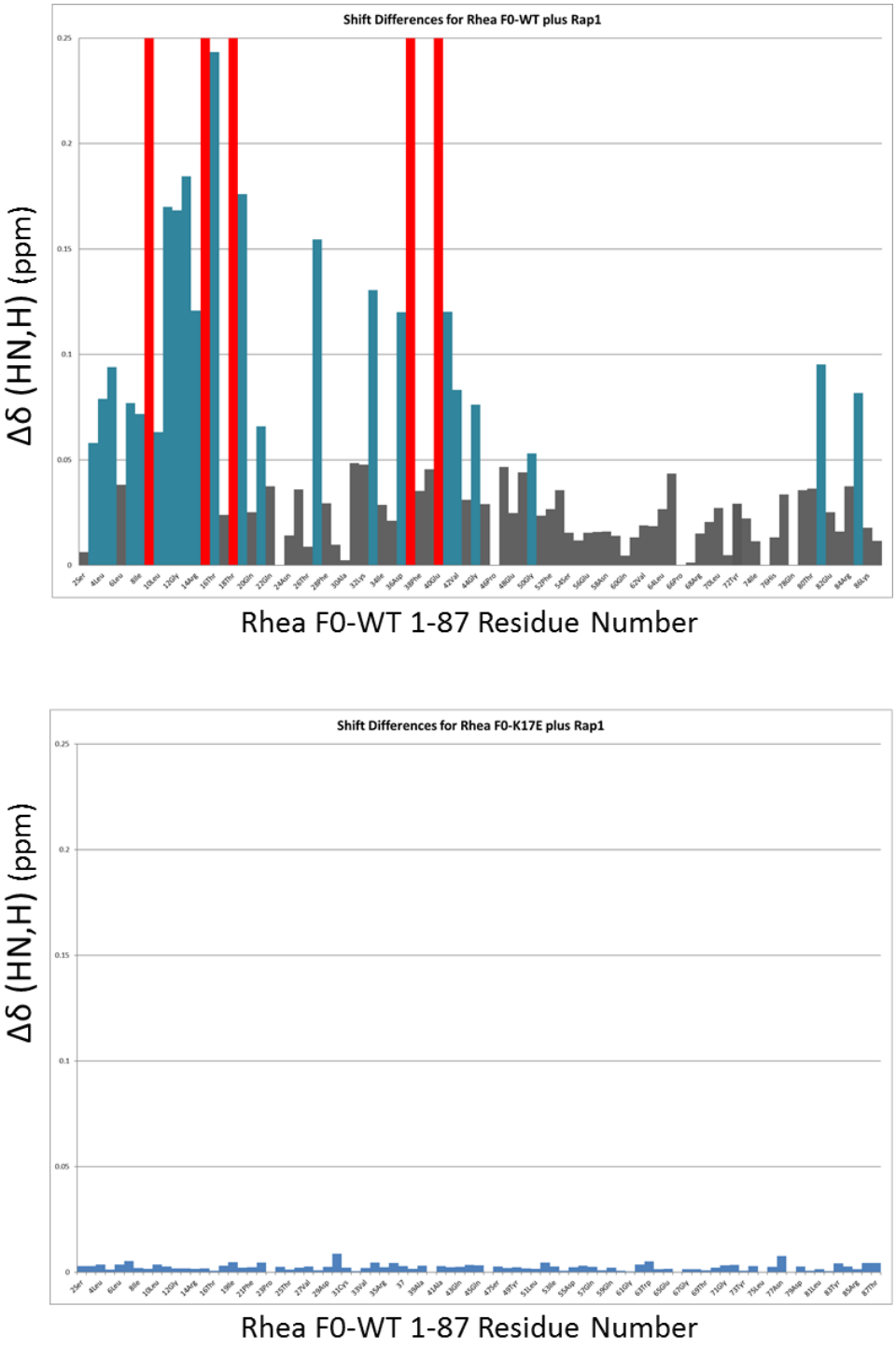


Figure 54: Chemical Shift Mapping of the Rhea F0-WT and F0-K17E shifts induced by Rap1 *Measuring the distances between the peaks of the Rhea alone, and in the presence of Rap1, allows us to quantify the shift differences. Top, the shift differences of wild-type are large on addition of Rap1. Shifts above 0.05 ppm were designated "medium shifts" and coloured in teal for the purposes of shift mapping. Peaks that disappeared from the spectra, were designated "large shifts" and coloured red. Bottom, the same analysis applied to the K17E mutant, minimal shift changes were observed confirming that there was no interaction with Rap1.*

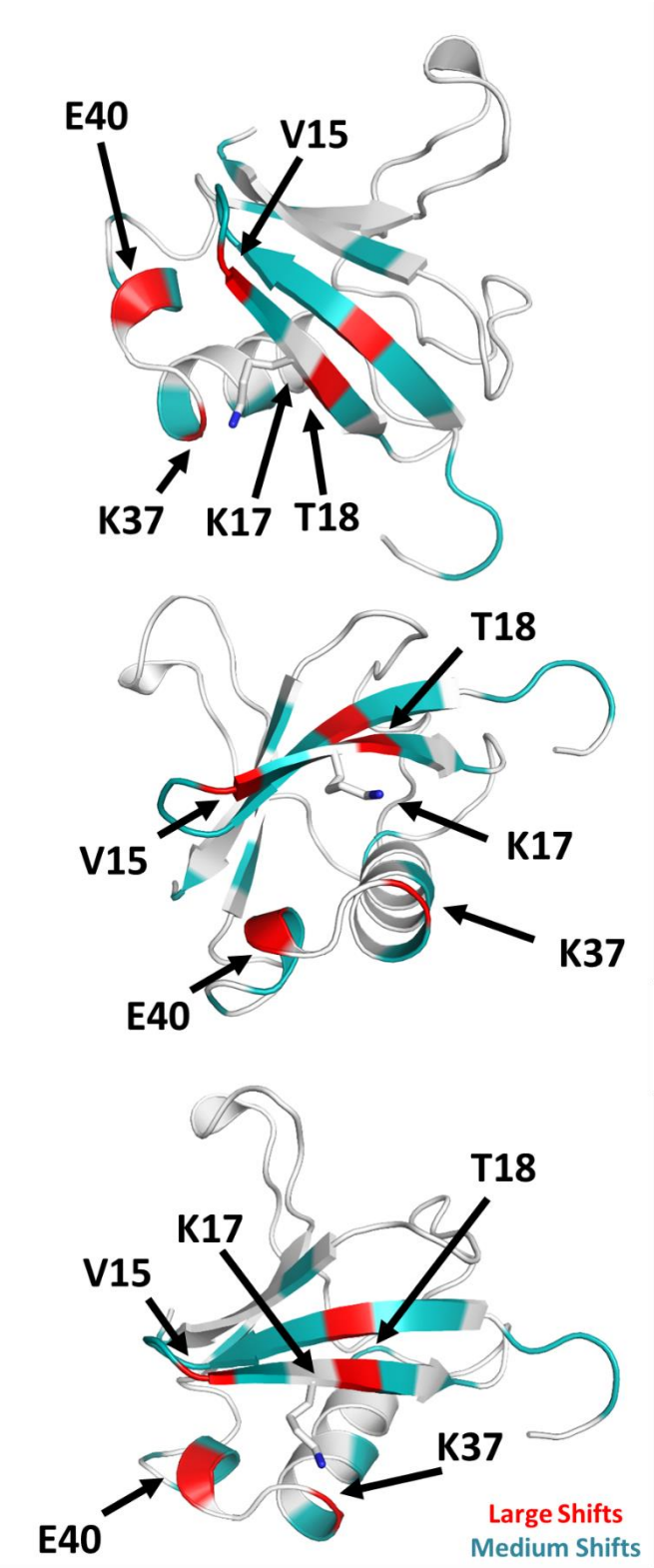


Figure 55: Chemical Shift mapping for Rhea F0-WT in the presence of Rap1
Using the quantitative data from **Figure 54**, *medium and large shifts were coloured onto the Rhea F0 structural model generated using PHYRE. The large shifts are labelled, as is the K17E residue. The shifts map to a distinct region of the Rhea F0 domain, indicating the Rap1 binding surface. This binding surface is similar to those seen with other RA domains interacting with GTPases. The large shifts are all close in tertiary structure, despite distance in the primary sequence (Kelley LA et al. 2015).*

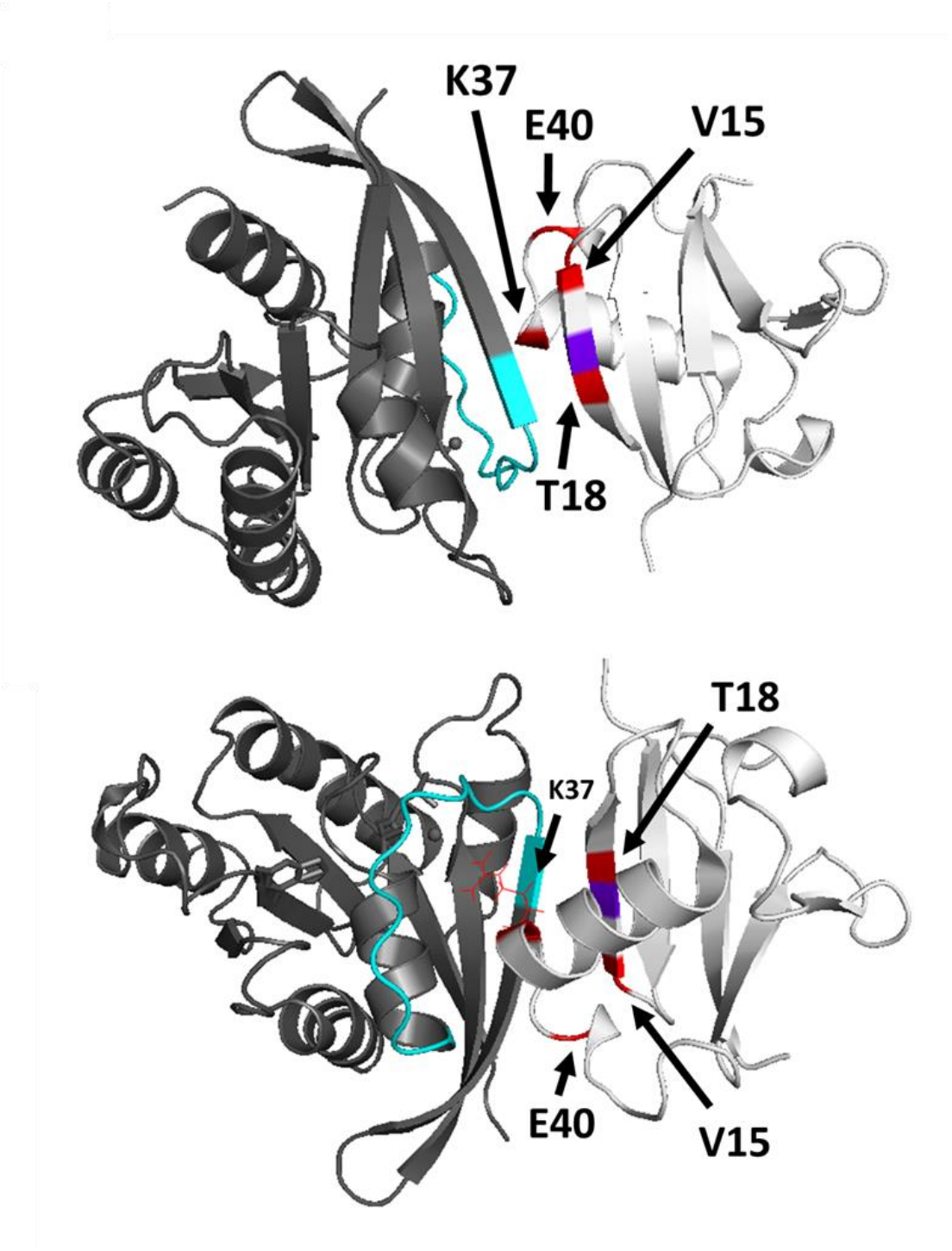


Figure 56: Model for Rap1:Talin F0 Binding
Model for the predicted binding of Rap1 (dark grey) to talin F0 domain (light grey). The residues that displayed the largest shifts in the Rhea F0 have been coloured onto the mouse talin F0. This was done by sequence alignment of Rhea and Talin. The Rap1 switch 1 domain has been coloured in cyan, and the F0 K17 in dark blue. The side chain structure for K37 has been shown due to its close proximity to the switch domain of Rap1. The drosophila shifts (Figure 55) are coloured onto this mouse model using sequence alignments between the mouse and drosophila talin F0 (Figure 58). The model above was created with FoldX by Dr Goult.

As a control, after running the NMR experiments, a gel was run with a fraction from each Shigemi NMR tube, to determine the concentration of Rap1 used in the experiments, as well as to ensure that the ratios of both Rhea and Rap1 were identical in each experiment. As shown in (Figure 57), the proteins were all added at identical ratios. Despite aiming for a 5x excess of Rap1 to Rhea, it was confirmed by ImageJ that the Rap1 was in fact present in a fourfold excess.

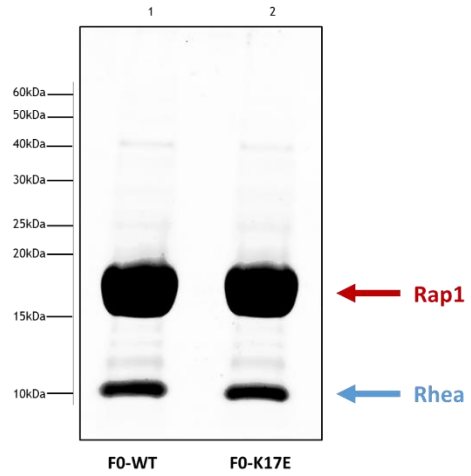


Figure 57: Gel of the NMR-Shift Experiments

Gel of the NMR experiments determining if Rhea F0-WT bound to Rap1 (lane 1), and if the Rhea F0-K17E bound to the Rap1 (lane 2). The Rap1 was added in ~4x excess to the Rhea as measured from the band density using ImageJ (Schneider, Rasband and Eliceiri, 2012).

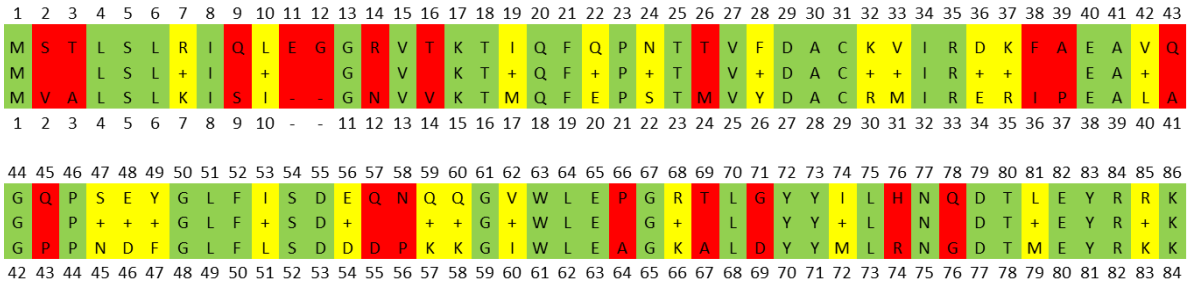


Figure 58: Sequence alignment of Fly Rhea F0 with Mouse Talin F0

The Fly Rhea F0 sequence is shown (top) with the mouse Talin F0 sequence aligned (bottom). Green residues are identical, red are different and yellow are similar amino acids (i.e. substitution of polar to polar, basic to basic etc.) (Altschul et al., 1990).

Section 4: Conclusions

4.1 Summary

In this study, a direct interaction between Rap1 and Rhea, the fly homolog of talin, has been demonstrated for the first time. Biochemical and biophysical characterisation have allowed us to map the Rap1 binding site on Rhea, and confirm that a K17E substitution, on Rhea, completely abolishes Rap1 binding. These findings have strong implications for RIAM-independent integrin activation, that could change our understanding of focal adhesion formation.

Over the course of this study, optimal conditions to express and purify Rhea F0-WT and the F0-K17E mutant have been elucidated. After optimisation, both proteins express well in E.coli BL21(DE3) cells and are efficiently purified by Ni-NTA based affinity chromatography. For both the wild-type and mutant protein, maximum yields of ~50-60 mg/litre were obtained. It was confirmed with circular dichroism that the full spectrum measurements, of the wild-type and mutant Rhea F0, were identical, as were their T_m values. This confirms that substitution of the K17 residue did not interfere with structural integrity, which was supported by the NMR data.

Optimal Rap1 protein was achieved using the mouse Rap1b 1-166 truncated construct, in CK600K cells, missing its hyper-variable C-terminal region. This region is highly conserved to that in fly (90% identical). It was demonstrated using NMR, for the first time, that the drosophila Rhea F0 domain is able to bind to Rap1b. By mutating a single basic lysine 17 residue, to an acidic glutamic acid residue, with a K17E mutant, the Rhea F0 protein is no longer able to bind to Rap1b, even when a four-fold excess of Rap1b is present. Triple resonance NMR experiments enabled the chemical shift changes upon Rap1b binding to be mapped and revealed the location of the Rap1b binding site on the wild-type Rhea F0. The largest chemical shift changes were observed for V15, K17, T18, K37 and E40, which are in the centre of this binding surface. To support these findings we generated Structural models of Rap1:Rhea F0. These structural models of Rap1:Rhea F0 binding agree with our findings, and these residues with the largest

shifts all are seen to make close contact with the Rap1, and confirms that Rhea interacts within close proximity of the Switch I domain of Rap1.

Together this work confirms that the fly talin, Rhea, can form a direct interaction with Rap1b and that a K17E mutant abrogates this interaction. The K17E mutant has been shown to cause lethality in flies and this work provides biochemical validation that this is likely due to disruption of the Rap1:Rhea binding.

4.2 Discussion

The biochemical data from this study will contribute towards a paper including work from our collaboration with Guy Tanentzapf, at the University of British Columbia, Vancouver, who has demonstrated that the Rhea K17E mutation is embryonic lethal in drosophila. These experiments are being performed using CRISPR insertion of the mutant into the endogenous talin gene. The observed phenotypes of these non-viable embryos, in addition to the embryonic stage at which death occurs, will allow further conclusions to be drawn about the nature of the mutation. Strengthening the evidence from the embryology data, will be experiments to attempt recovery of the Rhea mutant phenotype. If the K17E mutant can kill a fly embryo, then a rescue with the functional, wild-type protein will confirm that the observed phenotype is a direct consequence of the mutation.

Our structural work has proposed a biochemical explanation for the observed phenotypes: Rhea K17E is unable to bind to the Rap1 protein and this interaction could be key to embryonic viability. It is already known that drosophila Rap1 deletion mutants are unable to survive to adulthood, with migration defects observed during gastrulation (a time of extensive reorganisation in the early embryo, giving rise to the three germ layers) (Asha et al., 1999). It will be interesting to compare the phenotypes of Rap1 deletion mutants to those seen with the K17E mutants. Observing defects in cell migration, cell shape, or ovary and eye disk morphogenesis (Asha et al., 1999) will heavily implicate Rap1 involvement in the lethality resulting from the Rhea K17E mutation.

Rap1 evidently has other roles in the cell, including cell-cell junction formation (Knox, 2002), which will likely contribute to phenotypes observed in Rap1 deletion mutants. We would therefore expect that some of the Rap1 deletion phenotypes

would not be observed, if purely targeting the Rap1:Rhea pathway. What is currently unknown is how this pathway links to the classic pathway of integrin activation through Rap1 dependent recruitment of talin via RIAM (Shattil, Kim and Ginsberg, 2010., Goult et al., 2013). If the embryonic lethal phenotype is a result of interaction between Rap1 and Rhea, then the classic pathway is unable to recover or compensate for the lack of this interaction. This could lead to implications that the adhesion is double-dependent on Rap1, both through the classical pathway, whereby RIAM serves as an intermediary and also directly through the FERM domain. Further study will be required to fully understand the relative roles, and interplay between, these different pathways and at which stage each one is most important. One way to explore this, might be with carefully designed deletion and substitution mutants of talin, RIAM and Rap1 in embryos.

Interestingly, it is already shown that the F0 domain is required for integrin activation in a context dependent manner. Bouaouina et al (2008) show that talin F3 alone can activate β3 integrins, but fails to activate β1 integrins. This was despite the fact that the F3 domain bound both integrins with similar affinity. Their key finding, perhaps relating to Rap, is that larger fragments of the FERM domain, containing F0, are able to activate the beta1 integrins. Their proposal is that the F0 domain could be key for orientating the talin protein, by some as of yet unknown mechanism, in a way that allows integrin activation. This experiment shows us that there are key processes occurring that are poorly understood and more complex than our current understanding. The data in this report coupled with the findings of Bouaouina et al, allows us to propose a model whereby Rap1 is involved in orientating the talin molecule, in distinct integrin contexts. It would be interesting to design experiments to observe how the FERM domain interacts with integrin, following the K17E substitution. It could be that the F3 domain binds to specific beta-integrin tails, but requires Rap1 to anchor correctly to the membrane, to allow correct integrin activation (see Figure 59).

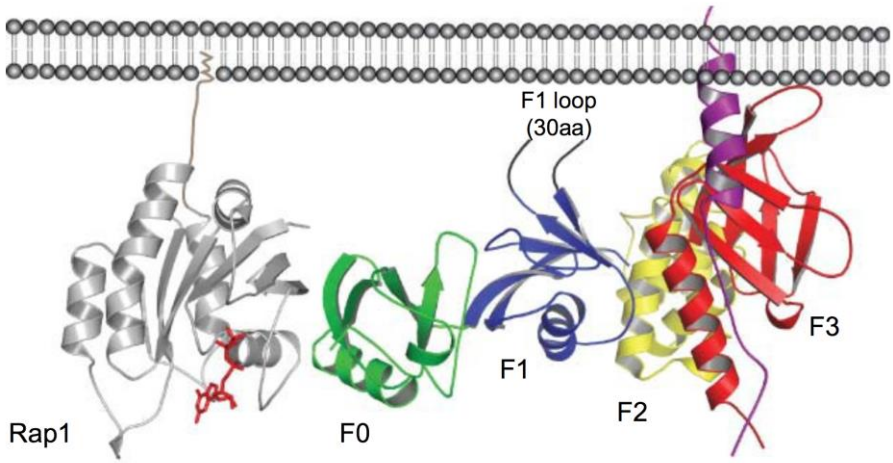


Figure 59: Rap1 may serve to anchor and orientate the FERM domain at the membrane, allowing integrin activation (unpublished figure, Goult 2016).

The possibility of Rap1 binding to talin was first proposed and demonstrated by Goult et al in 2010. Sequence alignments of the Rhea F0 domain were shown to have two distinct regions of conservation; on opposite faces of the F0 domain. The first region defines the F0-F1 interface, whereas the second was observed to be similar to the RalGDS RA (Ras-associating) site (Huang et al, 1998); RalGDS is a structural homolog of the talin F0 domain. One question that arose from the discovery of this talin RA site, related to the possibility of Rap1 binding being non-specific; i.e. could other GTPases bind to the site. Goult et al (2010) took one of the closest-related protein to Rap1 (H-Ras) and demonstrated that it fails to bind to the F0 domain. Interestingly however, the membrane targeting sequences of various Ras isoforms (H, K, R and N) have been able to increase integrin activation via RIAM (Lee et al., 2008). This suggests further roles for the GTPases in focal adhesion assembly and warrants further investigation; particularly when we consider the fact that a single amino acid substitution can abolish Rap1 binding, leaving the possibility of other GTPases binding to the Rhea RA site.

Another consideration comes from studies of RIAM function in platelets. A fascinating study (reviewed: Plow and Qin, 2015) has investigated the previous conceptions that RIAM is crucial for integrin activation in platelets. The findings reveal that RIAM-null mice, created through homologous recombination, exhibit normal integrin function and activation (Stritt et al., 2014). The authors did controls to test that RIAM was not present in the RIAM-deletion cells, and demonstrated that platelet aggregation (dependent on αIIbβ3 integrin) was identical in the wild-type and mutant cells. They also used antibodies, that

selectively bind to integrin in its active form, to show that levels of integrin activation were also identical in the RIAM deficient mice. The implications for this study – that platelets do not require RIAM for normal integrin function – are profound: in some cell types, the classic integrin activation pathway does not apply. This in turn means that integrin must be activated through other pathways, which adds weight to our argument that Rap1 could be involved in a similar process, through direct interaction with the Rhea FERM domain. It would be interesting to take this RIAM-null model, where integrin activation is unaffected, and design studies to further asses the role of Rap1 and Rhea in this context. It would be expected that a K17E version of talin in these models might result in complete loss of adhesion as both Rap1 dependent routes would be knocked out. It is my opinion that we may be able to best understand these complex protein interactions, when specific pathways can be selectively switched off, potentially making platelets a vital platform from which to study integrin activation. Platelets however, are difficult to generate and manipulate, arising from a complex megakaryocyte maturation process in vivo (Thon and Italiano, 2010). These findings do not rule out the possibility that there are other proteins able to perform the role of RIAM in platelets, so extensive study would be required. Additional interest comes from Pico- (the fly homolog of RIAM) deletion mutants, which have no strong phenotype in the embryo according to internal discussions with our collaborators. This again suggests RIAM-independent integrin activation. This is controversial however, with more sever Pico-deletion phenotypes being reported by some (Lyulcheva et al., 2008).

4.3 Further Work

There is much work to be done which directly continues from our current progress. Firstly, in collaboration with Marie Anderson in the Goult Lab, crystallography trials are underway to solve crystal structures of Rap1b:talin2/talin1 (see Figure 60).

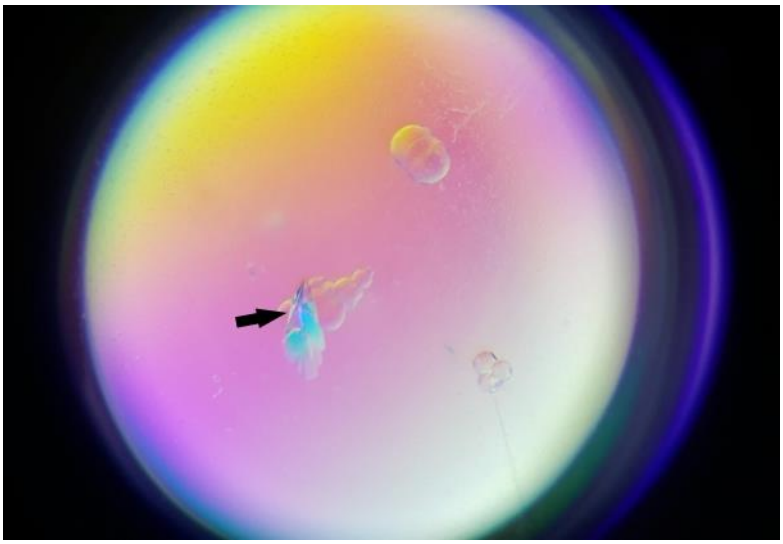


Figure 60: Crystallography is being carried out on Rap1b bound to Talin2
Crystals of Rap1b:Talin2 (arrow) have already been prepared in collaboration with Marie Anderson. These structures have yet to be solved, as of writing. Crystals can be seen growing in the phase separation.

Crystal structures of the Rhea FERM domain and fly Rap1 homolog would further our understanding of how these interactions are occurring. With more time, NMR titration experiments would have been performed to allow us to quantitate the binding of Rap1 and Rhea and calculate the binding constants. This would have been performed by repeating the NMR shift experiments, but using a range of increasing Rap1b concentrations.

Work also needs to be done on determining the cause of the double band of Rap1b observed during purification (see Figure 48). An initial thought was that translational errors or protein cleavage could be forming two versions of Rap1. The Mass Spectroscopy data however, encompassed fragments from across the same regions of the Rap protein in each band. Unless the cleavage is occurring at the N terminus, it can be concluded that this is not occurring. More likely, is the possibility that the distinct bands represent the nucleotide-bound state of the protein. It is possible that one band could represent the GTP/GDP bound state, and the other the apo form. Alternatively, one band could represent the GTP bound state, and the other the GDP bound state. Based on the fact that a source of external GTP was not supplied to the E.coli, it seems more likely that the Rap1 protein pool fails to reach saturation in the cell, and therefore we have some protein present in the apo form. Additionally, the Rap1 has intrinsic GTPase

activity, which could have converted all GTP to GDP by the time the protein was purified. This leads to an extension of work where the intrinsic rate of GTP hydrolysis is studied in Rap1. This could be extended further by measuring how these rates change in the presence or absence of different GAPs and GEFs. An important optimisation to the experiments would be to purify the protein and strip out any GTP/GDP with EDTA and then supply an external source to saturation. To overcome the intrinsic hydrolysis, a non-hydrolysable GTP analogue (such as GNP-pnp) could be used in subsequent experiments. The process of nucleotide exchange has been documented previously (Gorzalczany et al., 2000). This would likely resolve the double-band issue, and would allow us to more accurately assay the differences, between the active and inactive forms of Rap1, in Rhea binding.

To strengthen our data, we could perform experiments that show that the Rhea F0-K17E mutation does not affect integrin binding. This would involve expressing the entire FERM domain and measuring its binding constant for integrin, and then repeating this with the K17E mutation present. My current theory however, is that the F0 domain is far away enough from the F3 domain, and FERM membrane binding residues, to have little effect on integrin binding. Additionally, experiments performed by Dr Goult have shown that integrin binding can still occur when labelling different FERM domains with (GST); a large 25 kDa affinity tag.

Another series of experiments that could be performed, are substitutions of different amino acids on Rhea F0. Then we could assess the ability of the protein to bind to Rap1 again. Interestingly, K37 on Rhea is in close proximity with the key switch I domain of Rap1 (see Figure 56, Figure 62A); it can therefore be predicted that a substitution of this amino acid would drastically affect the ability of Rap1 to bind. Interestingly, K37 is one of the key residues that displayed a large shift on the Rhea F0-WT when in the presence of Rap1. Dr Goult has already performed this experiment with the mouse talin F0, by substituting R35 (the conserved K37 residue in Rhea (see Figure 58) for glutamate, and shown that this abolishes Rap1 binding (Unpublished results). We could also approach this experiment from the opposite angle and perform triple resonance experiments on Rap1, allowing us to map the Rhea F0-WT induced chemical shifts. This would allow us to identify potential targets on the Rap1, to mutate, that would prevent Rhea binding. A limitation with this experiment however, would be that Rap1 likely has a larger

number of binding partners and downstream effectors, making embryological studies more difficult to interpret.

Further cell biology and embryology experiments could investigate subcellular localisation of Rhea and Rap1, and particular cell types, to explore if this pathway can be specifically isolated from the RIAM pathway. Likewise, if mutations can be identified that disrupt the RIAM binding site on talin they could be used to tease apart these two different pathways. This could conclude if these interactions are occurring in the membrane, at focal adhesions, or in the cytosol.

Ideally, there is a desire to complete the optimisation of the true drosophila Ras-like protein 3 (the Rap1b homolog), which should prove straight forwards following the optimisation of the mouse construct. However, sequence alignments of residues 1-166 of mouse Rap1b with the corresponding region of drosophila Ras-like protein 3, show 90% identity between the proteins, with most changes being semi-conserved (see Figure 61). Additionally, the main differences between the two proteins are not at the binding interface (see Figure 62).

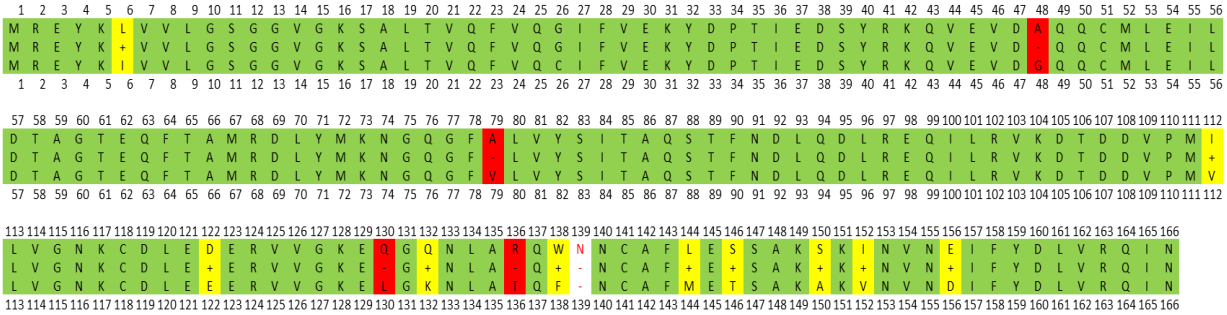
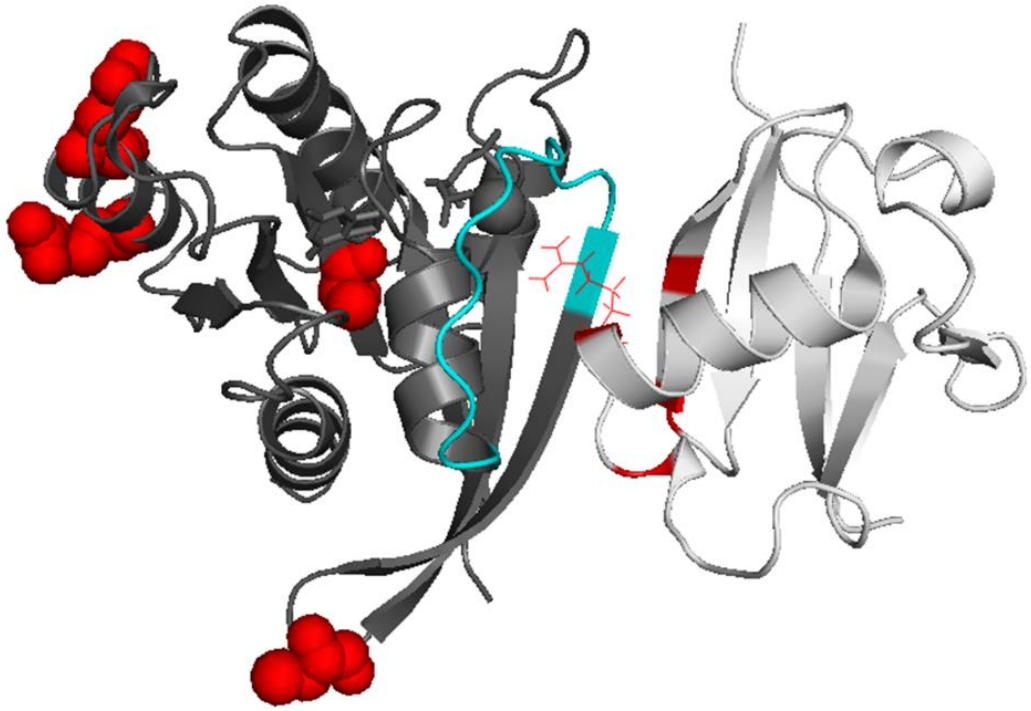
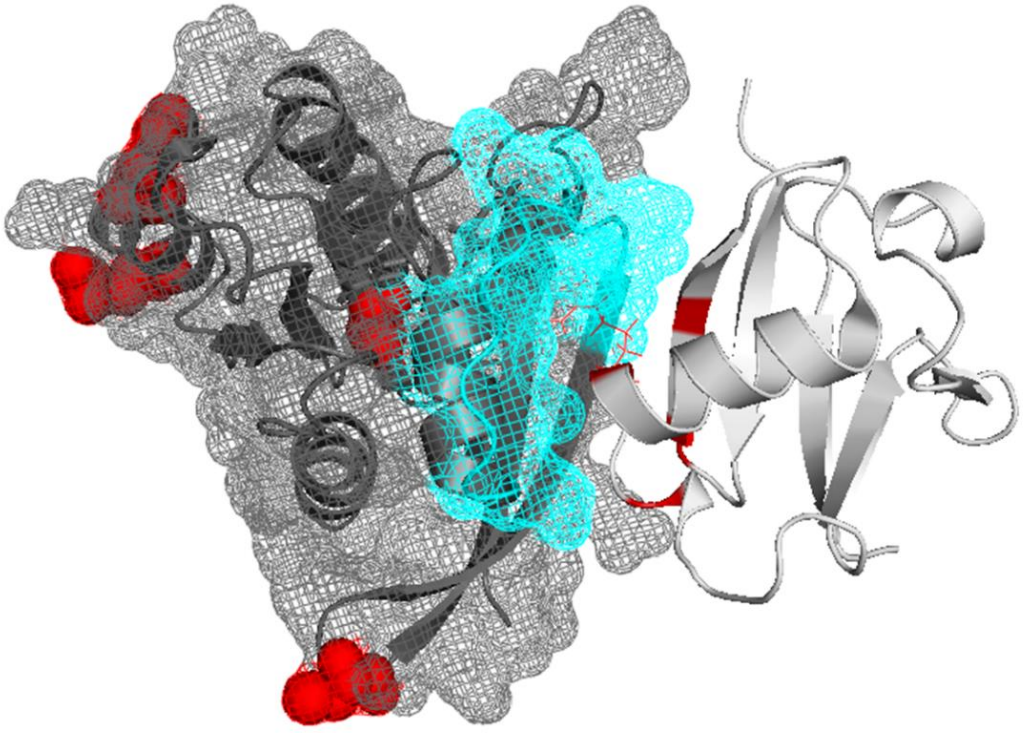


Figure 61: Mouse and Drosophila Rap1 are highly conserved

The mouse Rap1b sequence is shown (top) with the drosophila Ras-like protein 3 sequence aligned (bottom). Green residues are identical, red are different and yellow are semi-conserved amino acids (i.e. substitution of polar to polar, basic to basic etc.) (Altschul et al., 1990).



A



B

Figure 62: Differences between mouse Rap1b and drosophila Ras-like Protein 3 are not near the binding interface for Rhea F0

A) Rap1 is coloured in dark grey with the structural differences between it and Ras-like Protein 3 being shown with red spheres. The switch I domain is shown in light blue. Rhea F0 is shown in light grey.

B) For clarity, the mesh surface of Rap1 is shown, to illustrate the spatial separation of Rap1 and Ras-like protein 3 differences.

Another potential area for study, is the comparison between the binding of Rap1b to the F0 domains of Talin1, Talin2 and Rhea Isoform B. Interestingly, the mouse talin1 F0 domain comprises 85 amino acids. Mouse talin2 and Rhea Isoform B however, each have 87 amino acids. The F0 domains of mouse talin1 and mouse talin2 share 76% identity. The F0 domains of mouse talin1 and Rhea Isoform B share 51% identity, compared to the 56% identity between talin2 and Rhea. These two additional amino acids (at position 10-11), present (but not conserved) in both talin2 and Rhea Isoform B, could have a profound effect on the binding affinity of these two proteins, to Rap1b, when compared to talin1. This is a question that the lab has begun to answer: could the interaction between Rap1 and talin2 have tighter affinity and thus could a direct interaction with Rap1 play a larger role in talin2 function?

Our current studies have created a robust pipeline for the expression and purification of Rap1 and further work needs to be done to extend this progress to other GTPases. Work began on expressing Rab25, a more distant relative of Rap1, which belongs to the Rab sub-family of Ras GTPases. The intention was to study the Rab25 in the context of its ability to bind to the beta1 cytoplasmic tails of alpha5beta1 integrins, in order to understand its role in integrin recycling pathways. This interaction was first discovered by Caswell et al in 2007. Due to time constraints, this study remain unfinished, but the data obtained was very promising; it was found that full length His- and GST-tagged Rab25 constructs (with pet15b and pGEX4TI expression systems respectively), were largely insoluble in BL21(DE3) E.coli cells. This is in spite of high levels of overexpression (Figure 63). The Rab25-GST construct had a marginal soluble fraction, with which some preliminary pulldown experiments were performed with beta1 integrin (not shown). These results appeared to confirm Caswell's findings that Rab25 and beta1 integrin bind, but due to time constraints were not confirmed with appropriate controls.

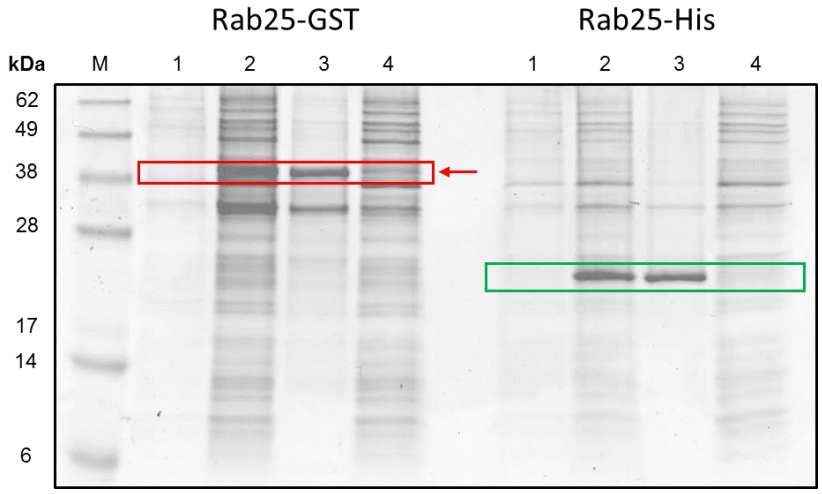


Figure 63: Expression of Rab25 was mostly insoluble

M, Marker, 1. Pre-Induction, 2. Post-induction Whole Cell, 3. Insoluble Fraction, 4. Soluble fraction. Rab25-GST location is shown in red, a small amount of soluble protein is indicated by the red arrow. Rab25-His location is shown in green.

By extending our findings from Rap1, the next step in this process would be to use an untagged construct of Rab25, missing the hypervariable C-terminal domain, in the CK600K cell line with similar induction conditions and expression systems.

To conclude, many aspects of focal adhesion complex formation, and integrin activation, are poorly understood, despite significant progress in these areas of research. It is likely that there are many new interactions that are yet to be documented, that could completely change our understanding of these processes. The currently described pathways also need much further study to truly understand, and allow us to apply this knowledge for therapeutic gain. With cell migration being so key to development and invasive cancer, this is an area of critical importance for research. The findings of our study shows that this novel Rap1b:Rhea interaction is evidently essential for adhesion, the implications of which are important enough to warrant considerable investigation.

Bibliography

Albert Einstein College of Medicine, (2012). *1H NMR spectrum of a protein*. [image] Available at: http://www.bioc.aecom.yu.edu/labs/girvlab/nmr/course/COURSE_2012/Lecture_BasicPrinciples_20120111.pdf [Accessed 1 Aug. 2016].

Alberts, B. (2002). Molecular biology of the cell. New York: Garland Science.

Allen, G. and Cooper, A. (1999). *Thermodynamics of Protein Folding and Stability, Volume 2*. Burlington: Elsevier.

Altschul, S., Gish, W., Miller, W., Myers, E. and Lipman, D. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), pp.403-410.

Anthis, N., Wegener, K., Ye, F., Kim, C., Goult, B., Lowe, E., Vakonakis, I., Bate, N., Critchley, D., Ginsberg, M. and Campbell, I. (2009). *The structure of an integrin/talin complex reveals the basis of inside-out signal transduction. EMBO J*, 28(22), pp.3623-3632.

Asha, H., Wang, M., Hariharan, I. and de Ruiter, N. (1999). *The Rap1 GTPase functions as a regulator of morphogenesis in vivo. The EMBO Journal*, 18(3), pp.605-615.

Aszodi, A., Legate, KR., Nakchbandi, I., Fassler, R. (2006) *What mouse mutants teach us about extracellular matrix function*. Annu Rev Cell Dev Biol 22: 591–621

Bailey, C. (2009) *The Roles of Rap1 in Cancer Metastasis and Pancreatic Islet Beta Cell Function*. PhD Thesis. Department of Pharmacology & Cancer Biology Duke University

Bouaouina, M., Lad, Y. and Calderwood, D. (2008). The N-terminal Domains of Talin Cooperate with the Phosphotyrosine Binding-like Domain to Activate 1 and 3 Integrins. *Journal of Biological Chemistry*, 283(10), pp.6118-6125.

Birnbaumer, L. (2007). The discovery of signal transduction by G proteins. A personal account and an overview of the initial findings and contributions that led to our present understanding. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1768(4), pp.756-771.

Bodenhausen, G. and Ruben, D. (1980). Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chemical Physics Letters*, 69(1), pp.185-189.

Bos, J. (2005). Linking Rap to cell adhesion. *Current Opinion in Cell Biology*, 17(2), pp.123-128.

Calderwood, D., Campbell, I. and Critchley, D. (2013). *Talins and kindlins: partners in integrin-mediated adhesion. Nature Reviews Molecular Cell Biology*, 14(8), pp.503-517.

Campbell, I. and Humphries, M. (2011). *Integrin Structure, Activation, and Interactions. Cold Spring Harbor Perspectives in Biology*, 3(3), pp. 4994

Caswell, P., Spence, H., Parsons, M., White, D., Clark, K., Cheng, K., Mills, G., Humphries, M., Messent, A., Anderson, K., McCaffrey, M., Ozanne, B. and Norman, J. (2007). Rab25 Associates with α5β1 Integrin to Promote Invasive Migration in 3D Microenvironments. *Developmental Cell*, 13(4), pp.496-510.

Chen, JM., Monaco, R., Manolatos, S., Brandt-Rauf, PW., Friedman, FK., Pincus MR. (1997). Molecular Dynamics on Complexes of ras-p21 and Its Inhibitor

Protein, rap-1A, Bound to the ras-Binding Domain of the raf-p74 Protein: Identification of Effector Domains in the raf Protein. *J Protein Chem*. 16 (6), p619-29.

Cherfils, J. and Zeghouf, M. (2011). Chronicles of the GTPase switch. *Nature Chemical Biology*, 7(8), pp.493-495.

Cherfils, J. and Zeghouf, M. (2013). Regulation of Small GTPases by GEFs, GAPs, and GDIs. *Physiological Reviews*, 93(1), pp.269-309.

Chrzanowska-Wodnicka, M., White, G., Quilliam, L. and Whitehead, K. (2015). *Small GTPase Rap1 Is Essential for Mouse Development and Formation of Functional Vasculature*. *PLOS ONE*, 10(12), p.e0145689.

Edreira, M., Li, S., Hochbaum, D., Wong, S., Gorfe, A., Ribeiro-Neto, F., Woods, V. and Altschuler, D. (2009). Phosphorylation-induced Conformational Changes in Rap1b. *Journal of Biological Chemistry*, 284(40), pp.27480-27486.

Ellis, S., Goult, B., Fairchild, M., Harris, N., Long, J., Lobo, P., Czerniecki, S., Van Petegem, F., Schöck, F., Peifer, M. and Tanentzapf, G. (2013). Talin Autoinhibition Is Required for Morphogenesis. *Current Biology*, 23(18), pp.1825-1833.

Felder, S., Zhou, M., Hu, P., Ureña, J., Ullrich, A., Chaudhuri, M., White, M., Shoelson, S. and Schlessinger, J. (1993). SH2 domains exhibit high-affinity binding to tyrosine-phosphorylated peptides yet also exhibit rapid dissociation and exchange. *Molecular and Cellular Biology*, 13(3), pp.1449-1455.

Goldberg, J. (1998). *Structural Basis for Activation of ARF GTPase*. *Cell*, 95(2), pp.237-248.

Gorzalczany, Y., Sigal, N., Itan, M., Lotan, O. and Pick, E. (2000). Targeting of Rac1 to the Phagocyte Membrane Is Sufficient for the Induction of NADPH Oxidase Assembly. *Journal of Biological Chemistry*, 275(51), pp.40073-40081.

Goult, B., Bate, N., Anthis, N., Wegener, K., Gingras, A., Patel, B., Barsukov, I., Campbell, I., Roberts, G. and Critchley, D. (2009). *The Structure of an Interdomain Complex That Regulates Talin Activity*. *Journal of Biological Chemistry*, 284(22), pp.15097-15106.

Goult, B., Skinner, S., Fogh, R., Boucher, W., Stevens, T., Laue, E., and Vuister, G. (2015). Structure calculation, refinement and validation using CcpNmr Analysis. *Acta Cryst D Biol Crystallogr*, 71(1), pp.154-161.

Goult, B., Bouaouina, M., Elliott, P., Bate, N., Patel, B., Gingras, A., Grossmann, J., Roberts, G., Calderwood, D., Critchley, D. and Barsukov, I. (2010). *Structure of a double ubiquitin-like domain in the talin head: a role in integrin activation*. *EMBO J*, 29(6), pp.1069-1080.

Goult, B., Zacharchenko, T., Bate, N., Tsang, R., Hey, F., Gingras, A., Elliott, P., Roberts, G., Ballestrem, C., Critchley, D. and Barsukov, I. (2013). *RIAM and Vinculin Binding to Talin Are Mutually Exclusive and Regulate Adhesion Assembly and Turnover*. *Journal of Biological Chemistry*, 288(12), pp.8238-8249.

Greenfield, N. and Fasman, G. (1969). Computed circular dichroism spectra for the evaluation of protein conformation. *Biochemistry*, 8(10), pp.4108-4116.

Jaffe, A. and Hall, A. (2005). RHO GTPASES: Biochemistry and Biology. *Annual Review of Cell and Developmental Biology*, 21(1), pp.247-269.

Huang, L., Hofer, F., Martin, GS., Kim, SH. (1998) Structural basis for the interaction of Ras with RalGDS. Nat Struct Biol 5: pp422–426

Johnson, B. and Hecht, M. (1994). Recombinant Proteins Can Be Isolated from E. coli Cells by Repeated Cycles of Freezing and Thawing. *Bio/Technology*, 12(12), pp.1357-1360.

Karnoub, A. and Weinberg, R. (2008). Ras oncogenes: split personalities. *Nature Reviews Molecular Cell Biology*, 9(7), pp.517-531.

Kelley LA et al. The Phyre2 web portal for protein modelling, prediction and analysis. Nature Protocols 10, 845-858 (2015).

Kelly, S., Jess, T. and Price, N. (2005). How to study proteins by circular dichroism. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1751(2), pp.119-139.

Kim, E. and Choi, E. (2010). Pathological roles of MAPK signalling pathways in human diseases. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1802(4), pp.396-405.

Kitayama, H., Sugimoto, Y., Matsuzaki, T., Ikawa, Y. and Noda, M. (1989). A ras-related gene with transformation suppressor activity. *Cell*, 56(1), pp.77-84.

Knox, A. (2002). Rap1 GTPase Regulation of Adherens Junction Positioning and Cell Adhesion. *Science*, 295(5558), pp.1285-1288.

Kupzig, S., Deaconescu, D., Bouyoucef, D., Walker, S., Liu, Q., Polte, C., Daumke, O., Ishizaki, T., Lockyer, P., Wittinghofer, A. and Cullen, P. (2006).

GAP1 Family Members Constitute Bifunctional Ras and Rap GTPase-activating Proteins. *Journal of Biological Chemistry*, 281(15), pp.9891-9900.

Lee, H., Lim, C., Puzon-McLaughlin, W., Shattil, S. and Ginsberg, M. (2008). RIAM Activates Integrins by Linking Talin to Ras GTPase Membrane-targeting Sequences. *Journal of Biological Chemistry*, 284(8), pp.5119-5127.

Lyulcheva, E., Taylor, E., Michael, M., Vehlow, A., Tan, S., Fletcher, A., Krause, M. and Bennett, D. (2008). Drosophila Pico and Its Mammalian Ortholog Lamellipodin Activate Serum Response Factor and Promote Cell Proliferation. *Developmental Cell*, 15(5), pp.680-690.

Malumbres, M. and Barbacid, M. (2003). Timeline: RAS oncogenes: the first 30 years. *Nature Reviews Cancer*, 3(6), pp.459-465.

Milburn, M., Tong, L., deVos, A., Brunger, A., Yamaizumi, Z., Nishimura, S. and Kim, S. (1990). Molecular switch for signal transduction: structural differences between active and inactive forms of protooncogenic ras proteins. *Science*, 247(4945), pp.939-945.

Minikel, E. (2015). *Circular Dichroism reference spectra for poly-L-lysine*. [image] Available at: http://www.cureffi.org/2015/12/27/circular-dichroism-on-recombinant-prp/ [Accessed 28 Jul. 2016].

Nerz-Stormes, M. (2016). *The Basics Nuclear Magnetic Resonance Spectroscopy*. [online] Brynmawr.edu. Available at: http://www.brynmawr.edu/chemistry/Chem/mnerzsto/The_Basics_Nuclear_Magnetic_Resonance%20_Spectroscopy_2.htm [Accessed 6 Jul. 2016].

Pai, E., Kabsch, W., Krengel, U., Holmes, K., John, J. and Wittinghofer, A. (1989). Structure of the guanine-nucleotide-binding domain of the Ha-ras oncogene product p21 in the triphosphate conformation. *Nature*, 341(6239), pp.209-214.

Pannekoek, W-J. and Bos, J. (2014). Rap Signalling. In: Wittinghofer, A *Ras Superfamily Small G Proteins: Biology and Mechanisms 1: General Features, Signaling*. -: Springer. p233.

Pascal, S. (2008). *NMR primer*. Chichester: IM Publications.

Pizon, V., Chardin, P., Lerosey, I., Olofsson, B., Tavitian, A. (1988). Human cDNAs rap1 and rap2 homologous to the Drosophila gene Dras3 encode proteins closely related to ras in the 'effector' region. *Oncogene*. 3 (2), p201-4.

Plow, E. and Qin, J. (2015). The role of RIAM in platelets put to a test. *Blood*, 125(2), pp.207-208.

Prior, I., Lewis, P. and Mattos, C. (2012). A Comprehensive Survey of Ras Mutations in Cancer. *Cancer Research*, 72(10), pp.2457-2467.

Prior, I. and Hancock, J. (2012). Ras trafficking, localization and compartmentalized signaling. *Seminars in Cell & Developmental Biology*, 23(2), pp.145-153.

Rattle, H. (1995). *An NMR primer for life scientists*. Fareham, Hants: Partnership Press.

Rojas, A and Valencia A. (2014). Evolution of the Ras Superfamily of GTPases. In: Wegener, K., Partridge, A., Han, J., Pickford, A., Liddington, R., Ginsberg, M. and

Campbell, I. (2007). *Structural Basis of Integrin Activation by Talin*. *Cell*, 128(1), pp.171-182.

Shattil, S., Kim, C. and Ginsberg, M. (2010). *The final steps of integrin activation: the end game*. *Nature Reviews Molecular Cell Biology*, 11(4), pp.288-300.

Schneider, C., Rasband, W. and Eliceiri, K. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, 9(7), pp.671-675.

Stewart, M. (2007). Molecular mechanism of the nuclear protein import cycle. *Nature Reviews Molecular Cell Biology*, 8(3), pp.195-208.

Stritt, S., Wolf, K., Lorenz, V., Vogtle, T., Gupta, S., Bosl, M. and Nieswandt, B. (2014). Rap1-GTP-interacting adaptor molecule (RIAM) is dispensable for platelet integrin activation and function in mice. *Blood*, 125(2), pp.219-222.

Thon, J. and Italiano, J. (2010). Platelet Formation. *Seminars in Hematology*, 47(3), pp.220-226.

Tsygankova, O., Wang, H. and Meinkoth, J. (2013). *Tumor Cell Migration and Invasion Are Enhanced by Depletion of Rap1 GTPase-activating Protein (Rap1GAP)*. *Journal of Biological Chemistry*, 288(34), pp.24636-24646.

Tucker, J., Sczakiel, G., Feuerstein, J., John, J., Goody, R. S., & Wittinghofer, A. (1986). Expression of p21 proteins in Escherichia coli and stereochemistry of the nucleotide-binding site. *The EMBO Journal*, *5*(6), pp.1351–1358.

Turcan, S. and Chan, T. (2013). MAPping the genomic landscape of low-grade pediatric gliomas. *Nature Genetics*, 45(8), pp.847-849.

van Dam, T., Bos, J. and Snel, B. (2011). Evolution of the Ras-like small GTPases and their regulators. *Small GTPases*, 2(1), pp.4-16.

Vetter, I. and Wittinghofer, A. (2001). The Guanine Nucleotide-Binding Switch in Three Dimensions. *Science*, 294(5545), pp.1299-1304.

Vetter, I. (2014). The structure of the G domain of the Ras Superfamily. In: Wittinghofer, A *Ras Superfamily Small G Proteins: Biology and Mechanisms 1*. London: Springer. p27.

Walker, F., Kato, A., Gonez, L., Hibbs, M., Pouliot, N., Levitzki, A. and Burgess, A. (1998). Activation of the Ras/Mitogen-Activated Protein Kinase Pathway by Kinase-Defective Epidermal Growth Factor Receptors Results in Cell Survival but Not Proliferation. *Molecular and Cellular Biology*, 18(12), pp.7192-7204.

Wegener, K., Partridge, A., Han, J., Pickford, A., Liddington, R., Ginsberg, M. and Campbell, I. (2007). *Structural Basis of Integrin Activation by Talin*. *Cell*, 128(1), pp.171-182.

Wennerberg, K. (2005). The Ras superfamily at a glance. *Journal of Cell Science*, 118(5), pp.843-846.

Wittinghofer, A *Ras Superfamily Small G Proteins: Biology and Mechanisms 1*. London: Springer. p3-24.

Ye, X., Shobe, J., Sharma, S., Marina, A. and Carew, T. (2008). *Small G proteins exhibit pattern sensitivity in MAPK activation during the induction of memory and synaptic facilitation in Aplysia*. *Proceedings of the National Academy of Sciences*, 105(51), pp.20511-20516.

Zhang, H., Chang, Y., Brennan, M. and Wu, J. (2013). *The structure of Rap1 in complex with RIAM reveals specificity determinants and recruitment*

mechanism. *Journal of Molecular Cell Biology*, 6(2), pp.128-139.

ZHANG, W. and LIU, H. (2002). MAPK signal pathways in the regulation of cell proliferation in mammalian cells. *Cell Res*, 12(1), pp.9-18.

Zerial, M. and McBride, H. (2001). Rab proteins as membrane organizers. *Nature Reviews Molecular Cell Biology*, 2(2), pp.107-117.

NMR Information Sources

1D NMR physics background, on the basic principles of NMR, is from a series of video tutorials produced by Magritek Ltd and delivered by Sir Paul Callaghan.

Main triple resonance NMR information comes from a series of lecture notes, for a graduate NMR course, entitled “Triple resonance experiments and structure determination”, from Mark Girvin and Sean Cahill at the Albert Einstein College of Medicine, NY, 2012 who cite the following authors: Cavanagh, J. (2007). *Protein NMR spectroscopy*. Amsterdam: Academic Press. Evans, J. (1995). *Biomolecular NMR spectroscopy*. Oxford: Oxford University Press.