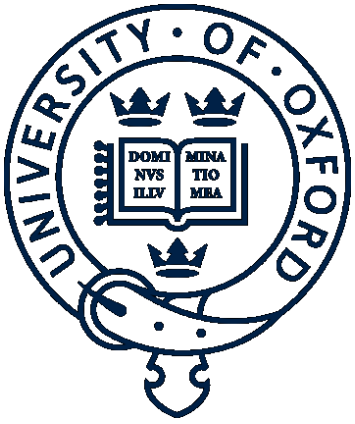
Structural Studies in Cell Adhesion and Division

Luke Alexander Yates

Nuffield Department of Clinical Medicine, Division of Structural Biology

Merton College

University of Oxford



A Thesis submitted in partial fulfilment of the requirements of *Doctor of Philosophy*

Hilary Term 2012

ABSTRACT

Cell adhesion is a critical process that allows the organisation and functioning of tissues in three-dimensions. However, the replenishing of cells, *via* cell division, within tissues is equally important for functioning complex life. Both cell adhesion and division are tightly controlled processes and rely on a complex network of signals that, as yet, are not wholly understood. This Thesis presents a structural analysis of several proteins involved in these processes.

In the case of cell adhesion, we have made use of high-throughput (HTP) cloning and expression screening technologies in the Oxford Protein Production Facility (OPPF) for the study of the Kindlin protein family – a recently discovered set of proteins essential for integrin-mediated cell adhesion. As a direct result of the HTP pipeline used we were able to determine the high resolution crystal structure of a single domain, the Pleckstrin Homology Domain, from the isoform Kindlin-1. Deletion of this domain in the full-length protein resulted in impaired integrin activation *in vivo*. This structure, in combination with molecular dynamics simulation demonstrated that, unlike other well characterised PH domains, the binding of secondary messenger lipids (phosphoinositides) is dictated by a, previously unseen, salt bridge that occludes the putative binding site. Mutation of the salt bridge alters the binding characteristics of this domain *in vitro*. In addition to the PH domain, we have also studied and biophysically characterised full-length Kindlin-3, a blood cell specific isoform. By optimising baculovirus-infected Sf9 cell expression systems we were able to obtain, for the first time, sufficient quantities of protein for characterisation. Furthermore, by using small-angle X-ray scattering (SAXS) in solution we were able to determine a low resolution solution structure of Kindlin-3, revealing a linear arrangement of its FERM domain - a novel conformation known only otherwise in talin. We characterised the interaction of full-length

Kindlin-3 with β-integrin cytoplasmic tails using nuclear magnetic resonance spectroscopy, which confirmed that a direct interaction with a membrane distal NPxY motif occurs, and demonstrated the importance of a preceding Serine/Threonine rich region in peptide binding.

In the case of cell division, we have determined the crystal structure of the cell cycle checkpoint control related protein, Cid1, a terminal uridine tranferase from *Schizzosaccharomyces pombe*, alone and in complex with UTP. Structural and biochemical analysis of Cid1 identified a novel Uridine selection mechanism that is suggested to be conserved in metazoan ZCCHC enzymes involved in *let*-7 miRNA biogenesis, which are important for proliferation, differentiation and cell fate. We have also demonstrated that Cid1 is an RNA binding protein, a property essential for activity that employs a novel mechanism of RNA binding in the absence of RNA binding motifs.

The structural work undertaken in this thesis has focussed on two distinct, but interwoven, aspects of cell biology and has significantly added to both fields of research. Excitingly, this has opened many new avenues of investigation and, in the case of Cid1, has the strong potential to lead to the development of novel anticancer therapies.

This Thesis is dedicated to my Wife Samantha, the most courageous person I know, and to God; because we are "fearfully and wonderfully made; your works are wonderful, I know that full well" (Psalm 139:14)

.

DECLARATION OF WORK

The work described herein is my own, except where collaborative projects have contributed experimental results that are described in this Thesis. These findings are explicitly stated prior to the chapter and in the text, where appropriate.

Luke Alexander Yates

Merton College, University of Oxford

ACKNOWLEDGEMENTS

I am very thankful for the support and guidance from my supervisor Dr. Robert Gilbert and for the opportunity to study such stimulating projects. Dr. Gilbert's confidence in my ability has been encouraging and made this a hugely rewarding endeavour. I am particularly thankful for the compassion and generosity that Robert has shown during the last few years in some very difficult circumstances. I am equally grateful for the compassion and support from the Division of Structural Biology and the Nuffield Department of Medicine.

I would like to take this opportunity to thank Professor Yvonne Jones and Professor Anne Willis for agreeing to the examination of this Thesis.

Many thanks go to the many collaborators for creating exciting research projects and colleagues for discerning discussions. I would like to thank; Craig Lumb and Prof. Mark Sansom for the molecular dynamic simulations; Dr. Anna Fuezery, Dr. Michele Erat, and Prof. Iain Campbell for the integrin NMR work and Kindlin-3 SAXS data acquisition; Dr. Sophie Fluerdepine, Dr. Olivia Rissland and Dr. Chris Norbury for the Cid1 expression plasmid and the extensive Cid1 functional work; Dr. Louise Bird, Dr. Ray Owens and the OPPF team for guidance in HTP cloning experiments; all of these collaborations have been inspiring and I am grateful to every member for their patience and expertise. I am thankful to all my friends and colleagues in STRUBI for an enjoyable work environment. I would like to thank, in particular, Dr. Erika Mancini and her group, along with Dr. Kay Grünewald and his group for thought-provoking discussions during group meetings, Dr. Tom Walter for protein crystallisation expertise and for general crystallisation advice, Dr. Karl Harlos for crystal handling and freezing, Dr. Luigi De Colibus, for crystallography training and advice, Dr. Geoff Sutton and Weixen Lui for tissue culture training and guidance, and various members of STRUBI for help with X-ray data collection at the Synchrotron. In particular I would like

to thank the Gilbert group members, past and present, for making me feel welcome and pointing me in the right direction when needed.

I would like to thank Dr. Mathew Coleman for his friendship and sharing the burden of commuting and for stimulating discussions during our time in the car together. I also thank Dr. Christina Mayer and Phil West for their support during this shared endeavour. I am also grateful to James Newton and his wife, Dr. Ruth Newton, for their friendship.

I am very grateful to my parents, Darren and Shuk, my brothers, Matthew and Andrew and family-in-law, Alan, Ros, and Ian, for their encouragement, and continual love and support that has ultimately allowed me to pursue my doctoral research.

Finally a very special thank you to my Wife Samantha for her unwavering love, her constant belief in my ability and for her continuous support - despite facing great suffering. I dedicate this thesis to Samantha.

TABLE OF CONTENTS

ABBREVIATIONS

List of abbreviations used in this thesis

ATP	Adenosine Triphosphate
AUC	Analytical ultracentrifugation
CT	Cytoplasmic tails
DLS	Dynamic light scattering
DNA	Deoxyribonucleic acid
E. coli	*Escherichia coli*
GST	Glutathione S-transferase
h	Hour
HBS	HEPES buffered saline
HEK	Human embryonic kidney
IPTG	Isopropyl-β-D-thiogalatopyranoside
kDa	Kilodalton
LB Broth	Luria Bertani Broth
M	Molar (Moles/Litre)
mg	Milligram
min	Minutes
ml	Millilitre
mM	Millimolar
MW	Molecular weight
MWCO	Molecular weight cut-off
nM	Nanomolar
nm	Nanometer

PBS	Phosphate buffered saline
PH Domain	Pleckstrin Homology Domain
PtdInsP	Phosphatidylinositol phosphate
PUP	Poly(U) polymerase
R_g	Radius of gyration
RNA	Ribonucleic acid
rpm	Revolutions per minute
SAXS	Small angle X-ray scattering
SDS-PAGE	Sodium dodecylsulphate – polyacrylamide gel electrophoresis
SEC	Size Exclusion Chromatography
Sf	*Spodoptera frugiperda*
TB	Terrific Broth
TBONEX	Terrific Broth Overnight Express
TUT	Terminal Uridylyl Transferase
UTP	Uridine Triphosphate

Standard single letter amino acid codes were used

CHAPTER 1

Introduction

Since its advent, structural biology has illuminated a vast number of biological processes in great detail; one of the most notable is the structure(s) of the ribosome, which has significantly advanced our understanding of translation (reviewed in Schmeing and Ramakrishnan, 2009). What is more, structural biology and in particular crystallography, has led to significant advances in medicine, with the most direct example being the structure-based inhibitor design of two successful and clinically used drugs zanimivir (Relenza) (von Itzstein *et al.,* 1993) and oseltamivir (Tamiflu) (Kim *et al.,* 1997) for the treatment of influenza infection. These drugs were designed as a result of the structure determination of neuraminidase in complex with its natural ligand – sialic acid (Varghese *et al.,* 1992). Today, there is a demand for fast and cost-effective protein structure determination together with in-depth functional characterisation, especially for rational structure-based drug design. As such, structural genomics (SG) initiatives were generated around the world in order to address this demand, focussing their attention on medically relevant targets. With this has come the development of high-throughput (HTP) molecular biology techniques and protein crystallisation methods, which have been enhanced by the generation of liquid handling

robotic technology. HTP technologies and strategies developed by structural biology laboratories (extensively reviewed by Gräslund *et al.,* 2008), such as the Oxford Protein Production Facility (OPPF), has resulted in the design of versatile, precision ligation-independent-cloning for high-throughput (HTP) pipelines, together with HTP expression screening methods (Berrow *et al.,* 2007), and nanolitre crystallisation screening techniques (Walter *et al.,* 2005) – that, despite its scale can still yield protein structures, for which many examples exist, including the structures determined in this Thesis (see Chapter 4, for example). Moreover, automated tracking, imaging and data storage of crystallisation screens over the course of the experiment's lifetime, developed and implemented in the OPPF crystallisation laboratory, permits easy timescale analysis of the experiments and likely crystallisation 'hits' (Mayo *et al.,* 2005). This has resulted in an improved scenario, whereby large numbers of protein targets can be cloned and expressed, and the most promising targets crystallised with minimal material. Furthermore, specialised synchrotron beamlines and X-ray detectors (e.g. Pilatus 6M detectors) have also contributed to structural biology by significantly reducing the time for crystallographic data collection, which has been further compounded by automated data processing pipelines (e.g. Xia2) (Winter, 2010).

This thesis is concerned with two fundamental aspects of cell biology, (1) integrin-mediated cell adhesion and (2) cell cycle control in cell division, and will use structural biology and its associated techniques to study proteins involved in these processes. Therefore, this introduction will be divided in half, in the first instance providing a brief overview of cell adhesion and the role of the integrins followed by a brief overview of cell division.

1.1. A Very Brief Introduction to Cell-Matrix Adhesion

Cell-to-cell and cell-extracellular matrix (ECM) adhesion represents a fundamental mechanism featured in all metazoans as they allow cells to attach to each other or an extracellular substratum, organising the cells into a three-dimensional and functional whole (Geiger and Yamada, 2011). This three-dimensional organisation has allowed the formation of tissues and organs but more fundamentally the emergence of multicellular life. In complex life the formation of cell adhesions together with their dynamic regulation is essential for embryogenesis, haemostasis, immune cell function, wound healing, cell proliferation and migration (Geiger and Yamada, 2011; Svennson *et al.,* 2009; Moser *et al.,* 2009; Montanez *et al.,* 2008; Ziegler *et al.,* 2008; Hynes, 2002), but its aberrant regulation plays a role in human disease, including; cancer invasiveness and metastasis, thrombosis and stroke, inflammation and infection (Geiger and Yamada, 2011; Moser *et al.,* 2009, Ziegler *et al.,* 2008). Cell adhesions not only allow the linking and organisation of cells together and to the surrounding matrix but also act as signalling conduits conveying accurate information about the cellular and local extracellular environment (Geiger and Yamada, 2011). Moreover, signalling at cell adhesions can influence the life of the cell, which can directly result in a number of cellular outcomes including cell migration, proliferation, differentiation and survival (Geiger and Yamada, 2011; Patla *et al.,* 2010). In addition to cells responding to the surrounding matrix, they can actively remodel the ECM resulting in a reciprocal, bi-directional cell-ECM interaction with the restructured matrix (Geiger and Yamada, 2011).

The extracellular matrix (ECM) serves as the substratum to which cells adhere through cell-ECM adhesions and is initially constructed and subsequently remodelled by those same adhesions (Alberts *et al.,* 2008). The molecular and structural diversity of the ECM and its role in regulating cell adhesion is highly complex and is described in detail in several

physiological instances, for example, in the basement membrane (see Yurchenco, 2011; for instance). Therefore, a complex description of the ECM and its influence in cell-ECM adhesion is summarised below and is neither highly detailed nor exhaustive but serves to acknowledge the many facets of cell-ECM adhesion.

The ECM is highly variable, depending on the type of matrix, and its composition influences the chemical environment surrounding cells, which plays an important role in the cell's response to the pericellular environment. Interestingly, specific ECM immobilised growth factors and morphogens (e.g. transforming growth factors, heparin binding epidermal growth factors or morphogens, such as sonic hedgehog) associate with the matrix and form microenvironments that stimulate adherent cells, suggesting that intracellular signalling from cell adhesion to the ECM occurs *via* the recruitment of proteins to the adhesion site or the direct stimulation of growth factor receptors (Geiger and Yamanda, 2011; Munger and Sheppard, 2011). Additionally, the so-called "dimensionality" of ECM constituents resulting in microenvironments or specific topographic arrangements/spacing of substrate directly affects the organisation, biological activity, and morphology of adherent cells upon it (see Geiger *et al.,* 2009, for instance). Furthermore, in addition to the chemical composition of the ECM and its apparent dimensionality a third key facet of the ECM must be considered – that is its mechanical properties (Geiger and Yamanda *et al.,* 2011; Geiger *et al.,* 2009). Several important mechanical features of ECM including; rigidity, porosity and cross-linking, and anisotropy, are major regulatory effectors of cell proliferation rate, stem cell fate, gene expression, cell migration and tissue shape and orientation (Geiger and Yamanda, 2011; Klein *et al.,* 2009). Therefore, the 'stiffness' of a highly organised matrix is primarily responsible for the complex cellular architecture of tissues and organs (Butcher *et al.,* 2009). This concept is simplified in cancer, for example, where the matrix that surrounds tumours often contains high concentrations of collagen resulting in a rigid and isotropic tissue – i.e. a

palpable tumour which is significantly 'stiffer' than normal tissue (Geiger and Yamanda, 2011; Butcher *et al.,* 2009).

In this introduction and the majority of the thesis concerned with cell adhesion, a major emphasis is placed on a class of proteins that mediate cell-matrix adhesion, namely the integrins and therefore integrin-mediated cell adhesion is discussed in depth (see below). However, it must be acknowledged that there are multiple adhesive interactions that are mediated by receptors other than integrins but these proteins and their role in cell adhesion are beyond the scope of this work and therefore will not be discussed further.

1.1.1. A Brief Overview of the Integrins

As discussed previously, cell adhesion and more specifically cell–matrix adhesion, is a critical process in a plethora of physiological phenomena, including; cell proliferation and migration, embryogenesis, tissue integrity, immunity, haemostasis and wound healing (Ziegler *et al.,* 2008; Svensson *et al.,* 2009). The major cell surface receptors that cells and tissues employ in cell-matrix adhesion, which recognise and utilise a functional extracellular matrix, are the integrins.

Integrins were first characterised as early as 1986 (Tamkun *et al.,* 1986) and are a family of heterodimeric transmembrane glycoproteins comprised of a non-covalently associated α and β subunit (Campbell and Humphries, 2011). The integrins serve as the bridging machinery that links the intracellular cytoskeleton and the extracellular matrix (ECM), and are so-called 'integrators' of the cytoskeleton and matrix, a function from which they are named (Giancotto and Ruoslahti, 1999). Integrins integrate the ECM and the cytoskeleton through focal adhesions (FA), which are integrin containing multiprotein complexes that allow the mechanical linking of the ECM and the termini of actin filaments

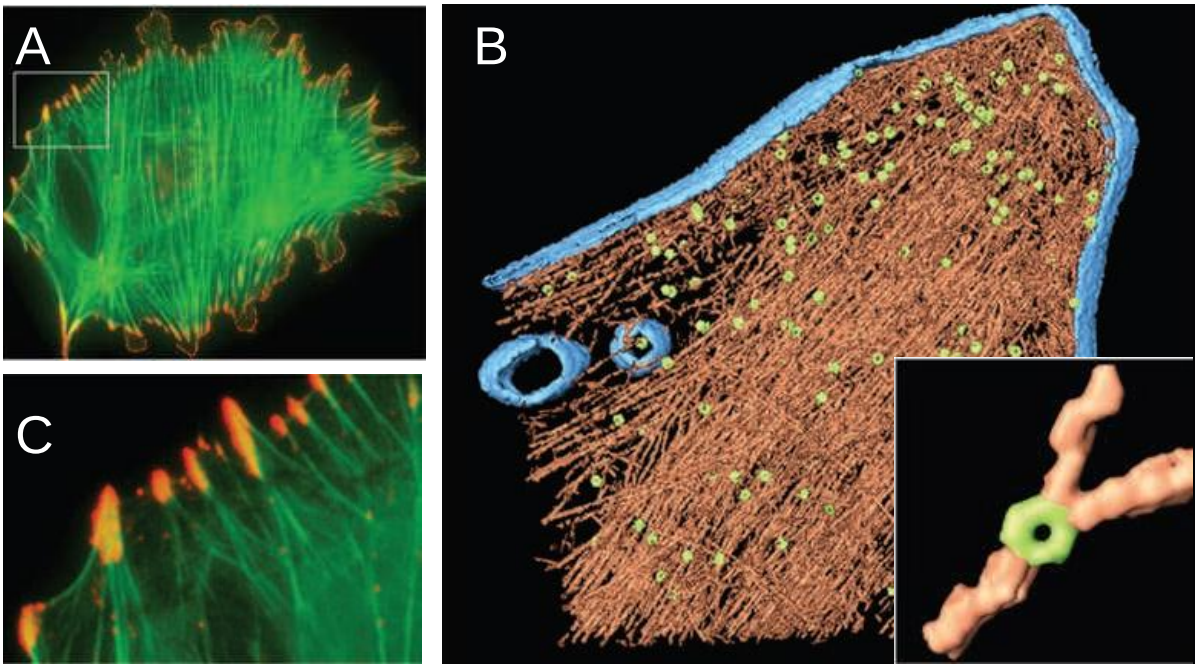


Figure 1.0 – Ultra-structure of the Integrin Focal Adhesion. (A) Fluorescence microscopy of focal adhesions and associated actin filaments shown in greater detail in (B), showing the filamentous actin network (green) and focal adhesions sites (red). (C) The architecture of focal adhesion sites as seen by cryo-electron tomography. Surface-rendering view of the focal adhesion site from the direction of the substrate toward the cell. Actin (brown) and membranes (blue) are shown along with a number of uniformly oriented particles that are adhesion related (green) and are located between the termini of the actin bundles and the membrane. An enlarged view of an individual particle and the associated filaments is also shown (inset). Figure contents was taken and modified from (Patla *et al.* 2011).

(Patla *et al.*, 2010). Focal adhesions are dynamic complexes composed of a large number of actin-associated adaptor proteins and signalling molecules, allowing the physical linking of F-actin and the plasma membrane (and ECM through integrins) together with the ability to regulate cell survival, proliferation and differentiation (Patla *et al*. 2010). Integrin gene deletion studies in many organisms have demonstrated their importance in developmental biology and their primary role in cell-matrix adhesion (Legate *et al.,* 2009). Knock-out studies of integrins from the nematode *Caenorhabditis elegans,* which possesses a single β-subunit and 2 α-subunits, clearly demonstrates integrin function in cell-ECM adhesion, with the loss of the β-subunit (β-PAT3) resulting in a paralyzed, arrested elongation at two-fold

(PAT) phenotype, hallmarked by the detachment of muscle from the body wall (Legate *et al.*, 2009).

1.1.2. The Structure of Integrins

In mammals there are 18 α-subunits (usually denoted by Roman numerals) and 8 β-subunits (denoted using Arabic numerals) that assemble into 24 integrins in mammals, with each integrin heterodimer exhibiting different ligand binding properties and tissue expression patterns (Hynes, 2002). Owing to the many isoforms of these receptors and some functional redundancy, characterisation of some gene knock-out phenotypes in higher eukaryotes has been complicated (Legate *et al.*, 2009). The integrin subunits are constructed from several domains strung together by flexible linkers and possess a unique architecture (see Figure 1.1). Additionally, landmark crystal structures of $\alpha_V\beta_3$ integrin heterodimer ectodomain (Xiong *et al.*, 2001), in complex with the Arg-Gly-Asp ligand (Xiong *et al.*, 2002), along with the subsequent $\alpha_{IIb}\beta_3$ (Zhu *et al.*, 2008) and the more recent $\alpha_X\beta_2$ (Xie *et al.*, 2010) have been solved. In addition to the large multidomain extracellular portion, each subunit possesses a single transmembrane helix and, typically, a short flexible cytoplasmic tail which is important for regulation.

The α-subunit of the integrin heterodimer (~1000 amino acids) is comprised of four extracellular domains; a seven-bladed β-propeller, a thigh and two calf domains. However, nine of the eighteen integrin isoforms possess an additional I domain (denoted as α-I), inserted between blades 2 and 3 of the propeller domain, which is similar in structure to the von Willebrand A domain. The thigh and calf domains both have immunoglobulin-like, β-sandwich folds (Figure 1.1). The β-subunit, on the other hand, possesses seven extracellular

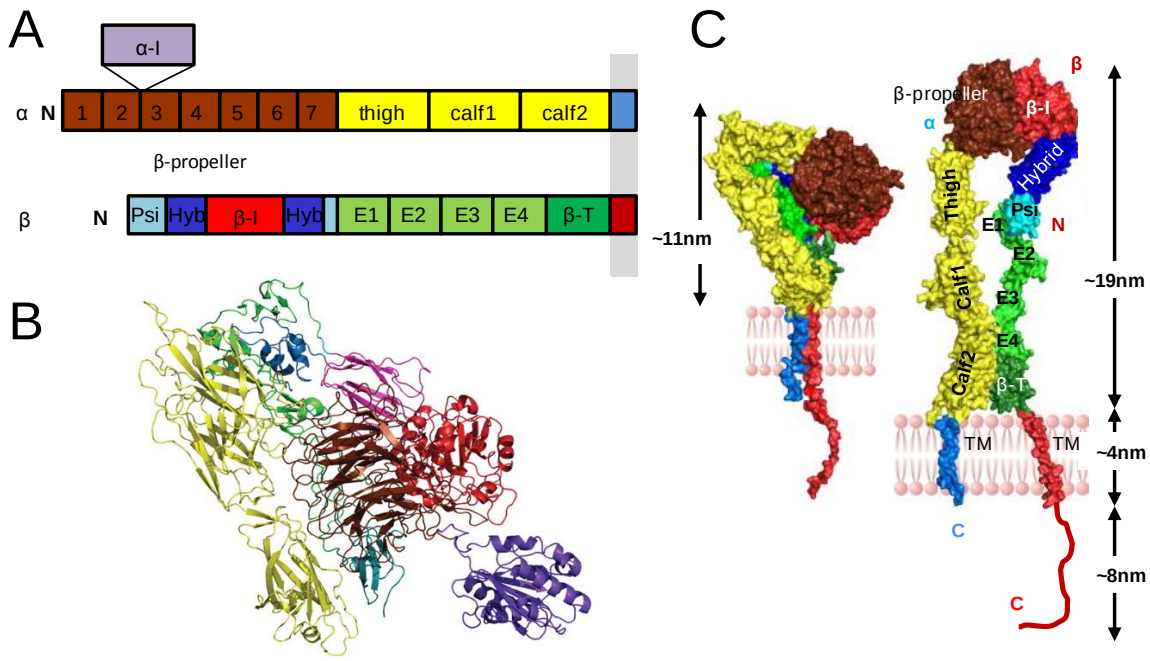


Figure 1.1 – The Integrin structure. (A) Domain structure of an α and β subunit ectodomains. Some α-subunits possess an additional I-domain inserted into the β-propeller domain. The plasma membrane is shown as a gray bar. (B) Structure of αxβ2, colour coded as in (A) (PDB: 3K6S). The αx possesses an I-domain and is coloured accordingly. (C) Surface cartoon representation of bent and upright conformations of the integrin with approximate dimensions. The domains are labelled; Psi, Plexin/Semaphorin/Integrin domain and TM domain, transmembrane domain. Figure (B) was taken and modified from (Campbell and Humphries, 2011). Figure (C) was taken and modified from (Anthis and Campbell, 2011).

domains; an I domain (denoted β-I), a hybrid, and a PSI (Plexin–Semaphorin–Integrin) domain, four epidermal growth factor (EGF) domains and a β-tail domain. Interestingly, the β-I domain is inserted into the hybrid domain, which is, in turn, inserted into the PSI domain resulting in complex interdomain connections. The overall topology of the integrin ectodomain and its domain organisation is shown in Figure 1.1.

1.1.3. Integrins and Matrix Binding

The integrin ectodomain is responsible for the binding of the integrins to matrix ligands and structural studies of integrins have revealed the molecular detail of ligand binding. In integrins without an I-domain in the α-subunit, the extracellular ligand binds at a site that is located at the interface between α- and β-subunits (Xiong *et al.,* 2002), where it interacts with the β-propeller domain of the α-subunit and β-I domain of the β-subunit and is dependent upon a cation occupied 'metal-ion-dependent adhesion site' (MIDAS) within the β-subunit (Moser *et al.,* 2009). There are three closely linked metal ion binding sites that are formed by loops from the β-I domain of the β-subunit. Mg^{2+} bound to the central MIDAS is critical for ligand binding as it directly co-ordinates the acidic side chain common to all integrin ligands (Zhu *et al.,* 2008). The Mg^{2+} binding site is also flanked by two Ca^{2+} binding sites (Zhu *et al.,* 2008). Of the calcium ion binding sites, that site adjacent to the MIDAS (denoted as the ADMIDAS) is considered inhibitory, with the other calcium binding site considered synergistic and therefore denoted as a synergistic metal ion binding site (SyMBS). In the case of integrins that possess an α-I domain, the ligand binding site is located within this domain and is also dependent on a Mg^{2+} or Mn^{2+} occupied MIDAS.

1.1.4. Integrin Activation

The interaction of integrins with the extracellular matrix at the cell's surface must be controlled by signalling from within the cell that regulates the integrin's affinity for its ligands. Thus integrins exist in low-, intermediate-, or high-affinity states (Moser *et al.,* 2009) that are highly regulated. It has been observed that in all crystal structures of intact integrin ectodomains solved to date, the integrins were crystallised in a "bent" conformation (Xie *et*

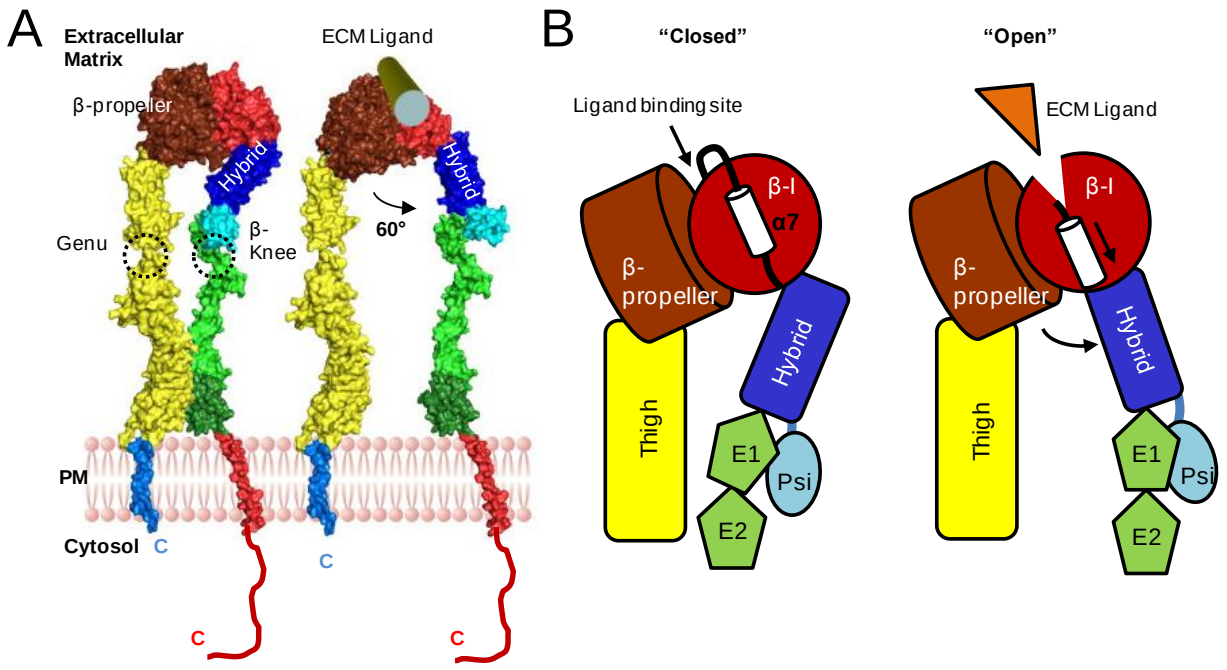


Figure 1.2 – Conformation of the Integrin Ectodomain. (A) Surface cartoon representation of possible intermediate (left) and high-affinity (right) state. The integrin 'legs' have separated along with the transmembrane (TM) domains and cytoplasmic tails and the swing out of the hybrid domain results from the binding of the extracellular ligand. (B) An illustration of the piston movement of helix α7 in the I domain of the β-subunit and the swing out of the hybrid domain as seen by crystallography (Zhu *et al.* 2008). The movement of α7 results in an open conformation that is capable of binding to the ligand. For integrins with an additonal α-subunit I-domain (not shown), the ligand is bound by α-I, but the movement of α7 is still occurs as the conformational change in α-I is transferred to I domain of the β-subunit resulting in the hybrid swing out. Figure (A) was taken and modified from (Anthis and Campbell, 2011).

al., 2009; Zhu *et al.*, 2008; Xiong *et al.*, 2002; Xiong *et al.*, 2001), which places the ligand binding site in close proximity to the transmembrane domain and therefore the outer leaflet of the plasma membrane (Xiong *et al.*, 2002). It is thought that integrins are in a low affinity state when the ectodomain is in this bent conformation. However, there are regions of interdomain flexibility in both α- and β-subunits that may allow other conformations to exist. In the α-subunit there are two main regions of flexibility; the β-propeller-thigh linker and the "genu" or knee between the thigh and calf domains. The β-subunit is, in general, more flexible (Xie *et al.*, 2009), particularly the EGF1-EGF2 linker (β-knee), which bends the β-

subunit analogously to the α-subunit and is similarly placed, allowing the entire ectodomain to hinge at the knees thereby extending the structure. Furthermore, studies of integrins by electron microscopy (EM) demonstrate observable flexibility of the ectodomains (Ye *et al.*, 2010; Zhu *et al.*, 2008) that together with the existence of conformation-dependent epitopes on the integrin that are recognised by conformation-specific antibodies (e.g. PAC1) suggests that multiple, physiologically relevant, conformations exist (Campbell and Humphries, 2011). This has led to the wide acceptance that integrins undergo conformational changes *in vivo* that result in intermediate- and high-affinity integrins (Moser *et al.*, 2009). Structural studies of the integrin ectodomain have led to the propostion of an upright high-affinity ectodomain structure (Xie *et al.*, 2009; Zhu *et al.*, 2008; Xiong *et al.*, 2001), however, the conformational events surrounding this change in affinity and its relationship to this sort of extended structure are still disputed (Campbell and Humphries, 2011). Thus two models are proposed: the "switchblade" model, whereby a large conformational change takes place moving the integrins from bent to upright, and the integrins from low-affinity to high-affinity (Luo *et al.*, 2007), and the "deadbolt" model whereby more conservative conformational changes take place resulting in a high-affinity integrin in the absence of an extended integrin (Xiong *et al.*, 2003).

Although knowledge of the exact sequence of events leading to high affinity integrins from the cell interior is incomplete, there are several key conformational events that are known. Conformational changes in the β-I domain are a key determinant for ligand affinity (Campbell and Humphries, 2011). A conformational transition between "closed" and "open" states of the β-I domain has been observed in co-crystal structures of αIIbβ3 headpieces bound to ligands and, with combined mutational analysis, have suggested that closed and open conformations correspond to low- and high-affinity for the ligand, respectively (Xiao *et al.*, 2004). The conformational changes are linked to a crankshaft-like motion of a helix in the

β-I domain (α7), which is displaced toward the hybrid domain and, in turn, results in the swing-out of the hybrid domain by 60° with respect to the β-I domain (see Figure 1.2). Accompanying the rod-like motion of the helix α7 in the β-I domain is the upward movement of α1 (Xiao *et al.,* 2004). The motion of these two helices promotes backbone movements of the loops that contain the metal ions, and results in a high affinity β-I domain. Given the structural basis for integrin's affinity for its ligand, it is likely that the unbending of the integrin to an extended state is possibly a pre-requisite to hybrid domain swing-out, given that a bent state prevents such motions (Campbell and Humphries, 2011).

The transmembrane (TM) domains and cytoplasmic domains also play a key role in the transition of integrins from a low- to a high-affinity state (Moser *et al.,* 2009; Wegener and Campbell, 2008). A structure of the integrin $\alpha_{IIb}\beta_3$ transmembrane heterodimer has been solved, within a model membrane environment using phospholipid bicelles, by NMR (Lau *et al.,* 2009). The TM helix of the β-subunit is tilted with respect to the α-helix by ~25° (Lau *et al.,* 2009). Interestingly, the length of the TM domain of the β-subunit is a conserved 30 residues (24 in the TM of the α-subunit) and is therefore too long for a canonical TM helix, therefore implying that the tilt, seen in the structure, accommodates all 30 residues in the membrane, and is also conserved (Moser *et al.,* 2009). The structure demonstrates the formation of inter-helical glycine packing interactions and an unusual backbone reversal of the α-subunit that packs two consecutive phenylalanines against the β-subunit (Lau *et al.,* 2009). Furthermore, the unusual reversal promotes the formation of a salt-bridge between an aspartate and arginine, of the α-subunit and β-subunit, respectively that together form an inner leaflet clasp. A plethora of structural, cellular, biophysical and mutational studies has demonstrated that this intersubunit clasp is responsible for integrin activation (Campbell and Humphries, 2011; Lau *et al.,* 2009; Larjava *et al.,* 2008; Wegener and Campbell, 2008). This

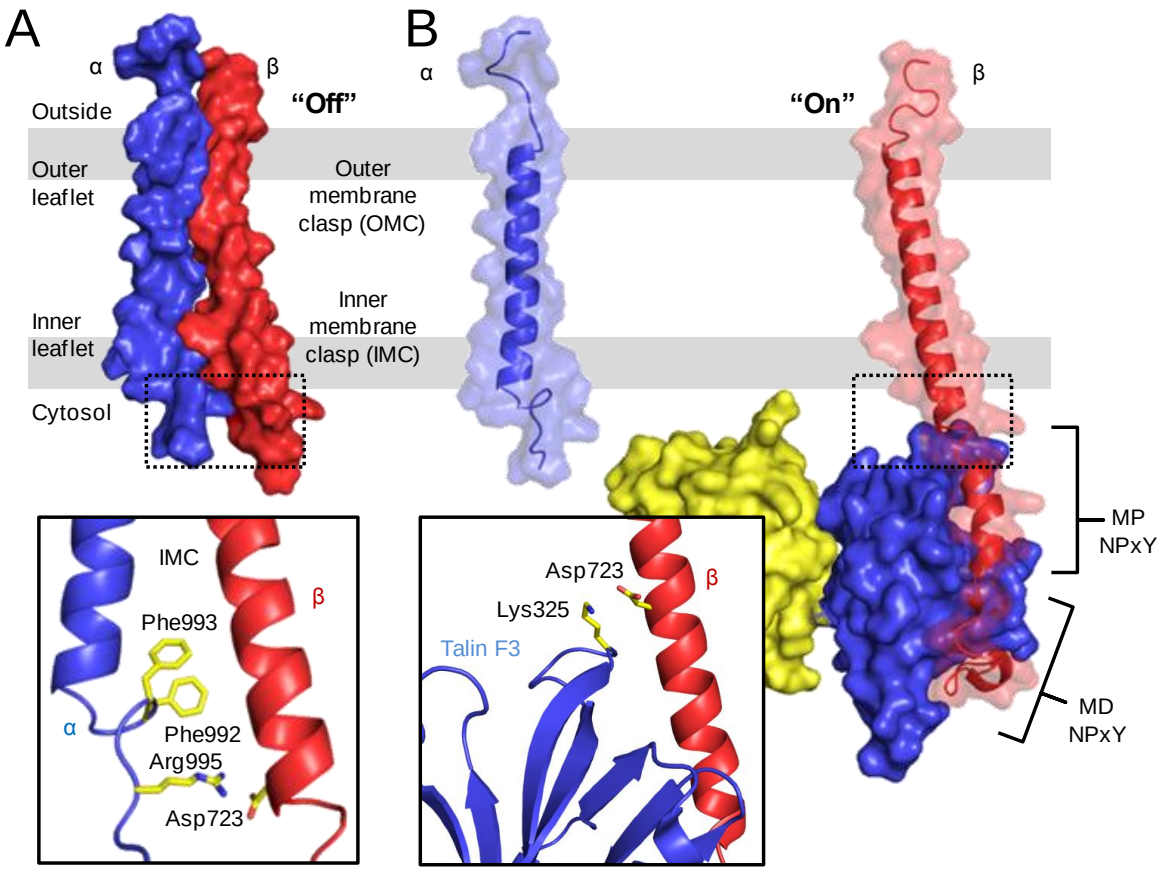


Figure 1.3 – Integrin Activation at the Membrane. (A) The transmembrane (TM) domain of the inactive integrin heterodimer with both α (blue) and β (red) integrin TM helices shown. The approximate position of the lipid bilayer is shown by grey bars. This dimerisation of the TM helices is mediated by an outer membrane clasp (OMC) and an inner membrane clasp (IMC). The IMC is shown in detail with chain reversal allowing two consecutive Phenylalanines (yellow) to partition into the membrane. (B) The complex formed between the β-cytoplasmic tail (red) and the F2 (yellow) and F3 (blue) sub-domains of the talin head illustrates the activated integrin. The membrane proximal (MP) and membrane distal (MD) regions important for integrin function of the tail are labelled. The Asp^{995}-Arg^{723} salt bridge that clasps the integrin heterodimer with Arg^{995} replaced by Lys^{325} when engaged with the F3 subdomain of Talin. *(B)* was constructed from a composite of PDB coordinates 2K9J and 3G9W in PyMol (Delano Scientific).

interaction between TM helices maintains the integrin in a resting state, such that the extracellular domain is in a 'bent' low-affinity conformation (Luo *et al.*, 2007, Larjava *et al.*, 2008). The disruption of the clasp results in the separation of the integrin subunits, which also

leads to the activation of the integrin, to a high-affinity state (Luo *et al.*, 2007, Larjava *et al.*, 2008).

Activation of integrin *via* disruption of the 'clasp' (termed 'inside-out signalling') is regulated by the flexible cytoplasmic tail of the integrin β-subunit – more specifically by the binding of proteins (Moser *et al*, 2009). One such protein that binds to the cytoplasmic tail of the β-subunit and regulates integrin affinity is talin (Tadokoro *et al.*, 2003). The large cytoplasmic protein talin (~250kDa) possesses a globular N-terminal head domain (THD, ~70kDa) which is, in turn, comprised of four subdomains; F0, F1, F2 and F3 (Shattil *et al.*, 2010). Inside-out signalling is mediated by the phosphotyrosine binding domain (PTB), or F3 subdomain of talin binding to one of two conserved NPxY motifs in the cytoplasmic tail of the integrin β-subunit. This binding event results in the alteration of the cross angle of the TM helices, and thus destabilizes the stacking interactions (see Figure 1.3) but also results in the disruption of the conserved salt bridge clasp (Anthis, *et al.*, 2009; Wegener *et al*. 2007).

Gene ablation studies suggested that talin is essential for integrin activation (Tadokoro *et al.*, 2003), however, recent studies on integrin activation have suggested talin is not wholly sufficient and that the protein Kindlin is also essential for activating integrins (Moser *et al.*, 2009; Moser *et al.*, 2008; Montanez *et al.*, 2008). More recent studies have suggested that Kindlins are, in fact, co-activators of integrins (Larjava *et al.*, 2008). A number of proteins have been reported to bind directly to the integrin cytoplasmic tails, yet only talin and Kindlins can regulate integrin's affinity for its ligand (Moser *et al.*, 2009).

1.1.5. Integrin Signalling

Integrin-mediated cell adhesion is not simply the physical attachment of the cell to extracellular substratum at its surface, but must also be able to regulate cellular signalling if the interactions between the cell and surrounding extracellular environment are to have any meaningful biological outcomes (Campbell and Humphries, 2011). Thus integrins are also responsible for the modulation of cellular processes, for example; cell migration/retraction, survival, cell cycle progression and regulation of differentiation pathways, in addition to, and as a result of, adhesive interactions (Legate *et al.*,2009). Integrins are an unusual family of membrane proteins as signalling events are considered bi-directional – i.e. cellular signalling activates binding of integrins to extracellular matrix components (inside-out signalling) and in turn the receptor: ligand complex elicits a cellular response (outside-in signalling) (Hynes *et al.*, 2002, Montanez *et al.*, 2008, Legate *et al.*, 2009). Unlike growth factor receptors, integrins possess no enzymatic activity; instead signalling is induced by the converging of adaptor and signalling proteins to the cytoplasmic side of the integrin and the subsequent radiation of signalling cascades to influence cellular processes. Interestingly, Zaidel-Bar *et al.*, (2007) mapped a 'functional atlas' of the integrin adhesome using comprehensive experimental data mining and generated a detailed complex network of 156 components linked together and modified by 690 interactions. It is now clear that Kindlins and talin are the most important of the many integrin-binding proteins found in the cell. A detailed introduction to Kindlin biochemistry will be provided at the beginning of Chapter 3.

1.2. A Very Brief Introduction to Cell Division

The cell cycle is the most fundamental biological mechanism for the continuity of life – its sole aim, to accurately duplicate and divide a single cell into two. Whilst single cellular life depends on cell division for reproduction (bacteria and yeasts, for example), multi-cellular life relies on cell division for a functioning organism. Eukaryotic cells have evolved a complex network that regulates cell division through multiple cellular pathways and checkpoints. This multi-layered system of cell cycle control exists to enhance the fidelity of cell division and to allow appropriate and required cell division in response to extra- and intra-cellular signals. In humans, for example, the improper control of the cell cycle can lead to inappropriate growth of tissues and therefore result in tumours, some of which are cancerous and life-threatening. Conversely, improper control of cell cycle may also prevent an immune response to infection or prevent wound healing.

In a simplified form, the animal cell cycle consists of two major phases responsible for the execution of two basic events in cell division; (1) accurate chromosomal DNA replication (Synthesis or S phase), followed by (2) the partitioning of the replicated chromosomes into nuclei of two genetically identical daughter cells (Mitosis or M-phase) (Norbury and Nurse, 1992). In most animal cells additional *gap* phases termed G1 and G2 are inserted between the M and S phase and the S and M phase, respectively. These gaps serve to allow cells to double their other contents; otherwise cells would get smaller with each round of division. Importantly during these gap phases, although not necessarily throughout them, an integration of information is allowed in order to determine the readiness of the cell to commit to S or M phase (Norbury and Nurse, 1992). Cells have evolved a series of checkpoints that prevent cells from entering a new phase until they have completed the preceding one, thus ensuring proper progression through the cell cycle (Malumbres and

Barbacid, 2001). Our understanding surrounding the molecular basis of cell cycle checkpoint control has advanced substantially. However, this introduction will not provide an exhaustive description of the mechanism of cell division *per se*, but will hopefully provide a brief and focussed introduction into the cell cycle and its checkpoints.

1.2.1. A Brief Overview of Molecular Mechanisms of Checkpoint Control

Most of the cells that comprise adult metazoans are considered quiescent and in a non-proliferative state. Mitogenic signals can allow cells to re-enter the cell cycle but require the inactivation of the retinoblastoma protein (pRB) and the transcription of multiple genes required for DNA replication during S phase or genes involved in chromosomal segregation in M phase (Bueno and Malumbres, 2011). Mitogenic signals result in the expression of the mitogenic sensors, D-type cyclins, which are usually synthesised as a result of RAS/RAF/MAPK signal transduction cascades (Malumbres and Barbacid, 2001). Subsequently, these D-type cyclins form heterodimeric complexes with, and activate, the cyclin-dependent kinases (CDKs), for example CDK4 and CDK6, that control early cell cycle events (Bueno and Malumbres, 2011; Malumbres and Barbacid, 2001). Interestingly, cyclin D1 synthesis increases in G1 and remains elevated throughout the cell cycle (Malumbres and Barbacid, 2001). These CDKs, which are serine/threonine kinases, bound to their D-type cyclin regulatory subunits phosphorylate and inactivate retinoblastoma protein – a transcriptional repressor that binds to and inhibits E2F family transcription factors (Bueno and Malumbres, 2011; Malumbres and Barbacid, 2001). Additional CDK substrates in G1 progression include p107 and p130, both members of the retinoblastoma family (Malumbres and Barbacid, 2001). The ‘restriction-point’ is a point of no return, in which a cell commits to

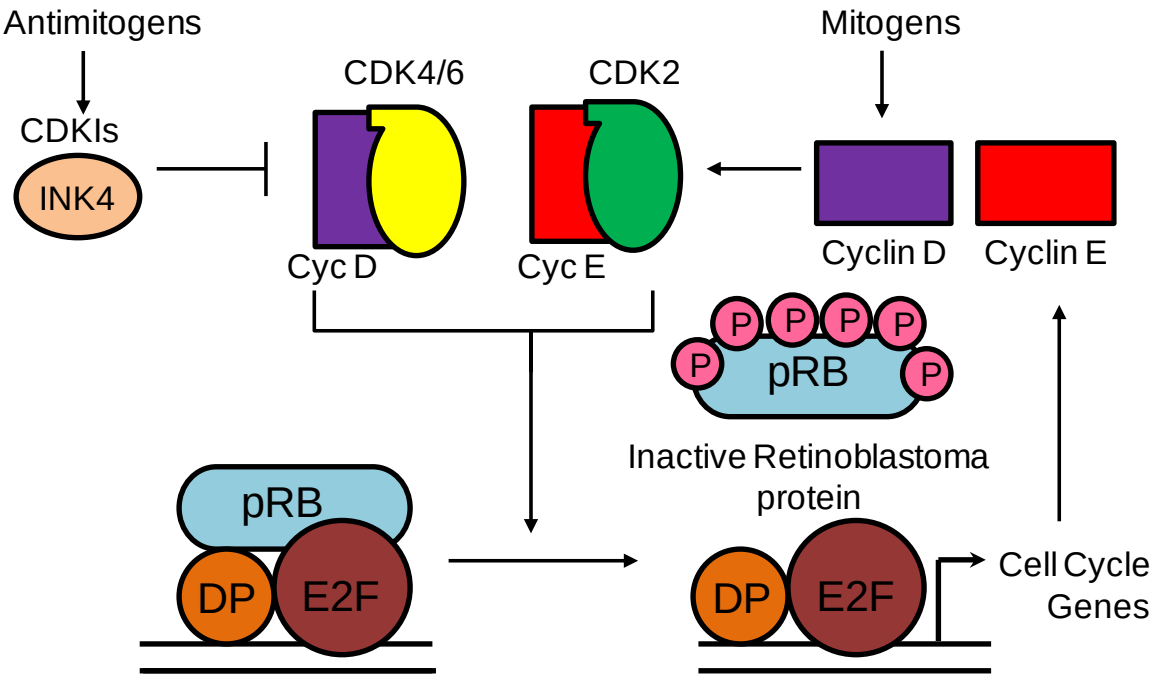


Figure 1.4 – Cyclin dependent Kinase Control in G1 Phase. (A) A simplified model of cyclin dependent kinase (CDK) phosphorylation of retinoblastoma protein (pRB) that results in the expression of DNA synthesis machinery. Mitogenic signals results in the synthesis of cyclin D that activated CDK4 and CDK6, which phosphorylate the repressor retinoblastoma protein (pRB). Inactivation of pRB allows E2F-DP-dependent expression of cell cycle genes including E-type cyclins. Expression of Cyclin E activates CDK2 and ensures full phosphorylation of pRB and therefore cell cycle gene expression. Antimitogenic signals results in the expression of CDK inhibitors, such as the INK4 family, that act to inhibit the activity of the CDK-Cyclin complexes.

the cell cycle, by ensuring that it has met its burden of mitogen-induced events (Pardee, 1974). It has therefore become commonplace to consider the G1 phase as having two stages, an early mitogen-dependent and a late mitogen-independent stage which are separated by the restriction point. Several CDKs are involved in progressing the G1 phase and driving the cell's transition from G1 to S phase, mainly; CDK4, CDK6 and CDK2 (Malumbres and Barbacid, 2005). However CDKs 4/6 are thought to drive early G1 phase progression with CDK2 driving late G1 phase and initiating the transition to S phase (Morgan *et al.*, 1997). The phosphorylation-dependent inactivation of pRB results in the induction of E2F responsive genes that are required for cell cycle progression, for example, the cyclin E1 gene

(see below), or genes involved in DNA synthesis. Whilst cyclin dependent kinases 4/6 are activated by D-type cyclins in the early G1 phase, CDK2 is sequentially activated by E-type and A-type cyclins, and is required for G1/S transition (Malumbres and Barbacid, 2005).

Whilst CDKs can be activated by cyclins, they can also become inhibited by CDK inhibitors (CDKIs). CDKIs can be expressed as a result of anti-proliferative signalling. There are two types of CDKI: the INK4 family and the WAF/KIP family that are comprised of four and three family members, respectively (Sherr and Roberts, 1999). The CDKIs exert their inhibitory activity by binding to CDKs, for example the INK4 family CDKIs bind to CDK4/6 and prevent the association of the activating D-cyclins, whereas the WAF/KIP family CDKIs form a heterotrimeric complex with the cyclin-CDK heterodimer and distorts its active site (Bueno and Malumbres, 2011; Sherr and Roberts, 1999).

Once the cell is committed to S phase and the DNA replication machinery is activated, S-phase progression is monitored by various checkpoints that act to slow or arrest S phase in the presence of aberrant DNA replication, or DNA damage (Bueno and Malumbres, 2011). For instance, stalled DNA replication is signalled through 'checkpoint Rad' proteins, i.e. the *S. pombe* protein Rad26, and Rad3 the orthologue of ATM (Ataxia Telangiectasia Mutated) /ATR (ATM and Rad3 related) (Read and Norbury, 2002). ATR (Rad3) is activated during S the phase of every cell cycle to regulate DNA replication at origins of replication, repair of damaged replication forks, and prevents premature entry into mitosis (Cimprich and Cortez, 2008). Replication stress results in the stalling of replication forks and is suggested to expose single stranded DNA (ssDNA) (Zou and Elledge, 2003). In one simplified model in humans, ssDNA is recognised by the ssDNA binding protein, Replication Protein A (RPA) which localises ATR to the stalled replication fork (Cimprich and Cortez, 2008, Zou and Elledge, 2003). Recognition of RPA-coated ssDNA by ATR is dependent on ATR-interacting protein

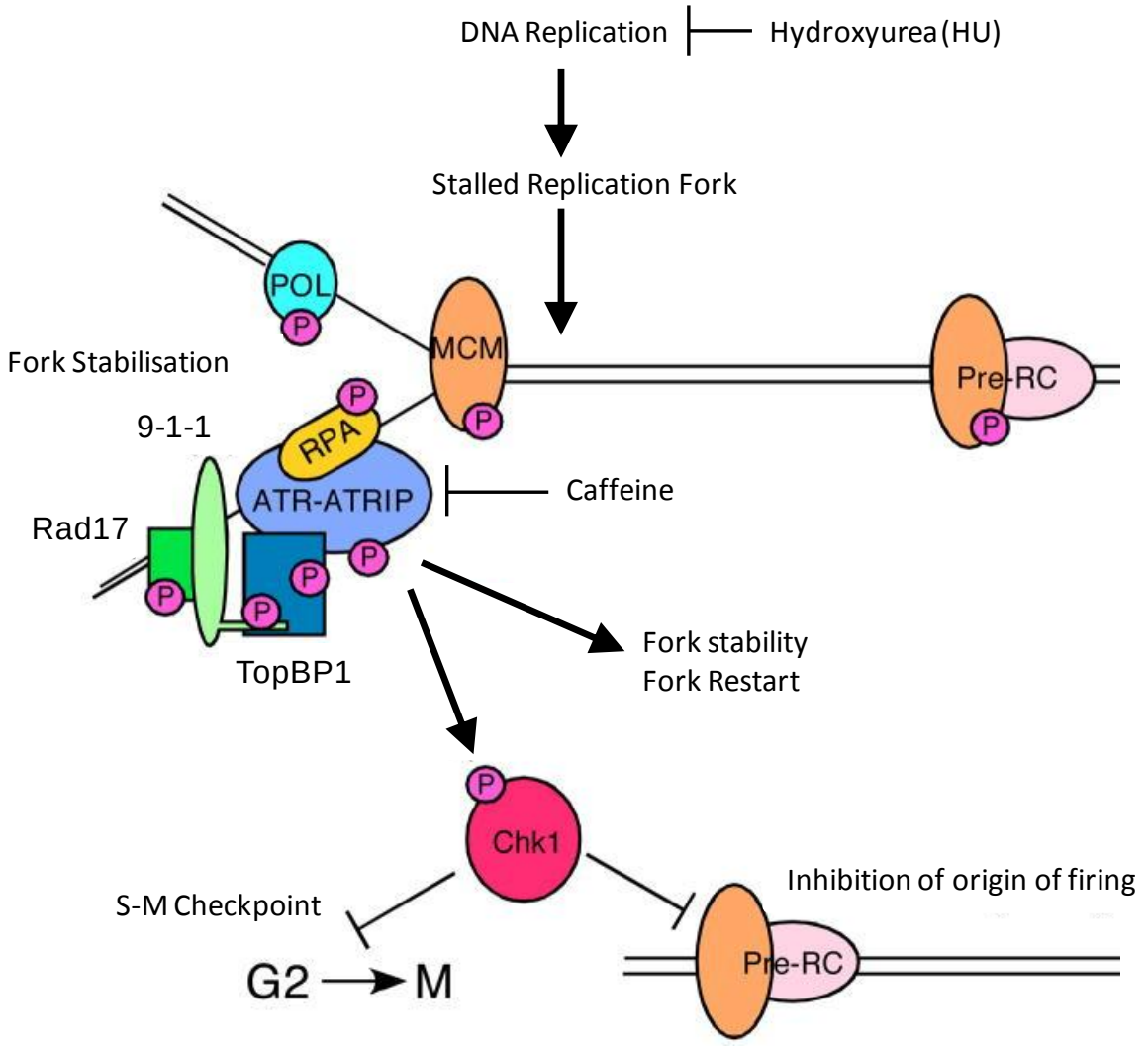


Figure 1.5 – Stalled DNA replication is signalled through 'checkpoint Rad' proteins ATM (Ataxia Telangiectasia Mutated) /ATR (ATM and Rad3-related). (A) The ribonucleotide reductase inhibitor hydroxyurea (HU) inhibits DNA replication by stalling replication forks. (B) Simplified model of replication fork stabilisation and cell cycle inhibition. The exposed ssDNA from the stalled replication fork is recognised and bound by Replication Protein A (RPA), which in turn localises ATR (Rad3 in *S. pombe*) ATR-interacting protein (ATR-IP) complex. Activation of ATR depends on the co-localisation of Rad9-Rad1-Hus-1 complex (9-1-1), which is loaded onto the primer-template junction *via* Rad17, and the ATR activator topoisomerase (DNA) II binding protein-1 (TopBP1; Rad4/Cut5 in *S. pombe*). Activated ATR phosphorylates Chk1 resulting in the restart of the stalled replication forks together with cell cycle arrest and slow replication at origins. Figure taken and modified form (Cimprich and Cortez, 2008).

(ATR-IP) (Zou and Elledge, 2003, Cortez *et al.*, 2001). However, activation of ATR is dependent on the independent co-localisation of the heterotrimeric Rad9-Rad1-Hus-1

complex (9-1-1 complex) that is structurally analogous to the replicative sliding clamp PCNA, and is loaded onto the primer-template junction in an ATP-dependent and protein clamp loader, Rad17, dependent manner (Kumagai *et al.,* 2006). The 9-1-1 complex brings the critical ATR activator topoisomerase (DNA) II binding protein-1 (TopBP1; Rad4/Cut5 in *S. pombe*) (Kumagai *et al.,* 2006) that binds to and activates ATR in an Rad-9 phosphorylation and ATR-IP dependent manner resulting in the phosphorylation of the cell cycle kinases Chk1 or Cds1 by ATR (Cimprich and Cortez, 2008, Rhind and Russell, 2000). Stabilisation of stalled replication forks during replication stress and activation of ATR and its effectors ultimately results in the restart of the stalled replication forks, together with cell cycle arrest and slow replication at origins (Cimprich and Cortez, 2008).

Once the genome is replicated in S phase, yielding two copies of each chromosome and the cell in G2 phase, the chromosomes must segregate into two daughter cells. CDK1, in complex with A- or B-type cyclins, is a major kinase responsible for the transition from G2 to M phase and since most cell cycle proteins, including this kinase, have been synthesised earlier in the cell cycle, entry into the mitotic phase depends on post-translational events (Bueno and Malumbres, 2011). CDK1 is held in an inactive state by inhibitory phosphorylation at two sites, by the kinase WEE1, which requires dephosphorylating by the phosphatase CDC25 for activation (Bueno and Malumbres, 2001). During incomplete DNA replication (discussed earlier) ATR/ATM is activated and ultimately causes the *in*activation of the mitotic phosphatase CDC25 *via* the kinase CHK1 and thus the inhibitory phosphorylation of the phosphatase CDC25 by CHK1 prevents the cell from entering into mitosis by maintaining the inactivating phosphorylation of CDK1 (Zeng *et al.,* 1998).

Additional kinases like the Polo-like kinases (PLK) and Aurora kinases, together with many other regulator proteins, are involved in the cytoplasmic and nuclear changes necessary for mitosis, for instance the establishment of the bi-polar spindle and the segregation of

chromosomes into two daughter cells (Bueno and Malumbres, 2011). Exit of mitosis requires the purging of the cell cycle regulating proteins. This is achieved proteolytically, in part, by the E3 ubiquitin-ligase, Anaphase-promoting complex/cyclosome (APC/C) that ubiquitinylates cell cycle regulators for the proteosome (Nakayama and Nakayama, 2006).

1.2.2. MicroRNAs and Cell Cycle Control

Since their discovery in nematodes, microRNAs (miRNAs) have been the focus of considerable attention due to their roles in development, stem cell biology, cell differentiation and cancer. MicroRNAs have now been linked to cell cycle regulation, targeting and regulating classical cell cycle control pathways, such as cyclin-dependent kinases (Cdks), for example (Chivukula and Mendell, 2008).

MicroRNAs (miRNAs) are short endogenous RNAs (~22nt) that modulate gene expression through complementary base-pairing with target mRNAs (Kim *et al.,* 2009) and are generated *via* two-step processing. In animals, miRNAs are transcribed as long precursors in the nucleus and are denoted as primary miRNAs (pri-miRNA) (Heo *et al.,* 2009). These pri-miRNAs are subsequently clipped into ~80 nucleotide hairpin-structured precursor-miRNAs (pre-miRNAs) by the RNaseIII enzyme complex, Drosha-Pasha/DGCR8 (Lehrbach *et al.,* 2009). Pre-miRNAs are subsequently exported into the cytoplasm by exportin 5 (Bohnsack *et al.,* 2004) and are further processed by another RNaseIII endonuclease, Dicer resulting in a mature miRNA duplex of ~22 nucleotides (Heo *et al.,* 2009) with, through an unknown mechanism, one strand associating with Argonaut proteins and loaded onto the microRNA-induced silencing complex (miRISC) (Hagan *et al.,* 2009). The binding of miRNA-protein complexes to mature mRNAs, through nucleotide base-pairing, inhibits translation and stability of the transcripts that results in the downregulation of the protein

product (Bueno and Malumbres, 2011). Numerous miRNAs have been identified with many expressed in a tissue- and developmentally-specific manner and their deregulation has been implicated in cancer initiation, progression and metastasis (Hagan *et al.,* 2009)

Many components involved in the control of the cell cycle directly control, or are the target of, miRNAs (Chivukula and Mendell, 2008). Evidence for the role of miRNAs in cell division came with their initial discovery, with the first miRNAs lin-4 and let-7 in *Caenorhabditis elegans*. The let-7 miRNA, a product of the *lethal-7* gene, was shown to be essential for stem cell differentiation and developmental timing in *C. elegans* by controlling cell fates during larval development (Reinhart *et al.,* 2000). Bioinformatic analysis of the mature miRNA let-7 sequence revealed highly conserved orthologues in metazoans (Roush and Slack, 2008). Additionally, multiple copies of the let-7 family miRNA genes are often present in the genome at different loci. Furthermore, other let-7 isoforms also exist and therefore it has become commonplace to denote let-7 miRNAs isoforms *via* a single letter suffix (i.e. let-7a) and multiple copies of particular isoforms at different genomic loci by an additional number (i.e. let-7a-1) (Roush and Slack, 2008).

The role of let-7 in cell cycle has been predominated by mechanistic understanding of its ability as a tumour suppressor. Early work in *C. elegans* revealed that miRNA-84, a let-7 family member regulates the worm homolog of RAS (Rat Sarcoma Viral Oncogene homologue), let-60 (Johnson *et al.,* 2005). In humans let-7-dependent downregulation is mediated by its targeting of the oncogenes, c-MYC, KRAS (Kristen Rat Sarcoma Viral Oncogene homologue) and NRAS (neuroblastoma RAS), with, for example, low let-7 expression correlating to high RAS expression in lung squamous cell carcinoma (SCC) (Johnson *et al.,* 2005), thereby providing evidence for its role in tumour suppression (Chivukula and Mendell, 2008). The role of let-7 in cell proliferation was evident when let-7 overexpression in cancer cells decreased cell cycle progression (Johnson *et al.,* 2007).

Furthermore, microarray data and reporter assays have identified cell cycle proto-oncogenes involved in G1 – S phase transition as direct targets for let-7-mediated downregulation, including *CDK6*, *CDC25A* and *CCND2* (Cyclin D2) (Johnson *et al.*, 2005).

The let-7 miRNAs deserve particular attention for two reasons; (1) there is now considerable evidence demonstrating the role of the let-7 miRNA family in cell cycle control (reviewed by Bueno and Malumbres, 2011) and (2) numerous studies have reported the posttranscriptional control of let-7 RNA processing, specifically *via* 3' uridinylation (Hagan *et al.*, 2009; Heo *et al.*, 2009; Lehrbach *et al.*, 2009) thus correlating RNA 3'uridinylation and cell cycle control. The second part of the thesis (Chapters 6-8) will describe the yeast protein Cid1 which is properly introduced at the beginning of Chapter 6 and which is a uridylyl transferase involved in S-M checkpoint control whose orthologues in humans ZCCHC6 and ZCCHC11 are key to let-7 miRNA turnover control (Heo *et al.*, 2009; Hagan *et al*. 2009).

1.3. An Overview of This Thesis

This work focuses on two fundamental aspects of cell biology, (1) integrin-mediated cell adhesion and (2) cell cycle control in cell division, both of which, if aberrant, lead to a number of human diseases. In this Thesis, we use a variety of structural techniques in order to gain an insight into two specific proteins essential for both processes. These proteins studied are; the integrin co-activators and focal adhesion protein family, the Kindlins, and the yeast polyuridylyl polymerase involved in S-M checkpoint cell cycle control, Caffeine Induced Death Suppressor-1 (Cid1).

In the first three results chapters of this work, a high-throughput structural genomics approach is utilised for the generation of recombinant Kindlins for structural characterisation and is presented first. Subsequently described is the crystal structure of the bacterially expressed Pleckstrin Homology (PH) domain from the Kindlin-1 isoform and its structural, biochemical and cell biological characterisation, which has revealed a role for the domain in focal adhesion assembly and has highlighted important residues for ligand binding. The biophysical characterisation of recombinant Kindlin-3 will then be described. Small angle X-ray scattering of Kindlin-3 with additional biophysical techniques reveals that the Kindlins possess a non-canonical FERM (4.1, Ezrin, Radixin, Moesin) domain that is linearly arranged, analogous to talin. In addition, nuclear magnetic resonance (NMR) studies of Kindlin-3 binding to integrin β-tails demonstrate sequence specific requirements for efficient binding.

In the later 2 chapters of this work, the first crystal structure of the *Schizosaccharomyces pombe* protein Cid1 will be presented. Structural analysis and its phylogeny reveal Cid-1 to be an ancient enzyme. The crystal structure of Cid-1 in complex with UTP is described subsequently. This, along with biochemical characterisation, has allowed identification of residues important for uracil discrimination and suggested a mechanism more closely related to mammalian terminal uridylyl transferases involved in the regulation of let-7 miRNA biogenesis. We also demonstrate an induced fit of Cid1 when bound to UTP. We show that Cid-1 is capable of efficient RNA binding in the absence of a detectable RNA binding motif/module. Structural and biochemical analyses reveal a novel mechanism for RNA binding. This ability is unique to this uridylyl transferase and is critical for catalytic activity.

CHAPTER 2

Materials and Methods

With the advent of high throughput DNA sequencing it is possible to determine the complete genomes of many whole organisms with these genomic sequences now publically available for many bacteria, and a host of eukaryotic model organisms, including yeasts, fly, chicken, rat, mouse and human. Some of the biologically relevant products from genes, namely proteins, that allow life cannot currently be predicted from sequence alone and therefore each gene cannot be immediately assigned a function or biological role without extensive experimental determination. Since the function of the gene product is determined by its organisation in three-dimensions, the determination of the structure provides insights into biological function that, in turn, provide the placement of the protein within a particular cellular pathway and also provide a basis for predicting the biological functions of highly related proteins as in protein families. To date, this process of defining a protein's structural and biological function requires extensive experimental determination of each protein in isolation. The production and purification of proteins is still a prerequisite to their *in vitro* study and structure determination. The majority of the research featured in this thesis is reliant on the over-expression and purification of proteins for structural and functional characterisation.

This Chapter describes the materials and methods employed during the course of my research, with a particular emphasis on purification of overexpressed proteins and their biophysical characterisation using structural techniques. Methods used in collaborative work will be acknowledged where appropriate. An account of the general theory behind each technique will be given prior to the detailed documentation of protocols for the associated technique that have led to specific results described in later chapters. Although many

techniques are now standardised and well rehearsed by myself, these 'standard' methods have, in some cases, been adapted in order to overcome some specific difficulties.

2.1. General Molecular Biology Techniques

A variety of molecular biology techniques have been employed in this thesis. However the numerous bacterial cultivation protocols, media recipes and details of 'standard' techniques will not be presented here. Instead the reader is directed to these protocols either described in "Molecular Cloning: A Laboratory Manual" (Sambrook and Russel, Cold Spring Harbor Laboratory Press, 3rd edition, 2001) or procedures/recipes detailed in manuals accompanying commercially available cloning/expression cells and kits where appropriate.

2.2 High-Throughput Cloning and Expression Screening

Efforts to define the three-dimensional structures of all gene products have been addressed by structural genomic consortia around the world. Structural genomics has developed highly parallelised pipelines and workflows in order to increase productivity and lessen the expense of protein production, crystallisation and structure determination. The Oxford Protein Production Facility (OPPF) has employed miniaturised and automated protocols for high-throughput protein production, purification and crystallisation. In many cases, this pipeline is capable of taking a gene of interest and producing purified recombinant protein for biophysical characterisation and structure determination. The use of the OPPF pipeline in this thesis is dependent on a number of staff at the OPPF. Therefore, protocols for the cloning and expression trials described here that are part of the standard OPPF pipeline procedure were performed in conjunction with, and under the guidance of, Dr Louise Bird.

2.2.1. Oligonucleotide Primer Design

Oligonucleotides with sequences complimentary to the template cDNA were designed using the OPINE (Albeck *et al.,* 2006) primer design utility. Thus ninety eight forward and reverse primers complementary to 5' and 3' cDNA sequences of the desired protein construct were designed with a uniform melting temperature of ~60°C and an optimised GC content of 40% - 60% for parallelised cloning. Within the OPINE primer design suite the desired primers are appended with 15 nucleotides that are homologous to the cloning site within the target vector for the subsequent cloning procedure. Primers were synthesised by Eurofins MWG Operon and were delivered in 96-well format. Lyophilised primers were diluted in 10mM Tris-HCl, pH 8.0, to a stock concentration of 100µM and stored at -20°C for short term or -80°C for long term storage.

2.2.2. DNA Amplification by Polymerase Chain Reaction (PCR) for High-throughput Cloning

The reaction mixtures (usually 50 µl) for 96 PCR reactions were prepared on ice in a 96-well thin-walled skirted PCR plate and were composed from the mixtures listed in Table 2.1. Primers were diluted 1:10 before PCR reactions (final concentration of 10µM), before being added to each reaction mixture. The commercially available *Thermococcus kodakaraensis* (KOD) high fidelity DNA polymerase (Novagen) and KOD in complex with a monoclonal antibody, termed hot-start (KOD-HS) (Novagen), were used to amplify the template DNA. Approximately, 100ng of template cDNA was added to each reaction mixture. PCR products were obtained after 30 cycles of PCR using a PCR thermal cycler and the thermal cycle procedure listed in Table2.2.

Table 2.1 – PCR Reaction Mixture

Component	Quantity
Forward Primer	3µl (10µM)
Reverse Primer	3µl (10µM)
Template DNA (mini-prep)	2 µl (50ng/µl)
PCR Master Mix (see below)	41µl

PCR Component Master Mix					
Number of (50µl) Reactions	25	50	75	105	210
10 X KOD HiFi Buffer #2 (µL)	125	250	375	525	1050
10 X dNTP mix (µL)	125	250	375	525	1050
25mM $MgCl_2$ (µL)	50	100	150	210	420
KOD DNA Polymerase (2.5U/µl)	10	20	30	42	84
Sterile Water (µL)	740	1480	2220	3108	6216
Total Volume (µL)	1060	2100	3150	4410	8820

Table 2.2 – PCR Thermal Cycling Steps

1	94°C	2 minutes (for Hot-start polymerases only)
2	94°C	30 seconds
3	60°C	30 seconds
4	68°C	2 minutes (for products 200bp - 5000bp)
5	Go to step 2	Repeat for 29 Cycles
6	72°C	2 minutes
7	4°C	End

PCR products were analysed by gel electrophoresis. Samples of PCR product (5µl) were diluted with loading buffer (0.25% (w/v) Bromophenol Blue in 30% (v/v) Glycerol) and DNA products separated by electrophoresis in a 2% agarose TBE gel. Electrophoresis of samples was achieved using 100mV for 40 minutes. The separated DNA in the gel was visualized on a transilluminator. PCR reactions were repeated with KOD, KOD Hot-Start and BioXact (Novagen) polymerase enzymes, using 35 cycles with an annealing temperature of 50 ºC and extension temperature of 68ºC, where initial PCR reactions failed.

2.2.3. Purification of PCR Products

To remove methylated template DNA from the PCR products the reaction mixtures were treated with 1µl Dpn1 enzyme (NEB) by incubation for 1 hour at 37ºC. The PCR products were purified using AMPure (Agencourt) magnetic beads, which bind to PCR products >100nt. Each PCR reaction was mixed, by pipette, with 90µl of the magnetic beads, until a homogenous solution was obtained. The reaction mixtures were incubated for 5 minutes at room temperature. The magnetic beads were separated from solution via small magnets, in a time dependent manner and the resulting cleared supernatant was aspirated and the pellet resuspended and incubated in 200µl 70% (v/v) ethanol for 30 seconds. The beads were separated from solution as described earlier and supernatant aspirated. This process of washing was repeated twice. After the final wash the pellets were air dried for 10 -20 minutes until residual ethanol evaporated. The PCR products were eluted from the beads *via* the addition and re-suspension of beads in 40µl of 10mM Tris-HCl, pH 8.0. The eluants were carefully transferred and stored at -80ºC.

2.2.4. High-throughput Cloning of PCR Products

The high throughput cloning used by the OPPF pipeline employs ligation independent cloning (LIC) techniques using the commercially available InFusion™ (Clontech) enzyme, which is capable of incorporating PCR products into the target vector *via* homologous recombination. 100ng of linearized pOPINF vector was mixed with 10ng PCR product and diluted with RNase free water to a final volume of 10µl. The DNA-vector mix was added to lyophilized In-Fusion™ components and dissolved to homogeneity by pipette mixing. The reaction was sealed and incubated at 42ºC for 30 minutes in a thermal cycler. The reactants were immediately diluted in 40µl TE. The diluted reaction mixtures were used immediately for transformation of the chemically competent *E. coli*, Omnimax (Invitrogen). 5µl of diluted reactants were incubated with 50µl aliquots of Omnimax competent cells on ice for 20 minutes. The *E. coli* was heat shocked at 42ºC for 30 seconds and immediately returned to ice for a further 2 minutes. Transformant outgrowth was achieved by the addition of 450µl of LB per transformation reaction and incubating for 1 hour at 37ºC without shaking. 10µl of the *E. coli* were plated onto 1ml agar plates supplemented with, 50µg/ml ampicillin, 1mM (final concentration) IPTG, 1:1000 dilution 20% (w/v) X-Gal in DMF; duplicates were also plated but using a 1:10 dilution of *E. coli* culture. Agar plates were incubated at 37ºC overnight. X-gal was used for blue-white screening to discern transformants possessing inefficiently linearized parental plasmid from those possessing plasmid with PCR product. White colonies were picked from plates and inoculated into 3ml power broth supplemented with 50µg/ml of ampicillin. The inoculated medium was incubated at 37ºC overnight with shaking at 200 r.p.m.. Cultures were used for plasmid purification and creating glycerol stocks. A 100µl sample of each culture was used for glycerol stocks by dilution in equal volume of 30% (v/v) glycerol in LB. The glycerol stocks were stored at -80ºC until use. The remaining cultures were centrifuged at 5000 x g for 15 minutes. The resulting supernatant was decanted and the

pellets were stored at -20ºC until use. Frozen pellets were thawed and resuspended in "resuspension buffer P1" (Qiagen) for DNA purification via Qiagen Biorobot 8000, using a similar protocol to that found in the Qiagen miniprep handbook using a vacuum manifold. Purified plasmids were tested for PCR product insert by PCR verification, using a T7 forward primer **(**5'-TAATACGACTCACTATAGGG-3'**)** and the DNA specific reverse primer, as for the high-throughput PCR. PCR verification was performed using KOD DNA polymerase as described earlier.

2.2.5. Site Directed Mutagenesis

Substitution mutations of amino acids were performed in several projects to assess the contribution of specific amino acids toward the protein's function. In both the Kindlin-1 PH domain project and the Cid1 project site directed mutagenesis was undertaken. In the Cid1 project the site directed mutagenesis was performed by S. Fleurdepine, Sir William Dunn School of Pathology, University of Oxford, as part of a collaboration.

2.2.6. Site Directed Mutagenesis of Kindlin PH Domain

A single point mutation (Glu^{416}>Ala) was introduced with PCR based mutagenesis, with the primer 5'-AAACCTTAGAGGCTGCGCAATTGTGCCAGATGTGA-3' and the complementary primer 5'-TCACATCTGGCACAATTGCGCAGCCTCTAAGGTTT-3', using Phusion-Hot Start DNA polymerase (Finzymes) and the pOPINF-*KINDLIN1-PH* construct as a template. The PCR reaction (50µl) was performed following the manufacturer's instructions accompanying the DNA polymerase. Subsequently, template DNA was digested with DpnI for 1 hour at 37°C. The PCR product was checked using 1% (w/v) agarose gel

electrophoresis and used for transformation of *E. coli* for amplification using standard methods. DNA sequencing (Source Bio, Oxford, UK) was performed to verify the introduction of the mutant.

2.2.7. Site Directed Mutagenesis of tCid1

tCid1 mutants were generated by PCR-based site-directed mutagenesis using Pfu Turbo DNA polymerase (Stratagene) and pGEX6P-1 tCid1 (Rissland *et al.,* 2007) as template. Standard PCR reactions were performed with 18 cycles of 50 s at 95°C, 50 s at 55°C and 9 min at 68°C. The products were digested with Dpn I to remove the parental methylated strand and 1/25 of the digestion was transformed into XL1Blue competent cells. The combination of mutation sites was generated by successive site-directed mutagenesis procedures on previously mutagenised templates. All constructs were verified by sequencing.

2.3. Heterologous Protein Expression

Historically proteins were obtained and isolated from whole organisms using specific tissues or fluids with a naturally high quantity of endogenous target protein. The production of large quantities suitable for biophysical characterisation, especially for the determination of the protein's three-dimensional structure, is achieved nowadays by the use of molecular biology and numerous expression host organisms that overexpress the target protein. In essence the target protein, which is encoded within plasmid DNA, is introduced into the host and the protein of interest is expressed. The workhorse for general protein production is *E. coli*, which is engineered to recognise and utilise the plasmid DNA. In addition to bacteria, several alternative expression hosts are becoming increasingly popular for production of

eukaryotic and viral proteins, especially those that require post-translational modification. These hosts include, insect cells, human or other mammalian cells, and yeasts. However each host requires specific sequence determinants within the plasmid DNA. The two principal expression hosts used in this thesis are *E. coli* and insect cell systems and these will therefore be described in greater depth. Therefore, methods for the expression of proteins studied in this thesis will be dealt with in turn and described in the context of the expression host used.

2.3.1. Expression in *Escherichia coli*

In cases where the target protein does not require extensive post-translational modifications, it is appropriate to begin any expression trials in a bacterial host. Typically an inducible *E. coli* system is preferred whereby the expression of the recombinant protein is initiated *via* a change in the culture environment. Usually, isopropyl-β-D-thiogalactopyranoside (IPTG)-inducible *E. coli* systems are employed, whereby a *lac*-controlled operon that encodes bacteriophage T7 RNA polymerase contained in the host's genome is expressed in a galactose or IPTG-dependent manner. Therefore, plasmid DNA encoding a protein of interest is introduced into these bacteria, whereby the target sequence is linked to a T7 promoter, thus allowing efficient transcription of the desired gene in a T7 dependent manner. A large number of *lac*-inducible T7 RNA polymerase-containing *E. coli* expression strains are widely available from multiple commercial sources.

2.3.2. Expression Screening of Constructs in *E. coli*

Purified plasmid DNA was used for the transformation of chemically competent *E. coli* expression strains using 1µl of plasmid DNA. After outgrowth the transformants were

cultivated on LB agar plates supplemented with 50µg/ml ampicillin, in all cases, and additional strain specific antibiotics, overnight at 37ºC. Single colonies of transformed bacteria were picked and inoculated into 1.5ml power broth supplemented with antibiotics and incubated overnight at 37ºC with shaking at 200 r.p.m.. For autoinduction expression screening, the overnight cultures were diluted (1:50) into 3ml terrific broth overnight express media (TBONEX) supplemented with appropriate antibiotics and incubated at 37ºC for 4 hours and then at 18ºC for a further 18 hours, with shaking at 200 r.p.m. Expression screening with classical IPTG induction using 3ml terrific broth supplemented with appropriate antibiotics were incubated at 37ºC for 4 hours, or until mid-logarithmic phase ($O.D_{600nm}$ ~ 0.5), and protein production induced by the addition of 0.5mM (final concentration) IPTG. Induced cultures were incubated at 18ºC for 18 hours with shaking at 200 r.p.m.. Protein expression is usually assessed at this stage by SDS-PAGE of whole cell lysate from induced cultures *versus* control cultures. In the case of high throughput expression screening, 1.5 ml of each culture was transferred and was harvested by centrifugation at 6000 x g at 4ºC for 30 minutes. The supernatant was decanted and the pellets frozen at -80ºC before a Ni-NTA binding screen. The pellets were thawed and resuspended in 210µl of 50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, 1% v/v Tween 20, pH 8.0 supplemented with 400U/ml DNase type 1 and 1mg/ml lysozyme allowing 30 minutes for bacterial cell lysis before centrifuging at 5000 x g at 4ºC for 30 minutes to clarify the lysate. The resulting supernatant was carefully transferred and incubated with 20µl of Ni-NTA magnetic beads, which were incubated at room temperature for 30 minutes with shaking at 600 r.p.m.. The Ni-NTA beads were separated from the supernatant via magnets, with the resulting supernatant carefully aspirated. The beads were resuspended in 200µl of 50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, 0.05% v/v Tween 20, pH 8.0 with shaking at 600 r.p.m.. for 5 minutes. The supernatant was separated and aspirated as described, the bead

pellet was resuspended as described previously and this process repeated twice. Finally the protein was eluted from the beads by resuspension in 50µl of 50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, 0.05% v/v Tween 20, pH 8.0. The protein composition of the eluants was analysed by SDS-PAGE.

2.3.3. Protein Expression of Kindlin-1 PH domain

Recombinant mouse Kindlin-1 PH domain was, unusually, constitutively expressed; therefore *Escherichia coli* Rosetta (pLysS/RARE) expression strains carrying recombinant pOPINF-*KINDLIN1-PH* plasmids were cultured in Luria-Bertani (LB) media supplemented with ampicillin (50µg/ml) and chloramphenicol (34µg/ml) at 37°C overnight with shaking at 200 r.p.m.. Cultures were diluted at a ratio of 1:1000 into Terrific broth (TB) supplemented with ampicillin and chloramphenicol and incubated at 37ºC with shaking at 200 r.p.m. until saturation. Cells were harvested by centrifugation at 5000 x g for 20 minutes at 4ºC. The supernatant was decanted and the pellet was transferred and frozen at -80°C until use.

2.3.4. Protein Expression of tCid1

The Cid1 project is a collaboration with the research group of Dr. Chris Norbury (Sir William Dunn School of Pathology, University of Oxford, U.K.). We obtained plasmid DNA for expression in *E. coli* from the Norbury laboratory. Thus, BL21 (DE3) pLysS *E. coli* (Novagen) were transformed with pGEX-6P-*tCID1*, which encodes N-terminally GST tagged and N-terminally truncated (Δ31) Caffeine induced death suppressor protein-1 (tCID1) from *Schizosaccharomyces pombe*. Transformants were selected for and cultivated on agar plates supplemented with 50µg/ml Ampicillin and 34µg/ml Chloramphenicol, and incubated at

37°C. Single colonies were picked and inoculated into 10ml LB supplemented with 50µg/ml Ampicillin and 34µg/ml Chloramphenicol and incubated at 37°C overnight with shaking at 200 r.p.m.. Overnight cultures were diluted 1:100 into fresh LB supplemented with 50µg/ml Ampicillin and 34µg/ml Chloramphenicol and incubated at 37°C with shaking at 200 r.p.m.. Cultures were incubated at 37°C until mid-logarithmic phase (A_{600nm} = 0.5), where protein expression was induced *via* the addition of IPTG to a final concentration of 100µM. Induced cultures were incubated at 24°C for 12-18 hours with shaking at 200 r.p.m.. Cells were harvested by centrifugation at 5000 x g for 20 minutes at 4ºC. The supernatant was decanted and the pellet was transferred and frozen at -80°C until use

2.3.5. Expression in Insect Cells

Baculovirus expression systems are becoming increasingly popular and an important tool for the production of recombinant protein for X-ray crystallography. Despite being more experimentally demanding the baculovirus expression systems offer several advantages over *E. coli* one of which is a near native environment for proteins of eukaryotic origin e.g. appropriate chaperones or post-translational modification enzymes. The Baculovirus infects insect cells and for recombinant protein expression the baculovirus used in this work is the *Autographa californica* nuclear polyhedrosis virus (AcNPV). In nature the AcNPV, which infects *Autographa californica* (alfalfa lopper) insect larvae, requires the polyhedrin protein to form an occluded form whereby the virons are encapsulated in a crystalline protein matrix thus providing the necessary protection upon its release. In cultured cells the formation of occlusion bodies is not required for replication and is therefore dispensible. In the case of expressing foreign proteins the polyhedrin protein gene can be replaced in a recombinant AcNPV with the gene for the protein of interest. AcNPV can infect other Lepidopteron

species and for the purposes of recombinant protein expression army worm *Spodoptera frugiperda* cells are used.

2.3.6. Sf9 Cell Culture

Insect cell culture maintenance were carried out using standard protocols (Merrington *et al.,* 1997). *Spodoptera frugiperda* cells (Sf9) adapted to Sf-900 II SFM (Invitrogen) were cultured in GIBCO^{TM} Sf-900 II media (Invitrogen) supplemented with 100µg/ml penicillin and 100µg/ml streptomycin. Cells were grown in flasks at 27°C with shaking at 100 r.p.m. and maintained at a cell density range between $1x10^6$ cells/ml – $1x10^7$ cells/ml, by splitting and diluting the cell culture with fresh Sf900 II media. Cell density was approximated *via* counting cell numbers in a sample volume using a haemocytometer and light microscopy.

2.3.7. Expression Screening in Baculovirus Infected Sf9 Cells

Recombinant baculovirus was generated as described by (Merrington *et al.,* 1997). Briefly, Sf9 cells were cultured in GIBCO^{TM} Sf900 II media (Invitrogen) supplemented with 100µg/ml penicillin and 100µg/ml streptomycin in suspension to a cell density of $1x10^6$ cells/ml and were <3 days old. Approximately $1x10^6$ Sf9 cells were seeded per 35mm well of a 6-well dish in 2ml of Sf900 II media supplemented with 100µg/ml penicillin and 100µg/ml streptomycin and were incubated at room temperature until the cells had adhered into a monolayer. Recombinant baculovirus generation was performed by co-transfecting bacmid DNA and plasmid DNA into 35mm monolayer culture. For each transfection, 5µl of purified mini-prep plasmid DNA encoding the protein of interest that also possesses the ORF1629 baculovirus elements was mixed with 5µl of purified $BAC10_{KO:1629}$ in 100µl of Sf900 II SFM without antibiotics (Solution A). Separately and per transfection, 6µl of CellFECTIN Reagent

(Invitrogen) was diluted into 100µl of Sf900 II SFM without antibiotics (Solution B). The two solutions (A and B) were mixed and incubated at room temperature for 20 minutes to form a lipid-DNA complex. The lipid-DNA complexes were diluted with 800µl Sf900 II SFM without antibiotics and used to replace the media of the 35mm Sf9 monolayer. The cells were incubated in a humidified incubator at 27°C overnight and a further 1ml of Sf900 II SFM without antibiotics was added to each monolayer culture. The cells were incubated at 27°C for a further 5 days. Recombinant virus was harvested from the culture medium (2ml) and Sf9 cell debris clarified by centrifugation at 500x*g* for 5 minutes at room temperature. The virus was stored at 4°C in the dark until use.

2.3.8. Baculovirus Amplification

For the amplification of viral stocks, a suspension culture or monolayer culture of Sf9 cells cultured in Sf900 II SFM supplemented with 100µg/ml penicillin and 100µg/ml streptomycin was infected with viral stock at a multiplicity of infection (MOI) of 0.1 using the following formula;

$$\text{inoculum required (ml)} = \frac{\text{desired MOI (pfu/ml) x (total number of cells)}}{\text{titre of viral inoculum (pfu/ml)}}$$

assuming a conservative expected viral titre from baculovirus generation of 1 x 10^7 (pfu/ml). The virus was harvested 3 days post infection either by separating the cells from the media in suspension cultures by centrifugation at 1000x*g* for 5 minutes at room temperature, or by decanting the media from a monolayer culture and clarifying by centrifugation at 1000x*g* for 5 minutes at room temperature. A viral titre of 2x10^8 pfu/ml (or an amplification rate of 100pfu/cell) is expected.

2.3.9. Protein Expression of Full Length Kindlin-3 in Sf9 Cells

Baculovirus generation and insect cell culture maintenance were carried out using standard protocols (described earlier). Virus-containing supernatant was harvested between 5 and 6 days after transfection. Virus was amplified prior to expression in Sf9 cell suspension in Sf900 II supplemented with 100µg/ml penicillin and 100µg/ml streptomycin at an estimated multiplicity of infection (eM.O.I.) of approximately 0.1. Amplification requires optimum aeration and therefore the amplification culture did not exceed a 20^{th} of the flask volume. Kindlin-3 expression was achieved by infecting suspension cultures of Sf9 at a density of 2 x 10^6 cells/ml in Sf900 II media supplemented with 100µg/ml penicillin and 100µg/ml streptomycin and 1% foetal calf serum (FCS) with amplified recombinant virus at an MOI of 1. Infected cells were harvested 72 hours post-infection by centrifugation at 1000x*g* and stored at -20°C until use, or -80°C for long term storage.

2.3.10. Expression Screening in Cultured Mammalian Cells

Transfection of human embryonic kidney (HEK) 293 derivatives HEK293T (expressing simian virus 40 large T antigen) were transfected with purified DNA as described in (Aricescu *et al.,* 2006). Briefly, HEK293T cells were seeded onto 6-well plates in Dulbecco's modified eagles medium (DMEM, Sigma) supplemented with non-essential amino acids (Invitrogen), L-glutamine and 10% (v/v) foetal bovine serum (FBS, Invitrogen). The cells were allowed to adhere and recover for 6-12 hours by incubating at 37°C in a humidified incubator with 5% CO_2. After cells had adhered, the media was replaced with DMEM supplemented with 2% (v/v) FBS and the cells were transfected with a solution of 5µg of plasmid DNA and 10µl Lipofectamine 2000 (Invitrogen) in 500µl DMEM (Sigma). Transfected cells were incubated at 37°C with 5% CO_2 in a humidified incubator for 3 days

before expression testing. Expression of recombinant protein was assessed using western blotting.

2.4. Protein Purification

In order to biophysically characterize proteins in isolation it is essential that the protein under investigation is highly purified, especially for the study of protein in a crystalline state by X-ray diffraction. With the development of molecular biology, it has become commonplace to create recombinant proteins that possess a fusion partner protein or polypeptide to allow the selective purification of the recombinant protein *via* chromatographic techniques. However, despite highly specific tags proteins may not be sufficiently pure for biophysical study and therefore additional chromatographic procedures must be employed to purify the protein further using the individual physical and chemical characteristics of the protein e.g. molecular weight, surface charge distribution. In this thesis several proteins were purified and studied and each protocol was developed on a case-by-case basis. The protocols used for each protein are detailed below.

2.4.1. Purification of Kindlin-1 Pleckstrin Homology (PH) domain

Frozen cell pellets of Rosetta (DE3) pLysS/Rare expressing recombinant Kindlin-2 PH domain were thawed and resuspended in 50mM Tris-HCl, pH 7.5, 20mM imidazole, 500mM NaCl, 0.2% (v/v) Tween-20, at a ratio of 30ml buffer to 10g of cell pellet, on ice. The resuspended cell pellets were disrupted by sonication and the debris separated by centrifugation at 30,000 x g at 4ºC for 1 hour. The resulting supernatant was decanted and applied at a rate of 0.5ml/min to a pre-equilibrated HisTrap FF (GE Healthcare, 5mL) at

room temperature using an Äkta FPLC system. Once supernatant was applied the HisTrap FF column was washed with 10 column volumes of 50mM Tris-HCl, pH 7.5, 500mM NaCl, 20mM imidazole, and the protein eluted in steps using 50mM Tris-HCl, pH 7.5, 500mM NaCl, 20mM-500mM imidazole. The eluant was fractionated and collected. The protein composition of each fraction was assessed by SDS-PAGE. Fractions containing the PH-domain were pooled and concentrated using a centrifugation filter device (Millipore) with a molecular weight cut-off (MWCO) of 10KDa. The concentrated protein solution was applied to a pre-equilibrated Sephadex S75 (16/60) size exclusion chromatography (SEC) column (GE healthcare) at 4ºC using an Äkta FPLC at a rate of 1ml/min. SEC was performed using 20mM Tris-HCl, pH 7.5, 200mM NaCl, 0.5mM TCEP. The eluant was monitored by absorbance at 280nm and fractionated for collection. Fractions containing homogeneous protein were pooled and concentrated for crystallisation.

2.4.2. Purification of Full-Length Kindlin-3

Frozen baculovirus-infected insect cell (Sf9) pellets expressing recombinant Kindlin-3 were thawed and resuspended with 50mM NaH_2PO_4, pH 7.4, 500mM NaCl, 10mM Imidazole, 1% (v/v) Tween-20, supplemented with EDTA-free protease inhibitor cocktail (Roche) on ice prior to purification. Complete cell lysis was achieved by incubating the resuspension with the detergent. The recombinant protein was purified using three chromatographic steps after the lysate was clarified by centrifugation at 48,000 x g for 1 hour at 4°C. The supernatant was incubated with pre-prepared and equilibrated Nickel sepharose for 1 - 2hours at 4°C with end-over-end rotation. The beads were collected and washed using the batch method, whereby 10 bed volumes of 50mM NaH_2PO_4 pH 7.4, 500mM NaCl, 10mM Imidazole buffer were used to wash the beads. The protein was eluted and collected

using 1-3 bed volumes of 50mM NaH_2PO_4 pH 7.4, 500mM NaCl, 500mM Imidazole. The protein composition of the eluant was assessed by SDS-PAGE, in the first instance, and western blotting using an anti-pentaHis antibody. The eluant containing Kindlin-3 was pooled and buffer exchanged into 20mM Tris-HCl, pH 7.5, 200mM NaCl *via* a series of dilutions into buffer and sample concentrations using a centrifuge protein concentrator (Millipore) with a 50kDa molecular weight cut-off (MWCO). The buffer-exchanged protein solution was applied onto a pre-equilibrated HiTrap heparin HP column (5ml column volume, GE Healthcare). The protein that bound to the column was eluted using a linear NaCl gradient in the same buffer that increased from 0.2M NaCl to 1M NaCl, at a rate of 10mM/ml. The protein composition of the fractionated eluant was assessed by SDS-PAGE and western blotting and those containing Kindlin-3 pooled and concentrated using a centrifuge protein concentrator (Millipore) with a 50kDa MWCO to 0.5ml – 2ml sample size. Finally, the concentrated protein was polished and buffer exchanged using size exclusion chromatography (SEC). The purified protein was applied onto a pre-equilibrated Superdex S200 (16/60) or (10/30) (GE Healthcare) in 20mM tris-HCl, pH 7.5, 200mM NaCl, 1mM DTT at a rate of 1ml/min. The eluant from the column was fractionated and the protein elution monitored using the absorbance at 280nm. The fractions corresponding to a single absorbance peak that resulted from SEC were assessed by SDS-PAGE to determine homogeneity. The purified Kindlin-3 was assessed as >95% pure after this step. The protein was concentrated using a centrifuge protein concentrator (Millipore) with a 50kDa MWCO to ~15mg/ml in the gel filtration buffer and flash frozen in liquid nitrogen and stored at -80°C until used. The protein concentration was assessed spectroscopically using a calculated extinction coefficient (ε) of 109320 M^{-1} cm^{-1} (assuming all Cys residues are reduced).

2.4.3. Purification of tCid1

Frozen bacterial cell pellets were thawed and resuspended in modified HEPES buffered saline, which is 40mM HEPES, pH 7.0, 200mM NaCl supplemented with one EDTA-free protease inhibitor tablet, 2000U of DNAseI and 1mg of hen egg white lysozyme per 4g bacterial pellet (equivalent to 1 litre of culture). The suspension was cooled in a salt-ice bath and the cells disrupted *via* sonication (20 cycles of 5s pulses and 10s cooling periods). The lysate was clarified by centrifugation at 30,000 x g for 1 hour at 4°C and the resulting supernatent incubated with pre-equilibrated glutathione sepharose 4B FF beads (GE healthcare) at 4°C, with gentle mixing, overnight using the batch method. The beads, together with immobilized protein, were separated from the supernatent *via* a gravity flow column (BioRad), and the flowthrough collected for SDS-PAGE analysis. The beads were washed with 20 column volumes of modified HBS and the wash flow-through collected for SDS-PAGE analysis. The washed gluthatione sepharose beads, with adsorbed protein, were incubated with PreScission protease (GE healthcare) at 4°C for 48 hours thus liberating ~80% of tCID1 protein from its fusion partner *in situ*. After 3C treatment, the cleaved product, tCID1, was collected in the flow through. For completeness, the beads were washed with a minimal volume of modified HBS so that the remaining tCID1 protein from the column volume could also be collected. The presence of tCID1 in the eluants together with its purity, were assessed by SDS-PAGE. All eluants containing tCID1 were pooled.

Partially purified tCID1, batch purified and cleaved as previously described, was diluted in 40mM HEPES, pH 7.0, until the final NaCl concentration was equal to 50mM, and concentrated to 1ml using a centrifuge filter device with a 15kDa MWCO. The concentrated tCID solution in 40mM HEPES, pH 7.0, 50mM NaCl, was loaded onto a pre-equilibrated 5ml Heparin HP column (GE Helathcare) *via* an Äkta FPLC at a rate of 1ml/min at room

temperature. The column was washed with a low salt buffer (40mM HEPES, pH 7.0, 50mM NaCl), at a rate of 1ml/min until the A_{280} returned to the baseline and then the protein was eluted using a linear NaCl salt gradient from 0.05M to 2M at a rate of 20mM/1ml at 1ml/min. The flowthrough and salt gradient eluants were fractionated into 1ml samples and protein composition assessed by SDS-PAGE.

Finally, purification of tCID1 to homogeneity was achieved by size exclusion chromatography. Concentrated tCID1 solution was applied onto a pre-equilibrated S75 16/60 *via* an Äkta FPLC at a rate of 1ml/min. Migration of tCID1 was achieved by applying, at a rate of 1ml/min, 20mM HEPES, pH 7.0, 200mM NaCl, 1mM $MgCl_2$. The eluant was fractionated into 2ml fractions and protein composition of those fractions corresponding to a peak from a chromatograph was assessed *via* SDS-PAGE.

2.5. Protein Crystallisation

The determination of the three-dimensional structures of proteins by X-ray crystallography requires the purification of proteins to homogeneity and the formation of highly ordered crystals. The formation of crystalline protein from protein in solution is entropically costly; however these entropic costs are often offset by the enthalpic benefit of forming crystal contacts. The crystallisation of biological macromolecules in solution does not occur spontaneously. Crystallisation relies on the solubility of the protein in solution and the use of other molecules (precipitants) to drive the protein out of solution to supersaturation, i.e. beyond the protein solubility limit, and in the event of a nucleation event causes a phase separation where the protein separates from solution (precipitates) and, in highly favourable circumstances, self assembles into highly ordered crystals. The formation of a supersaturated protein solution is thermodynamically unstable and results in protein

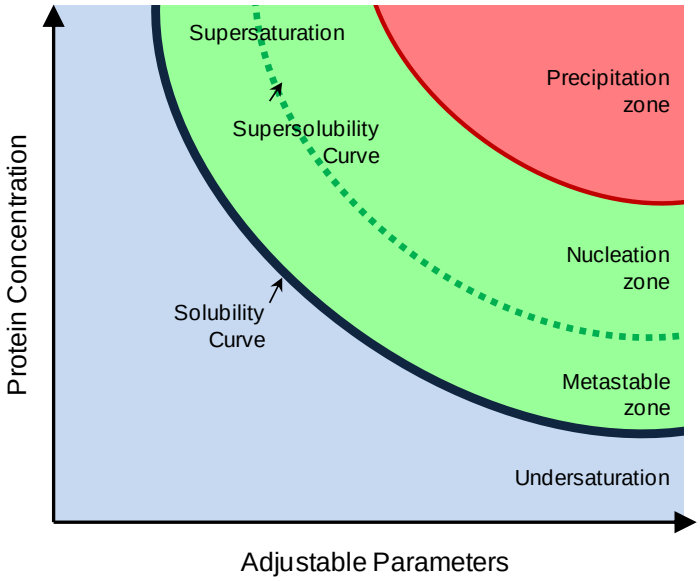


Figure 2.1 – Illustration of the protein solubility phase diagram. Adjustable parameters include precipitant or additive concentration, pH and temperature.

crystal nucleation or precipitation. A crystallisation diagram (Figure 2.1) conveniently illustrates the events that occur during a crystallisation experiment. However, whilst the solubility curve can be experimentally derived, the remainder of the crystallisation diagram is less well defined and is therefore a suggestion of events. The process of protein crystallisation, usually referred to as a bottleneck in crystallography, requires extensive and involved experimental effort to ascertain conditions that allow the protein to crystallise as the exact crystallisation conditions (if any) of proteins cannot be predicted.

There are innumerable reagents that can reduce protein solubility and therefore move a protein solution to a metastable zone. Practically, the search for suitable regents that allow protein crystallisation is relatively straightforward by the use of crystallisation cocktails. Whilst the purpose of the crystallisation cocktail is to reduce solubility, by containing a precipitant, additional components are also present that are beneficial for stabilisation or growth of crystals, including buffering agents, which affect the distribution of charges on the protein surface and therefore protein interactions in crystallisation. The simplest method of protein crystallisation is to mix the protein with the crystallisation cocktail. This method,

appropriately called the *batch method*, allows the direct formation of a supersaturated solution and crystal formation; another approach is crystallisation by dialysis, but the method used exclusively in this thesis is vapour diffusion crystallisation. More accurately described as *vapour diffusion assisted batch* crystallisation, this protein crystallisation method is the most widely used crystallisation method and is frequently implemented in sitting drop or hanging drop techniques and is easily implemented in high throughput crystallisation workflows. The vapour diffusion method makes use of the slow movement of water vapour and other volatile compounds from the batch crystallisation solution (protein solution mixed with crystallisation cocktail) by diffusion into a reservoir solution of higher precipitant concentration. This movement of water from the crystallisation experiment drives the protein crystallisation solution beyond supersaturation. Therefore, in a vapour diffusion experiment, a batch crystallisation solution (protein solution mixed with crystallisation cocktail) is equilibrated against a reservoir of crystallisation cocktail at a higher concentration (typically for screening two-fold) in a closed system.

2.5.1. Crystallization Screening

In response to the almost infinite combinations of chemicals that reduce protein solubility and may allow crystallisation, crystallisation screening schemes have been developed that explore common parameters for crystallisation such as precipitant type and concentration, pH, protein concentration, and temperature. These screens can be extended by the presence of small molecules that influence protein stability or crystal growth. Sparse matrix formulations, which are commercially available, allow the efficient screening of a wide range of popular and effective protein precipitants, including salts (e.g. ammonium sulphate, potassium phosphate, lithium sulphate), polymers (e.g. low or high molecular

weight polyethylene glycols) and organic solvents (e.g. 2-propanethanol, 2,4-methylpentanediol) at a range of pHs.

In this thesis, protein crystallisation screening was performed in the Oxford Protein Production Facility crystallisation laboratory using high throughput workflows developed therein (see Mayo *et al.*, 2005, for instance). Therefore crystallisation screening experiments were performed in a 96-well format. Thus, 100µl aliquots of commercial sparse matrix crystallisation screens, formatted for automated 96-well experiments (listed in Table 2.3), were dispensed into the reservoir chamber of CrystalQuick 96 Well crystallisation plates (Greiner Bio-one) using a Hydra 96 micro-dispenser (Matrix Technologies Ltd, UK) liquid handling robot. 100nl of protein solution was subsequently dispensed as sitting drops onto the raised platform of each crystallisation well in the 96-well crystallisation plate, using a Cartesian Technologies Microsys MIC400 nanodrop crystallization robot (Genomic Solutions). Each sitting drop of protein solution was diluted with an equal volume of crystallization solution (reservoir solution) using the same robot. Finally each 96-well plate was sealed with VIEWseal (Greiner Bio-one) adhesive covers.

The completed crystallisation 96-well plate experiments were stored either at 20°C, in the Homebase automated storage vault (The Automation Partnership, Ltd, UK) or at 4°C, in a Rock Imager 1000 (Formulatrix). Each crystallisation 96-well plate experiment was stored for a maximum of 2 years and each sitting drop was imaged over the course of the experiment using an Oasis 1700 imaging system (Vecco Instruments) at 20°C or the in-built imaging system of the Rock Imager 1000 (Formulatrix) at 4°C.

Table 2.3 – 96-Well Crystallization Screens Formulation

In-House Screen	Commercial Screens	Manufacturer
Block 1	Crystal Screen I	Hampton Research
	Crystal Screen II	Hampton Research
Block 2	Wizard Screen I	Emerald Biosystems
	Wizard Screen II	Emerald Biosystems
Block 3	PEG/Ion	Hampton Research
	Grid Screen 6000	Hampton Research
	Grid Screen Ammonium Sulphate	Hampton Research
Block 4	Natrix	Hampton Research
	Crystal Screen Cryo	Hampton Research
Block 5	Grid Screen PEG/LiCl	Hampton Research
	Grid Screen Sodium Chloride	Hampton Research
	Grid Screen MPD	Hampton Research
	Quick Screen	Hampton Research
Index	Index Screen (Index HT)	Hampton Research
SaltRX	SaltRX	Hampton Research
PACT	PACT Premier	Hampton Research

Commercial Screens are supplied as 24-, 48- or 96-solution kits

2.5.2. Crystallization Optimisation

The formation of crystals (or crystalline material) may not be sufficient for the determination of a protein 3-dimensional structure. Therefore two distinct procedural stages are usually met in protein crystallization (1) coarse screening to identify leads or promising condition, followed by (2) the fine screening of chemical space around the promising condition – or optimization. The most popular and easily explicable method is the grid

expansion method. This technique is the successive sampling of crystallization space using finer grids of the major crystallization cocktail components and pH using the initial lead condition as the starting position. The optimization techniques used in this thesis are described below.

2.5.3. Additive Screening

The use of small molecules in addition to the components of a lead crystallization cocktail is a routine strategy in the OPPF crystallization laboratory for the optimization of crystal growth and represents a type of grid expansion. The Additive Screen (Hampton Research, CA, U.S.A.), is comprised of 96 small molecules that influence protein solubility, protein-protein interactions in crystal growth, or solvent behaviour/characteristics; for example, organic solvents used to decrease the dielectric constant of the solution.

Therefore a standard crystallisation plate experiment was prepared using a single crystallisation solution (lead crystallisation reagent) for each optimisation. The reservoir solution was dispensed using a Robogo (MWG Biotech) single channel liquid handling robot and the nanolitre sitting drops were generated using the Cartesian Technologies Microsys 4000 nanodrop crystallization robot (Genomic Solutions). The non-volatile additive reagents were dispensed as 100nl drops on to the pre-formed nanolitre sitting drops whereas the volatile additives were dispensed directly into the reserviour as a single routine implemented by the Cartesian Technologies Microsys 4000 nanodrop crystallization robot (Genomic Solutions). The Additive Screen reagents were pre-formatted into 96-well plates for use by the crystallization robot by diluting the original screen 5-fold. As with the standard crystallization procedure each 96-well plate was sealed with VIEWseal (Greiner Bio-one) adhesive covers and stored at either at 20°C, in the Homebase automated storage vault (The

Automation Partnership, Ltd, UK) or at 4°C, in a Rock Imager 1000 (Formulatrix) for imaging.

2.5.4. Three-row Optimisation

The stepwise dilution of the crystallisation solution is a popular and extensively used optimisation technique implemented by the OPPF crystallisation laboratory and is used to sample crystallisation space at successively diluted crystallisation condition (and precipitant) concentrations. In this method the original crystallisation reagent is diluted with sterile water in 1% - 3% steps thus adjusting the overall concentration from 100% to 67% (at the most dilute). Therefore the crystallisation solution reservoir is dispensed and concomitantly serially diluted from 100% to 67% in 3% steps, for example, in rows F-H of the 96-well crystallisation plate using a Robogo (MWG Biotech) single channel liquid handling robot and a dilution routine. Nanolitre sitting drops are generated using this 3-row dilution screen and the Cartesian Technologies Microsys 4000 nanodrop crystallization robot, which in this instance optimises the crystallisation buffer-to-protein ratio dispensing and mixing nanolitre crystallisation sitting drops at 1:1, 1:2, 2:1 (protein: reservoir) ratios. As with the standard crystallization procedure each 96-well plate was sealed with VIEWseal (Greiner Bio-one) adhesive covers and stored for imaging.

2.5.5. Micro and Macro Seeding

Seeding is a commonly used method of balancing spontaneous nucleation and crystal growth, whereby the introduction of exogenous nucleation, *via* seeding, promotes crystal growth in a supersaturated solution that is incapable of spontaneous nucleation. The

introduction of suboptimal crystal fragments into a sitting drop, for instance, allows epitaxial growth on existing crystal surfaces. Two seeding techniques were used in this thesis; macroseeding and microseeding. For macroseeding, a Cat's whisker was used to dislodge small crystalline seeds from suboptimal crystals not suitable for diffraction studies (for example needles or clusters) that were introduced into pre-equilibrated crystallization sitting drops *via* 'streaking' the Cat's whisker through the drop. In microseeding (described in Walter *et al.,* 2008), suboptimal crystals were harvested and crushed into tiny fragments using a Seed Bead kit (Hampton Research), according to the manufacturer's instructions. The crystal fragments, or microseeds, were diluted in crystallization solution supplemented with protein and serially diluted. The dilution series of microcrystals was dispensed onto pre-formed sitting drops and the plate experiment equilibrated at either 20°C or 4°C.

2.5.6. Crystallisation of Kindlin-1 PH Domain

Crystallisation screening was performed as described earlier using 100 nl sitting drops of protein solution (5mg/ml - 12mg/ml in 20mM Tris-HCl, pH 7.5, 200mM NaCl, 0.5mM TCEP) diluted with an equal volume of crystallisation solution. Crystals appeared in several polyethylene glycol (PEG) -based conditions in 2-3 weeks at 4°C only. A single condition [20% (w/v) PEG3350, 200mM sodium phosphate] yielded the largest crystals and was optimised by Three-row optimisation. Optimal crystals grew within 16 days, with the largest crystals ~1mm in the longest dimension. Crystals were cryo-protected with mother liquor supplemented with 25%-30% (v/v) glycerol and directly frozen in liquid nitrogen prior to data collection.

2.5.7. Crystallisation of tCid1

tCid1 (in 20mM HEPES, pH 7.0, 50mM NaCl, 1mM DTT) was mixed with a 5-mer of DNA (5'AAAAA3'), UTP and $MgCl_2$ (1:1.2:8:8 molar ratio) and heated to 37°C for 30 minutes prior to crystallisation. Crystals of tCid1 in crystal form I were crystallised in 15% (v/v) Glycerol, 25.5% (w/v) PEG4000, 170mM Ammonium Acetate, 85mM tri-Sodium Citrate, pH 5.6 at room temperature within 5 days and grew to full size (~140µm x ~70µm x ~10µm) within 2 weeks in nanolitre sitting drops by vapour diffusion. Crystals could be optimised by the addition of organic volatiles from an additive screen (Hampton Research). Crystals of tCid1 in another crystal form (crystal form II) were grown in 15% (v/v) Glycerol, 25.5% (w/v) PEG4000, 170mM Ammonium Acetate, 85mM tri-Sodium Citrate, pH 5.6, in nanolitre sitting drops using both microseeding (Walter *et al.*, 2008) and additive screening in tandem. Crystals appeared in sitting drops containing taurine or L-proline after 6 weeks having grown to full size (~400µm x ~70µm x ~70µm). Despite the presence of both DNA and UTP neither was observed in the resulting crystal structures (see Chapter 7). Therefore in order to obtain the UTP complex crystals of crystal form I were soaked in mother liquor supplemented with 10mM UTP, for up to 1 hour at room temperature. Crystals of crystal form I, even optimized, were multiple thin plate-like crystals and were grown in the presence of glycerol, therefore no additional cryoprotectant was required for mounting and freezing in lN_2. Crystals of crystal form II were large single crystals, which required additional glycerol for adequate cryoprotection and were bathed in mother liquor supplemented with 25% (v/v) glycerol.

2.6. Principles of X-ray Diffraction by Protein Crystals

The mathematics of X-ray diffraction by protein crystals and structure determination is well established and therefore an account of the principles of these processes is briefly given below. However, the reader is directed to a number specialised texts (Sherwood and Cooper, "X-rays, Proteins and Crystals"; Rupp "Biomolecular Crystallography: Principles, Practice and Application to Structural Biology"; Hammond, "The Basics of Crystallography and Diffraction"; Messerschmidt, "X-ray Crystallography of Biomacromolecules"; and Sands, "Introduction to Crystallography") for an in depth description of crystallography and, in particular, with the exception of Rupp "Biomolecular Crystallography: Principles, Practice and Application to Structural Biology", the provision of mathematical proofs for the equations used in protein crystallography, which will be assumed in this Chapter.

2.6.1. Protein Crystals, Symmetry and Space Groups

As discussed earlier, the formation of protein crystals is a prerequisite to the determination of the three-dimensional structure of proteins and the process by which crystals are obtained is described in the preceding section. A protein crystal can be simply recreated conceptually by considering protein crystals as comprised of a *motif* ('unit of pattern') that is regularly repeated in space defined by a conceptual array of points, the *lattice*, that describe the geometrical relationship between the motifs. The lattice points in three dimensions can be divided into parallelepipeds, denoted as the *unit cell*. Repetition of the unit cell through translational operations from one lattice point to another generates the lattice and thus the crystal. The dimensions of the unit cell are defined by three independent lengths (edges), *a, b* and *c* and three angles α, β and γ between these edges. Seven distinct crystal systems are

derived from the characteristic symmetry (point group symmetry) and therefore follow specific angle and vector conditions of the cell parameters. The symmetry elements of the crystal systems combined with lattice centring give rise to 14 distinct types of 3-dimensional lattices termed *Bravais* lattices. For protein crystals the motif, termed the *asymmetric unit,* can, through space group symmetry operations, generate the unit cell. The symmetry operators that define the asymmetric units' relationship to one another in the unit cell is used to define the 65 chiral space groups (different arrangements of motifs in the crystal structure) of protein crystals distributed among 11 point group symmetries that and are described in the *International Tables of Crystallography*. Symmetry operators permitted in other (non protein) crystals (mirror planes, for instance) yield 230 space groups (32 point groups).

2.6.2. X-ray Diffraction by Protein Crystals

The diffraction of X-rays from crystals can, mathematically speaking, be treated as if they were being reflected from a plane leading to diffraction maxima, or 'reflections'. The geometry of this reflection event means that the diffracted X-rays make an angle of 2θ with the incident beam but also an angle θ with the reflection plane. Therefore 2θ is the experimentally measured angle between the incident beam and the detected diffracted X-rays. In a crystal X-rays are diffracted, simultaneously, from multiple parallel planes, whereby the X-rays must travel distance *d* between the planes. As electromagnetic waves possess wavelength and phase, when the inter-planar spacing (*d*) is a half-integral number of the incident wavelength then the reflected X-rays will remain in phase and yield constructive interference and are therefore detected. If the X-rays become 'out of phase' then these beams will interfere destructively and will therefore not be observed. *Bragg's* law provides a

relationship between the spacing between parallel planes and reflected X-rays that yield constructive interference:

$$2d\sin\theta = n\lambda \quad \text{(Eq. 2.1)},$$

where λ is the wavelength *d* is the inter-plane spacing and *n* is the order of reflection (an integer). A family of parallel planes can be constructed throughout the crystal that intercepts the lattice points – denoted as lattice planes. It is conventional in crystallography to describe the orientation of the lattice planes by means of their *Miller* Indices, *h, k* and *l*, which are integers. The Miller indices are the number of intercepts of lattice planes with unit cell vectors *a, b* and *c*, which in a general case intercepts the *a, b* and *c* axis at a/*h*, b/*k*, c/*l*. Therefore a second form of Bragg's law is used with Miller indexing,

$$2d_{hkl}\sin\theta = \lambda \text{ (Eq. 2.2)},$$

where the *n* term (order of reflection) is described by the Miller indices. Therefore the Miller indices (*hkl*) denote the plane in the crystal whereby Bragg's law is satisfied.

The crystal lattice (Bravais lattice) in Fourier space can be described by a reciprocal lattice – whereby a reciprocal lattice vector is normal to the lattice plane and its magnitude is equal to $1/d_{hkl}$. The *Ewald* sphere (a sphere of radius 1/λ and centred at distance of 1/λ from the origin of the reciprocal lattice), or *Reflecting* sphere, is an expression of Bragg's law and its relationship to the reciprocal lattice. This construction is the interpretation of the reciprocal lattice and the physical event of X-ray diffraction. A scattering vector between where the incident beam exits the Ewald sphere and where the diffracted beam exits the Ewald sphere is parallel to the reciprocal lattice vector d^*_{hkl} and therefore its magnitude describes the inter-plane spacing ($1/d_{hkl}$). What is more, the relationship between a vector notation expression of Bragg's law and the reciprocal lattice vector is:

$$\frac{|s - s_0|}{\lambda} = d^*_{hkl} = ha^* + kb^* + lc^*$$

(Eq. 2.3)

where, s is the reflected beam and s_0 is the incident beam and therefore $|d^*_{hkl}| = 1/d_{hkl}$ where d is the inter-plane spacing and λ is the X-ray wavelength, and $ha^* + kb^* + lc^*$ a reciprocal lattice vector. The Ewald construction enables the determination of the lattice vector to which the diffraction maximum corresponds. This allows indexing (discussed later) of diffraction spots with integers *h, k* and *l* corresponding to the appropriate reciprocal lattice vectors and thus the real lattice. Therefore, a 2-dimensional image of diffraction is considered a projection of the reciprocal lattice. In general, only a small number of reciprocal lattice points will satisfy Bragg's law and therefore the crystal (and the reciprocal lattice) must be rotated with respect to the incident beam so that different reciprocal lattice points satisfy the diffraction condition. Therefore additional, successive, diffraction images must be recorded where different reciprocal lattice points are brought into the diffraction position.

Structure determination is a two-part process; (1) determination of the unit cell parameters from the diffraction pattern geometry followed by (2) the determination of its lattice type and the arrangement of atoms from the (relative) intensities of the diffraction spots. Therefore the radiation of the reflected X-ray radiation is proportional to $|F(hkl)|^2$, i.e.,

$$I(hkl) \propto |F(hkl)|^2 \quad \text{(Eq. 2.4)}$$

where $I(hkl)$ is the intensity of reflected X-rays and $F(hkl)$ is the structure factor, which is the sum of all individually scattered waves by discrete scattering centres, or atoms, in the unit cell, giving the relationship;

$$F(hkl) = \sum_{n=1}^{n=N} f_n \exp[2\pi i(hx_n + ky_n + lz_n)] = |F(hkl)|\exp[i\alpha(hkl)]$$

(Eq.2.5)

where f_n is the scattering coefficient for the *n*th atom with fractional co-ordinates (x_n y_n z_n) and α(*hkl*) is the phase angle. The component *i* is the imaginary term, where $i = \sqrt{-1}$. The value *f* depends upon the type of atom and the Bragg angle θ. The structure factor F(hkl) is also related to the Fourier transform of real space electron density ρ at positions (*xyz*);

$$F(hkl) = \int_{x=0}^{1}\int_{y=0}^{1}\int_{z=0}^{1} \rho(xyz)\exp[2\pi i(hx + ky + lz)]\,dxdydz$$

(Eq.2.6)

where ρ(*xyz*) is the real space electron density. From this equation the real space electron density can be calculated;

$$\rho(xyz) = \sum_{h}\sum_{k}\sum_{l} F(hkl)\exp[-2\pi i(hx + ky + lz)]$$

(Eq. 2.7)

where equation 2.7 is a Fourier expansion of the electron density function. Unfortunately, the measurable data (diffraction intensities) yield only $|F(hkl)|^2$ which does not routinely lead to structure determination given that Eq.2.7 requires the structure factor F(*hkl*) (Eq. 2.5), whereas in actuality the intensities only give $|F(hkl)|^2$. Given that the square root of the intensities are the absolute values $|F(hkl)|$ of the structure factor (Eq. 2.5), the phase angles α(*hkl*) are lost in the diffraction experiment. This is known as the *phase problem* in crystallography and methods that solve this apparent impasse will be discussed later.

2.7. Data Reduction of Diffraction Data

2.7.1. Indexing

The unit cell dimensions are determined from the reciprocal lattice parameters from the diffraction pattern, which are identified by their miller indices, and the space group deduced from the symmetry of the diffraction. However, particular translational symmetry elements in the crystal lattice, for example screw axes, result in reflections that are said to be *forbidden*, or *extinguished*, or, more commonly termed, *systematically absent*. These special reflection conditions are described in *The International Tables of Crystallography* and help deduce the space group further. In some cases the exact space group cannot be determined by the diffraction pattern alone but is known only after the structure is solved. The diffraction spots were initially indexed by auto-indexing algorithms in *MOSLFM* (Leslie and Powell, 2007) and the Bravais lattice type chosen based on the lowest penalty scoring and highest symmetry lattice. Systematic absences were used to determine the exact space group space in the CCP4 program POINTLESS using the integrated and merged reflections (see below).

2.7.2. Integration of Reflections

The size of the crystal, or number of reflecting planes, has a direct effect on the broadening and intensity of the reflected beam. The Ewald construction considers an infinite lattice (and reciprocal lattice) with infinitely sharp lattice points. However, finite size of a real crystal leads to a broadening of the reflected beam and Scherrer's equation (see Hammond "The Basics of Crystallography and Diffraction"). The reflection from a crystal of finite thickness results in a distribution of intensity, or a broadened reflection. Additional factors also affect spot shape, but these are not discussed here. Both the reflection intensity and the

peak breadth are proportional to the crystal size (this relationship is best described by way of example; see Hammond "Basics of Crystallography and Diffraction"). The crystal size, or thickness, independent value is the area under the diffraction peak, or the integrated reflection and is a measure of total energy of reflection and a measure of $|F_{hkl}|$. Integration was performed after auto-indexing of the diffraction images, where the reflections are integrated using either summation integration or profile fitting in *MOSFLM* or XDS, respectively.

2.7.3. Scaling of Reflections

The measured intensity *I*(hkl) is proportional to the square of the structure factor $|F(hkl)|$. However, systematic effects, such as radiation damage, affect this proportionality. Therefore, in order that we might obtain $|F(hkl)|^2$ from *I*(*hkl*), the measured intensities must be internally consistent and therefore placed onto a common relative scale. This process requires that the symmetry of the diffraction (point-group) be known (dealt with during indexing) and the reflections be integrated. The scaling process compares recorded reflections with symmetry-related reflections and determines a scale-factor that is used to minimize the differences. The scaling process therefore produces scale factors for the correction of intensities and averaging of error estimates and reflection intensities of equivalent reflections. Scaling also provides an estimate of the quality of the data after scaling and averaging, i.e. a measure of internal consistency. Traditionally the measure of internal consistency of scaled data is provided by the merging R-factor (R_{merge}) which is defined as;

$$R_{merge} = \frac{\sum_{\mathbf{h}} \sum_{i}^{N} \left| I_{i(\mathbf{h})} - \overline{I_{(\mathbf{h})}} \right|}{\sum_{\mathbf{h}} \sum_{i}^{N} I_{i(\mathbf{h})}}$$

(Eq. 2.8)

where $\overline{I_{(\mathbf{h})}}$ is the averaged intensity of reflection **h**. This gives a value of the remaining error of multiple observations of the same reflection. However, this quality indicator has the disadvantage that the R value increases with redundancy, despite the merged data being improved by the averaging of a given reflection, or symmetry related reflection, observed multiple times. The use of a precision term $1/(N-1)^{1/2}$ into the merging R value accounts for the increase in the precision of intensities with the merging of more observations. Therefore, the redundancy related precision-indicating merging R-value (R_{pim}) is perhaps a more realistic measure of data quality and is defined as;

$$R_{pim} = \frac{\sum_{\mathbf{h}} \left(\frac{1}{N-1}\right)^{1/2} \sum_i^N \left| I_{i(\mathbf{h})} - \overline{I_{(\mathbf{h})}} \right|}{\sum_{\mathbf{h}} \sum_i^N I_{i(\mathbf{h})}}$$

(Eq. 2.9)

where the precision-indicating R value reduces with redundancy *N*.

The corrected intensity on a relative scale can be placed onto an quasi-absolute scale by applying a 'Wilson plot'. This is particularly useful for integrated intensities of very weak reflections that are below, or close to the background noise and therefore may be negative. The Wilson plot considers the thermal motion of scattering atoms in the crystal, which is defined as a temperature factor, or B-factor.

2.7.4. Resolution Limit of Diffraction

Theoretically, the maximum attainable resolution for a given X-ray wavelength is λ/2. For synchrotron radiation used in structure determination this would yield a resolution of ~0.5Å, which would more than suffice for the determination of protein structures at atomic

resolution. The thermal motion of scattering atoms, i.e. the constant vibration of an atom about its equilibrium position in the crystal, results in diffraction intensity falloff. This becomes more apparent at higher diffraction angles (higher resolution) and thus limits the number of possible measurable diffraction spots that can be recorded. The structure factor equation can be modified to take into account this thermal motion of atoms, thus providing;

$$F(hkl) = \sum_{n=1}^{n=\mathrm{N}} f_n \exp[2\pi i(hx_n + ky_n + lz_n)] \exp\left[-B_n(\sin\theta/\lambda)^2\right]$$

(Eq. 2.10)

where B_n is proportional to the mean square displacement of atom *n*.

The resolution of a diffraction experiment is popularly limited by the ratio between the estimated mean intensity *I* and estimated mean standard deviation of the measurements, σ*I*. Typically this *I*/σ*I* value is 2 or above.

2.7.5. Collection of X-ray Diffraction Data

Diffraction data collection for structure determination was performed using synchrotron radiation. The methods described below refer to the experimental methods used for data collection. Tables of data collection statistics can be found in Chapters 4 and 6, for the Kindlin-1 PH domain and tCid1, respectively.

2.7.6. Kindlin-1 PH domain Data Collection

Diffraction data from a single native Kindlin-1 PH domain crystal were collected at Diamond Light Source (Didcot, UK) on beamline IO3. Diffraction data were indexed and integrated using *MOSFLM* (Leslie and Powell, 2007) and were subsequently merged and scaled using the CCP4 program SCALA (Evans, 2006). The crystal form belongs to the C-centred orthorhombic space group $C222_1$ with unit cell dimensions: a=45.7Å, b = 84.1Å, c= 95.9Å, $\alpha = \beta = \gamma = 90^{o}$ and possesses one molecule per asymmetric unit (solvent content of ~49% with a ~20kDA protein, as determined by Matthews' coefficient).

2.7.7. tCid1 Data Collection

X-ray diffraction data from the two types of native tCid1 crystals together with the UTP soaked crystals were collected under cryo-temperature (~100K) conditions using single wavelength synchrotron radiation at Diamond Light Source (Didcot, UK). Due to the marked difference in size of the crystals it was appropriate to collect diffraction data using two different beamlines. In the case of the thin plate-like crystals the specialised microfocus beamline I24, Diamond Light Source, was used for data collection of native and UTP-soaked crystals with a beamsize of 10μm x 10μm. Data collection was performed by collecting several overlapping wedges of data from adjacent positions within a single mounted crystal. Diffraction data were collected using a Pilatus 6M detector using 0.2° oscillation per image. In the case of the large rectangular rod-like crystals the standard macromolecular crystallography beamline I04 at Diamond Light Source was utilised for data collection at a single wavelength.

The data were initially processed with XDS and SCALA implemented in Xia2 (Winter, 2009). The subsequent merging statistics were used in order to create an optimised subset of data to be merged. Optimal datasets were comprised of a total of ≥ 90 degrees of diffraction data from three to four positions on the mounted crystal. Data from each sweep were combined, merged and scaled in Xia2 (Winter, 2009) to give a final working native dataset at a resolution of 3.20Å and a UTP soaked dataset at a resolution of 3.02Å. In the case of the large crystals, several compete datasets to ~2.6Å were collected from a single frozen crystal at adjacent positions within the mounted crystal. Each dataset was indexed and integrated in *MOSFLM* (Leslie and Powell, 2007) and merged and scaled using the CCP4 program *SCALA* (Evans, 2006). The two crystal forms belonged to different space groups; C2 (form I) and $P2_12_12_1$ (form II), which contained 4 and 2 molecules per asymmetric unit, respectively

2.8. Structure Determination

2.8.1. Methods of Structure Determination

Knowing the electron density ρ(xyz) at every point is equivalent to knowing the crystal structure and therefore $F(hkl)$ could be evaluated by Eq.2.6. Conversely, if $F(hkl)$ is known then we can calculate electron density according to Eq.2.7. As discussed earlier the phase information is lost during a diffraction experiment therefore the experimental reflection intensities represent the square of the structure factor amplitudes ($|F(hkl)|^2$) only. Therefore the phase component of $F(hkl)$ (Eq. 2.5) must be known for the electron density, and therefore crystal structure, to be known. Several methods exist for determining the phase component of the structure factor equation. These methods include, single and multiple

isomorphous replacement (SIR/MIR), multi-wavelength and single-wavelength anomalous dispersion (MAD/SAD) and Molecular Replacement (MR). The former two methods (SIR/MIR or MAD/SAD), rely on the use of heavy atoms or anomalously scattering elements, usually metals, that are either associated to the protein (SIR/MIR) or incorporated into the protein in the form of selenomethionine or by association (e.g. Mn^{2+}) (SAD/MAD). Neither of these methods was used in this Thesis and thus precludes their further discussion here. A brief account of molecular replacement (MR) is given below as this forms the primary method by which the crystal structures presented in this thesis were solved.

2.8.2. Phasing by Molecular Replacement

Molecular replacement (Rossmann, 2001) is a method for phasing the crystal structure when a search model that is structurally similar, or partly similar, to the target molecule in the crystal is available. Molecular replacement requires a search for the proper placement (orientation and location) of the search model in the 'to-be-determined' crystal structure. This placement of the search model (probe) can be achieved using a 6-dimensional search that in practice is reduced to a computationally efficient 3-dimensional rotational search and a 3-dimensional translational search, once the correct rotation is found. This is accomplished by using Patterson methods, whereby the rotational element of the search involves searching distance-vector space for the best intra-molecular vector correlation. The translation search is accomplished using inter-molecular vector correlation. The Patterson function is defined as;

$$P(u,v,w) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l}|F(hkl)|^2\cos[2\pi(hu+kv+lw)]$$

(Eq. 2.11)

where V is the unit cell volume and $|F(hkl)|^2$ is the observed amplitudes. Simply, the Patterson function is an autocorrelation function that is essentially a map of all atom-atom vectors in the crystal and can be calculated with experimentally observed amplitudes. Molecular replacement techniques nowadays have been improved by the implementation of maximum likelihood statistics in their target functions.

2.8.3. Structure Solution of Kindlin-1 PH domain

The structure of the Kindlin-1 PH domain was solved by molecular replacement (MR) using *PHASER* (McCoy *et al.*, 2007) with the PDB coordinates at accession code 2YS3 (Li *et al.*, unpublished) as a search model. The model was converted to a poly-alanine chain in the CCP4 program CHAINSAW (CCP4, 1994) to reduce model bias. Separate MR calculations in *PHASER* were performed using individual models from the (20 model) NMR ensemble. A single model yielding an optimal *Z*-score (RF*Z*=6.3, TF*Z*=6.9) and log likelihood gain (LLG=59) was used as a molecular replacement solution. The MR calculation located a single molecule in the asymmetric unit, consistent with estimates from Matthews' coefficient values.

2.8.4. Structure Solution of tCid1

The Apo structure belonging to space group $P2_12_12_1$ (crystal form II) was solved by molecular replacement using *PHASER* (McCoy *et al.,* 2007) and 2B54 as a search model. Prior to this MR calculation, the search model was modified using the CCP4 program *CHAINSAW* (Stein, 2008) to yield a mixed model of polyalanine and conserved residues based upon a CLUSTALW sequence alignment between the target (Cid1) and the model sequence. The MR calculation located two molecules in the asymmetric unit consistent with the number of molecules in this crystal form suggested by Matthews' coefficient. After molecular replacement the model was edited in Coot (Emsley and Cowtan, 2004) using the electron density as a guide, where secondary structure was clearly visible. The updated model was used as a search model in *PHASER*, which improved the translation function (TFZ) and rotation function Z-scores (RFZ) and log-likelihood gain (LLG) markedly (initial MR, RFZ=5.5, TFZ=5.3, LLG=75; updated model MR, RFZ=15.2, TFZ=31.7, LLG=519). The structures of Apo and UTP-bound tCid1 (in space group C2) were solved by MR in *PHASER* using the previously determined structures. Molecular replacement solutions consisted of 4 molecules in the asymmetric unit, consistent with suggestions based upon the Matthews' coefficient.

2.9. Model Building and Refinement

Given an initial set of phases, either experimentally or computationally derived, the structure factors (Eq. 2.5) can readily be transformed to electron density by means of Eq. 2.7. The next step is to construct a stereochemically plausible atomic model that matches the electron density. Model building and refinement are tightly linked and form a two-step

iterative process of (1) real space model building and refinement and (2) reciprocal space refinement. In real space model building, individual amino acids are fitted into the electron density and the local geometry (i.e. bond lengths, bond angles etc.) is adjusted in real space to bring the model to reasonable geometry. This is cycled with global reciprocal space restrained refinement whereby the model parameters (x,y,z of atoms, B-factors due to thermal vibration of the atoms about its equilibrium position means) together with overall parameters (overall B-factors, bulk solvent correction) are refined against the experimental data. Thus the target is to minimize the residual between the structure factor amplitudes of the model and those of the experimental data. Thus the overall agreement between the model in the asymmetric unit and the experimental data is measured by the R-factor;

$$R = \frac{\sum |F_o(hkl) - F_c(hkl)|}{\sum F_o(hkl)}$$

(Eq. 2.12)

where $F_o(hkl)$ are the observed amplitudes and $F_c(hkl)$ are the calculated amplitudes from the model. Usually, 5% of reflections are excluded and set aside into a test set. The calculation of the cross-validation R-factor, R_{free}, from the 5% of reflections set aside is a measure of model accuracy against the data and is thus used to evaluate the quality of the model.

Minimizing the residual between the experimental and the calculated amplitudes is usually achieved using maximum likelihood target functions in refinement programs. The improved model, and therefore improved available phases, allows the calculation of electron density and therefore allows further model building.

2.9.1. Model Building and Refinement of Kindlin-1 PH domain

After a MR solution was located the model phases were used in the automatic model building program ARP/wARP (Langer *et al.,* 2008). Model building in Coot (Emsley and Cowtan, 2004) was continued from the map output by ARP/wARP and the model was refined using a translation libration screw (TLS) refinement protocol and individual B-factor refinement in Phenix (Adams *et al.,* 2010). TLS groups were determined using the TLS motion determination server (TLSMD) (Painter and Merritt, 2006). Water molecules were added in Coot and updated during rounds of refinement in Phenix. The refinement resulted in an R_{free} and R_{work} of 21.3% and 18.2%, respectively. The crystal structure was validated using MOLPROBITY (Chen *et al.,* 2010) with 98.3% of the refined residues in favoured regions and 100% in allowed regions. The final model contains residues 369 – 497 (mouse *FERMT1* numbering). Weak electron density was observed for residues 382 – 388 which is why they were not built. Tables of refinement statistics can be found along with data collection statistics in Chapter 4.

2.9.2. Model Building and Refinement of tCid1

Model building of the Apo tCid1 structure in space group $P2_12_12_1$ was initially performed after MR calculations in density modified electron density maps. NCS averaging, solvent flattening and histogram matching were performed in the CCP4 program *PARROT* using the MR search model and electron dentistry map for density modification. Model extension and rebuilding with real space refinement was performed using COOT in the density modified electron density maps. Electron density maps were also improved by refining the partial protein model and free atoms in ARP/wARP (Langer *et al.,* 2008).

Refinement of a near complete model was performed in autoBUSTER, using two-fold NCS restraints and individual B-factors. The final model was refined using a decomposition TLS refinement protocol in conjunction with individual B-factor refinement and 2-fold NCS restraints. Each chain was divided into 7 TLS groups according to TLSMD (Painter and Merritt, 2006). Final refinement yielded an R_{work} and R_{free} of 19.3% and 23.7%, respectively. The Ramachandran plot, as determined by MOLPROBITY (Chen *et al.,* 2010) demonstrated that 96.2% of refined residues were in favoured regions and 100% of all residues were in allowed regions.

The structure of Apo tCid1 in another crystal form (C2) was solved using the previously determined structure. In this space group the model was refined using 4-fold NCS restraints and individual B-factor refinement in autoBUSTER (Global Phasing) yielding good R factors (R_{work} and R_{free} of 17.8% and 21.4%, respectively). Final refinement using TLS with a single TLS group per chain yielded a final R_{work} and R_{free} of 17.2% and 20.7%, respectively. Protein model geometry was assessed during rounds of model building and refinement using MOLPROBITY, which scored the final model in the 100^{th} percentile against structures of equivalent resolution (Chen *et al.*, 2010). The Ramachandran plot as determined by MOLPROBITY shows 98.6% of all residues in favoured regions and 100% of all residues in allowed regions.

The UTP bound structure was determined using rigid body refinement of the Apo(I) crystal structure and the UTP-soaked data set. Unbiased electron density (Fc-Fo) for a single UTP molecule per chain was clearly visible. Coordinates for the UTP molecules were taken from HIC-UP database (Kleywegt, 2010) and were fitted in the corresponding electron density in COOT. AutoBUSTER (Global Phasing) was used for refinement analogous to the Apo crystal structure. A single TLS group was used for each chain and yielded a final R_{work} and R_{free} of 17.6% and 20.1%, respectively. The Ramachandran plot for the complex was

determined by MOLPROBITY and shows 98.1% of all residues in favoured regions and 100% in allowed regions. MOLPROBITY also scored this structure in the 100th percentile against structures of comparable resolution.

Tables of refinement statistics for each structure can be found along with data collection statistics in Chapter 7.

2.10. Protein Characterisation by Other Biophysical Methods

2.10.1. Small Angle X-ray Solution Scattering

Whilst X-ray crystallography provides atomic detail of proteins in a low energy conformation imposed by a lattice, small angle X-ray scattering (SAXS), on the other hand, provides structural information of a macromolecule in solution at lower resolution (50Å – 10Å). This technique and its parallels with X-ray crystallography are expertly reviewed by (Putnam *et al.,* 2007) and the theoretical basis of X-ray scattering is discussed at length by (Koch *et al.,* 2003) and therefore a limited account of the technique is given here.

The principle of SAXS is to scatter X-ray photons elastically off molecules in solution (proteins, nucleic acids, etc., in biological SAXS) and to record the scattering intensity as a function of the scattering angle. SAXS is considered a contrast method where the difference between the average electron density of macromolecules in solution and the bulk solvent gives rise to the scattering intensity. The scattered intensity *I* is recorded as a function of the momentum transfer (Q):

$$Q = (4\pi sin\theta)/\lambda$$

(Eq. 2.13)

where 2θ is the scattering angle, λ is the wavelength of the incident X-ray beam and the units of Q are the inverse of Å or nm. The random orientations of the macromolecules in solution results in a radially symmetric (isotropic) intensity distribution, i.e. proportional to the scattering of a single molecule averaged over all orientations. The scattering profile may be written as:

$$I(Q) = 4\pi \int_0^{D_{max}} p(r) \frac{\sin(Qr)}{Qr} \mathrm{d}r$$

(Eq. 2.14)

where D_{max} is the maximum distance in the scattering molecule and p(r) is the pair distribution function, or pair-density distribution function, which is a spherically averaged autocorrelation function. The p(r) function (a spherically-averaged Patterson space structure) can be obtained from the indirect Fourier transform of the scattering profile and is zero when r=0 and also when $r \geq D_{max}$, the maximum dimension of the scattering particle. The p(r) obtained from indirect Fourier transform programs (i.e. GNOM), can be used to calculate *ab initio* reconstitutions of the scattering molecule.

2.10.2. Small Angle X-ray Solution Scattering (SAXS) of Kindlin-3

SAXS data were collected at the EMBL beamline X33 at the DORIS III storage ring, DESY (Hamburg, Germany) (Roessle *et al.,* 2007). The measurements were carried out at 288 K in 20mM HEPES, pH 7.5, 300mM NaCl, 10mM DTT. A MAR345 image plate at a sample-detector distance of 2.43 m and wavelength λ of 0.93 Å, covering the momentum transfer range $0.04 < Q < 5.1$ nm^{-1} was used. The data were processed using standard procedures by the program package PRIMUS (Konarev *et al.,* 2003). SAXS datasets for

Kindlin-3 were collected at the following concentrations: 8.7, 4.35, 2.18 and 1.09 mg/ml. The reduced, rotationally-averaged SAXS profiles were analyzed using the program GNOM (Svergun, 1992). A series of alternative possible maximum dimensions for Kindlin-3 were used to locate a D_{max} defined by the data which also showed a good fit to the experimental scattering curve. Distance distribution functions p(r) were then used to calculate reconstituted models *ab initio* with the program GASBOR (Svergun *et al.,* 2001). Models calculated starting from the premise of an unknown overall shape were consistently elongated, prolate structures and therefore a prolate assumption was used in calculating the structures presented. For those samples showing no or limited inter-molecular scattering (the two lowest concentrations), twenty models were calculated *ab initio* each time, and then they were sorted in pairwise groups of ten, aligned, averaged and trimmed to a consensus core structure using the DAMAVER package (Volkov and Svergun, 2003) running its constituent programs DAMSEL, DAMSUP, DAMAVER and DAMFILT in turn and making use of the route supcomb20 for the alignment procedure which allows enantiomer testing. This produced two independent scattering bead models from each dataset which were then aligned to each other in CHIMERA (Pettersen *et al.*, 2004), converted to CCP4 envelopes using the program GAP (Grimes *et al.,* 1998) and correlated in Fourier space using the program WellMAP (J.F. Flanagan and R.J.C.Gilbert, unpublished). The resolution achieved was estimated based on the 0.5 Fourier shell correlation criterion commonly used in electron microscopy (Grigroieff, 2000), here adapted for SAXS data. The application of this approach to estimate the resolution of the SAXS structures is consistent with the most conservative such methods, in which two wholly independent models from a dataset are compared after calculation (Grigorieff, 2000). The maps were then filtered in WellMAP to the determined resolutions. A theoretical sedimentation coefficient and radius of gyration for the SAXS model was computed using the program HYDRO++ version 10 (Garcia de la Torre *et al.,* 2007). Plots of

I(Q) including a Kratky plot (Kratky *et al.*, 1951) were drawn using ProFit software (Uetikon-am-See, CH). Fits of atomic models to the scattering envelope were performed using the program VEDA (http://mem.ibs.fr/VEDA/).

2.10.3. Analytical Ultracentrifugation

Analytical ultracentrifugation (AUC) is a technique that allows the monitoring of the sedimentation of macromolecules in a centrifugal field and permits their hydrodynamic and thermodynamic characterization in solution and without interaction with a solid phase (i.e. a matrix as in size exclusion chromatography). The technique and its instrumentation has been reviewed by (Lebowitz *et al.*, 2002) and (Howlett *et al.*, 2006) to which the reader is directed. However, a very brief account of the principles of AUC will be given below.

Simply put, the application of a centrifugal force onto a protein solution causes a depletion of macromolecules at the meniscus and results in the formation of a moving concentration boundary that travels toward the bottom of the centrifugation cell as a function of time. The *sedimentation co-efficient*, s, of a macromolecule is given by the *Svedberg* equation;

$$s = \frac{u}{\omega^2 r} = \frac{MD(1 - \bar{\nu}\rho)}{RT}$$

(Eq. 2.15)

where the sedimentation co-efficient, s, is related to, *u*, the observed radial velocity of the macromolecule, $\omega^2 r$ is the centrifugal field defined by ω, the angular velocity of the rotor and *r*, the radial position. The sedimentation co-efficient is also related to the molar mass *M* of the protein and to the diffusion co-efficient *D*. $\bar{\nu}$ is the partial specific volume, ρ is the density of

the solvent, and *R* is the gas constant. The unit of s is given in Svedberg units, which correspond to 10^{-13}sec.

2.10.4. AUC of the Kindlin-1 PH domain

AUC was performed using a Beckmann Optima XL-I analytical ultracentrifuge, in both velocity and equilibrium modes. Velocity data were collected at 40,000 rpm using interference optics and analysed using the software SedFit (by Schuck, http://www.analyticalultracentrifugation.com) in c(s,f/f_0) mode. Equilibrium data were collected using absorbance and interference optics, taking pairs of scans 2 hours apart to ensure equilibrium had been reached, and were analysed using UltraSpin (Altamirano *et al.*, 2001), for three concentrations, at 15,500 rpm allowing extrapolation to correct for sample non-ideality. All AUC experiments were performed at 20°C.

2.10.5. AUC of Kindlin-3 and integrin β-tails

AUC was carried out using a Beckman (Palo Alto, CA) Optima XL-I AUC operating in velocity mode, with 12 mm sector-shaped centrepieces. Data were collected using 280 nm absorbance optics at a speed of 40,000 rpm with one scan being taken every 6 minutes for a total of 50 scans. The 11^{th} to 30^{th} scans were determined as providing the best resolution of the sedimenting species using initial analyses in ProFit software and the data were analysed using the c(s,f/f_0) method whereby the distribution of apparent sedimentation coefficients shown by the sample is plotted over a third dimension, the population of species with diffusion coefficients of specific magnitudes as influenced by the frictional ratio of the

species. Data analysis was carried out using SedFit software with a resolution in s of 100 and a resolution in f/f_0 of 10. SedFit was also used to compute the hydrodynamic radius, the frictional coefficient, and axial ellipsoidal ratio from the AUC data. The frictional coefficient is a unitless quantity which is the ratio between a species' theoretical behaviour given its weight, partial specific volume and the experimental conditions, and its actual sedimentation behaviour. Plots are presented as generated using ProFit software, fitted with a Gaussian function to determine peak centres (which are the species sedimentation coefficients) and peak variances (which are proportional to the diffusion coefficient of the species). All samples were in 20 mM Tris pH 7.5, 200 mM NaCl. Efforts were made to include a stoichiometric amount of each protein when mixtures were studied.

2.10.6. Dynamic light scattering

DLS measurements were conducted on a Protein Solutions DynaPro instrument at 22°C. Approximately, 1µg of protein was diluted in 20mM Tris-HCl, pH 7.5, 100mM NaCl, and 1mM DTT and centrifuged at 13,000rpm for 5 minutes before placing into the cuvette.

2.11. Methods for Protein Interaction

2.11.1. Protein interactions with lipids

PIP Strips™ (Echelon Biosciences Inc.) membranes were incubated with 1% (w/v) non-fat milk in PBS, for 1 hour at room temperature. This blocking solution was replaced with recombinant Kindlin-1 PH domain, or mutant PH domain, or Kindlin-3 protein solution (1.0µg/ml in 1% (w/v) non-fat milk and PBS) and incubated for 2 hours at room temperature.

The protein solution was decanted and the membrane washed with PBS, for 15 minutes, for a total of 3 washes. The membrane was incubated with a mouse anti-penta-His antibody (1:2000 dilution) in 1% (w/v) non-fat milk in PBS for 1 hour at room temperature. The primary antibody solution was discarded and the membrane washed as described earlier with PBS and 0.1% (v/v) Tween-20. The membrane was incubated with goat anti-mouse HRP-conjugated antibody in PBS with 1% (w/v) non fat milk for 1 hour at room temperature. The secondary antibody solution was discarded and the membrane washed as before. Subsequently, the membrane was incubated with Amersham™ ECL™ western blotting detection reagent (GE Healthcare) and then exposed to Amersham™ Hyperfilm ECL™ (GE Healthcare) for variable times (30s – 20minutes) at room temperature.

2.11.2. Size Exclusion Chromatography

Size exclusion chromatography was performed using a Superdex S200 10/30 (GE Healthcare) equilibrated with 10mM sodium phosphate, pH 7.4, 100mM NaCl at room temperature. Proteins and their mixtures were applied to the column with buffer at a rate of 0.5ml/min and an elution profile recorded using absorbance at 280nm. All samples were loaded at equivalent molarities and complex mixtures combined in a stoichiometric amount except for β_{1A} integrin tail mixtures, where a 3x molar excess was used. All samples that were injected were equivalent in volume. Elution profiles were plotted for analysis using ProFit software.

2.12. Comparative Structural Analysis

A comparison between structures that are related *via* their 3-dimensional structure, but demonstrate limited sequence similarity, or divergent sequences, is important for understanding the evolutionary relationship of protein folds.

The program SHP is used for such structural comparisons, in which probability calculation gap-penalties were used in pairwise superposition of proteins. The summed structural correlation, or the sum of probability of equivalence, between pairs of residues for the proteins compared are converted into an estimate of evolutionary distance using the expression,

$$D = -\log(P - 2/\langle N \rangle - 2)$$

(Eq. 2.16)

where D is the evolutionary distance, P is the sum of probabilities and <N> is the mean number of residues in the two molecules (Riffel *et al.*, 2002; Stuart *et al.*, 1979).

2.12.1. Comparative Structural Analysis of the Kindlin-1 PH Domain

A representative selection of 32 PH domains, including Kindlin-1 in this parenthesis, with diverse functions were selected from a list of structural homologues compiled using the protein structure database webserver DALI (Holm *et al.*, 2010). Superimposition of all PH domains was performed using SHP (Riffel *et al.*, 2002; Stuart *et al.*, 1979). A single chain (A) was selected from structures containing more than one chain in the asymmetric unit and the first model only from NMR ensembles. In the case of multidomain structures, the PH domain was excised from the most complete molecule in the asymmetric unit. The

phylogenetic tree was calculated using an evolutionary distance matrix. The tree representation was generated using the programs FITCH and DRAWTREE as part of the PHYLIP package (Felsenstein, 1997). Structural alignments were inspected manually in Chimera (Pettersen *et al.,* 2004) before the phylogenetic tree was constructed.

2.12.2. Comparative Structural Analysis of Cid1

For the structural phylogenetic analysis a selection of noncanonical poly(A) polymerase crystal structures, including the tCid1-UTP complex, canonical nuclear poly(A) polymerases and DNA polymerases, all belonging to the polβ-nucleotidyltransferase family were selected from a list of structural homologues compiled using the protein structure database webserver DALI (Holm *et al.,* 2010). Superimposition was performed using SHP (Riffel *et al.,* 2002; Stuart *et al.,* 1979). A single chain was selected from structures containing more than one chain in the asymmetric unit, including tCid1. The phylogenetic tree was calculated using an evolutionary distance matrix. The tree representation was generated using the programs FITCH and DRAWTREE as part of the PHYLIP package (Felsenstein, 1997). Structural alignments were inspected manually in Chimera (Pettersen, 2004) before the phylogenetic tree was constructed.

2.13 Additional Methods

Methods described in this section were performed in collaboration with other laboratories and are acknowledged where appropriate. The observations resulting from the use of these methods are discussed in the following chapters and are credited in the text where appropriate.

2.13.1. Integrin Activation Assay

The integrin activation assay was performed by N. Brahme in the laboratory of Prof. D.A. Calderwood, Department of Pharamcology, University of Yale, Connecticut, US.

Integrin activation assays were conducted and quantified as previously described (Goult *et al.*, 2009b). Briefly, CHO cells stably expressing αIIbβ3 integrins were transiently co-transfected with GFP- and DsRed-tagged constructs using PEI (Polysciences Inc.). Twenty four hours later, cells were suspended and stained with PAC1 (BD Biosciences) and Alexa 647 fluorophore-conjugated anti-mouse IgM secondary (Invitrogen) in the presence ("inhibited") and absence ("native") of 10mM Ethylenediaminetetraacetic acid. Total surface integrin levels were measured by staining with D57 (O'Toole *et al.*, 1994) and Alexa 647 fluorophore-conjugated anti-mouse IgG secondary (Invitrogen). Mean fluorescence intensities (MFI) of PAC1 and D57 binding was calculated using FlowJo FACS analysis software. The integrin activation index (AI) of GFP- and DsRed-positive cells in each experimental condition was quantified as $AI = (Native_{MFI} - Inhibited_{MFI})/D57_{MFI}$. Statistical significance was calculated using the student's t-test.

2.13.2. Molecular Dynamics Simulations

The MDS were performed by C.N. Lumb in the laboratory of Prof. M.S.P. Sansom, Department of Biochemistry, University of Oxford, UK.

MODELLER (Sali *et al.*, 1993) was used in combination with the MMTSB Toolset (Feig *et al.*, 2004) to place the missing region in the Kindlin-1 PH domain crystal structure, which comprises residue numbers 382 – 388 with sequence Pro-Lys-Lys-Leu-Met-Leu-Lys. A total of 200 models were generated. These were clustered based on the conformations

adopted by the flexible loop formed by residues 382-388. The top cluster was extracted on the basis of the average MODELLER score, and the highest scoring model from this cluster was used as a starting point for the molecular dynamics (MD) simulations. MD simulations were performed with GROMACS version 4.0.5 (Hess *et al.*, 2008), using both the united-atom GROMOS96 43al forcefield (van Gunsteren *et al.*, 1996) and the all-atom OPLS-AA/L forcefield (Kaminski *et al.*, 2001) with a timestep of Δt = 2 fs. The protein was placed in a cubic box with sides of length 60 Å and solvent molecules were subsequently added. The SPC (Hermans *et al.*, 1984) water model was used in conjunction with the GROMOS 43a1 forcefield, while the TIP4P (Jorgensen *et al.*, 1983) water model was used with the OPLS-AA/L forcefield. Chloride ions were added to neutralize the net charge of the system. Temperature was maintained at 310 K by coupling to a Berendsen thermostat (Berenedsen *et al.*, 1984) with a coupling constant of τ_T = 0.1 ps. Pressure was kept constant at 1 atm using a Parrinello-Rahman barostat (Parrinello and Rahman, 1981; Nose and Klein, 1983), with isotropic pressure coupling, a coupling constant of τ_p = 1.0 ps and a compressibility of 4.6 x 10^{-5} bar^{-1}. Bond lengths and angles were constrained using the LINCS algorithm (Hess *et al.*, 1997). Electrostatics interactions were treated with the particle mesh Ewald (Darden *et al.*, 1993) method using a cutoff of 10 Å.

2.13.3. Molecular Docking

The molecular docking into the Kindlin-1 PH domain structure was performed by C.N. Lumb in the laboratory of Prof. M.S.P. Sansom, Department of Biochemistry, University of Oxford, UK.

AutoDock version 4.2 (Morris *et al.*, 1998; Huey *et al.*, 2007) was used for ligand docking. The ligand structures used for molecular docking were extracted from high

resolution X-ray crystal structures deposited in the Protein Data Bank (PDB). The headgroups of six of the seven physiological phosphoinositide (PtdIns) isoforms were used in the docking calculations: Ins(1,3)P_2 (PDB code: 1JOC) (Dumas *et al*, 2001); Ins(1,4)P_2 (PDB code: 1I9Z) (Tsujishita *et al.*, 2001); Ins(1,3,4)P_3 (PDB code: 1Z2P) (Miller *et al.*, 2005); Ins(1,3,5)P_3 (PDB code: 1ZVR) (Begley *et al.*, 2006); Ins(1,4,5)P_3 (PDB code: 1MAI) (Ferguson *et al.*, 1995) and Ins(1,3,4,5)P_4 (PDB code: 1UNQ) (Milburn *et al.*, 2003). Although in total there are seven naturally occurring PtdIns isoforms in eukaryotic cells, to the best of our knowledge no structures with Ins(1,5)P_2 bound have yet been reported. Structures were prepared for docking by adding Gasteiger partial charges and ligand torsions using AutoDockTools. The protein was kept rigid and grid maps were calculated using 80 x 80 x 80 grid points with a spacing of 0.375 Å. Grids were centred such that they encompassed the whole Kindlin-1 PH domain. A Lamarckian genetic algorithm was used for ligand conformational searching and our docking parameters were based on those used in (Li *et al.*, 2004). For each run 100 trial dockings were performed with a population size of 300. Translation step ranges were 1.5 Å and rotation step ranges were 35 º. Elitism, mutation rate, crossover rate and local search rate were set to 1, 0.02, 0.8 and 0.06 respectively, with 25 million energy evaluations and 27000 generations. Starting positions and conformations were randomized, and clusters were evaluated using a tolerance of 1.5 Å root-mean-square deviation (r.m.s.d). To validate the docking results blind dockings were also performed using SwissDock (www.swissdock.ch) (Grosdidier *et al.*, 2011a) using the accurate/default parameter preset (Grosdidier *et al.*, 2011b). Ligands were converted to the SYBYL MOL2 format prior to docking using PRODRG (Schuttelkopf and van Aalten, 2004).

2.13.4. Nuclear Magnetic Resonance (NMR) Spectroscopy

The NMR Spectroscopy was performed by A.K. Fuezery in the laboratory of Prof. I.D. Campbell, Department of Biochemistry, University of Oxford, UK.

All NMR experiments were recorded at 25 °C on a Bruker 500 MHz spectrometer equipped with a z-axis gradient CryoProbe. All spectra were processed with NMRPipe (Delaglio *et al.*, 1995) and analyzed with Sparky (Goddard and Kneller, 2000). Samples were prepared in NMR buffer (18 mM HEPES pH 7.0, 247 mM NaCl, 2 mM TCEP) with 10 % D_2O and 0.1 mM DSS. Samples contained 50 μM β tail (β_{1A} or β_{1D}) and either 0 μM or 50 μM Kindlin-3. Amide ^{1}H and ^{15}N resonance assignments for free β_{1A} were obtained by transferring assignments previously obtained at pH 6.1 (Anthis *et al.*, 2009). While ^{1}H and ^{15}N resonance assignments at pH 6.1 were also available for free β_{1D}, the ^{15}N-HSQC spectrum at pH 7.0 was different enough to make direct transfer of all the assignments difficult. Therefore, an additional ^{15}N-HSQC of free β_{1D} was recorded at pH 6.4 to facilitate the assignment of ^{1}H and ^{15}N resonances at pH 7.0. Kindlin-induced changes in amide peak intensities of the β tails (I/I_o) are reported as the peak height in the presence of Kindlin (I) divided by the height of the same peak in the absence of Kindlin (I_o).

2.13.5. Electrophoretic Mobility Shift Assay (EMSA)

EMSA was performed by S. Fleurdepine in the laboratory of Dr. C. J. Norbury, Sir William Dunn School of Pathology, University of Oxford, UK.

RNA Binding assays were carried out in 10 μL. 10,000 - 30,000 cpm of 5' end-labeled RNA (corresponding approximately to 0.2 pmol) were boiled for 1 min and subsequently incubated for 2 min on ice. 5X buffer was added to final concentrations of 20

mM Tris HCl (pH 8), 150 mM KCl, 2 mM $MgCl_2$, 2.5 mM DTT, 100 µg/mL BSA, 5 % glycerol. The indicated amounts of proteins were added in 1 µL. Binding reactions were incubated at RT for 5 min or on ice for 15 min and then loaded on a non-denaturing Tris-glycine (25 mM- 250 mM) 5% polyacrylamide gel, pre-run for 15 min. Gels were run for 80 min at 170 V (10 V cm^{-1}), dried and exposed to a storage phosphor screen. Radioactivity was scanned with a Fuji FLA-5000 PhosphorImager and results analysed with AIDA image analysis software. Oligonucleotides used for EMSA are listed in Supplementary Table 4.

2.13.6. *In vitro* RNA nucleotidyl transferase assays

Nucleotidytransferase assays were performed by S. Fleurdepine in the laboratory of Dr. C. J. Norbury, Sir William Dunn School of Pathology, University of Oxford, UK.

Polymerisation reactions were set up as the binding reactions except that UTP, or when appropriate CTP, GTP or ATP, was added at a final concentration of 0.5 mM. Reactions were incubated for 15 min at room temperature and stopped by adding 20 µL of STOP buffer (20 mM Tris-Cl [pH 7.5], 0.1 M NaCl, 10 mM EDTA). The RNA was precipitated by adding 2.5 volume of ethanol. Pellets were resuspended in 6 µl of 8M urea loading buffer and separated by 8M urea 8% polyacrylamide gel electrophoresis. Gels were run for 80 min at 40 W, dried and exposed to a storage phosphor screen.

CHAPTER 3

3.1. The Kindlin Family of Proteins Are Co-Activators of Integrins

3.1.1. An Introduction to the Kindlins

The Kindlin family of proteins has recently emerged as a crucial component of focal adhesion assembly and they are therefore now thought essential for complex life. Kindlins are considered to be co-activators of integrins alongside talin (Moser *et al.,* 2009). Integrin-mediated cell adhesion connects the cell surface to the extracellular matrix (ECM) in higher eukaryotes. It is a critical and common process in a plethora of physiological phenomena, which include tissue integrity, embryogenesis, bone metabolism, haemostasis, and immunity. Integrin-mediated cell adhesion is activated through inside-out signal transduction *via* the binding of talin and Kindlin to the integrin β-subunit cytoplasmic tails (CTs).

Kindlins are multi-domain cytoplasmic proteins of approximatey 75kDa. The Kindlin family of proteins is hallmarked by the possession of a bipartite C-terminal FERM (*4*.1 band, *e*zrin, *r*adixin, *m*oesin) domain which is interrupted by a pleckstrin homology (PH) domain at the centre of its F2 subdomain [see Figure 3.1] (Meves *et al.,* 2009, Siegel *et al.,* 2003). The Kindlin FERM domain is considered homologous to the talin FERM domain, although the talin FERM domain is N-terminal and does not possess a pleckstrin homology domain (Goult *et al.,* 2010). Furthermore, Kindlins also possess an N-terminal F0 domain which is homologous to the Talin C-terminal F0 domain – a ubiquitin-like fold (Goult *et al.,* 2010).

The Kindlin family (consisting of 3 members in mammals; Kindlin-1, Kindlin-2 and Kindlin-3) exhibit paralogue-specific tissue expression patterns with Kindlin-1 expressed primarily in the epidermis, but also to a lesser extent the colon, stomach, and kidneys (Ussar *et al.,* 2006); Kindlin-2 ubiquitously expressed but concentrated in striated and smooth

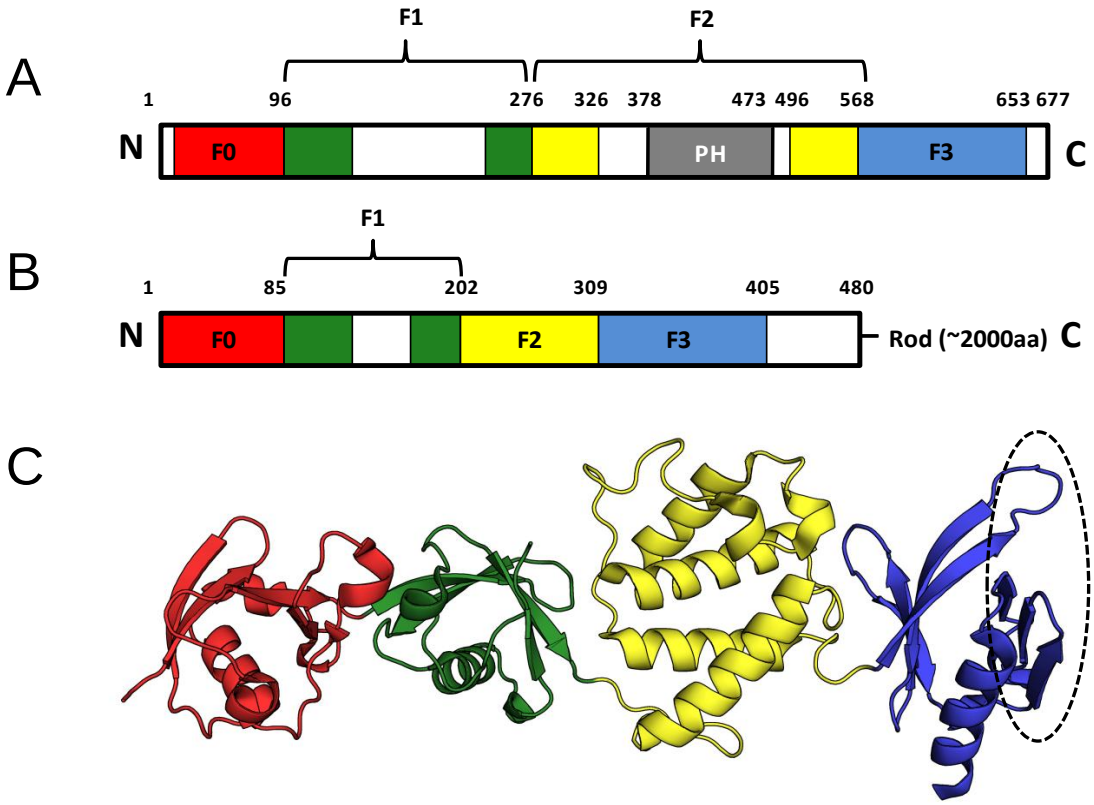


Figure 3.1 – Domain Organisation of Kindlin and talin. (A) Schematic of the domain organisation of the Kindlin proteins illustrating their hallmarking bipartite FERM (4.1, Ezrin, Radixin, Moesin) domain, comprised of subdomains F1 (greend), F2 (yellow) and F3 (blue), which is interrupted by a Pleckstrin Homology (PH) domain (grey) at the centre of the F2 subdomain. The Kindlins also possess an N-terminal F0 domain (red), with a known ubiquitin-like fold. (B) Domain organisation of talin-1 illustrating its N-terminal FERM domain and F0 domain, which are coloured as in (A). Talin-1 also possesses a C-terminal Rod domain of approximately 2000 amino acids (labelled). Domain boundaries for the Kindlin schematic are labelled and denote amino acid numbering from Kindlin-1. Domain boundaries of talin-1 are also denoted. Both are defined and described by (Goult *et al.* 2009b). (C) Cartoon representation of the crystal structure of talin-1 FERM domain [residues 1-400] (PDB code: 3IVF). The β-integrin cytoplasmic domain binding site is highlighted by a dotted oval. Each subdomain is coloured as in (B) and was rendered using PyMol (http://www.pymol.org/).

muscle (Ussar *et al.*, 2006); and Kindlin-3 was considered to be exclusively expressed in the haematopoietic tissues (Ussar *et al.*, 2006). It should be noted that the highest concentration of Kindlin-3 can be found in megakaryocytes (Ussar *et al.*, 2006), but more recent studies have suggested that functional protein is expressed in endothelial tissues (Bialkowska *et al.*,

2010). Interestingly, Kindlin-2 is the only Kindlin protein to be expressed during embryonic development in mice (Meves, *et al.,* 2009). Several studies to identify binding partners of the Kindlins have suggested four cell-ECM adhesion proteins: [1] integrin-linked kinase (ILK) (Mackinnon *et al.,* 2002; Tu *et al.,* 2003; Montanez *et al.,* 2008), [2] migfillin (Tu *et al.,* 2003; Wu *et al.,* 2005; Lai-Cheong *et al.,* 2008), [3] β_1-integrin and [4] β_3-integrin cytoplasmic tails (Kloeker *et al.,* 2004; Shi *et al.,* 2007; Montanez *et al.,* 2008; Moser *et al.,* 2008). Additionally, Kindlin-1 has been shown to bind both α-actinin and focal adhesion kinase (FAK) (Has *et al.,* 2009). It is the interaction between the Kindlins and talin with the integrin β-subunit cytoplasmic tails that is of biological significance, as this results in modulation of integrin activation.

3.1.2. Discovery of Kindlins and Their Role as Integrin Activation Modulators

Prior to the completion of the human genome map, Moerman and co-workers discovered a *Caenorhabditis elegans* gene *unc-112* that, upon knockout, gave a paralysed, arrested elongation at twofold (PAT) phenotype, defined by defective embryonic wall-muscle formation in nematodes (Rogalski *et al.,* 2000). GFP-tagged Unc-112 protein co-localized with β-integrin (β-PAT3) and mutants were demonstrated to alter focal adhesion assembly with mutations in UNC-112 phenocopying the UNC-52/perlecan, PAT2/α_2 (α_2-integrin) and PAT3/β_1 (β_1-integrin) mutation phenotypes (Rogalski *et al.,* 2000). Thus, Rogalski and colleagues discovered an ~80kDa protein involved in integrin mediated cell-adhesion, that on sequence analysis, revealed a FERM domain-containing protein homologous to talin, which was later described as Kindlin-1 (Siegel *et al.,* 2003). However, a human orthologue of Unc-112 protein, Mitogen induced gene-2 (MIG-2), was discovered by Wick and colleagues (Wick *et al.,* 1994), some 6 years prior to the nematode work, during an extensive analysis of

novel proteins involved in G_0/G_1 to S phase progression, in fibroblasts, induced with serum. A more comprehensive study on MIG-2 demonstrated its participation in ECM-adhesion structures in mammalian cells (Tu *et al.,* 2003). Cell biological studies using small-interfering RNA (siRNA) confirmed the role of MIG-2 in the control of cell shape and HeLa cell spreading on fibronectin (Tu *et al.,* 2003). Furthermore, yeast-two-hybrid screens suggested the binding of MIG-2 to the LIM domain-containing protein Migfillin - a protein that tethers focal adhesions to the actin cytoskeleton *via* filamin (Tu e*t al.* 2003).

Mutations in an Unc-112 orthologue gene KIND1 in humans, encoding Kindlin-1, was suggested to be responsible, through familial genetic studies, for the inherited skin fragility disease Kindler syndrome (Jobard *et al.,* 2003, Siegel *et al.,* 2003). This provided evidence for the role of Kindlin-1 in integrin mediated cell-ECM junctions and represented the first *bona fide* skin fragility disorder where disrupted focal adhesions were the causative factor (Siegel *et al.,* 2003). Initially, the Unc-112 human orthologue was denoted as Kindlerin (Jobard *et al.,* 2003) and later denoted as Kindlin-1 (Siegel *et al.,* 2003). Due to this discovery, the earlier name 'MIG-2', an established cell-biology term, was changed to Kindlin-2 due to the 62% amino acid sequence identity to the Kindler syndrome-disease causing protein Kindlin-1 and to avoid confusion with the structurally and functionally unrelated *C. elegans* gene mig-2 (Meves *et al.,* 2009, Siegel, *et al.,* 2003). Another Unc-112 related protein, MIG-2B, was discovered during screening for B-cell specific plasma membrane proteins in Chronic Lymphocytic Leukaemia (CLL) (Boyd *et al.,* 2003), and shares 49% primary protein sequence identity with Kindlin-1. MIG-2B is now known as Kindlin-3 to denote its place in the Kindlin protein family (Meves *et al.,* 2009).

The high degree of sequence similarity of the FERM domain of Kindlin-1 compared to that of talin raised questions about the molecular function of Kindlin-1 and its role in integrin mediated focal adhesions. Kloeker and co-workers provided the first evidence that

Kindlin-1 (denoted Kindlerin) forms a complex with β_{1A} and β_3 integrin cytoplasmic domains (Kloeker *et al.,* 2004). Given the extent of the sequence conservation of the Kindlin-1 FERM domain, particularly the C-terminal region, and talin FERM domain, particularly the PTB domain (F3 subdomain), it was proposed that Kindlin-1 acts analogously to talin in the binding of integrins (Kloeker *et al.,* 2004) – that is the binding of the talin PTB to integrin cytoplasmic tails. The cytoplasmic tails (CTs) of the integrin β-subunits that regulate integrin affinity for extracellular ligands can be subdivided into two parts; a membrane proximal and a membrane distal region. Each contains a conserved NPxY or NxxY motif, respectively (Anthis *et al.,* 2011). The N-terminal F3 subdomain of talin is known to bind to the membrane proximal NPxY (NPKY^{747} in β3) motif resulting in enhanced integrin activation (Anthis *et al.,* 2011, Wegener *et al.,* 2007, Ma *et al.,* 2008). At the molecular level, the interaction of the talin F3 subdomain with the membrane proximal NPxY motif results in the disruption of a clasp that maintains integrins in a low-affinity state (Wegener *et al*. 2007). However, the distal NxxY (NITY^{759} in β3) motif and an intervening T/S (TS^{752}T) cluster are also critical for integrin activation, thus mutational and biochemical analysis indicate that Kindlin-1 and Kindlin-2 directly interact with this membrane distal motif (Harburger *et al.,* 2009). Previously, genetic and biochemical studies suggested that talin is essential for integrin activation and represented a convergent and final common step in integrin activation (Tadokoro, *et al.,* 2003). However, recent studies on integrin activation, that is an integrin with high-affinity for its extracellular ligand, have suggested that talin is not solely sufficient and that the Kindlins are essential for integrin activation (Montanez *et al.,* 2008, Moser *et al.,* 2009). The first demonstration that talin was not wholly and exclusively responsible, as previously thought, came from *in vivo* and *in vitro* studies using knockout mice models which demonstrated inactive integrins in platelets and leukocytes regardless of normal talin protein levels (Moser *et al.,* 2009, Moser *et al.,* 2008). Furthermore, Kindlin-2 knockout blastocysts

demonstrate peri-implantation lethality due to inactive β_1 integrins (Montanez *et al.,* 2008) and overexpression studies of Kindlin-2 and talin results in $\alpha_{IIb}\beta_3$ integrin activation in Chinese Hamster Ovary (CHO) cells (Ma *et al.,* 2008). To date, a consistent model for Kindlin-mediated integrin activation has yet to be determined because the exact sequence of events in talin and Kindlin β-integrin cytoplasmic tail binding that elicit integrin activation is not known.

3.1.3. Kindlin Isoforms and Human Disease

As discussed previously, the Kindlin isoforms are expressed in a tissue specific manner and as a result their loss-of-function, or absence, results in particular phenotypes and diseases. Therefore an overview of Kindlin isoforms and their associated tissues and diseases is given in Table 1.1.

Loss-of-function mutations within the *KIND1* (Kindlin-1) gene in humans result in the rare recessive genodermatosis Kindler syndrome (KS), characterized by transient skin blistering, skin atrophy, variegated hyper-pigmentation and in some rare cases squamous cell carcinoma (Ussar *et al.,* 2008; Arita *et al.,* 2007; Siegel *et al.,* 2003; Jobard *et al.,* 2003). Kindler syndrome was first described by T. Kindler in 1954, where a 14 year old patient presented with traumatic bulla formation (blistering), congenital poikiloderma (thinning of skin and variegated pigmentation) and progressive atrophy of the skin (Kindler, 1954). Subsequent clinical presentations of this nature were classified as a subset of symptoms within the epidermolysis bullosa spectrum (Meves *et al.,* 2009). It wasn't until the discovery of KIND1 mutations that the disease Kindler syndrome was recognized as a separate disease due to the unique molecular findings (Siegel *et al.,* 2003; Jobard *et al.,* 2003). Additionally,

Table 3.1. - Kindlin Expression, Localisation and KO Phenotypes

Kindlin-1	Kindlin-2	Kindlin-3
Localisation: epithelial cells, especially epidermal keratinocytes	**Localisation:** ubiquitous but concentrated in smooth and striated muscle	**Localisation:** mostly haematopoietic cells, highest abundance in megakaryocytes
KO Phenotype: Skin fragility and abnormalities, ulcerative colitis (neonatal lethality in mice due to intestinal epithelia shearing from ulcerative colitis)	**KO Phenotype:** embryonic lethal (Knockdown: abnormal cardiac ultrastructure, disrupted intercalated discs)	**KO Phenotype:** bleeding (abnormal platelet aggregation), immune defects, erthrocytopenia, abnormal erythrocyte shape, osteopetrosis,
Human Diseases: Kindler Syndrome – skin blistering, ulcerative colitis, premature aging, squamous cell carcinoma, poikiloderma	**Human Diseases:** Cardiomyopathies?	**Human Diseases:** Leukocyte Adhesion Deficiency type-III (LAD-III) – Glanzmann's thrombasthenia-like bleeding disorders, recurrent infections, osteopetrosis

due to the expression of a Kindlin-1 splice-variant in gastrointestinal tissues, KS-associated ulcerative colitis can also present in patients with KS (Kern *et al*. 2007). Interestingly, Kindlin-1 transcription is significantly up-regulated in 70% of colon and 60% of lung carcinomas in humans (Weinstein *et al.*, 2003).

The loss of Kindlin-2 in murine models resulted in embryonic lethality, particularly before implantation of the blastocyst (Moser *et al.*, 2009). However, knock-down of Kindlin-2 in zebrafish has revealed an important role for Kindlin-2 in cardiac myocyte differentiation and morphogenesis (Dowling *et al.*, 2008b). Localisation and morpholino knock down studies of Kindlin-2 in zebrafish suggest a role for Kindlin-2 in costameres and intercalated disc formation (Dowling *et al.*, 2008a). Furthermore, knockdown Kindlin-2 hearts display disrupted intercalated discs and improper attachment of myofibrils to adhesion membrane

complexes, with the hearts exhibiting ventricular hypoplasia, suggesting a secondary role for Kindlin-2 in cardiac function (Dowling *et al.,* 2008a). Importantly, Kindlin-2 possesses a nuclear localization sequence (NLS) (Meves *et al.,* 2009) and can be observed in the nucleus of smooth muscle cells (Kato *et al.*, 2004). What is more, Kindlin-2 is known to interact with the LIM domain containing protein, migfillin, that has been shown to shuttle between the cytoplasm and the nucleus and interact with the cardiac homeobox transcription factor CSX/NKX2-5, which is important for vertebrate cardiac development (Dowling *et al.,* 2008a; Akazawa *et al.,* 2004). It has been suggested that Kindlin-2 may have a role in cardiac disease in humans as several mutations have been found in genes encoding intercalated disc proteins in patients presenting with dilated cardiomyopathy and arrhythmogenic right ventricular cardiomyopathy (Dowling *et al.,* 2008a; van Tintelen *et al.*, 2007, Perriad *et al*., 2003). Moreover, mutations in the known Kindlin-2 binding partner integrin-linked kinase (ILK) have been reported in patients with dilated cardiomyopathy (Knoll et al. 2007).

Studies show Kindlin-3 knock-out mice (*KIND3*$^{-/-}$) suffer from severe bleeding due to inactive platelet integrins, which also results in fatal anaemia because of additional erythrocytopenia (Moser *et al.,* 2009). Bleeding disorders in murine *KIND3*$^{-/-}$ embryos gave initial insight into the role of Kindlin-3 in thrombosis (Moser *et al.,* 2008). Further studies using *KIND3*$^{-/-}$ platelets taken from chimeric mice demonstrated that platelets could not spread normally, but adherence could still be achieved by the addition of exogenous thrombotic agents, for example Mn^{2+}, which forces integrin activation and platelet adherence to fibronectin (Moser *et al.,* 2008). Mutations in Kindlin-3 in humans result in the bleeding disorder, leukocyte adhesion deficiency (LAD) type-III, due to inactive leukocyte and platelet integrins (Svensson *et al*. 2009). Individuals with LAD-III suffer from Glanzmann's thrombasthenia-like bleeding disorders and recurrent infections, both of which are life threatening (Svensson *et al.,* 2009). The abnormal bleeding, or rather the lack of clotting,

phenotype in LAD-III resembles that of the human bleeding disorder, Glanzmann's thrombasthenia, which is caused by mutations in $\alpha_{IIb}\beta_3$ (in either subunit) resulting in no integrin activation (Meves *et al.,* 2009). Additionally, a specific human β_3-integrin mutation, Serine^{752} to Proline which lies within the putative Kindlin-3 binding site, also results in Glanzmann's thrombasthenia (Meves *et al.,* 2009, Moser *et al.,* 2008). The recurrent infection phenotype results from improper leukocyte migration, again due to inactive integrins, specifically integrin β_2. More recently, biochemical and cell biological data have revealed the role of Kindlin-3 in podocyte formation in osteoclasts, which is essential for bone resorption (Schmidt *et al.,* 2011). This, in part, explains the observed osteopetrosis found in Kindlin-3 deficient mice and in human LAD-III sufferers (Schmidt *et al.,* 2011, Moser *et al.,* 2009). Furthermore, Kindlin-3 is also an essential factor for the erythrocyte skeleton and therefore red blood cell (RBC) function, despite an absence of integrins in these cell types (Krüger *et al.,* 2008).

Clearly, the Kindlins play vital roles in wider physiological events that are essential for metazoan survival and development. What is most unclear is how the three Kindlin isoforms bind to, and activate, a range of integrins. Clearly the Kindlin isoforms are not redundant and therefore extensive mechanistic insight is required to clarify how each Kindlin recognises its range of integrin ligands and how that, in turn, results in integrin activation.

3.1.4. Kindlins and Outside-In Signalling

Kindlin-3 deficient platelets, as described earlier, display defective adherence due to inactive integrins (Moser *et al.,* 2008). However, *KIND3*$^{-/-}$ platelets with exogenously activated integrins display improper cell spreading but normal adherence to fibronectin (Moser *et al.,* 2008). This was echoed by additional studies on Kindlin-2 and endodermal

cells, in which loss of Kindlin-2 impaired integrin activation (inside-out signalling) as expected, but also disrupted focal adhesion assembly when integrins were exogenously activated resulting in abnormal cell spreading (Montanez *et al.,* 2008). It has therefore been suggested that Kindlins play a role in integrin outside-in signalling, as signal transducers, which elicits cellular responses to integrin ligand binding i.e. cell spreading. The discovery of Kindlin-2 binding to integrin-linked kinase (ILK), a crucial component of focal adhesions, may play a significant role as ILK is known to participate in integrin outside-in signalling (Legate *et al.,* 2009). Thus, Kindlins seem to play a role in signalling as result of integrin extracellular ligand binding that, perhaps through ILK, results in adaptive cellular outcomes i.e. proliferation or cell spreading.

3.2. High-throughput Cloning and Expression Strategies of murine Kindlin-1 and Other Isoforms

3.2.1. Summary

As discussed above, the Kindlin family of proteins has recently emerged as a critical component of focal adhesions and therefore as essential for complex life. We set out to obtain a structural understanding of how this novel class of proteins functions using crystallography. Sufficient protein quantity and quality is a pre-requisite for structural studies. Kindlin-1 was already known to be recalcitrant to over-expression in *E. coli* and therefore a high-throughput (HTP) structural genomics strategy was employed for construct generation and screening of Kindlin proteins. Initial recombinant Kindlin-1 expression screening was performed in *E. coli* with the majority of constructs (36 of 48) yielding soluble expression. However, expression in *E. coli* typically produced low quantities (estimated <0.3mg/L culture) but we found that

PH domain constructs could be expressed in *E. coli* in milligram quantities. To improve yield, a second round of HTP cloning and expression screening was employed using other Kindlin isoforms (Kindlin-2, and -3) in addition and extending the variety of tags. Expression screening in *E. coli* in the first instance did not provide a satisfactory quality or yield of protein; therefore additional hosts were screened with a subset of constructs focussing on His-tagged constructs. Transient mammalian expression in HEK293T cells proved unsuccessful. Baculovirus-infected Sf9 cell expression systems demonstrated a marked improvement over the other hosts tested, yielding milligram quantities of protein. Here, we have established appropriate expression conditions for obtaining recombinant Kindlin protein for further biophysical studies. This represents significant progress in the Kindlin biology field as the limited production of recombinant protein in other laboratories has been reported to impede biochemical functional studies (Harburger *et al.,* 2009).

3.2.2. Kindlin-1 (*FERMT1)* Cloning and Expression

Kindlins are novel multi-domain adaptor proteins and are a recently discovered integrin activator, essential in cell adhesion. Previous attempts to express adequate quantities of soluble recombinant protein for extensive biophysical studies, including structure determination, have been extremely difficult (Prof. Iain Campbell, pers. comm.). Therefore, a high-throughput cloning and expression approach was used in order that we might generate sufficient quantities of these proteins, and/or truncation mutants containing combinations of their sub-domains, for structural studies.

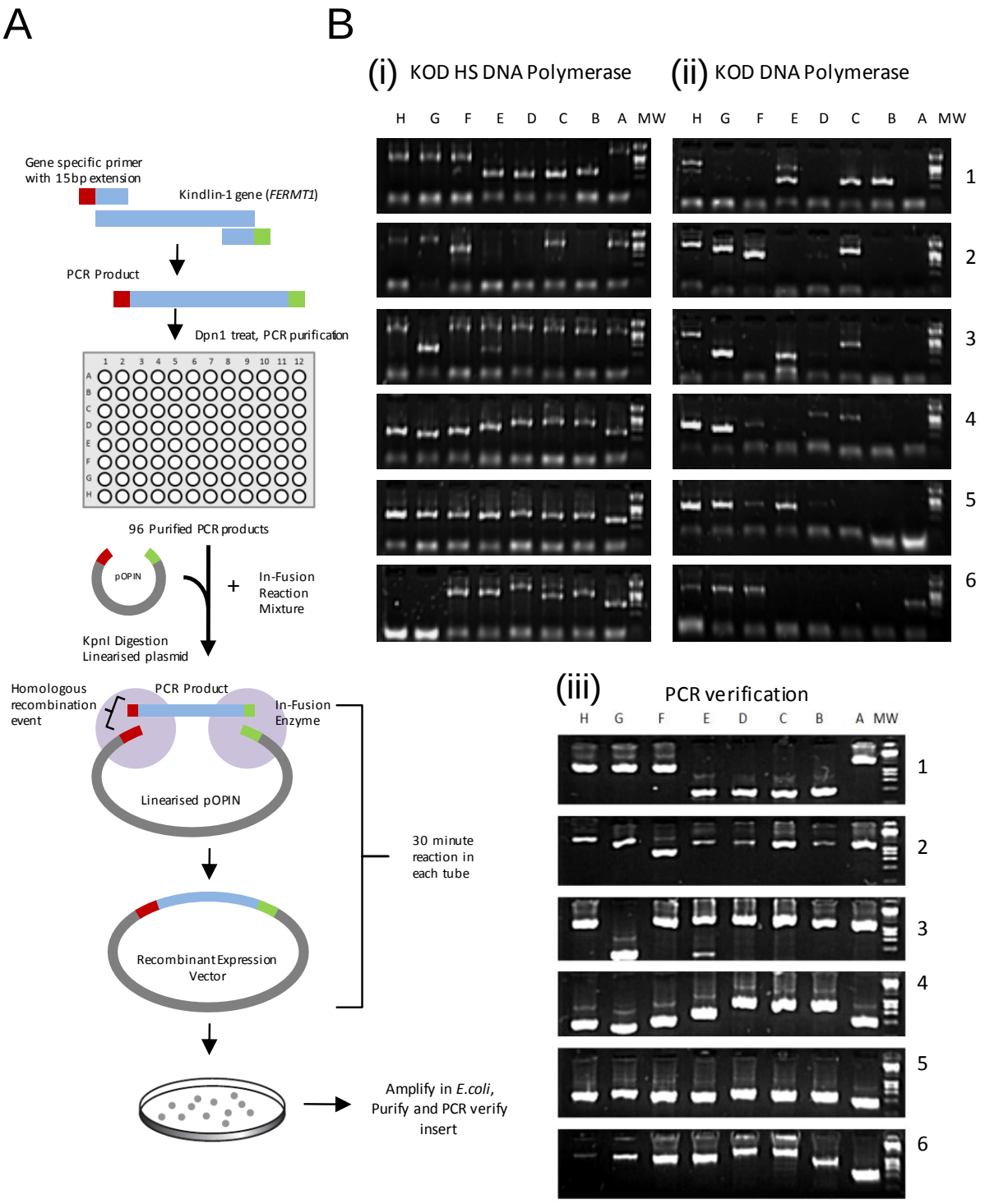
Owing to the sequence conservation across all Kindlin isoforms it was decided that Kindlin-1, the classificatory founding member, was an appropriate target in the first instance. Based on predicted domain boundaries from the annotated gene sequence (UniProt database)

and disorder prediction (www.oppf.ox.ac.uk/RONN) we designed 48 constructs of full length Kindlin-1 and other truncation mutants possessing different combinations of domains (see Table 3.2). We generated 48 *FERMT1* fragments by PCR from template murine Kindlin-1 (*FERMT1* gene; a kind gift from N.Anthis, Biochemistry, University of Oxford), in duplicate, using several high-fidelity DNA polymerases under stringent annealing conditions. Forward and reverse primers for the PCR amplification were designed, using the web based tool OPINE (www.oppf.ox.ac.uk), to possess a uniform annealing temperature of 60˚C thus allowing a single plate-based PCR reaction to be performed with a single cycling method. Furthermore, a 15bp extension to the primer, homologous to a specific site in the target vector, would allow the use of ligation independent cloning (LIC) (see Figure 3.2). In this instance, PCR amplification of fragments of *FERMT1* for expression plasmid construction was achieved using the high-fidelity KOD DNA polymerase from the *Thermococcus kodakaraensis* where greater accuracy is conferred by the 3'-5' exonulcease (proof-reading) activity. PCR reactions were assessed by agarose gel electrophoresis with a 1 kilobase ladder standard used to assess the appropriately sized products (Figure 3.2). In the first instance so-called 'Hot-Start' KOD DNA polymerase (EMD4Biosciences, US), complexed with two monoclonal antibodies to reduce polymerase and exonuclease activity at ambient temperatures, generated >90% of the 48 products expected compared with the ~50% successful reactions using the un-complexed KOD DNA polymerase. PCR products that were absent were repeated as before with additional reactions using Bio-X-Act™ DNA Polymerase (Bioline), with a reduced annealing temperature of 55ºC. A single plate containing a full complement of the 48 PCR products was constructed prior to further experiments maintaining their original PCR reaction positions – a positional arrangement that was retained throughout the remainder of the plate-based experiments.

Table 3.2 – Kindlin-1 Construct Design for HTP Cloning and Expression Screening

OPPF Number	Residue Range	Domains Composition	MW (Da)	Position
5498	2 – 677	F0 – F1 - F2' - PH - F2" - F3	78950	A1
5500	564- 677	F3	15312	B1
5501	564 – 660	F3	13398	C1
5502	564 – 662	F3	13641	D1
5503	564 – 671	F3	14740	E1
5504	285 – 677	F2' - PH - F2" - F3	46903	F1
5505	264 – 677	F2' - PH - F2" - F3	49512	G1
5506	259 – 677	F2' - PH - F2" - F3	50229	H1
5507	255 – 677	F2' - PH - F2" - F3	50717	A2
5508	252 – 677	F2' - PH - F2" - F3	51166	B2
5509	238 – 677	F2' - PH - F2" - F3	52577	C2
5510	227 – 677	F2' - PH - F2" - F3	53778	D2
5511	217 – 677	F2' - PH - F2" - F3	54802	E2
5512	285 – 564	F2' - PH - F2"	33878	F2
5513	131 – 677	F1" - F2' - PH - F2" - F3	64281	G2
5515	157 – 677	F1" - F2' - PH - F2" - F3	61178	H2
5517	185 – 677	F1" - F2' - PH - F2" - F3	58303	A3
5519	208 – 677	loop - F2' - PH - F2" - F3	55790	B3
5520	2 – 660	F0 - F1 - F2' - PH - F2" - F3	77036	C3
5521	2 – 671	F0 - F1 - F2' - PH - F2" - F3	78378	D3
5522	6 – 677	F0 - F1 - F2' - PH - F2" - F3	78606	E3
5523	24 – 677	F0 - F1 - F2' - PH - F2" - F3	76654	F3
5524	364 – 479	PH	15532	G3
5526	2 – 564	F0 - F1 - F2' - PH - F2"	66056	H3
5527	364 – 498	PH	17754	A4
5529	380 – 671	PH - F2" - F3	35626	B4
5530	380 – 677	PH - F2" - F3	36197	C4
5531	380 – 660	PH - F2" - F3	34283	D4
5532	364 – 564	PH - F2"	25049	E4
5533	364 – 509	PH	18839	F4
5534	368 – 479	PH	15073	G4
5535	368 – 509	PH	18381	H4
5536	368 – 498	PH	17296	A5
5537	475 – 660	F2" - F3	23332	B5
5538	475 – 671	F2" - F3	24675	C5
5539	475 – 677	F2" - F3	25246	D5
5540	479 – 660	F2" - F3	22901	E5
5541	479 – 671	F2" - F3	24243	F5
5542	479 – 677	F2" - F3	24815	G5
5543	498 – 677	F2" - F3	22633	H5
5544	509 – 677	F2" - F3	21521	A6
5545	2 – 380	F0 - loop - F2'	45198	B6
5546	2 – 659	F0 - F1 - F2' - PH - F2" - F3	76908	C6
5547	2 – 662	F0 - F1 - F2' - PH - F2" - F3	77279	D6
5548	285 – 659	F2' - PH - F2" - F3	44861	E6
5549	285 – 660	F2' - PH - F2" - F3	44989	F6
5550	285 – 662	F2' - PH - F2" - F3	43246	G6
5551	285 – 671	F2' - PH - F2" - F3	46332	H6

Figure 3.2 – (Opposite). (A) Schematic of HTP In-Fusion cloning together with (B) agarose gel electrophoretic analysis of (i) *T. kodakaraensis* hot start high fidelity DNA polymerase and (ii) *T. kodakaraensis* DNA polymerase PCR products using *KIND1* (*FERMT1*) as template cDNA. PCR products were DpnI treated to remove methylated template and purified, using Agencourt AMPure® beads and the manufacturer's instructions, before adding linearised pOPIN vector and In-Fusion enzyme. The 96 newly generated recombinant vectors were used in the transformation of cloning-grade *E.coli for* DNA amplification. HTP purification was performed using plate based Qiaprep Miniprep Kit and robotic liquid handling according to manufacturer's instructions. Detection of the incorporated PCR fragments into the amplified and purified pOPIN vector was verified by PCR and agarose gel electrophoresis (**c**) using gene specific reverse and T7 specific forward primers, thus presence of PCR product suggests incorporation. Original PCR plate positions were maintained at all times and lanes of agarose gels (**i**), (**ii**) and (**c**) were labelled according to their original position within the 96-well plate experiment and relate to Table 3.2 column 5 (Position). Electrophoresis was achieved with 1.5% (w/v) agarose gel using 100mV.


A
B
(i) KOD HS DNA Polymerase
(ii) KOD DNA Polymerase
H G F E D C B A MW
H G F E D C B A MW
1
2
3
4
5
6
Gene specific primer with 15bp extension
Kindlin-1 gene (FERMT1)
PCR Product
Dpn1 treat, PCR purification
96 Purified PCR products
pOPIN
+
In-Fusion Reaction Mixture
KpnI Digestion Linearised plasmid
Homologous recombination event
PCR Product
In-Fusion Enzyme
Linearised pOPIN
30 minute reaction in each tube
Recombinant Expression Vector
Amplify in E.coli, Purify and PCR verify insert
(iii) PCR verification
H G F E D C B A MW
1
2
3
4
5
6

Kindlin-1 DNA fragments generated by PCR were fused into the pOPIN vector, pOPINF, using homologous recombination catalysed by the InFusion enzyme (Clontech) - in a single plate-based experiment. The target vector, pOPINF, encodes an N-terminal hexa-histidine (His) tag interrupted by a 3C cleavage sequence so that the tag can be removed during downstream processing. An N-terminal his-tag was selected in the first instance as genetic, biochemical, and cellular data revealed that the most C-terminal domain is responsible for integrin activation, through protein-protein interactions, and a C-terminal His-tag could potentially interfere with future biophysical binding studies. Additionally, the pOPINF vectors, and related vectors in the suite, possess a β-galactosidase gene that is interrupted upon insertion of a PCR product. Therefore, blue-white screening using X-gal and IPTG supplemented agar plates, after transformation of *E. coli* with In-Fusion reaction products, was used to assess success rate. We observed that all bacteria for all forty eight reactions exhibited >90% white colonies, indicating that all the PCR products were successfully incorporated into the pOPINF vector. Those white colonies for each plasmid were propagated in liquid culture and the DNA extracted and purified using standard molecular biology techniques adapted for a liquid handling robot. The presence of incorporated DNA fragments into the final plasmid was verified using PCR, i.e. using a gene of interest sequence specific primer and a general plasmid specific primer (T7 promoter specific). The presence of appropriately sized products indicated that >98% of the In-Fusion reactions were successful.

Small scale expression studies were performed using Rosetta pLysS/Rare BL21 *E. coli* in liquid culture, transformed using purified plasmid DNA. All transformations and subsequent cultures were performed maintaining the original PCR positions (see Table 3.2). The study was parallelized to explore protein expression with classical IPTG induction methods, and the auto-induction media TBONEX (Novagen). Soluble protein expression was

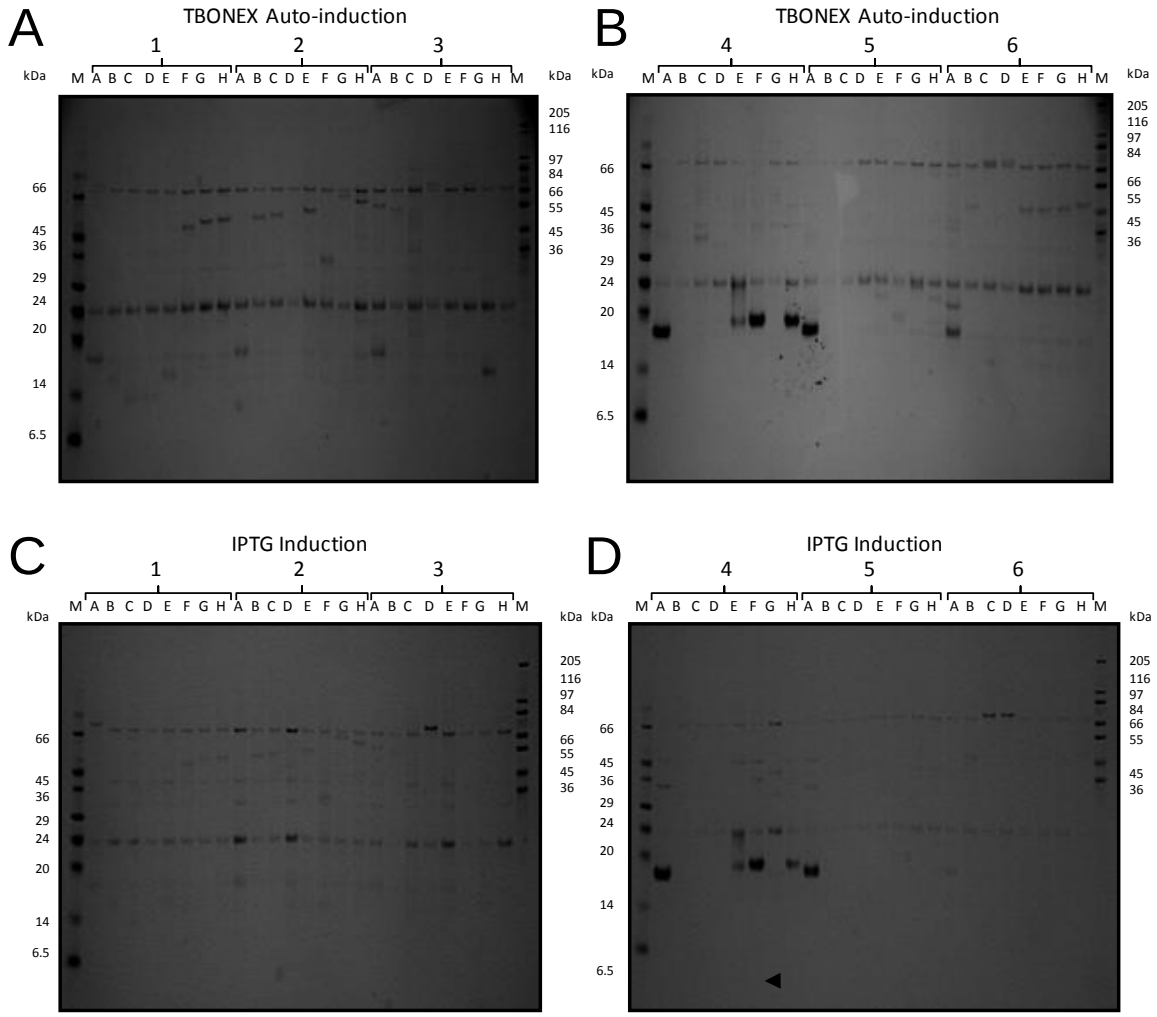


Figure 3.3 – SDS-PAGE analysis of soluble expression screening. Small (3ml) cultures of transformed Rosetta (DE3) pLysS/Rare were induced *via* (A, B) auto-induction media or (C, D) IPTG induction, and the cells lysed and clarified. His-tagged protein was purified using magnetic Ni-NTA beads. Lanes were labeled according to the original PCR positions within a 96-well plate. SDS-PAGE analysis was performed using a 10% Tris-Glycine gel, electrophoresis was performed at 100mV in MOPS buffer. Proteins were visualized *via* SimplyBlue Safe stain (Invitrogen). Lanes are labelled A – H and 1 – 6 and refer to PCR plate positions as in Figure 3.2 and are also listed in Table 3.2, column 5 (Position). Construct F4 (arrowed) was used for subsequent studies (see Chapter 4).

determined by SDS-PAGE analysis of purified his-tagged protein extracted *via* small scale immobilized metal affinity chromatography (IMAC) of clarified bacterial lysate. Such a study suggested that >75% of the constructs used to transform Rosetta (DE3) pLysS/RARE strain of bacteria yielded soluble expression (Figure 3) after induction. Further to this, auto-induction media – Overnight Express^{TM} media (Novagen) increased the intensity of the band in the Coomasie stained acrylamide gel, suggesting a greater yield of protein, assuming that pipetting errors weren't the cause of such observations.

The high-throughput (HTP) expression screening revealed that the Pleckstrin homology (PH) domain constructs, i.e. those constructs within the amino acid range 364 – 509, over-expressed in *E. coli* and were therefore attractive targets for large scale expression, purification and crystallisation trials, in the first instance. The largest of the PH domain constructs readily expressed and was purified, crystallised and the structure solved (see Chapter 4). All other constructs expressed minimally (estimated <0.5mg/L culture) and were comparable to one another (Figure 3.3). Owing to the severe lack of any biophysical data on Kindlins, it was agreed that any recombinant purified protein that could be obtained would be of great value and therefore full length constructs, despite minimal expression, were selected for large scale up experiments – up to 10L cultures. The several candidates for large scale expression were sequenced (Geneservices, Oxford) prior to large scale protein production.

3.2.3. A Highly Parallel Expression Optimisation Protocol

Despite attempts to express and purify near full-length Kindlin-1, reproducibly, low expression levels remained problematic. This was further exacerbated by proteolytic sensitivity of the protein and the overwhelming excess of host stress response proteins that adsorb to metal resins in IMAC, for example the *E. coli* peptidoylproline *cis-trans* isomerise

Table 3.3 – Kindlin Family Construct Design for HTP Cloning and Expression Screening

OPPF	Residue range	Domain Composition	pOPIN Vector/tag	MW (Da)	Position
6433	1-677	F0-F1'-loop-F1"-F2'-PH-F2"-F3	pOPINE/C-his	76450	**A1**
6436	96-677	F1'-loop-F1"-F2'-PH-F2"-F3	pOPINE/C-his	66000	**B1**
6440	286-677	F2'-PH-F2"-F3	pOPINE/C-his	45100	**C1**
6444	589-677	F3	pOPINE/C-his	14080	**D1**
6448	1-680	F0-F1'-loop-F1"-F2'-PH-F2"-F3	pOPINE/C-his	76780	**E1**
6452	98-680	F1'-loop-F1"-F2'-PH-F2"-F3	pOPINE/C-his	66110	**F1**
6456	288-680	F2'-PH-F2"-F3	pOPINE/C-his	45210	**G1**
6460	570-680	F3	pOPINE/C-his	14190	**H1**
6464	1-655	F0-F1'-loop-F1"-F2'-PH-F2"-F3	pOPINE/C-his	75130	**A2**
6468	99-665	F1'-loop-F1"-F2'-PH-F2"-F3	pOPINE/C-his	64350	**B2**
6472	265-665	F2'-PH-F2"-F3	pOPINE/C-his	46090	**C2**
6476	554-665	F3	pOPINE/C-his	14300	**D2**
5498	1-677	F0-F1'-loop-F1"-F2'-PH-F2"-F3	pOPINF/N-his	76450	**E2**
6437	96-677	F1'-loop-F1"-F2'-PH-F2"-F3	pOPINF/N-his	66000	**F2**
6441	286-677	F2'-PH-F2"-F3	pOPINF/N-his	45100	**G2**
6445	589-677	F3	pOPINF/N-his	14080	**H2**
6449	1-680	F0-F1'-loop-F1"-F2'-PH-F2"-F3	pOPINF/N-his	76780	**A3**
6453	98-680	F1'-loop-F1"-F2'-PH-F2"-F3	pOPINF/N-his	66110	**B3**
6457	288-680	F2'-PH-F2"-F3	pOPINF/N-his	45210	**C3**
6461	570-680	F3	pOPINF/N-his	14190	**D3**
6465	1-655	F0-F1'-loop-F1"-F2'-PH-F2"-F3	pOPINF/N-his	75130	**E3**
6469	99-665	F1'-loop-F1"-F2'-PH-F2"-F3	pOPINF/N-his	64350	**F3**
6473	265-665	F2'-PH-F2"-F3	pOPINF/N-his	46090	**G3**
6477	554-665	F3	pOPINF/N-his	14300	**H3**
6434	1-677	F0-F1'-loop-F1"-F2'-PH-F2"-F3	pOPINJ/GST	100470	**A4**
6438	96-677	F1'-loop-F1"-F2'-PH-F2"-F3	pOPINJ/GST	90020	**B4**
6442	286-677	F2'-PH-F2"-F3	pOPINJ/GST	69120	**C4**
6446	589-677	F3	pOPINJ/GST	38100	**D4**
6450	1-680	F0-F1'-loop-F1"-F2'-PH-F2"-F3	pOPINJ/GST	100800	**E4**
6454	98-680	F1'-loop-F1"-F2'-PH-F2"-F3	pOPINJ/GST	90130	**F4**
6458	288-680	F2'-PH-F2"-F3	pOPINJ/GST	69230	**G4**
6462	570-680	F3	pOPINJ/GST	38210	**H4**
6466	1-655	F0-F1'-loop-F1"-F2'-PH-F2"-F3	pOPINJ/GST	99150	**A5**
6470	99-665	F1'-loop-F1"-F2'-PH-F2"-F3	pOPINJ/GST	88370	**B5**
6474	265-665	F2'-PH-F2"-F3	pOPINJ/GST	70110	**C5**
6478	554-665	F3	pOPINJ/GST	38320	**D5**
6435	1-677	F0-F1'-loop-F1"-F2'-PH-F2"-F3	pOPINM/MBP	116470	**E5**
6439	96-677	F1'-loop-F1"-F2'-PH-F2"-F3	pOPINM/MBP	106020	**F5**
6443	286-677	F2'-PH-F2"-F3	pOPINM/MBP	85120	**G5**
6447	589-677	F3	pOPINM/MBP	54100	**H5**
6451	1-680	F0-F1'-loop-F1"-F2'-PH-F2"-F3	pOPINM/MBP	116800	**A6**
6455	98-680	F1'-loop-F1"-F2'-PH-F2"-F3	pOPINM/MBP	106130	**B6**
6459	288-680	F2'-PH-F2"-F3	pOPINM/MBP	85230	**C6**
6463	570-680	F3	pOPINM/MBP	54210	**D6**
6467	1-655	F0-F1'-loop-F1"-F2'-PH-F2"-F3	pOPINM/MBP	115150	**E6**
6471	99-665	F1'-loop-F1"-F2'-PH-F2"-F3	pOPINM/MBP	104370	**F6**
6475	265-665	F2'-PH-F2"-F3	pOPINM/MBP	86110	**G6**
6479	554-665	F3	pOPINM/MBP	54320	**H6**

Protein
Kindlin-1
Kindlin-2
Kindlin-3

Vector
pOPINE
pOPINF
pOPINJ
pOPINM

(SlyD) that binds to nickel resins (typical *E. coli* proteins that contaminate IMAC are reviwed in [Bolanos-Garcia and Davies, 2006]).

One strategy to try to significantly improve protein expression was to employ a further optimised and parallel cloning experiment using different tagging vectors from the pOPIN suite combined with different isoforms of murine Kindlins. Further to this, some domain boundaries were also redefined experimentally (Goult *et al*. 2009b) and therefore necessitated a redesign of Kindlin fragments (see Figure 3.1. for a schematic of the re-defined domain boundaries). Several candidate fragments and full length constructs were designed (see Table 3.3) based upon domain boundaries provided by Dr. B. Goult (University of Leicester) which, in turn, were based upon domain boundaries of a recently solved structure of the homologous protein talin (Goult *et al.,* 2010). This domain boundary redefinition was applied to Kindlin-1 initially; however, a simple sequence alignment between family members, via ClustalW, quickly aided construct redesign for the *FERMT* family. We designed 4 representative constructs, with each domain successively truncated in the N terminus → C terminus direction for each isoform, i.e. 12 different constructs, repeated to yield 4 groups of 12 constructs with each group possessing a different tagging method, i.e. C-terminal His_6, N-terminal His_6, GST and MBP (Table 3.3). The broad versatility of the In-Fusion cloning technique meant that we designed only two groups of 12 primer pairs (forward and reverse), each group with separate homology regions flanking the resulting PCR product, to yield the 48 constructs (Table 3.3 and Figure 3.4). All constructs with N-terminal tags share recombination sites (Kpn1 and HindIII restriction sites), therefore allowing the same PCR fragment to be infused into multiple vectors. Vectors encoding C-terminal His tags, N-terminal His tags, GST fusion partners, and MBP fusion partners were chosen for the infusion. The so-called solubility tags (GST and MBP) also possessed N-terminal His tags, allowing uniform nickel nitrilotriacetic acid (Ni-NTA) IMAC based screening. Additionally,

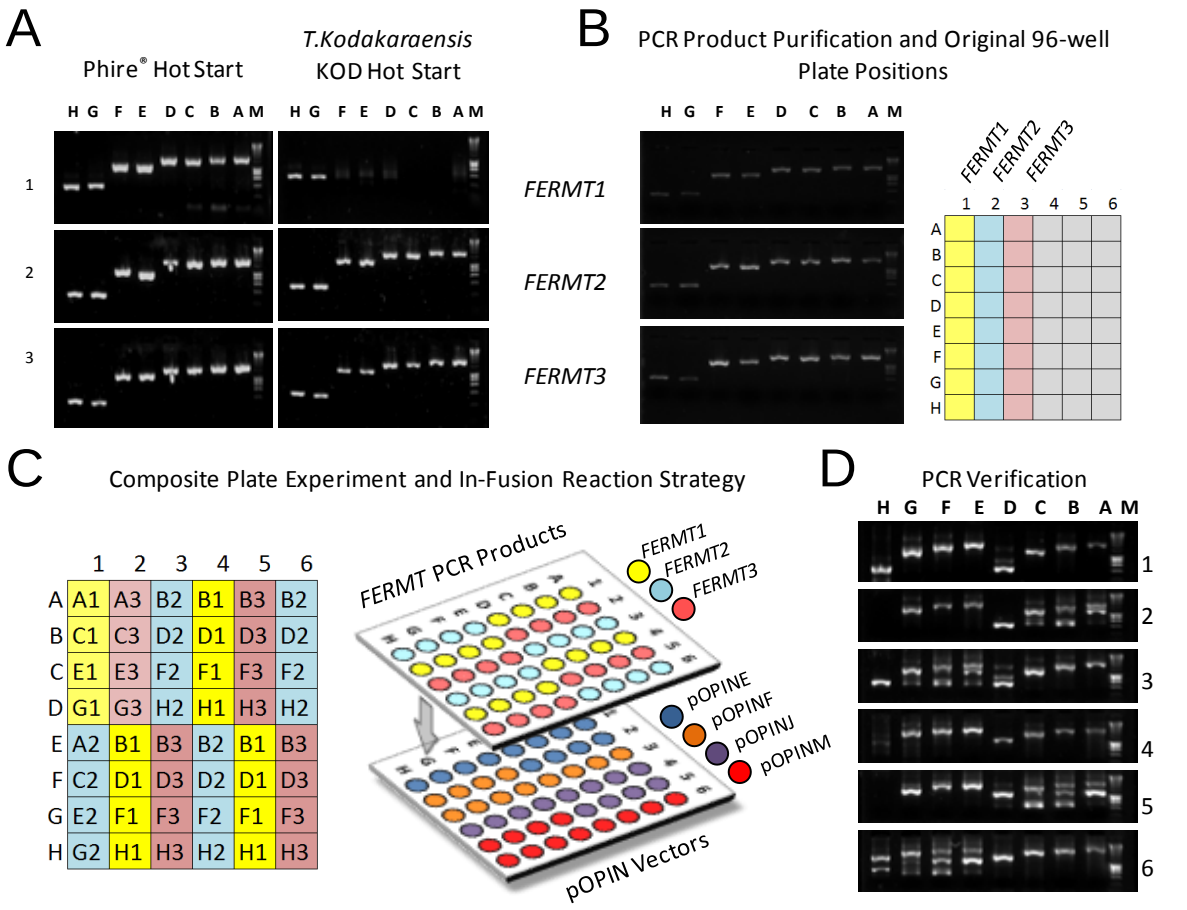


Figure 3.4 – High throughput cloning of Kindlin isoforms. (A) Agarose gel electrophoresis of PCR products generated using two high-fidelity polymerase, Phire and KOD, both of which are complexed with the 'Hot-Start' antibody for minimizing activity at ambient temperature. PCR products were purified after DpnI treatment and transferred to a composite 96-well plate. (B) Those PCR products with appropriate ends were positioned within a new 96-well plate for In-Fusion cloning, grid (C) illustrates the position of PCR products together with original positions from the PCR plate. A second 96-well plate was designed to possess appropriate pOPIN vectors so that one plate can be mixed with the second together with In-Fusion enzyme to yield correct plasmids (schematized in C). The constructs designed and final vector positions illustrated in (C) are tabulated and colour coded in this figure as in Table 3.3. (D) After In-Fusion reaction, presence of insert was verified by PCR.

the solubility tags also possessed a 3C cleavage site so that the fusion partner could be removed and the protein of interest liberated. Since only 24 PCR products were generated to yield 48 separate constructs, due to aforementioned redundancies, downstream plate experiments required careful planning and strategising in order to position appropriate PCR

products, to be incorporated, with appropriate destination vectors for a single plate-based In-Fusion experiment. Therefore, we generated two composite plates so that these 96-well plates could be 'superimposed' into a single plate experiment. One such plate was divided on the basis of the four different tagging systems and thus contained the appropriate pOPIN vectors in appropriate positions (see Figure 3.4). The second, complementary, composite 96-well plate was engineered to contain appropriate PCR fragments at specific positions (Figure 3.4). Template cDNA, for this HTP cloning experiment, of *FERMT1, FERMT2* and *FERMT3*, was kindly gifted from Prof. R. Fäessler (Max Planck Institute, Germany).

Small scale expression studies were performed using Rosetta (DE3) pLysS/Rare BL21 and B834 (DE3) methionine auxotrophic strains of *E. coli,* which were transformed using purified plasmid DNA. Again all transformations and subsequent cultures were performed maintaining their original PCR positions within the 96 well plate experiments. This study was parallelized to explore protein expression with classical IPTG induction methods, and auto-induction technology with each strain (Figure 3.5). Small scale IMAC purification experiments revealed that most full length Kindlin proteins did not benefit, in terms of expression levels, from the addition of a fusion partner. However some small fragments, for example, the integrin binding domain (F3), expressed only as a fusion protein. Additionally, several constructs failed to express in Rosetta BL21 strains, favouring the methionine auxotroph B834 (DE3). Several candidate constructs were sequence verified and used in large scale bacterial expression cultures.

In one case, the integrin binding domain of Kindlin-3, denoted as F3, MBP fusion protein was expressed in 10L bacterial culture, however, NiNTA affinity chromatography, followed by gel filtration, resulted in heavily contaminated protein. One particular contaminant was a 70kDa protein, suggested to be the *E. coli* chaperone, DnaK, which suggested that, the recombinant protein was incorrectly folded.

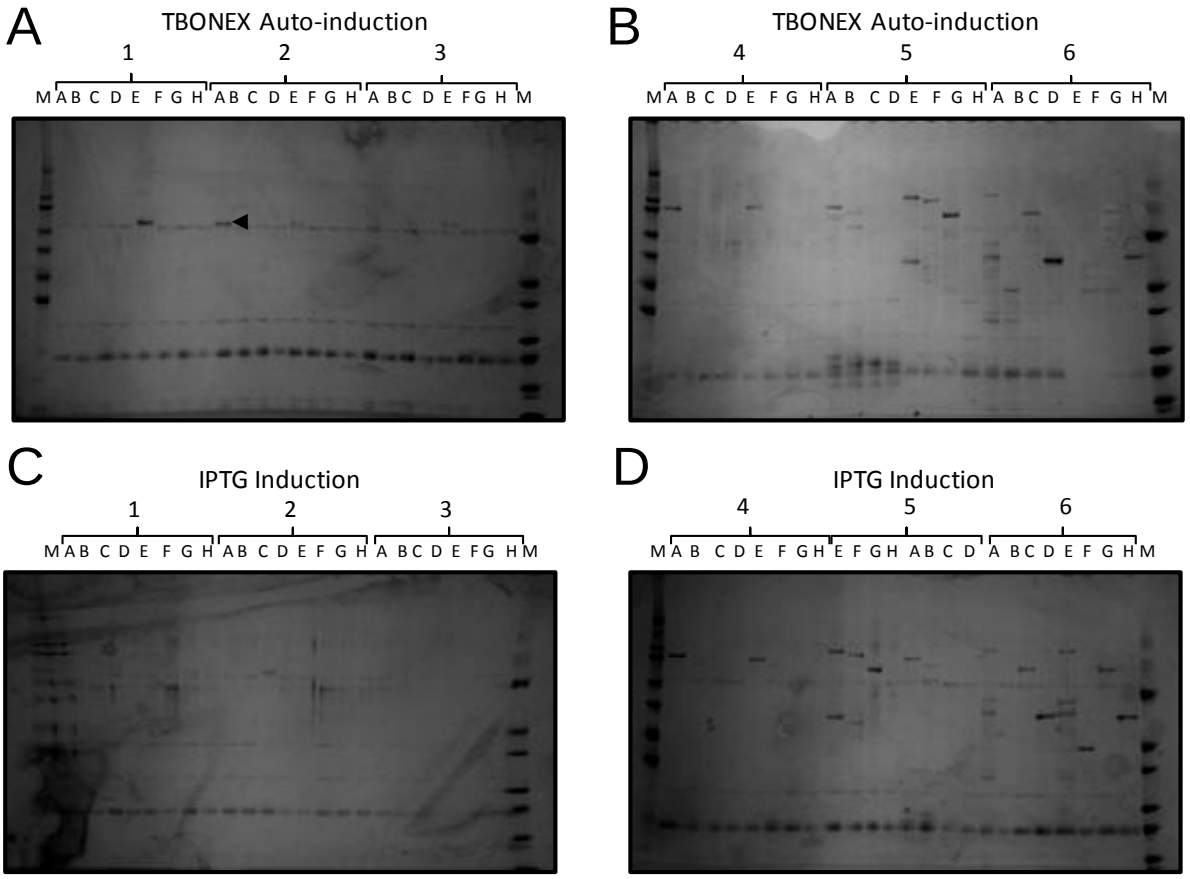


Figure 3.5 – SDS-PAGE analysis of soluble expression screening. Small (3ml) cultures of transformed B834 (DE3) pLysS (A-D) were induced *via* auto-induction media (A, B) or IPTG induction (C, D) and the cells lysed and clarified. His-tagged proteins were purified from the soluble fraction using magnetic Ni-NTA beads (see Chapter 2). Lanes were labeled according to their position within a 96-well experiment, which also coincides with the original cloning plate experiment (see Table 3.3 [column 6] and Figure 3.4). Construct A2 (arrowed) was used in subsequent studies (see Chapter 5). SDS-PAGE analysis was performed using a 10% Tris-Glycine gel, electrophoresis was performed at 100mV in MOPS buffer. Proteins were visualized *via* SimplyBlue Safe stain (Invitrogen).

This was confirmed after liberating the F3 domain from its fusion partner, *via* 3C digestion, which resulted in precipitation of the protein. The absence of expression of both amino- and carboxyl terminal His-tagged F3 constructs is suggestive of poor domain boundary definition,

or that the preceding subdomain (F2) stabilizes the protein, with the fusion partner of the recombinant protein acting analogously. Other cases where full-length constructs were also scaled up to tens of litres of expression cultures and the protein purified by NiNTA purification proved problematic as the proteins were co-purified with chaperones and a large excess of nickel binding proteins thus yielding very limited material for further purification or analysis.

3.2.4. Expression Strategies in Different Strains and Hosts

The optimisation of expression of Kindlin full-length constructs was vital for any biophysical characterisation, including crystallisation. After screening many solubility tagging systems and different isoforms from the Kindlin family, the next step was to utilize different strains of bacteria. Codon usage optimisation was assessed during initial HTP expression screening, *via* the implementation of Rosetta (DE3) pLysS/RARE expression strains, therefore the screening of other strains was primarily used to overcome the perceived toxicity related to the protein and/or vector DNA. If the resulting protein was indeed toxic this would certainly explain the low level of expression, as cell death would ensue after induction. Two strains C41 (DE3) pLysS and C43 (DE3), known for their tolerance of toxic and membrane proteins were used to ascertain whether protein toxicity played a role in low expression. SDS-PAGE analysis of whole cell lysate of cultures induced by the additive IPTG, compared with un-induced controls revealed no significant amounts of recombinant protein expressed.

Gene base composition has been reported to influence recombinant protein expression if the percentage of GC bases is beyond the physiological limits of the host organism (Baca and Hol, 2000). Observing GC and AT content of the *FERMT* genes, it was quickly

established that the GC content was not responsible for low yields of recombinant protein, given that the *FERMT1* gene possesses ~50% GC content and analysis of the *E. coli* genome indicates a minimum and maximum GC content of 27%-67%.

We also attempted large scale purifications of constructs designed and gifted from Dr. B. Goult (University of Leicester). In particular a construct encoding Kindlin-1 subdomains F2, PH and F3 was demonstrated to express, in high yield, as a GST fusion protein in *E. coli.* Therefore, several litres of expression cultures were used in a large scale purification using standard procedures with glutathione sepharose 4FF. The fusion protein was eluted using reduced glutathione and GST liberated using Tobacco Etch Virus (TEV) protease cleavage. Upon incubating the protein with TEV protease at 4˚C overnight, the liberated Kindlin-1 fragment precipitated (data not shown).

Based upon these observations it was clear that *E. coli* expression was problematic and that the low yielding expression was not improved by the addition of solubility tags or the use of other Kindlin isoforms. Mindful of the lack of structural and biophysical information on Kindlins, it was vital that milligram quantities could be made so that a comprehensive biophysical study of the proteins could be made. Due to the pOPIN vector suite properties, i.e. appropriate enhancers, a Kozak sequence and a polyadenylation sequence, it seemed sensible to transfer existing constructs for expression in eukaryotic hosts.

A sparse matrix of twenty four representative Kindlin family constructs (constructs A1 – H3, see Table 3.3) were selected for a highly parallel expression trial using eukaryotic expression hosts. As suggested by Gouaux and co-workers, transient expression in human embryonic kidney (HEK) cell lines offers a fast screening method for identifying promising expression construct leads that, in turn, could be transferred to baculovirus driven expression so that appropriate quantities of recombinant protein could be obtained (see Kawate *et al.* 2009, for example). Accordingly, transient mammalian expression was utilised as such a

screen with the detection of expression a simple western blot, analogous to methods established within the host Division of Structural Biology (Aricescu *et al.*, 2007). Interestingly, of the twenty four constructs tested, three constructs were detectable by SDS-PAGE followed by western blotting, with a further three constructs at barely detectable levels (data not shown). This screen was repeated to ensure that experimental flaws were not the cause of the lack of expression of the majority of the construct sparse matrix. Assuming that this screening method represented a predictable result in baculovirus infected Sf9 cells the six lead constructs were utilised in generating recombinant virions for expression trials. Unexpectedly, initial western blotting of these constructs revealed an expression profile unpredicted from the transient mammalian expression screen. An expression construct encoding Kindlin-1 residues 285 – 677 was poorly expressed and barely detectable in HEK293T cells, but was expressed highly compared to the selected panel of constructs transferred to baculovirus-infected Sf9 cells. Thus, transient expression in HEK293T was an ineffective fast screening method for detecting Kindlin constructs that could be readily and predictably transferred and scaled into baculovirus-infected Sf9 cells. In fact each expression host is, in this case, mutually exclusive and represents separate orthogonal hosts that should be explored empirically. Although, there are known examples of fast screening in mammalian cells followed by large scale protein production in baculovirus infected Sf9 cells (Kawate *et al.*, 2009, Shaffer *et al.*, 2009, Sobolevsky *et al.*, 2009), it must also be noted that in all of these examples the proteins of interest were membrane proteins – specifically membrane channels.

In accordance with this observation we used the sparse matrix of constructs to generate recombinant baculovirions for each construct using standard protocols (Zhoa *et al.*, 2003) and repeating the generation of some virions used in earlier experiments to maintain consistency. Recombinant baculovrius generation followed by small scale (2ml cultured Sf9

monolayer) expression trialling of each construct was assessed by SDS-PAGE followed by western blotting using an anti-penta-histidine antibody of whole cell extract of the baculovirus-infected Sf9 cells. Curiously, SDS-PAGE revealed an unusual phenomenon in that protein, visualised by staining with Coomasie-based reagents, did not electrophoretically migrate properly as expected, despite the molecular weight standard migrating optimally. Instead protein from the samples was immobilized at the stacking-separating acrylamide gel interface. Furthermore, western blot analysis revealed that the immobilized protein was his-tagged and therefore the recombinant Kindlin proteins – suggesting aggregation. Efforts were invested in optimising this type of analysis as this method was crucial for evaluating small-scale expression screening, but more importantly, represented the only accurate read-out at this scale. Ordinarily, the baculovirus-infected Sf9 cell monolayer used in small-scale expression trials was resuspended in phosphate-buffered saline (PBS) equalling a fourth of the volume of culture; i.e. 0.5ml in the case of small–scale experiments. Whole cell extract was diluted in standard laemmli SDS loading buffer and boiled for 10 minutes at >95˚C. Multiple methods were employed to improve migration of the protein samples, which included; reducing the current to 25mA during electrophoresis; adding detergents such as Triton X-100, polyoxyethylene sorbitan monolaurate (Tween-20) and 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS) during the cell resuspension, all of which did not improve electrophoretic mobility of the proteins as assessed by SDS-PAGE and western blotting. Finally, using 10% (w/v) sodium-dodecyl sulphate (SDS) in PBS to resuspend the Sf9 cell monolayer together with brief (10-20s) sonication pulses of each sample prior to SDS-PAGE resulted in proper migration. We observed expression of fifteen of the twenty-four constructs in baculovirus, with all twelve C-terminal hexa-his tagged proteins and three N-terminally hexa-his tagged constructs

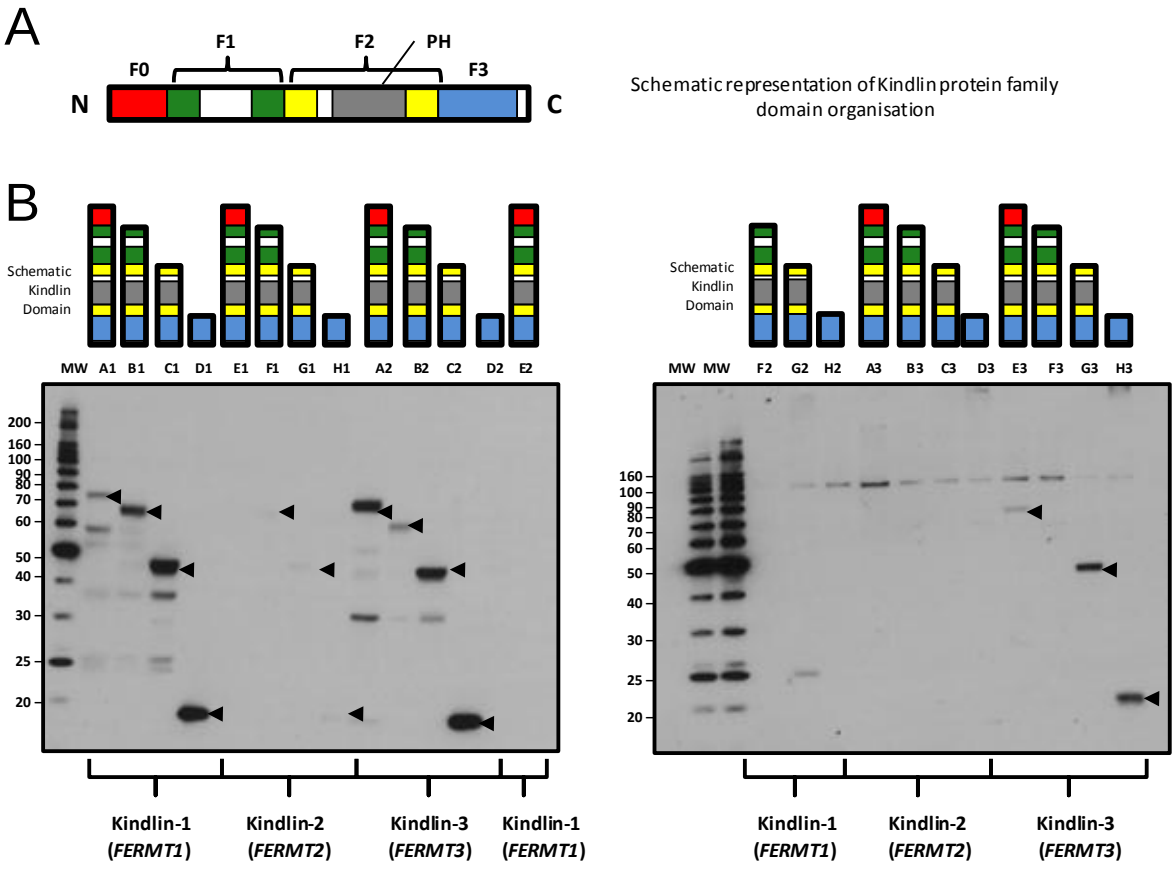


Figure 3.6 – Western Blot Analysis of Baculovirus-Infected *Spodoptera frugiperda* (Sf9) Expression Screening. (A) Schematic representation of the Kindlin domain organisation as illustrated in Figure 3.1. (B) Twenty four Kindlin family constructs were utilized in the transfection of twenty four small (2ml) monolayer cultures of Sf9 cells. Whole cell lysate was analysed by SDS-PAGE using a 4%-12% gradient gel (InVitrogen) 72 hours after infection and expression of recombinant proteins was detected using a primary mouse anti-pentaHis antibody probe. A schematic, as found in Figure 3.1, is shown above each lane and illustrates the domain composition of the construct tested. Additionally each lane (construct) is labeled according its position in the PCR experiment (see Table 3.3 for position). Immunoreactive bands that exhibit correct molecular weight of the desired construct are arrowed for clarity.

expressing to various extents (Figure 3.6). Interestingly, C-terminally His_6-tagged Kindlin-1 (construct A1) and Kindlin-3 (construct A2) constructs demonstrated expression in Sf9, which were used for further viral amplification and large scale expression experiments. Due to the predominantly erythropoietic classification of Kindlin-3 and its specific activity *in vivo*

it is, for the purposes of this work, considered separate and, accordingly, is discussed in greater depth elsewhere (see Chapter 5).

Recombinant virus encoding a C-terminally His-tagged Kindlin-1 was amplified and used in further large-scale protein production and purification. Interestingly, recombinant Kindlin-1 could not be efficiently extracted from Sf9 cells and its expression required some optimisation. Mild sonication was employed to disrupt the cells together with the use of detergents. Kindlin-1 was purified by virtue of its His-tag from the resulting soluble fraction after cell lysis but remained heavily contaminated. SDS-PAGE and western blotting revealed an improper migration of the protein, which along with the previous observation of heavy contamination could indicate aggregation inside the cell.

Small-scale expression trials of the Kindlin sparse matrix in Sf9 cells revealed that the Kindlin-1 constructs that were left for >5 days exhibited a unique proteolytic fragmentation pattern that was immunoreactive to anti-penta-His antibody during western blot analysis, i.e. the fragments were His-tagged. These were visualised with a prolonged exposure of the film to the enhanced chemiluminescent (ECL) substrate treated nitrocellulose membrane (Figure 3.6). As the hexa-Histidine tag was C-terminal we could readily identify that the fragments were obviously degraded in the N→C direction. Approximate molecular weights were suggested for each fragment based on the molecular weight standard (see Figure 3.7). Interestingly, the three Kindlin-1 constructs additional to the full-length, which were truncation mutants in the N→C direction, were not represented in the proteolytic fragmentation observed. Furthermore, the smallest construct, a supposed 14kDa C-terminal F3 subdomain did not correlate with the smallest proteolytic fragment, which was 25kDa (see Figure 3.7). Moreover, large-scale purification of the Kindlin-1 fragments revealed that they could not be extracted from the soluble fraction of the recombinant baculovirus infected Sf9 cells and were suggested to be insoluble – perhaps explaining their absence in the proteolysis

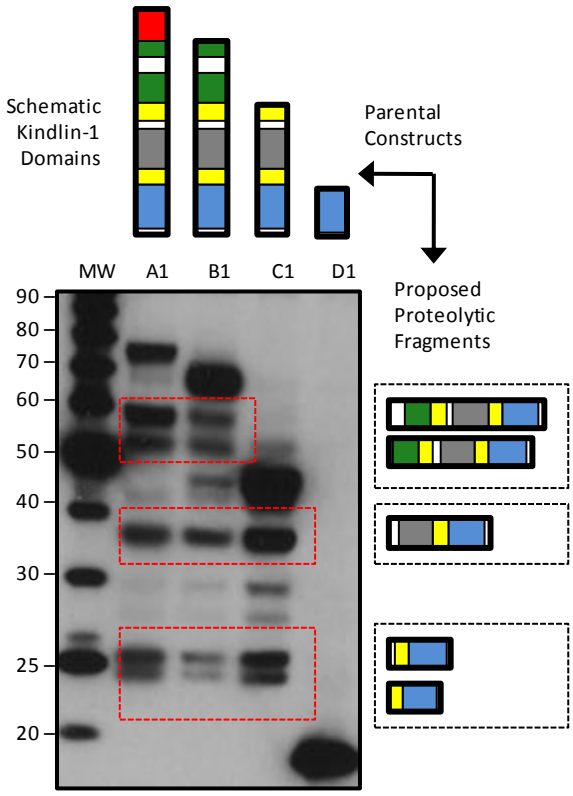


Figure 3.7 – Western Blot Analysis of Baculovirus Infected Sf9 Expression of Kindlin-1 Constructs 5 days Post Infection. Small scale Sf9 cell cultures were infected with recombinant baculovirus incorporating Kindlin-1 constructs from the sparse Kindlin construct matrix all with C-terminal His_6 tags. Cells were harvested after 5 days and protein expression assessed by western blot analysis as described before. Prolonged exposure (5 minutes) during blot development allowed visualization of proteolytic fragments with the C-terminal His_6-tag intact. The molecular weights of consistent fragments were estimated using the molecular weight marker and a domain composition proposed. As in Figure 3.6, a schematic of the domain boundaries that each construct possesses is displayed above the corresponding lanes and for the consistent fragments (right of blot).

pattern. The estimated molecular weights of each of the fragments were used to suggest subdomain composition; owing to the fact the degradation was N-terminal. The fragments approximately identified were cross correlated with constructs in hand generated from our initial HTP screening of Kindlin-1. Thus twenty four constructs from our initial HTP screen were selected to approximate the domain composition and approximate molecular weight of the proteolytic fragments. Based on the known high level expression of Kindlin-1 in

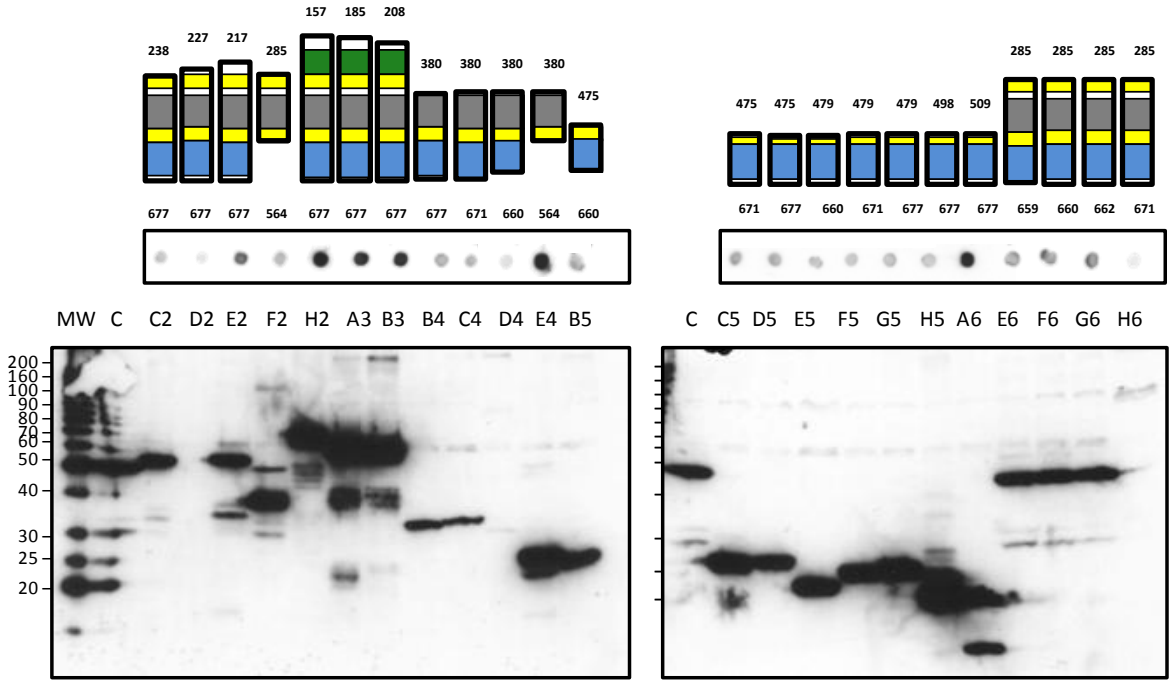


Figure 3.8 – Western Blot Analysis of Kindlin-1 Fragments. A selection of 24 Kindlin-1 constructs designed and generated in the initial round of HTP cloning and expression (see Figure 3.2 and Table 3.2) that contain similar domain composition to those fragments suggested during *in vivo* proteolysis in insect cells. Sf9 cells were infected with recombinant baculovirus and expression assessed by western blotting using an anti-pentaHis antibody probe. The soluble fraction of lysed Sf9 cells was also spotted onto nitrocellulose membrane and probed with the same anti-pentaHis antibody (shown above blot). A construct encoding GATA3, a 50kDa protein (a kind gift from S. Nurmohamed, Division of Structural Biology, University of Oxford) was used as a positive control (labeled C above the blot). As in Figure 3.6, a schematic of the domain boundaries, with amino acid limits, that each construct possesses is displayed above the corresponding lanes together with a position of the construct in the original 96-well cloning experiment (see Figure 3.2 and Table 3.2). The Schematic representation of the Kindlin domain organisation is analogous to that illustrated in Figure 3.1.

baculovirus infected Sf9 cells, virions for each of the constructs were generated and the expression assessed analogously to screening methods employed and described earlier. Twenty two of the twenty four constructs expressed and were detected by western blotting. Mindful of the fact that the Kindlin-3 truncation mutants were insoluble, these small–scale expression trials were resuspended in 1ml PBS and the resuspended cells aliquoted into 0.5ml

samples. Aliquots representing a complete trial were used for determining expression by western blotting. Again the addition of SDS, to a final concentration of 1% (w/v), previously described was important for proper migration during SDS-PAGE and subsequent western blot analysis. The second half were lysed and clarified using mild conditions, i.e. PBS with 1% (v/v) polyoxyethylene sorbitan monolaurate (Tween-20). The resulting supernatant was assessed for the presence of soluble his-tagged recombinant Kindlin-1 fragments by dot blotting. In short, the clarified lysate from each baculovirus infected Sf9 cell culture, prepared earlier, was 'spotted' directly onto a nitrocellulose membrane. Protein was detected using standard western blotting methods. Those samples where no detectable expression of His-tagged protein fragments was observed in western blotting analysis were used as internal controls, in this type of dot-blotting, to measure the level of cross reactivity of endogenous Sf9 cells. Furthermore, as with the western blot analysis, a positive control GATA3, known to express highly in baculovirus-infected Sf9 (a kind gift from Dr. S. Nurmohamed, Division of Structural Biology, University of Oxford, UK) was used for completeness. Despite the promising expression of twenty two of the constructs, only five were considered soluble by this method.

3.2.5. Discussion

The Kindlin family have proven somewhat problematic in bacterial expression hosts, with only a subset of constructs expressing in *E. coli* all of which possessed a single domain, the largest of which has subsequently been crystallised and the structure solved to high resolution (see Chapter 4). The use of solubility tags with these proteins has proven less successful than previously anticipated, with a recent report from the OPPF echoing such notions (Bird, 2011). Larger fragments containing the integrin binding domain have been much less successful in *E. coli* thus prompting the use of additional hosts. To that end, the

baculovirus-infected Sf9 cells proved much more successful with a subset of constructs. Interestingly, whilst many constructs were expressed, only a small number were deemed soluble. Full length Kindlin-3, the smallest member of the family, is the most attractive target for biophysical characterisation, of which milligram quantities can be obtained (see Chapter 5). Additional constructs still require further investigation, including an unexpected F3 subdomain construct with half of the preceding F2 subdomain, that is suggested to be soluble and initial expression and purification of this construct have proven this to be the case (data not shown). Overall, the Kindlin family of proteins will remain the focus of attention for many years and the increasing demand for structural studies will become great. Having identified the baculovirus expression system as the most useful, further experimentally led construct design, i.e. limited proteolysis, and expression screening in Sf9 cells alone may prove the most successful route to obtaining new structures. However, producing quantities of recombinant protein for functional studies is equally important and this type of expression screening has proven particularly useful in that.

CHAPTER 4

4.1. Pleckstrin Homology Domains Have Diverse Functionalities but a Common Fold

4.1.1. The PH domain Superfamily Fold

Structural biology, especially structural genomics programs, has led to the determination of many structures which have confirmed the concept of protein superfamilies and superfolds (Jones *et al.*, 1992) - that is proteins with minimal amino acid sequence homology and different functionalities but a common scaffold or core tertiary structure. The superfold provides a core module upon which a range of different functions can be imposed by varying surface regions and loops (Blomberg *et al.*, 1999). The pleckstrin homology (PH) domain is one such superfold.

The Pleckstrin Homology (PH) domain was originally discovered as an internally repeating motif of ~100 amino acids in the haematopoietic protein, *p*latelet and *l*eukocyte *C* *k*inase sub*str*ate prote*in* [pleckstrin] (Haslam *et al.*, 1993). Additionally, reports in the same year of a 100-120 residue stretch of sequence similarity in proteins involved in cytoskeletal organisation and cellular signalling, among other processes, were suggested to be an homologous domain (Mayer *et al.*, 1993, Musacchio *et al.*, 1993). As this motif/domain was first described by Haslam and colleagues its nomenclature, as a domain homologous to that found in pleckstrin, has remained. The completion of the human genome revealed that an estimated 252 different human proteins possess a PH domain, thus making it the 11th most common module in humans (Lemmon *et al.*, 2002).

Many structures of PH domains have now been solved from a variety of proteins, including that of Kindlin-1, which is described in this Chapter, and all share a common fold.

The tertiary structure of this module was first described for β-spectrin PH domain (Macias *et al.* 1994) and the N-terminal pleckstrin PH domain (Yoon *et al.,* 1994) by nuclear magnetic resonance (NMR) a year after the domain's discovery. The now well-documented PH superfold comprises of a 7-stranded β-sandwich (β1 - β7), or collapsed β-barrel, which is constructed from a 4-stranded amino-terminal anti-parallel β-sheet (β1 – β4) stacked near-orthogonally to a carboxy-terminal 3-stranded β-sheet meander (β5 – β7) [see Figure 4.1] (Lemmon *et al.,* 2002, Rebecchi and Scarlatta, 1998). The β-barrel is filled with hydrophobic residues and one end of the β-barrel is capped by a C-terminal α-helix, stabilizing the overall fold and is characteristic of all PH domains (DiNitto and Lambright 2006, Lemmon *et al.,* 2002, Rebecchi and Scarlatta, 1998) [see Figure 4.1]. The PH domain possesses 6 loops between the β-strands that are located at opposing ends of the β-barrel. The loops between strands β1-β2, β3-β4 and β6-β7 crown the open end of the β-barrel and are considered hypervariable (DiNitto and Lambright 2006; Rebecchi and Scarlatta, 1998), with these loops differing most in sequence and length in early PH domain alignments (Lemmon and Ferguson 2000). It is these variable loops which confer functionality and specificity in PH domains and by analogy to the immunoglobulin-like domains, as an example, form the basis for the ligand binding site (Lemmon and Ferguson 2000). Other early observations of this domain included the electrostatic polarisation of the domain, with a positively charged surface in close proximity to, and including, the variable loops (DiNitto and Lambright 2006, Lemmon and Ferguson, 2000; Rebecchi and Scarlatta 1998; Ferguson *et al.,* 1994; Macias *et al.,* 1994). Felisk and co-workers (1994), first described the, albeit weak, interaction of PH domains with lipid bilayers containing phosphatidylinsoitol phosphate (PtdInsP) lipids with the Pleckstrin PH domain binding to micelles containing PtdIns(4,5) bisphosphate with

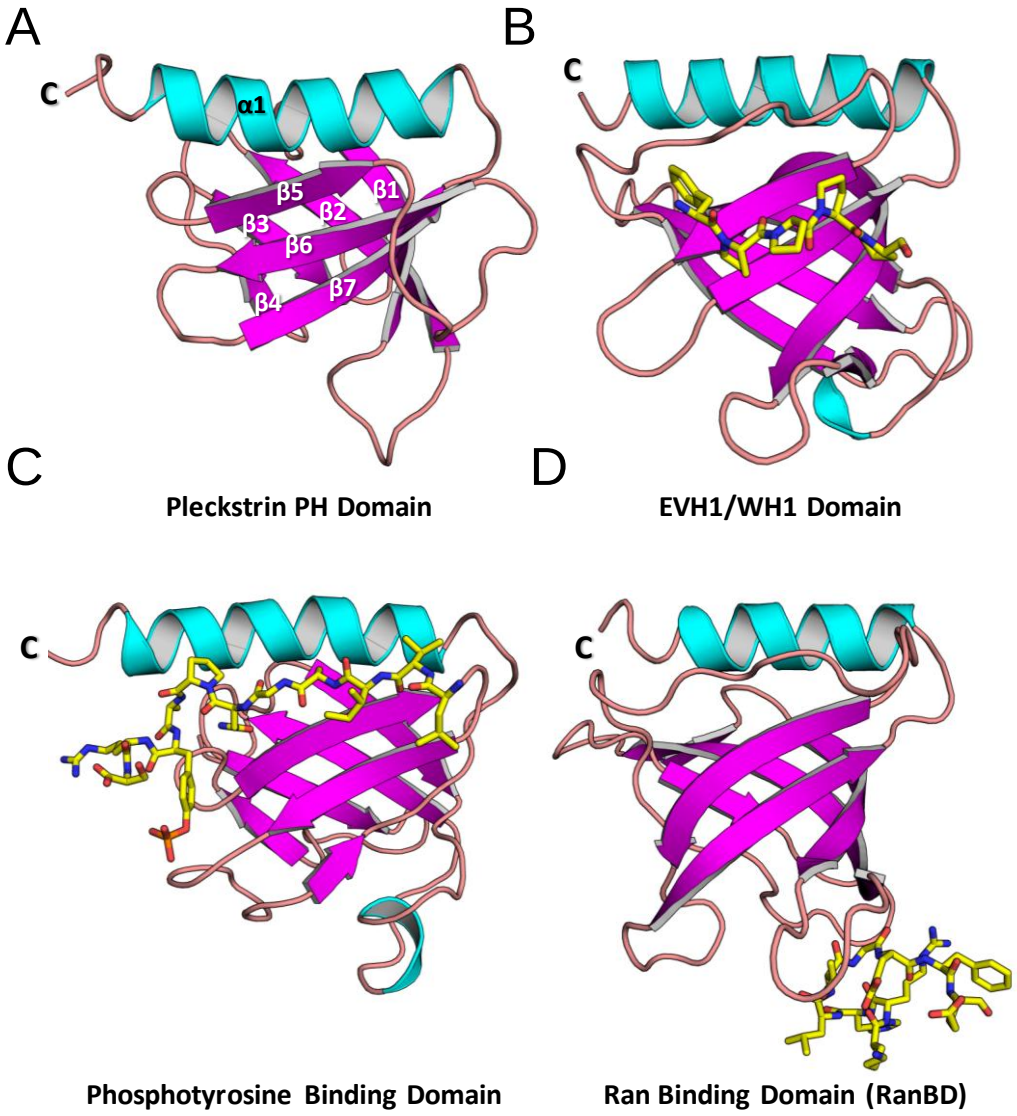


Figure 4.1. – A structural comparison of different modules with the PH domain-superfold. (A) The N-terminal PH domain from pleckstrin (PDB:1PLS) compared with domains that have distinct biological functions but adopt the same fold. Secondary structure elements are individually coloured and the prototypical 7-stranded β-barrel is labelled along with the C-terminal capping helix. (B) The EVH1/WH1 domain from Mena (PDB:1EVH) bound polyproline peptide and rendered as in (A). (C) The PTB (phosphotyrosine-binding) domain from IRS-1 (PDB:IRS1) bound phosphotyrosine-containing peptide of the insulin receptor which is representated as sticks. Secondary structure elements are highlighted as for (A). (D) The RanBD1 (PDB:1RRP) in complex with the Ran effector loop shown as stick representation. Again, secondary structure elements are rendered as in (A). The C-termini are labelled, but the N-termini are obscured from view in all cases. This Figure was generated with PyMol (https://www.Pymol.org).

micromolar affinity (Harlan *et al.*, 1994). Further studies using NMR revealed that the positively charged face of the PH domain mediated this interaction, with VL1, VL2 and VL3 all contributing to the PtdIns(4,5)P_2 binding. Insights from this work led to the proposal that PH domains were responsible for lipid association and lipid-dependent membrane targeting (Harlan *et al.*, 1994) – a function that is discussed further below.

The PH domain module has been found in mammalian proteins, proteins from *Drosophila melanogaster*, *Caenorhabditis elegans* and yeast, but not found in plants or, until recently, bacteria suggesting its early evolution (Bottomley *et al.*, 1998). The crystal structures of several members of the PFO8000 protein family, found in bacteria and archea, revealed the existence of a domain that was a structural homologue of the PH domain superfold and therefore suggests that the PH module had evolved earlier than previously thought (Xu *et al.*, 2010). In addition, the PH domain superfamily fold has been extended over the years to include other functionally distinct domains which share a common structure, these include; phosphotyrosine binding domains [PTB] (Zhou *et al.*, 1995), EVH1 (Enable/VASP homology 1) or WH1 (WASP homology 1) domains (Prehoda *et al.*, 1999), and the Ran-binding domain [RanBD] (Vetter *et al.*, 1999) (see Figure 4.1).

4.1.2. The Diverse Biological Functions of PH Domains

As discussed previously, the Pleckstrin Homology domain possesses a common tertiary structure but is involved in many different processes. The PH domain occurs in many multidomain proteins implicated in signal transduction, membrane trafficking and cytoskeletal rearrangement (DiNitto and Lambright, 2006). Examples of PH domain containing proteins, illustrating the versatile nature of this domain, include the protein kinases; Bruton's tyrosine kinase (Btk), β-adrenergic receptor kinase (β-ARK) and Protein

Kinase B (PKB, also known as Akt), all the Phospholipase C (PLC) isoforms, insulin receptor substrates (IRS)-1 and -2, Phosphoinositide 3-kinase (PI3-kinase), Son-of-sevenless (SOS), Rho GTPase-Activating Protein (RhoGAP) and Dynamin. Furthermore, Pleckstrin, as noted earlier, but also Phospholipase C-γ isoforms (PLC-γ) and T-lymphoma invasion and metastasis 1 (Tiam1) are examples of proteins that possess two PH domains (Bottomley *et al.,* 1998). What is more, due to the stable nature of the PH domain fold, several examples of a split PH domain have been reported (Wen *et al.,* 2008, Wen *et al.,* 2006, Yan *et al.,* 2005) – that is a PH domain interrupted by additional domains, but forming a complete canonical PH fold. The PH domain of α-synotrophin, as an example, is interrupted by a PDZ domain and was the first split PH domain to be structurally characterised (Yan *et al.,* 2005). Interestingly, the N-terminal half (PH_N) and the C-terminal half (PH_C) of the split PH domain fold *de novo* if individually expressed as unfolded half-domains and mixed in solution (Yan *et al.,* 2005). Comparisions of the five split PH domains that have been structurally characterised have revealed some salient features of this subclass of domains, that is (1) additional domains are inserted into the loop regions (e.g. β3-β4, in α-Syntrophin and PLC-γ; β6-β7, in ROCK and VPS36; and β5-β6, in PIKE) and (2) the additional domain(s) co-operate with the PH domain to form a supramodule with distinct biological functions (Wen *et al*. 2008, Yan *et al.,* 2005).

Initially, the pleckstrin homology domain was described as a lipid-dependent membrane-targeting module (Harlan *et al.,* 1994), which permits the membrane-association of the host protein in a phosphoinositide-dependent manner (Lemmon and Ferguson 2002). This was originally thought to be a function widely associated with PH domains; however, yeast genome-wide analysis of PH domains revealed that the majority of PH domains do not bind with high affinity or specificity to PtdInsP ligands (Yu *et al.,* 2004). It is now considered that of the ~234 PH domains in Humans, approximately 10% - 20% are known to bind specifically, and with high affinity, to PtdInsP lipids (Lemmon 2008, DiNitto and Lambright

2006). Several PH domains are well-described as lipid-dependent membrane-targeting domains (described below), but these constitute a small subset in the PH family. The function of the remaining fraction is still unclear (Lemmon 2008).

4.1.3. The PH domain as a Juxta-Membrane Targeting Module

The reported function of PH domains to bind to acidic lipids, particularly PtdInsP ligands, *via* its interaction with the positively charged face permitting targeting of the protein to the membrane (Harlam *et al.,* 1994), has been the focus of research into PH domains for a number of years (Lemmon 2008). A wide variety of biophysical techniques have been employed to rigorously study the binding of PH domains to PtdIns lipids, including Surface Plasmon Resonance (SPR) and Isothermal Titration Calorimetry (ITC) to determine accurate dissociation constants (DiNitto and Lambright 2006). To date, nearly all PH domains bind to acidic lipids in some capacity, but a relatively minor number of PH domains have been characterised as efficient, stereo-specific, PtdInsP lipid species-recognising domains, with only ~10% of all PH domains considered to possess this function (Lemmon 2008, DiNitto and Lambright 2006, Lemmon and Ferguson 2000). This subset of PH domains typically bind to cognate PtdInsP ligands with similar dissociation constants (K_d) ranging from the low nanomolar to the low micromolar (Lemmon 2008, DiNitto and Lambright 2006, Bottomley *et al.,* 1998). Well-characterised domains that demonstrate high-affinity stereo-selective PtdInsP binding include the Phospholipase C (PLC)-δ1 [PtdIns $(4,5)P_2$ specific] (Ferguson *et al.,* 1995), Bruton's tyrosine kinase (BTK) [PtdIns$(3,4,5)P_3$ specific] (Fukuda *et al.,* 1996, Salim *et al.,* 1996) and General receptor of Phosphoinositides-1 (Grp1) [PtdIns$(3,4,5)P_3$ specific] PH domains (Klarlund *et al.,* 1997, Kavran *et al.,* 1998, Klarlund *et al.,* 2000). Other PH domains with characterised preferences are listed in Table 4.1.

The PH domain of PLC-δ1 is sufficient to target the enzyme to the membrane that contains its substrate [PtdIns $(4,5)P_2$] (Paterson *et al.*, 1995). The structure of the N-terminal PH domain of PLC-δ1 in complex with its lipid headgroup $Ins(1,4,5)P_3$, for which it has a dissociation constant of 0.21μM and high stereo-selectivity (Lemmon and Ferguson, 2000), provided the first structural insights into its stereo-specificity (Ferguson *et al.* 1995).

Table 4.1 – Affinities of PH domains for PtdInsP Ligands[a]

PH domain	Ligand	K_D (μM)	Reference
PLC-δ1	$Ins(1,4,5)P_3$	0.21	Lemmon *et al.*,(2005)
	$PtdIns(4,5)P_2$	1.60	
Grp1 (2G Variant)	$Ins(1,3,4,5)P_4$	6.30	Kavran *et al.*, (1998)
(3G Variant)		4.47	Cronin *et al.*, (2004)
Grp1 (2G Variant)	$Ins(1,4,5)P_3$	0.035	Kavran *et al.*, (1998)
(3G Variant)		0.50	Cronin *et al.*, (2004)
BTK	$Ins(1,3,4,5)P_4$	0.040	Baraldi *et al.*, (1999)
PDK1	PtdIns(3)P	0.084	Curie *et al.*, (1999)
	$PtdIns(3,4)P_2$	0.0052	
	$PtdIns(4,5)P_2$	0.024	
	$PtdIns(3,4,5)P_3$	0.00016	
DAPP1	$Ins(1,3,4)P_3$	0.15	Ferguson *et al.*,
	$Ins(1,3,4,5)P_4$	0.043	(2000)
	$PtdIns(3,4,5)P_3$	0.003	
			Thomas *et al.*, (2001)
FAPP1	PtdIns(4)P	0.020	Dowler *et al.*, (2000)
PEPP1	PtdIns(3)P	0.325	Dowler *et al.*, (2000)

[a] Ligand Dissociation constants (K_D) as determined by ITC and SPR only.

Abbreviations; PLC, Phospholipase C; Grp1, General receptor of Phosphoinositides-1; BTK, Bruton's Tyrosine Kinase; PDK1, 3-Phosphoinositide-dependent protein kinase-1; DAPP1, Dual adaptor of phosphotyrosine and 3-phosphoinositides-1; FAPP1, Four-phosphate-adaptor protein-1; PEPP1, Phosphatidylinositol-(3) Phosphate-binding PH domain protein-1.

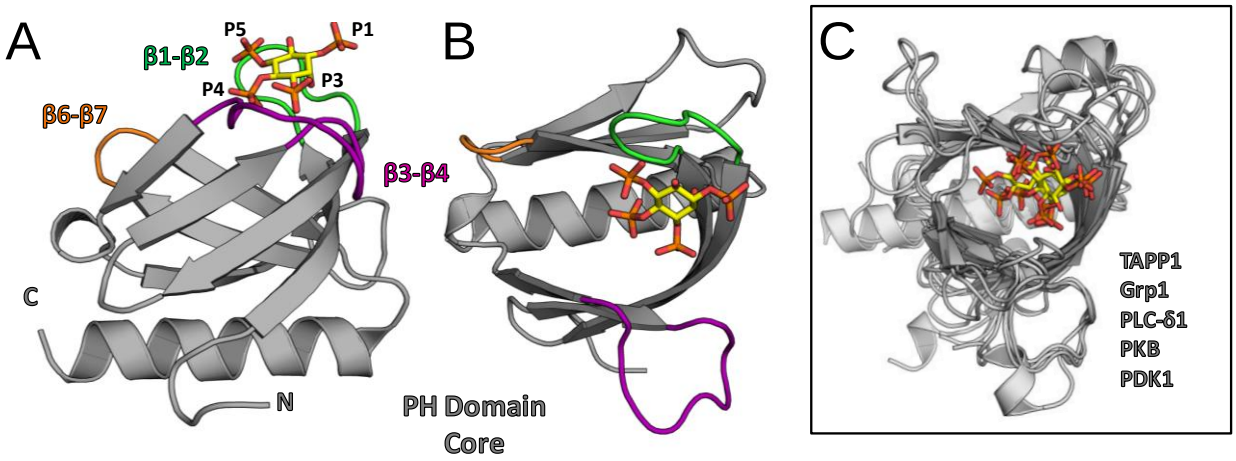


Figure 4.2 – Representative PH domain Ligand Binding. (A) Cartoon representation of Dual-adaptor of phosphoinositides and phosphotyrosine-1 (DAPP1, PDB Code: 1FA0) bound to inositol 1,3,4,5-tetrakisphosphate. The PH superfold (conserved structural core) is rendered in grey and the hypervariable loops are rendered green (β1-β2), purple (β3-β4) and orange (β6-β7). The structure is rotated about the plane of the page in (B). (C) Superimposition of TAPP1 (Tandem PH Domain-Containing Protein-1, PDB:1EAZ); Grp1 (General Receptor of Phosphoinositides, PDB:1FGY); PLC-δ1 (Phospholipase C-δ1, PDB:1MAI); PKB (Protein Kinase B or Akt, PDB:1UNQ); PDK1 (3-Phosphoinositide-Dependent Kinase-1, PDB:1W1D), demonstrating a common PtdInsP binding site. The ligand species, Ins(3,4)P_2, Ins(1,3,4,5)P_4, and Ins(1,4,5)P_3 are rendered as stick representation. This Figure was generated with PyMol (https://www.Pymol.org).

Crystallographic analysis of the PH domain in complex with IP_3 demonstrated that the lipid headgroup binds to the surface of the PH domain crowned by variable loops (VL) -1, -2 and -3 at the open end of the β-barrel (see Figure 4.2 as an example) (Ferguson *et al.*, 1995). The phosphates, on the other hand, are co-ordinated by amino acids from VL1 and 2 only (Lemmon and Ferguson, 2000, Ferguson *et al.*, 1995). Interestingly, the crystal structure of the intact PLC-δ1 revealed that the N-terminal PH domain was disordered and tethered to the enzyme *via* a flexible linker (Essen *et al.*, 1996), thus allowing the phospholipase to associate with the membrane when substrate is present and efficiently hydrolyse lipid substrate without a dissociation and subsequent re-binding event after each catalytic cycle (Lemmon and

Ferguson, 2000). The PH domain of PLC-δ1 binds with greater (8-fold higher) affinity to the soluble headgroup, Ins(1,4,5)P_3 [IP_3] over the membrane embedded PtdIns(4,5)P_2 thus displacing the PH domain from the membrane when cytoplasmic IP_3 is generated (Lemmon 2008, Hirose *et al.*, 1999). Interestingly, the N-terminal PH domain of the lipid metabolizing enzyme, PLC-δ1, was demonstrated to bind to PtdIns(4,5)P_2 and its soluble metabolite D-myo-inositol (1,4,5) trisphosphate [Ins(1,4,5)P_3] before PH domains were in fact discovered, with the N-terminal portion (a proteolytic fragment) of PLC-δ1 considered a PtdInsP specific binding site in a proteolysis study (Cifuentes *et al.*, 1993).

Following the functional description of PH domains as a phosphoinositide binding modules, Franke and colleagues (1995) reported that the serine/threonin protein kinase B (PKB, also known as AKT) was a downstream effector of platelet-derived growth factor (PDGF) receptor-activated Phosphatidylinositol (PI)-3 kinase signalling, and was dependent on its PH domain. Thus, the response (membrane targeting) of PKB as a result of PI-3 kinase activity derives from the specific recognition of the receptor-activated and transiently-generated PI-3K products; PtdIns(3,4)P_2 and PtdIns(3,4,5)P_3, by the PKB PH domain (Lemmon 2008, Lemmon and Ferguson 2000). Other PH domains are well-known effectors of lipid second messengers analogous to PKB, which includes those PH domains from Bruton's tyrosine kinase (BTK) and General receptor for phosphoinositides (GRP)-1 and thus recognize PI-3K products (3-phosphoinositides) with high affinity and specificity (Lemmon 2008). The specificity and affinity of these PH domains is striking and allows proper membrane targeting to membranes that contain the 3-phosphoinositides, despite there being 10-20 fold higher levels of PtdIns(4,5)P_2 (Lemmon and Ferguson 2000). These PH domains represent classical signal-dependent membrane targeting modules (Lemmon 2008) that are responsible for the membrane recruitment (and activation) of signalling molecules, i.e. the activation of kinases as a result of surface receptor activation. Mutations in the PH domains

that ablate 3-phosphoinositide binding result in severe signalling defects, as exemplified by the human diseases X-linked agammaglobulinemia when the BTK PH domain is mutated (Lindvall *et al.,* 2005). Conversely, mutations in the PH domain that promotes the constitutive membrane-association can result in cancer due to aberrant signalling, as suggested for PKB/AKT1 (Carpten *et al.,* 2007).

Several crystal structures of PH domains in complex with soluble PtdInsP headgroups have elucidated the structural determinants of stereo-selectivity and high-affinity binding, but all demonstrate a similar binding site for PtdInsP lipids in these PH domains (Figure 4.2) (Lemmon 2008, DiNitto and Lambright 2006). The open end of the β-barrel, together with variable loop 1 (VL1) between the first and second β-strands, forms the foundation of PtdInsP interactions (Lemmon 2008, DiNitto and Lambright 2006). Crystal structures of high affinity binders have highlighted a basic motif K-X_n-(K/R)-X-R (DiNitto and Lambright 2006) in which the basic residues Arginine (R) and Lysine (K) mediate interactions with the phosphate groups (Lemmon 2008, DiNitto and Lambright 2006), with corresponding residues co-ordinating phosphates at the 4 and 5 positions seen in the PtdIns(4,5)P_2 bound PLC-δ1 PH domain crystal structure. Additional interacting side chains from the binding site define the PH domain's preferred phosphorylation pattern and stereo-chemistry; however analysis of multiple PH domain crystal structures bound to PtdInsP headgroups has allowed robust rules for predicting ligand preference for ligand binding (Lemmon 2008; DiNitto and Lambright 2006; Ferguson *et al.,* 2000). Interestingly, crystal structures that co-ordinate inorganic phosphate and sulphate ions, from the crystallisation solution, demonstrate the basic motif's primary role as a co-ordinator of adjacent phosphates in PtdInsP ligands (DiNitto and Lambright 2006). The differences in orientation (and stereo-selectivity) of PtdInsP headgroups can be attributed to the pseudosymmetry of the inositol ring and the secondary structure of the variable loops (DiNitto and Lambright 2006).

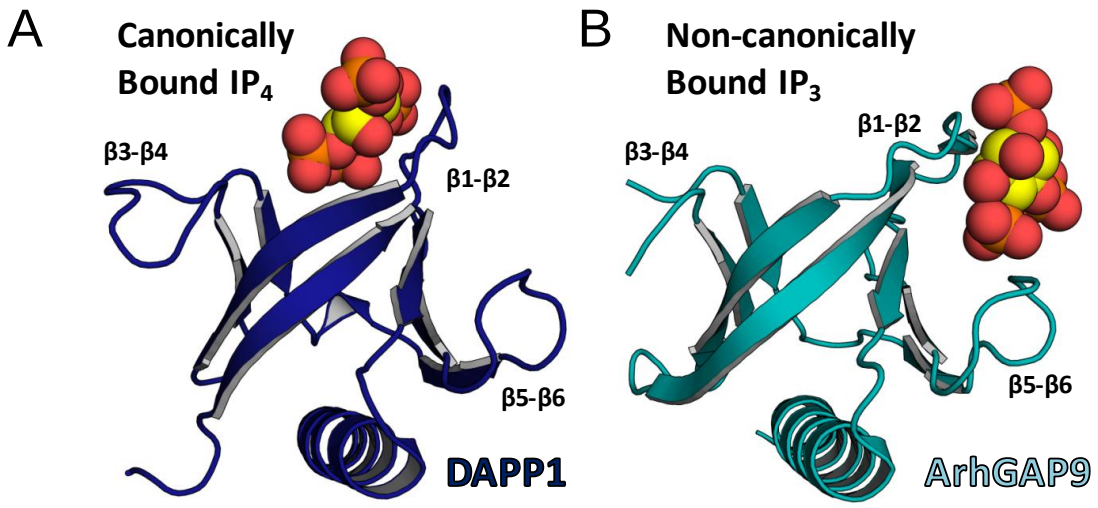


Figure 4.3. – Phosphatidylinositol Phosphate Binding Modalities in PH Domains. A structural comparison of PH domains of (A) Dual-adaptor of phosphoinositides and phosphotyrosine-1 (DAPP1; rendered dark blue) bound to inositol 1,3,4,5-tetrakisphosphate [Ins(1,3,4,5)P_4], which is considered a canonical binder, and (B) GTPase activating protein, ArhGAP9 (rendered cyan), bound to inositol 1,4,5-trisphosphate [Ins(1,4,5)P_3], which is considered a non-canonical binder of PtdInsPs. In the case of the canonical binding modality, the ligands sits within the open end of the β-barrel, co-ordinated by the so-called variable loops that crown the lip of the barrel (β1-β2 loop, β3-β4 loop and the β6- β7 loop [obscured from view]). In the non-canonical binding modality the ligand is packed against the C-terminal half of the β-barrel of the PH domain fold, however, one of the variable loops (VL1; β1- β2 loop) still co-ordinates the ligand, but using the opposite face of the loop compared to the canonical mode. Furthermore, the β5- β6 loop is also involved in the co-ordination of the ligand along with residues on the β-strands. This Figure was generated with PyMol (https://www.Pymol.org).

The binding of PtdInsP by PH domains at a site within the open end of the β-barrel is considered a 'canonical' mode of PtdInsP recognition (Figure 4.3) (Lemmon 2008). However, PH domains from the guanine nucleotide exchange factors (GEF), T-lymphoma invasion and metastasis 1 (Tiam1) and Rho GTPase-activating protein-9 (ArhGAP9) bind to PtdInsP ligands through a distinct site away from the so-called canonical binding site (Lemmon 2008, Ceccarelli *et al.*, 2007). This 'non-canonical' binding site is located on the

opposite side of the β1-β2 loop, against the outside of the 4-stranded β-sheet of the β-barrel (Ceccarelli *et al.*, 2007) and is related to the Ins(1,4,5)P_3 binding site in β-spectrin (Hyvonen *et al.*, 1995) (Figure 4.3). In β-spectrin, moderate affinity, but specific recognition of Ins(1,4,5)P_3 at the non-canonical binding site is mediated by two short motifs K-X_n-R-S-W, within the β1-β2 loop, and Y-X-K in the β5-β6 loop regions (Ceccarelli *et al.*, 2007; Hyvonen *et al.*, 1995). Comparisons between the non-canonical binding site of ArhGAP9 PH domain and β-spectrin reveal that the PH domain of ArhGAP9 displays a more extensive network of interactions, accounting for its greater affinity for I(1,4,5)P_3 [0.1μM in ArhGAP9 and ~40μM in β-spectrin] (Ceccarelli *et al.*, 2007). However, a conserved Tryptophan residue is positioned identically in both cases and engages the IP_3 ligand *via* its indole nitrogen to the inositol ring (Ceccarelli *et al.*, 2007).

4.1.4. The PH domain as a Protein-Protein Interaction Module

It was first hypothesised that PH domains possess a protein-interaction function after its identification in G-Protein Coupled Receptor (GPCR) Kinase 2 (GRK2, also known as β-adrenergic receptor kinase-1; β-ARK) within a region of the kinase that was previously known to interact with G-protein subunits (Musacchio *et al.*, 1993, Shaw, 1993). Glutathione S-Transferase (GST)-fused GRK2 PH domain co-precipitated with the βγ subunits of trimeric G-proteins suggesting a general G-protein binding function of PH domains (Touhara *et al.*, 1994). Furthermore, crystallographic analysis of this interaction has revealed that the PH domain of GRK2 engages the $G_{\beta\gamma}$ subunit *via* its β2-β3 loop and β3 and β4 stands together with its C-terminal helix, which extensively overlaps the Gα interface in the trimeric G-protein (Lodowski *et al.*, 2003). Despite the GRK2 PH domain engaging the $G_{\beta\gamma}$ subunits, it is suggested that the PH domain counters anionic lipids – co-ordinating phosphate groups *via*

the canonical PtdInsP binding site, thereby binding to anionic phospholipids and membrane-associated $G_{\beta\gamma}$ thus targeting GRK2 to the membrane and attenuating GPCR signalling (Lodowski *et al.,* 2003). Additionally, PH domains of PLC-β isoforms are weakly activated by an interaction with $G_{\beta\gamma}$ *via* an initial interaction with their PH domains but do not bind to the same extent as the extended PH domain of β-ARK (Wang *et al.,* 1998). Thus it is now understood that PH domains do not have a general functionality, but instead impose a variety of specific functions on the host protein depending on their own modification on a common scaffold.

Another well-characterised example of PH domains that mediate protein: protein interactions are the Guanine nucleotide Exchange Factors (GEF) of the Dbl family, which generally contain a Dbl homology (DH) domain followed by a PH domain (Rossman *et al.,* 2005, Blomberg *et al* 1999). The Dbl homology domain is a specific Guanine nucleotide exchange factor (GEF) for the Rho family GTPases and crystallographic analysis of the Dbs: Cdc42 complex reveals that the PH domain, in this example assists in the productive nucleotide exchange activity of Dbs *via* a direct interaction with Cdc42, particularly the switch region that allows GTP binding (Rossman *et al.,* 2002). Other DH-PH tandem modules have been solved for Tiam1 (Worthylake *et al.,* 2000) and Son-of-Sevenless (SOS)-1 (Soisson *et al.,* 1998), however the PH domain position with respect to the DH GEF domain is a determinant for the PH domains' role in nucleotide exchange (Rossman *et al.,* 2002). Several crystal structures of GEFs complexed with their cognate substrates have been solved to date. Examples include; RhoA:Dbs complex (Snyder *et al.,* 2002), Dbs:Cdc42 complex (Rossman *et al.,* 2002), and Rac1:Tiam1 (Worthylake *et al.,* 2000). Comparative analysis suggests that the PH domain exhibits distinct modalities of protein interaction with no common binding motif; however the protein interaction site is near the closed end of the β-barrel, away from the PtdInsP binding site (DiNitto and Lambright 2006).

A distinct mechanism of protein: protein interactions mediated by the PH domain has also been proposed. Further to the split PH domain (see section 5.1.), some have reported an intermolecular PH domain as a form of protein interaction (Rossum *et al.*, 2005) whereby two halves of the PH-fold, present on different interacting proteins, complement to form a functional PH domain (Lemmon, 2005). Thus, Rossum and colleagues (2005) posited the notion that a C-terminal partial PH domain (PH_c) of Phospholipase C- γ1 (PLC-γ1) interacts with a partial, previously "invisible", amino-terminal partial PH domain of the Transient Receptor Channel (TRP) 3C – regulating surface expression of TRP3C. Intriguingly, the complementation of the partial PH domains from TRP3C and PLC-γ1 results in a phosphatidylinositol binding module, particularly PtdIns (4,5) bisphosphate (Rossum *et al.*, 2005). The structure of this split intermolecular PH domain was tested by NMR; however the data failed to provide any evidence of such an interaction between PLC-γ1 and the TRP-3C calcium channel (Wen *et al.*, 2006).

4.1.5. The Kindlin-2 PH Domain

As discussed previously (see Chapter 3), Kindlins possess a PH domain at the centre of the F2 subdomain of their FERM domain. However, the exact role of the PH domain is unclear. Recently, the role of the Kindlin-2 PH domain was suggested by Wu and co-workers, whereby the PH domain preferentially binds to the PI3K products PtdIns(3,4,5,)$P_{3,}$ and PtdIns(3,5)P_2 with additional recognition of PtdIns(4,5)$P_{2,}$ and with the greatest affinity for PtdIns(3,4,5)P_3 (Qu *et al.*, 2011). PI3K is well known to play an important role in regulating integrin-mediated cell adhesion, with the integrin activating protein Kindlin-2 targeted to membranes containing PI3K products PtdIns(3,4,5)P_3 enriched membranes, possibly linking PI3K activation with integrin-ECM adhesion (Qu *et al.*, 2011). What is more, the PH domain was previously shown to impair integrin activation when deleted (Ma *et*

al., 2008), and further studies suggest that PtdInsP binding is essential for integrin activation; however the PH domain, but not PtdInsP binding, is essential for focal adhesion assembly (Qu *et al.,* 2011).

4.2. Structural and Functional Characterisation of Kindlin-1 Pleckstrin Homology (PH) Domain

4.2.1. Summary

Kindlin FERM (*4*.1 band, *ez*rin, *r*adixin, *m*oesin) domain is hallmarked by its interruption by a Pleckstrin Homology (PH) domain at the centre of the F2 subdomain. Previously, high-throughput (HTP) expression screening revealed that the PH domain could be produced in milligram quantities in bacteria (see Chapter 3). We expressed, purified and crystallized the Pleckstrin Homology domain of Kindlin-1 and determined its 3-dimensional crystal structure at a resolution of 2.1 Å by molecular replacement. The structure revealed an archetypal PH superfamily fold at its core and an additional C-terminal extension beyond the canonical fold comprised of a second α-helix. This α-helix is integral to the fold of the domain and is found only in the Kindlin protein family. *In vitro* lipid overlay assays indicate that the PH domain preferentially binds to inositol mono-phosphate lipid moieties, with apparent low affinity. An isoform-specific invariant salt-bridge in the crystal structure occludes the potential binding site for phosphatidylinositol phosphate ligands, but molecular dynamics (MD) simulations demonstrate a transient switching of the salt-bridge to an alternative bonding pattern thus resulting in an open binding site. Docking experiments demonstrate that the open conformation enables ligands to dock in the binding pocket. Removal of the salt-bridge by site-directed mutagenesis confers an additional stereospecific di-phosphate binding capacity. Moreover, integrin activation assays in mammalian cells

uncovered the importance of the PH domain in Kindlin-1-mediated integrin activation and its involvement in focal adhesion assembly. Here, we have described the structure of the Kindlin-1 PH domain and characterized its function as a weak inositol mono-phosphate binder, due to a previously unseen occluding isoform specific salt-bridge. Furthermore, we demonstrate that Kindlin-1 PH deletion mutants directly affect focal adhesion assembly and impair integrin activation in cells, consistent to that which has been reported previously for Kindlin-2 (Ma *et al.*, 2008).

4.2.2. Kindlin-1 PH Domain Expression, Purification, Crystallisation and X-ray Diffraction Analysis

The PH domain of Kindlin-1 (residues 364 – 509) was cloned as part of a HTP cloning and expression screen (see Chapter 3), but briefly, we cloned the PH domain of *M. musculus* Kindlin-1 (*FERMT1* cDNA was a kind gift from R. Faessler, MPI, Martinsreid, Germany) using the forward primer 5'-AAGTTCTGTTTCAGGGCCCGgaggacataactgacatccctaaact-3' and reverse primer 5'-ATGGTCTAGAAAGCTTTAttcaagactggaagccacca-3'into the pOPINF vector (Berrow *et al.*, 2007), which encodes a cleavable (3C protease) N-terminal His-tag under the control of T7 polymerase and the lac-operator. We subsequently used Rosetta (DE3) pLysS/RARE *E. coli* as an expression strain in the first instance. We discovered during small scale expression studies that SDS-PAGE analysis of whole cell extract demonstrated no appreciable difference between cultures grown in the presence or absence of IPTG. It was considered that the PH domain expressed constitutively in the Rosetta pLysS/Rare *E. coli*

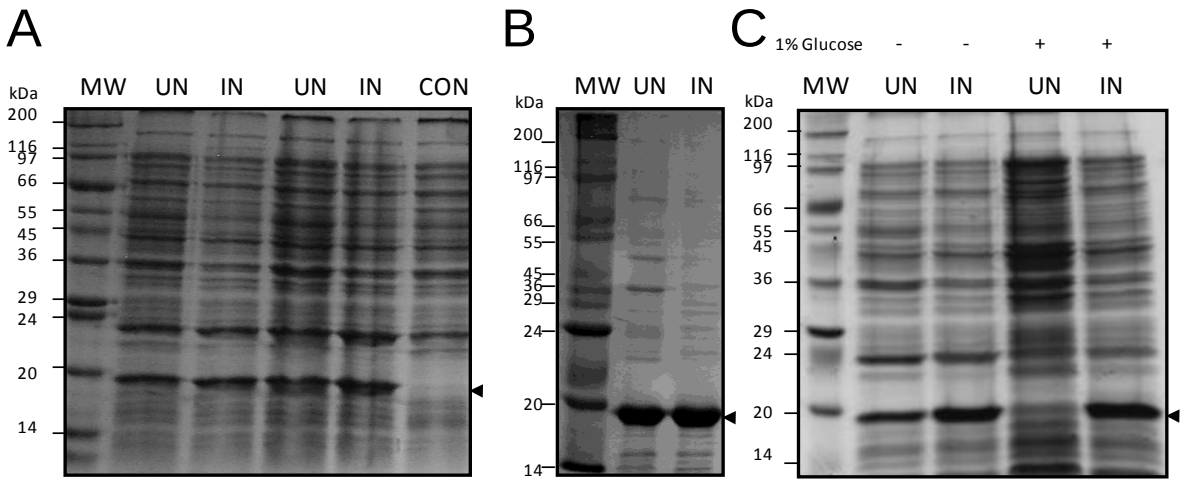


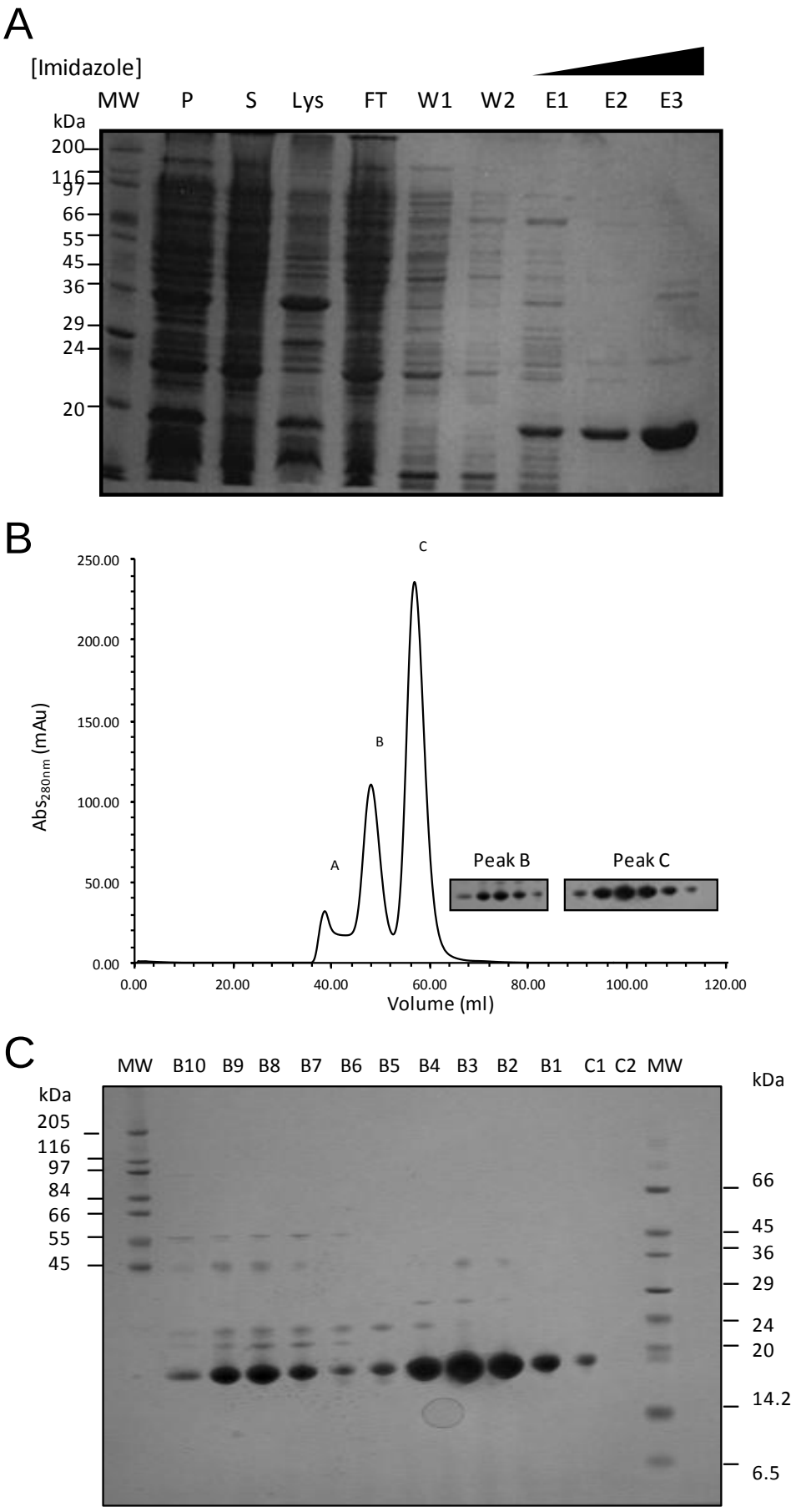
Figure 4.4 – SDS-PAGE Analysis of Constitutive Expression of Kindlin-1 PH domain in Rosetta (DE3) pLysS/RARE. (A) Constitutive expression of an ~18kDa protein (PH domain, arrowed) in the presence (lane labelled IN) or absence (lane labelled UN) of IPTG (1mM final) compared to *E. coli* not carrying recombinant vector (lane labelled CON), suggesting that the ~18kDa PH domain expresses in *E. coli* without IPTG induction. (B) Talon^{TM} Cobalt immobilised metal affinity chromatography (IMAC) of the soluble fraction from lysed cultures grown in the absence of (labelled UN) and the presence of (labelled IN) IPTG, thus demonstrating an 18kDa protein (arrowed), consistent with the PH domain being expressed in *E. coli* constitutively. IMAC was performed using a batch method. (C) Inhibition of protein production *via* the addition of 1% (w/v) glucose cultures grown in the presence (lane labelled IN) or absence (lane labelled UN) of IPTG, providing evidence for *Lac operon* controlled expression of an 18kDa protein (arrowed).

when whole cell extract of the same expression strain not carrying the pOPINF-PH domain plasmid was used as a negative control (Figure 4.4). Additional proof came from small scale immobilised metal affinity chromatography (IMAC) purification using Cobalt resin (Talon^{TM}, Clontech) and cultures grown in the presence and absence of IPTG. SDS-PAGE analysis of the proteins purified by IMAC revealed a ~18kDa protein, consistent with the PH domain construct we cloned, in both cultures and suggested the presence of a His-tagged protein (see Figure 4.4). Finally, the addition of 1% (w/v) glucose to the culture confirmed that the same 18kDa protein, that could be purified *via* Talon IMAC, was not present and its expression could be induced with the further addition of IPTG [1mM final concentration] (see Figure 4.4). Overall, these data indicate that expression of the PH domain is unusually leaky in

Rosetta *E. coli* and that the constitutive expression is comparable to IPTG induced expression. The addition of glucose suggests the transcription factor Catabolite Activated Protein (CAP) that is activated by cyclic adenosine monophosphate (cAMP) as the culprit in leaky expression from the *Lac* promoter, although a thorough investigation to confirm this is out of the scope of this work. Thus, large scale protein production was achieved by incubating 1 litre cultures in Luria broth (LB) at 37ºC until saturation (optical density at 595nm of ~4.0) in the presence of antibiotics only.

We initially purified the PH domain from the soluble fraction of bacterial cell lysate from 4 grams of bacteria (~1l of bacterial culture in LB) using nickel affinity chromatography due to the N-terminal His-tag. The PH domain was partially purified using a single stepped elution. However, SDS-PAGE analysis of the fractionated elution suggested that the elution protocol could be improved, due to obvious contamination. Accordingly, the purification was optimised using a stepped gradient elution using imidazole concentration increments of 50mM. The protein content of the eluate was assessed by SDS-PAGE, which suggested an improved purification procedure and those fractions appearing to contain the PH domain were pooled and concentrated. Purification of the PH domain to homogeneity was only achieved by using size exclusion chromatography (Figure 4.5). Three peaks were observed when monitoring the elution profile using a wavelength of 280nm. Of the three peaks (labeled A, B and C in Figure 4.5) peak A represents the void volume, and peaks B and C, suggest proteins of 60kDa and 40kDa, respectively, based on the volume of elution. SDS-PAGE analysis of the fractionated peaks B and C revealed that both contain an 18kDa protein consistent with the PH domain, therefore suggesting that the PH domain exists as oligomers with trimeric (Figure 4.5 - peak B) and a dimeric (Figure 4.5 - peak C) species present. SDS-PAGE analysis suggested that peak ‘C’ yielded >95% pure protein, which was used for subsequent studies. To investigate the oligomeric state of the PH domain further we used

Figure 4.5 - Purification of Kindlin-1 PH domain. (A) SDS-PAGE analysis of a typical Nickel IMAC of Kindlin-1 PH domain expressed in *E. coli*. The soluble (labelled S) and insoluble (labelled P) fractions of cell lysis were analysed together with the total cell lysate (labelled Lys). The soluble fraction was incubated with NiNTA beads using a batch method. The beads were removed from the soluble fraction, and analysed (labelled FT) and were subsequently washed, twice (labelled W1 and W2). The bound protein was eluted using a three step elution using increasing imidazole concentration (50mM labelled E1, 250mM labelled E2 and 500mM labelled E3). (B) Size exclusion chromatography (SEC) of immobilised metal affinity chromatography (IMAC) purified Kindlin-1 PH domain using a Sephadex S75 16/60 (GE Healthcare). Three peaks were observed when monitoring elution using 280nm wavelength. Of the three peaks (labelled A, B and C), A represents the void volume, both B and peak C contain the PH domain, as determined by SDS-PAGE analysis (shown inset), suggesting an oligomerisation of the PH. (C) SDS-PAGE analysis of the protein composition from SEC of peak B and C.


A
[Imidazole]
MW P S Lys FT W1 W2 E1 E2 E3
kDa
200
116
97
66
55
45
36
29
24
20
B
C
B
A
Peak B
Peak C
250.00
200.00
150.00
100.00
50.00
0.00
Abs280nm (mAu)
0.00 20.00 40.00 60.00 80.00 100.00 120.00
Volume (ml)
C
MW B10 B9 B8 B7 B6 B5 B4 B3 B2 B1 C1 C2 MW
kDa
205
116
97
84
66
55
45
kDa
66
45
36
29
24
20
14.2
6.5

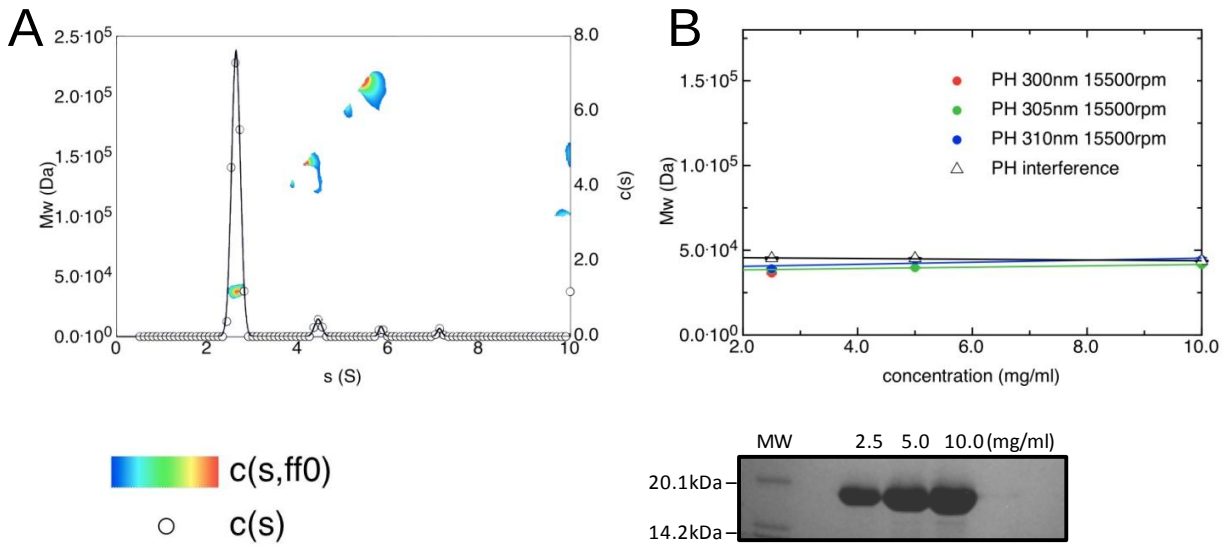


Figure 4.6 – Analytical Ultracentrifugation (AUC) Analysis of the Kindlin-1 PH Domain Oligomeric State. (A) Sedimentation velocity AUC data analysed by $c(s,f/f_o)$, the open symbol and Gaussian fit representing c(s) and the temperature contour plot $M_w(s)$. The Kindlin-1 PH domain has a sedimentation coefficient of 2.62S, and a weight between 30 and 40 kDa, which suggests it is a dimer given the monomeric protein is 18kDa. This is confirmed by sedimentation equilibrium analysis (B) using three concentrations (2.5mg/ml, 5.0mg/ml and 10mg/ml) of the Kindlin-1 PH domain (SDS-PAGE analysis of the protein shown below) and monitoring sedimentation using 3 wavelengths (300nm, 305nm and 310nm) and interference optics at 15,000 rpm. Sedimentation equilibrium analysis suggested a consistent MW of ~40,000Da using all three concentrations - indicating a dimeric species.

sedimentation velocity and equilibrium velocity analytical ultracentrifugation (AUC) so that a more accurate assessment could be made using proteins in solution and thus remove any influence that the chromatographic matrix imposes. We analysed the sedimentation velocity AUC data by $c(s,f/f_o)$, to reveal that the Kindlin-1 PH domain has a sedimentation coefficient of 2.62S, and a weight between 30 and 40 kDa (see Figure 4.6). Given that the PH domain is ~18kDa, this suggests it is a dimer in solution. Furthermore, sedimentation equilibrium analysis using three protein concentrations (2.5mg/ml, 5mg/ml and 10mg/ml) suggested that at all concentrations the molecular weight was calculated as ~40 kDa (see Figure 4.6).

As described earlier, homogenous protein (peak C in gel filtration – Figure 4.5) was concentrated and utilized in crystallization screening using commercially available crystallization reagents and vapor diffusion methods. Thus 100nl sitting drops of protein solution (5mg/ml - 12mg/ml in 20mM Tris-HCl, pH 7.5, 200mM NaCl, 0.5mM TCEP) were dispensed using a nanodrop crystallization robot (Cartesian Technologies) and diluted with an equal volume of crystallization solution (Figure 4.7) (Walter et al., 2005). The sitting drops were equilibrated against 100µl of crystallization reagent at 16˚C (289K) and were duplicated for equilibration at 4°C (277K). The His_6-tag was not removed prior to crystallization screening, but could be removed *via* treatment with 3C protease if necessary. Crystals appeared in several polyethylene glycol (PEG)-based conditions in 2-3 weeks at 4°C only. A single condition [20% (w/v) PEG3350, 200mM sodium phosphate] yielded several of the largest obelisk shaped crystals, with a dimension of ~400µm in the longest direction (see Figure 4.7). This condition was further optimized using a 3-way serial dilution protocol, with the crystallization condition [20% (w/v) PEG3350, 200mM sodium phosphate] ranging from 100% - 79% (diluted with sterile water) in one direction of the crystallization plate experiment together with three different protein: crystallization solution drop ratios i.e. 1:1 (protein : reservoir), 2:1 and 1:2, in another direction. Very large crystals grew in the majority of these conditions within 16 days at 4˚C, with the largest crystals ~1mm in the longest dimension (Figure 4.7). The crystals were flash frozen before their diffraction quality was tested using synchrotron X-ray radiation at Diamond light source (Didcot, UK). Initial frozen single crystals diffracted to 2.0Å resolution, however, ice-rings were present on the diffraction images, due to crystalline water, which obscured diffraction data. Therefore several crystals were soaked in different cryo-protectants to reduce the formation of ice crystals during the freezing process. Glycerol was deemed appropriate based on the quality of

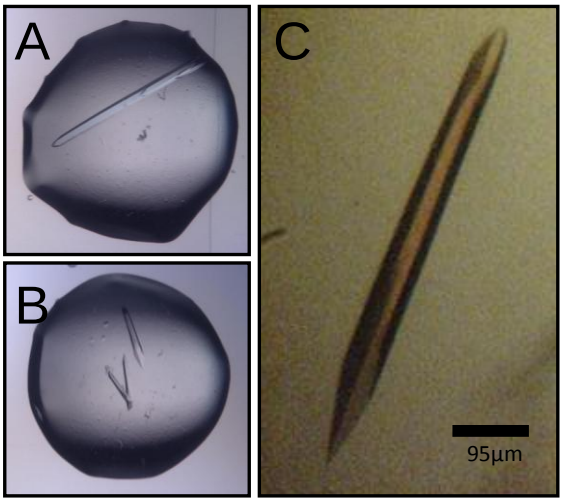


Figure 4.7 - Crystallisation of Kindlin-1 PH domain. Images of Nanoliter crystallisation experiments using sitting drops and the vapour diffusion method (A and B). Sitting drops were equilibrated against 100µl of crystallisation solution. The single crystal in (B) is ~1mm in the longest dimension. (C) Image of a single crystal of dimensions ~370µm x ~20µm x ~20µm, used for X-ray diffraction data collection. All crystals were grown in PEG 3350 based conditions at 277K.

diffraction that resulted from single frozen crystals and therefore the concentration of glycerol was optimized to improve the diffraction data quality. Thus, the crystals were bathed in mother liquor supplemented with 30% (v/v) glycerol and directly frozen in liquid nitrogen prior to data collection.

Diffraction data (see Figure 4.8 for diffraction image) from a single native Kindlin-1 PH domain crystal were collected under cryo-temperature conditions using single wavelength synchrotron radiation at Diamond Light Source (Didcot, UK) on beamline IO3. Indexing of the diffraction data was performed using *MOSFLM* (Leslie and Powell, 2007) and the automatic data reduction software Xia2 (Winter, 2009). The best autoindexing solution in MOSFLM suggested it belonged to the C-centered orthorhombic space group but the correct space group was properly determined once data were integrated and scaled. Scaling of indexed data was performed using SCALA in CCP4 (CCP4, 1994) or in the automatic data

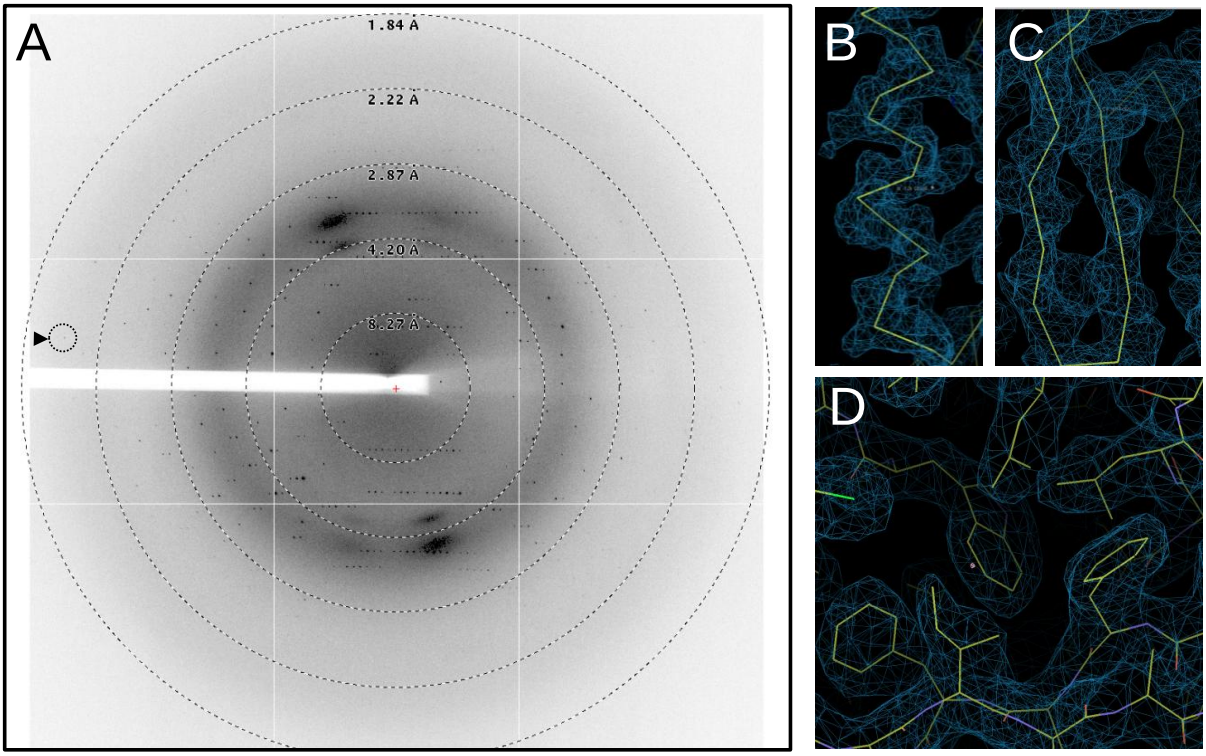


Figure 4.8 - (A) A diffraction image from a single cryo-frozen crystal at 77K using synchrotron radiation at beamline I03, Diamond Light Source, Oxford. The crystal was cryo-protected in 30% (v/v) glycerol before being exposed to X-ray radiation at a wavelength of 0.973Å. The diffraction image was collected on an ADSC Q315R detector. Resolution ranges are shown as dashed lines and are labelled on the diffraction image. Reflections were observed at 2.1Å (circled with dotted line). (B-D) High quality electron density map (2Fo- Fc) from Kindlin-1 PH Domain, solved using molecular replacement, displaying helical density for an (B) α-helix and (C) several strands of a β-sheet meander together with (D) electron density of large hydrophobic side chains.

reduction software Xia2 (Winter 2009), which successfully determined the correct space group as $C222_1$. The crystal possesses the unit cll dimensions, a=45.7Å; b = 84.1Å; c= 95.9Å; $\alpha = \beta = \gamma = 90^{o}$ and contains one molecule per asymmetric unit as determined by Matthews' coefficient (Matthews, 1968) which suggested a solvent of content of 49% with one protein molecule of ~18kDa protein. X-ray diffraction data collection statistics are summarized in Table 5.2. The structure of the Kindlin-1 PH domain was solved by molecular replacement (MR) using PHASER (McCoy *et al.*, 2007) in CCP4 (CCP4, 1994) using the PDB coordinates at accession code 2YS3 (Li *et al.*, unpublished) as a search model. Separate

MR calculations in PHASER were performed using individual models from the 20 model NMR ensemble. The models were converted to a poly-alanine chain in CHAINSAW (CCP4, 1994) to reduce model bias prior to MR. A single model, model 20, yielding an optimal translation and rotation *Z*-score (RF=6.3 and TF=6.9) and log likelihood gain (LLG=59) was subsequently used. MR calculations in space group C222 did not yield a solution, whereas MR calculation in space group $C222_1$ yielded a single solution. Phases estimated from this model were used in the automatic model building program ARP/wARP (Evrard *et al.*, 2007). Model building in Coot (Emsley and Cowtan, 2004) was continued in the resulting high quality electron density map from ARP/wARP (Figure 4.8) and the model was refined using a translation libration screw (TLS) refinement protocol and individual B-factor refinement in Phenix (McCoy *et al.*, 2007). TLS groups were determined using the TLS motion determination server (TLSMD) (Painter and Merritt, 2006). Water molecules and a single glycerol molecule were added in Coot and the water was updated during rounds of refinement in Phenix. The refinement resulted in an R_{free} and R_{work} of 21.3% and 18.2%, respectively. The crystal structure was validated using MolProbity (Chen *et al.*, 2010) with 98.3% of the refined residues in favoured regions and none in disallowed regions. The final model contains residues 369 – 497 (mouse *FERMT1* numbering) however, weak electron density was observed for residues 382 – 388 and therefore not built. The crystallographic data collection and refinement statistics are summarized in Table 4.2.

Table 4.2 – Data Collection and Refinement Statistics for Kindlin-1 PH Domain

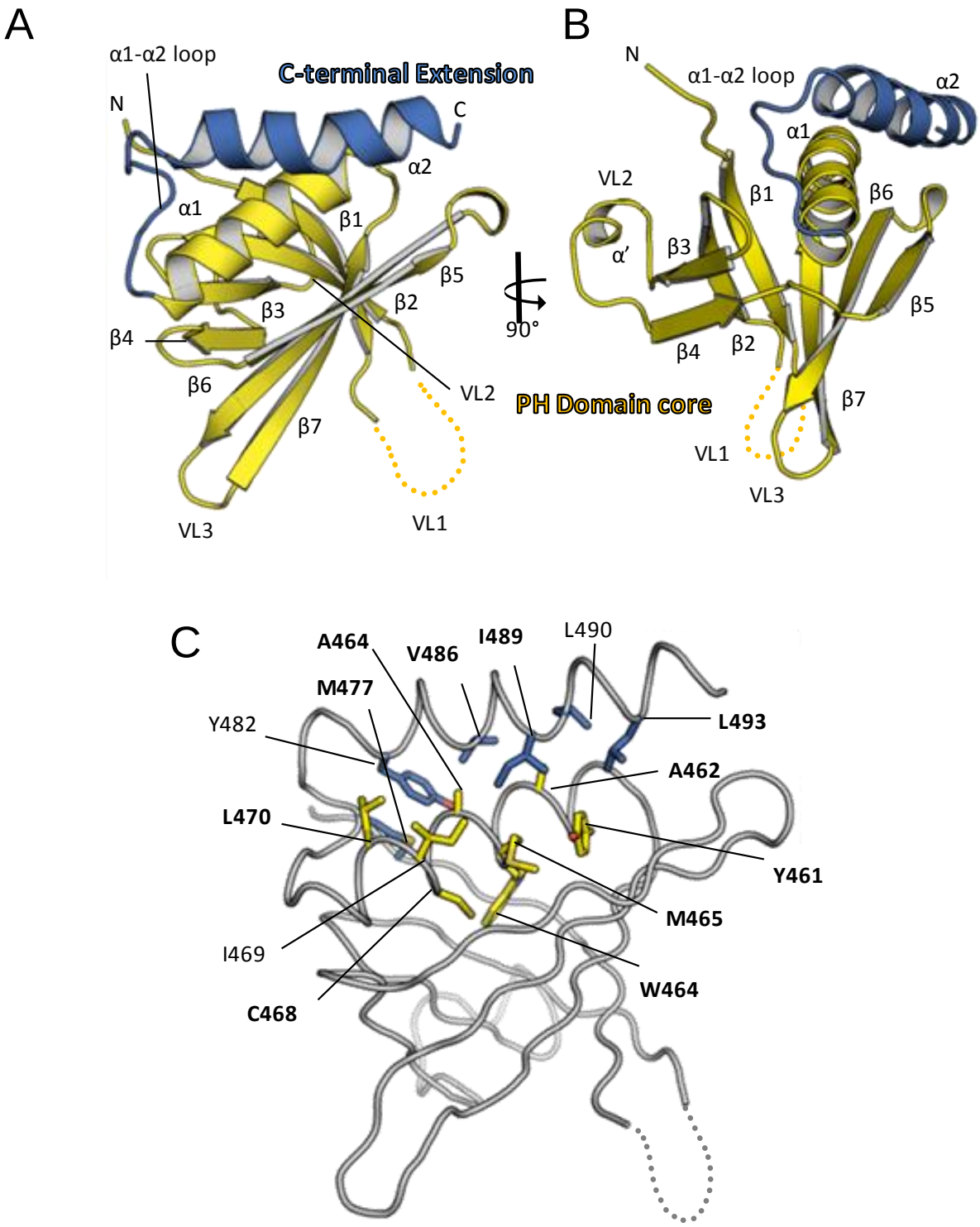
Data collection and refinement statistics	
Space Group	$C222_1$
Diamond Light Source Beamline	IO3
Wavelength/Energy (Å)	0.973
X-ray Detector	ADSC Q315r
Temperature (K)	~77.0
Unit cell (Å, °)	a = 45.29, b = 82.78, c = 95.99, α = β = γ = 90
Resolution (Å)	41.4 – 2.1 (2.21 – 2.1)
Completeness (%)	95.4 (95.2)
Redundancy	7.0 (7.3)
$I/\sigma I$	13.7 (3.5)
R_{merge} [b]	0.083 (0.565)
$R_{p.i.m.}$ [c]	0.223
Average Wilson B-factor (Å^2)	32.2
Model and refinement	
Resolution (Å)	40 – 2.1 (2.4 – 2.1)
Unique Reflections	10324 (1478)
R_{work} [c]	0.1821
R_{free} (using 5% of reflections)	0.2131
Model Contents	
Residues	370 – 496
Waters	49
Ligand	1 (glycerol)
Total number atoms	1040
Average B-factor (Å^2)	
Protein	42.4
Water	45.4
Glycerol	79.5
Root Mean Square Deviation from Ideal	
Bond lengths (Å)	0.008
Angles (°)	1.011
Ramachandran	
Favoured (%)	95.3
Allowed (%)	3.7
Generously allowed (%)	0.9
Disallowed (%)	0.0

[a] Values in parentheses account for the highest resolution shell.
[b] Merging *R* factor, *R*merge = $\sum |I_o - I| / \sum I_0$
[c] Precision indicating merging *R* factor, $R_{p.i.m.} = \sum_{hkl}[1/N-1]^{\frac{1}{2}} \sum_i \left|I_i(hkl) - \overline{I(hkl)}\right| / \sum_{hkl} \sum_i I_i(hkl)$

4.2.3. The Overall Structure of the Kindlin-1 PH Domain

The Kindlin-1 PH domain structure is comprised of 129 amino acids and exhibits an archetypal pleckstrin homology (PH) superfamily fold at its core that is a partly-open seven-stranded β-barrel capped at one end by a C-terminal α-helix. The amino terminal half of the PH domain is comprised of a four stranded anti-parallel β-sheet (β1 – β4) whereas the

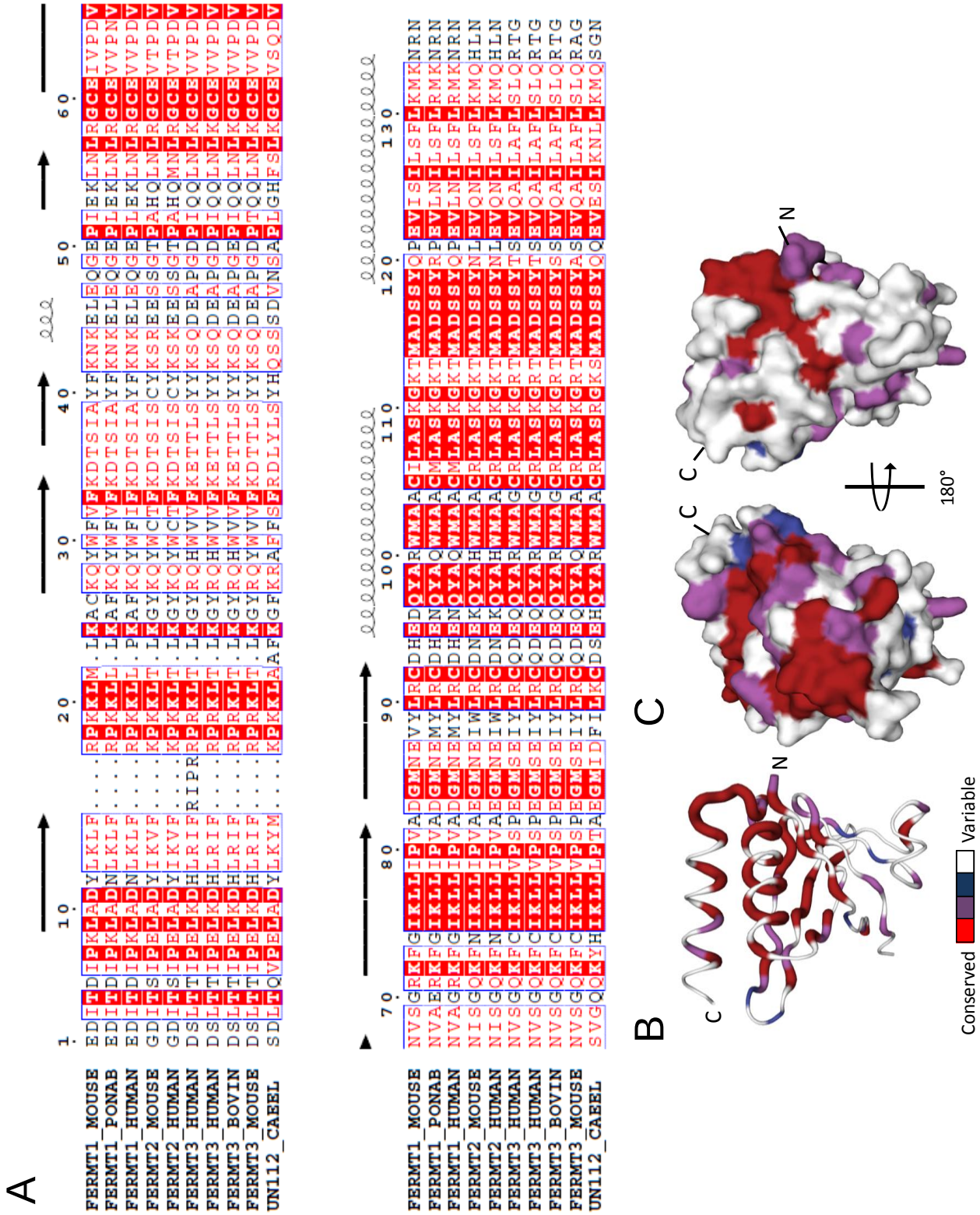
Figure 4.9 – (Opposite) Structure of Kindlin-1 Pleckstrin Homology (PH) domain. (A) The Kindlin-1 PH domain comprised of a core seven-stranded β-barrel (β1 – β7) capped with a two α-helices (α1 – α2). The open end of the β-barrel is decorated with variable loops (VL), with VL2 (β3 – β4 loop) forming a single-turn helix. The core archetypal PH domain fold is rendered in yellow, with the additional C-terminal extension α-helix, specifically found in Kindlin proteins, rendered in blue. (B) As in (A) but rotated 90° to show the β-barrel fold together with the variable loop regions. (C) Hydrophobic side chains that stabilise the stacking of the α1 and α2. Hydrophobic residues contributing to the core of stacking helices are shown as sticks. Residues are labelled according to mouse full length Kindlin-1 numbering (UniProt Accession code: P59113) and those bolded indicate invariant residues conserved across the Kindlin family in all species. All figures were rendered in PyMol (http://www.pymol.org).


A
α1-α2 loop
C-terminal Extension
N
C
α2
α1
β1
β5
β2
β3
β4
β6
VL2
β7
VL3
VL1
90°
B
N
α1-α2 loop
α2
α1
VL2
β1
β6
α'
β3
β5
β4
β2
β7
PH Domain core
VL1
VL3
C
A464
V486
I489
L490
M477
Y482
L493
A462
L470
Y461
M465
I469
W464
C468

remaining three strands form a second anti-parallel β-sheet (β5 – β7) that stacks nearly orthogonally against the first. This β-sandwich is filled with hydrophobic residues forming its core and is capped at one opening of the β-barrel (encircled by strands β1 and β7) by a predominantly hydrophobic C-terminal α-helix (α1). In PH domain folds, the helix that caps this opening of the β-barrel is typically amphipathic, with its hydrophobic residues buried in the core and polar residues exposed to the solvent (Lemmon *et al.*, 2002). However, in the case of the Kindlin-1 PH domain the majority of this helix is hydrophobic, with hydrophobic residues facing into and away from the β-barrel (Figure 4.9). This feature, which is conserved in all known Kindlin protein PH domains, stabilises the stacking of an additional C-terminal extension that is comprised of a single amphipathic helix (α2), which is not found in other PH domains (Figure 4.9). The interconnecting loop that lies between the α1 and α2 helices is highly conserved across all Kindlin isoforms (see Figure 4.10). An invariant methionine (Met^{477}), within the α1-α2 loop is involved in the stabilisation of the α1-α2 hydrophobic interactions (see Figure 4.9). The methionine side chain interacts with a tyrosine (Tyr^{482}) side chain from α2 and the hydrophobic core of the β-barrel fold. Sequence alignment between mammalian Kindlin isoforms and UNC-112, the sole Kindlin protein from *C. elegans*, indicates that those amino acids responsible for the hydrophobic properties of α1, and the amphipathicity of α2 that mediates the α1/α2 helix packing, are highly conserved with the α1/α2 loop invariant across all species (Figure 4.10). This suggests that this C-terminal extension is an integral part of the PH domain fold found in Kindlins.

Pleckstrin homology domains were first identified as phosphoinositide-binding modules that target proteins to the membrane in a phosphoinositide-dependent manner (Lemmon, 2008). The phosphoinositide binding affinity/specificity is governed by three variable loops (VL1, 2 and 3) that crown the open end of the β-barrel (β1/β2 [VL1], β3/β4

Figure 4.10 (Opposite) – Structure-based sequence alignment and projected conservation of Kindlin pleckstrin homology domains. (A) ClustalW multiple sequence alignment of the Kindlin-1 PH domain (residues 383-387, mouse numbering) with the Kindlin-1 crystal structure secondary structure shown above with labelling analogous to Figure 2. Invariant residues are shaded in red boxes, with residues that are conserved coloured red and variable residues shown in black. Multiple sequence alignment was prepared using ClustalW (83) and secondary structure superimposition prepared using ESPript (84). The sequences used are as follows: Kindlin-1 of mouse (UniProt Accession code: P59113), orang-utan (UniProt Accession code: Q5R8M5), and human (UniProt Accession code: Q9BQL6); Kindlin-2 of mouse (UniProt Accession code: Q8CIB5) and human (UniProt Accession code: Q96AC1); Kindlin-3 of mouse (UniProt Accession code: Q8K1B8), human isoform I (long) (UniProt Accession code: Q86UX7-1), human isoform II (short) (UniProt Accession code: Q86UX7-2), and cow (UniProt Accession code: Q32LP0); Unc-112 of the nematode worm (UniProt Accession code: Q18685). (B) Kindlin PH domain sequence conservation mapped onto a 3-dimensional structure of Kindlin-1 PH domain. Sequence conservation is shown with invariant residues coloured red, highly conserved rendered purple, semi-conserved rendered blue and variable rendered white. (C) Kindlin PH domain sequence conservation projected onto a surface representation of Kindlin-1 PH domain. Amino acid conservation is rendered as in (B). Structural representation of sequence conservation was performed using ESPript and rendered in PyMol


A
B
C
FERMT1_MOUSE
FERMT1_PONAB
FERMT1_HUMAN
FERMT2_MOUSE
FERMT2_HUMAN
FERMT3_HUMAN
FERMT3_BOVIN
FERMT3_MOUSE
UN112_CAEEL
C
N
180°
Conserved
Variable

[VL2], and β6/β7 [VL3] loops) (DiNitto and Lambright, 2006) (Figure 4.9). The β1/β2 loop is disordered in the Kindlin-1 crystal structure as suggested by a lack of electron density corresponding to this region, which suggests much conformational flexibility. This could also be inferred by the lack of ordered water molecules within that region. Additionally, at the β3/β4 loop Kindlin-1 forms a small single-turn helix, denoted here as α' (Figure 4.9) and similar to that of β-spectrin (Hyvonen *et al.*, 1995) and protein kinase B (PKB/Akt) (Thomas *et al.*, 2002) PH domains.

4.2.4. Structural Comparison of the Kindlin-1 PH Domain and Other PH Domains

A structural similarity search using the Kindlin-1 PH domain in DALI (Holm and Rosenstrom, 2010) indicated, not surprisingly, that it is similar to other PH domains and phosphotyrosoine binding (PTB) domains, with a Z score range of 17.6 – 4.9 for > 500 structures. The corresponding r.m.s.d. values for these aligned structures ranged from 1.8 Å to 4.2 Å using 60 to 124 Cα atoms with pairwise sequence identities ranging from 6% to 26% with a single structural homologue (human Kindlin-3 PH domain, Li *et al*. unpublished) sharing 59% sequence identity (Z-score = 17.2, r.m.s.d. = 2.5Å). The top five structural homologues to the Kindlin-1 PH domain, as determined by DALI structural alignment searches, are: human Kindlin-3 PH domain (PDB code; 2YS3), phosphoinositol 3-phosphate binding protein-1 (PEPP1) (PDB code; 1UPR, Z-score = 13.8, r.m.s.d. = 2.3Å, 24% identity),

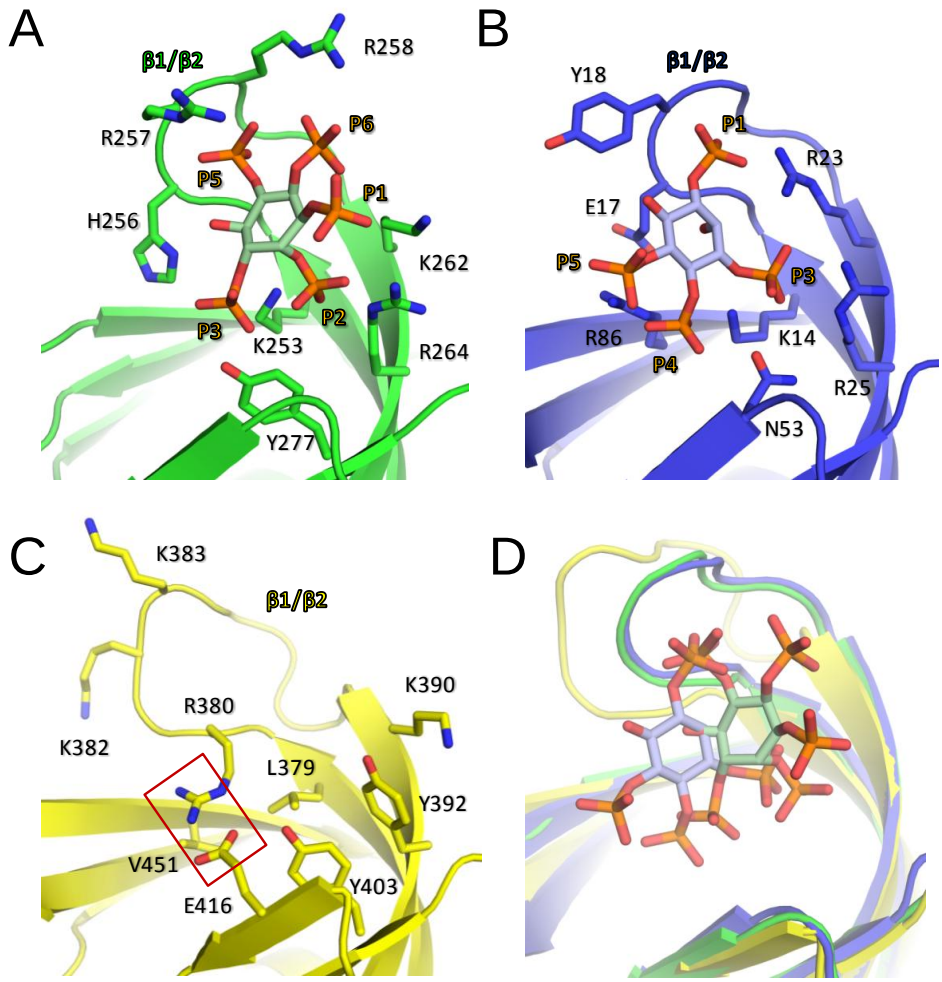
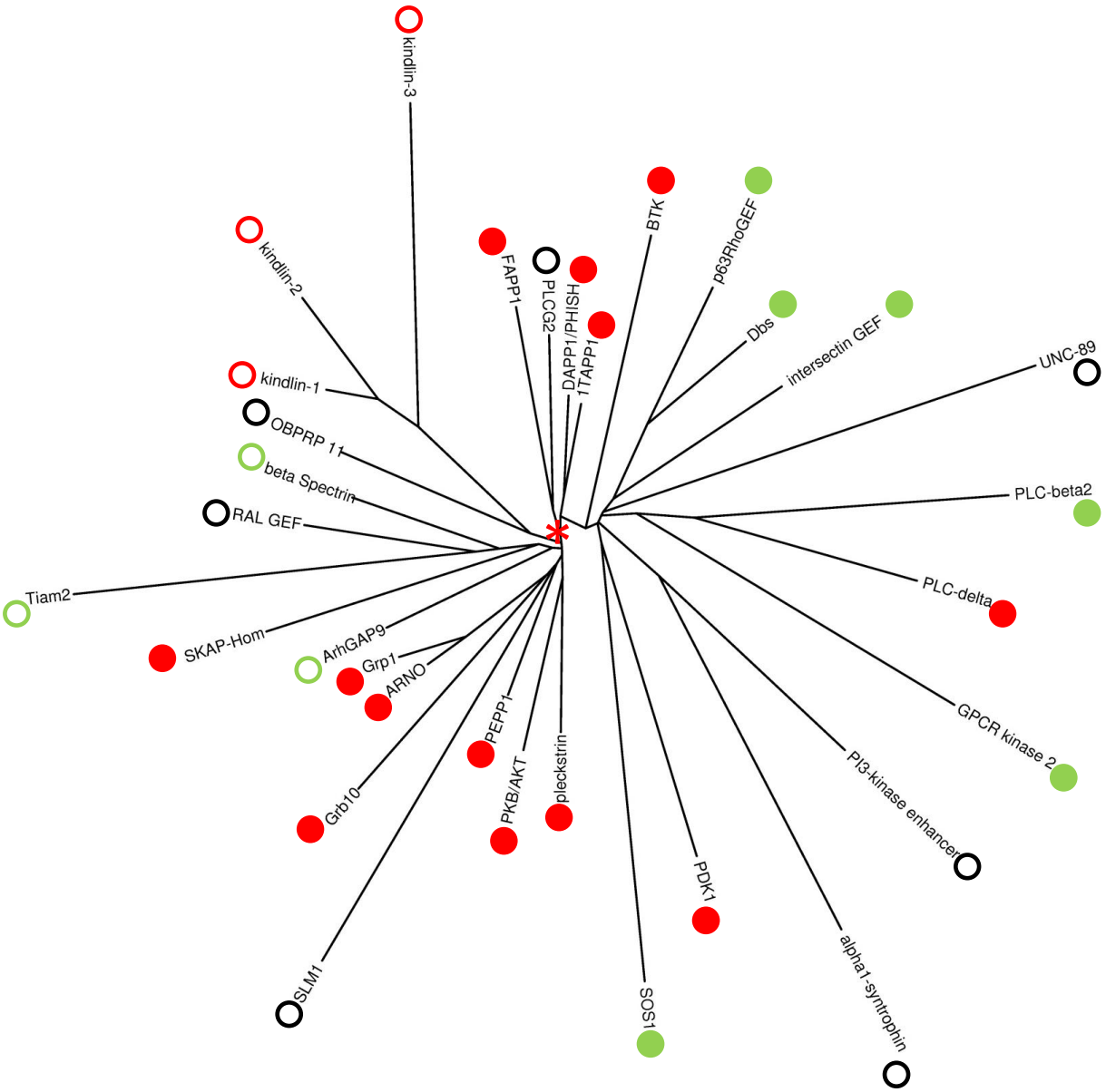


Figure 4.11 – Molecular detail of canonical PtdIns binding modalities. (A) Pleckstrin C-terminal PH domain crystal structure (rendered green) bound to D-myo-inositol 1,2,3,5,6-pentakisphosphate (IP_5) (PDB code: 2I5F) together with side chains of residues (labelled) that co-ordinate the phosphate groups of the ligand (Jackson *et al.* 2007). (B) Protein Kinase B (Akt) PH domain crystal structure bound (rendered blue) to inositol 1,3,4,5-tetrakisphosphate (IP_4) (PDB code: 1UNQ) together with side chains of residues (labelled) that co-ordinate the ligand. (C) Structure of Kindlin-1 PH domain with modelled β1-β2 loop (rendered yellow) not placed in the crystal structure (see Experimental Methods) together with side chains at equivalent positions to phosphate co-ordinating side chains of both pleckstrin and PKB. Note the salt bridge that occludes the canonical PtdInsP binding site (boxed in red). Loop β1-β2 is labelled in (A), (B) and (C) for spatial clarity. Each structure was superimposed using SHP (Stuart *et al.* 979) (D) and shows a similar binding site for the PtdIns ligand that is the opening of the β-barrel close to the β1-β2 loop, in both pleckstrin and PKB. The arrangement of basic residues that co-ordinate phosphates of PtdInsP ligand is similar in pleckstrin and PKB PH domains, but a lack of those residues in the Kindlin-1 PH domain structure is evident. Figures were rendered in PyMol (http://www.pymol.org).

the pleckstrin PH domain (PDB code; 2I5F, Z-score = 13.8, r.m.s.d. = 1.9Å, 15% identity), T-lymphoma invasion and metastasis 1/2 (Tiam1/2) PH domain (PDB code; 38AP, Z-score = 13.5, r.m.s.d. = 3.1Å, 13% identity) and human AKT3 kinase PH domain (PDB code; 2X18, Z-score = 13.3, r.m.s.d. = 2.3, 11% identity). Surprisingly, the Kindlin-2 PH domain NMR structure (Liu *et al.*, 2011) only had a Z-score of 12.8 (r.m.s.d. 2.8Å) and ranked 50th despite a sequence identity of 69%. Interestingly, some distant structural homologues include PTB-containing proteins such as SHC, with a Z-score of <5 but an r.m.s.d. of 2.8Å despite a sequence identity of only 7%. Structure comparison with PEPP1 and pleckstrin PH domain, the closest structural homologues according to DALI, suggests conformational deviations in the β1/β2 and β4/β5 loops. The crystal structure of the pleckstrin PH domain was complexed with D-myo-inositol 1,2,3,5,6-pentakisphosphate (IP_5) (Jackson *et al.*, 2007). Therefore sequence comparisons from the structural superimposition of Kindlin-1 PH and pleckstrin PH domains revealed that several basic residues required for the co-ordination of phosphates at the D2 and D3 position of IP_5 in pleckstrin are hydrophobic residues at equivalent positions in Kindlin-1 (Figure 4.11). Although residues within the Kindlin-1 PH β1/β2 loop (383-387) are considered conformationally flexible, sequence comparisons between the Kindlin β1/β2 loop and the pleckstrin β1/β2 loop suggest that two arginine residues in Kindlin may be able to co-ordinate phosphates analogously to pleckstrin. Furthermore, structural comparisons of the PH domain of protein kinase B (Milburn *et al.*, 2003), bound to inositol 1,3,4,5 tetrakisphosphate (PDB code: 1UNQ) suggests a similar arrangement of basic residues that co-ordinate phosphates of PtdInsP ligand to pleckstrin (Figure 4.11). In the case of Kindlin-1 equivalent residues are replaced with hydrophobic amino acids; however there are a limited number of basic residues that are equivalent and could potentially co-ordinate phosphates (Figure 4.11).

Figure 4.12 (Opposite) – A Structure-Based Phylogenetic Analysis of Kindlin-1 PH domains and other PH domains. The structures were superimposed and a pairwise alignment carried out using the program SHP (Stuart *et al.* 1979) to generate an evolutionary distance matrix. PDK1: protein dependent kinase 1 (*H. sapiens*, 1W1H); Pleckstrin (*H. sapiens*, 2I5F); PKB/AKT: protein kinase B/AKT (*H. sapiens*, 1UNQ); PEPP1 (*H. sapiens*, 1UPQ); SLM1: phosphatidylinositol 4,5 bisphosphate binding protein 1 (*S. cerevisiae*, 2NSU); Grb10: growth factor receptor-bound protein 10 (*H. sapiens*, 2HK0); ARNO: Arf nucleotide binding site opener (*M. musculus*, 1U27); Grp1: general receptor for phosphoinositides 1 (*M. musculus*, 1FGY); ArhGAP9: a rho GTPase activating protein (*H. sapiens*, 2P0D); SKAP-Hom: Src kinase associated adaptor protein 1 (*M. musculus*, 1U5E); Tiam-2: T-lymphoma invasion and metastasis-inducing protein 2 (*M. musculus*, 3A8P); RalGNF: Ral guanine nucleotide exchange factor (MS0666) (*M. musculus*, 2DTC): B-spectrin: beta spectrin (*M. musculus*, 1BTN); OBPRP11: oxysterol binding related protein 11 (*H sapiens*, 2D9X); Kind-1: Kindlin-1 (*M. musculus*) (this study); Kind-3: Kindlin-3 (*H. sapiens*, 2YS3); Kind-2: Kindlin-2 (*H. sapiens*, 2LK0); FAPP-1: four-phosphate-adaptor protein 1 (*H. sapiens*, 3RCP); PLCG1: 1-phosphatisylinositol-4,5-bisphosphate phosphodiesterase gamma-2 (*H. sapiens*, 2W2W); DAPP1/PHISH: dual adaptor of phosphotyrosine and 3-phosphoinositides (*H. sapiens*, 1FB8); TAPP1: tandem PH domain containing protein 1 (*H. sapiens*, 1EAZ); BTK1: Bruton's tyrosine kinase 1 (*H. sapiens*, 1B55); PIKE: phosphatidylinositol-3-kinase enhancer protein (*H. sapiens*, 2RLO); syntrophin: α-1-syntrophin (*M. musculus*, 2ADZ); p63RhoGEF: p63 Rho guanine nucleotide exchange factor (*M. musculus*, 2RGN); Dbs: guanine nucleotide exchange factor Dbs (*M. musculus*, 1KZ7); intersectin: guanine nucleotide exchange factor intersectin (*H. sapiens*, 1KI1); UNC-89: a muscle protein (*C. elegans*, 1FHO); PLC-beta: phospholipiase C-β2 (*H. sapiens*, 2FJU); PLC-delta: phospholipase C-δ (*R. Norvegicus*, 1MAI); GPCRK: G protein-coupled receptor kinase 2 (*H. sapiens*, 1BAK); SOS1: Ras exchanger son of sevenless (*H. sapiens*, 1AWE). Red filled circles indicate canonical PtdInsP binders, green open circles indicate non-canonical PtdInsP binders, open black circles indicate PH domains of unknown ligand tropism, and green filled circles indicate protein-binding PH domains and red open circles are used for the Kindlin domains. The root of the tree is marked by a red asterisk.


kindlin-3
kindlin-2
kindlin-1
OBPRP 11
beta Spectrin
RAL GEF
Tiam2
SKAP-Hom
ArhGAP9
Grp1
ARNO
Grb10
SLM1
PEPP1
PKB/AKT
pleckstrin
SOS1
PDK1
alpha1-syntrophin
PI3-kinase enhancer
GPCR kinase 2
PLC-delta
PLC-beta2
UNC-89
intersectin GEF
Dbs
p63RhoGEF
BTK
DAPP1/PHISH
1TAPP1
PLCG2
FAPP1

Structure comparisons by way of pairwise alignments between structural homologues suggest that the Kindlin-1 PH domain may have some capacity to bind to polyphosphate inositol lipids; however a wider structure comparison and analysis may provide a clearer notion as to its functional role. Given the diverse nature of PH domain function despite a common fold (DiNitto and Lambright, 2006) we performed a structural phylogenetic analysis (Riffel *et al.,* 2002) to investigate evolutionary relationships between PH domains (Figure 4.12). We used an all-pairs, pairwise structural superimposition of the Kindlin-1 PH domain and a large number of PH domain structures from the PDB in order to provide a structurally-based evolutionary position for the Kindlin-1 PH domain. Among the PH domains, proteins known to bind to inositol phosphate lipid ligands are seen to cluster together. Interestingly, murine T-lymphoma invasion and metastasis-1 (Tiam1) PH domain is positioned close to β-spectrin from the same species, both of which share the same modality of binding to PtdInsP ligands, i.e. non-canonically (Terawaki *et al.,* 2010), and are positioned away from those PH domains that bind to PtdInsPs in a canonical mode, such as Grp1 (Cronin *et al.,* 2004) from the same species. Furthermore, human ArhGAP9 PH domain, also considered β-spectrin-like in its mode of binding PtdInsP (Ceccarelli *et al.,* 2007), branches from the same common scaffold as Tiam1 and β-spectrin. Moreover, those PH domains that predominantly bind directly to Rho family G-protein ligands, such as Intersectin and Dbs (Snyder *et al.,* 2002), also cluster with structural paralogues positioned close to one another, with orthologous structures branching from the same common scaffold. Conversely, Tiam1, the PH domain of which is positioned along with β-spectrin-like PH domains, also binds to the Rho GTPase, Rac1 (Worthylake *et al.,* 2004). However, the PH domain of Tiam1, in this case, does not contact the GTPase directly (Worthylake *et al.,* 2000). It is quite apparent that a large number of PH domains are closely associated, thus dividing the tree in half at its root (see Figure 4.12). The closely-related PH domains in its left hand half are known to bind to

phophatidylinositol phosphate ligands and on closer inspection it seems that the vast majority possess micromolar (μM) dissociation constants (e.g. Akt/PKB, K_D = 1μM (Frech *et al.*, 1997); PEPP1, K_D = 0.32μM (Dowler *et al.*, 2000)), with some PH domains with nanomolar affinity (e.g. BTK, K_D = 0.04μM (Baraldi *et al.*, 1999; FAPP1, K_D = 0.04μM (Dowler *et al.*, 2000)) within this grouping. It was noted that some of those PH domains with nanomolar affinity were clustered, but this is more than likely an artefact primarily due the fact that they are all from *H. sapiens* and therefore represent structural paralogues. Among those PH domains in the, less related, opposing (right hand) half a small number of proteins possess μM dissociation constants, with the majority of those loosely-related PH domains preferring protein ligands. According to this structural analysis, Kindlin-1 is closely related, as expected, to Kindlin-2 and Kindlin-3 PH domain, but is unexpectedly structurally related to oxysterol binding protein related protein 11 (OSBP11), sharing a common ancestral protein module. Superimposition of Kindlin-1 and OSBP (Figure 4.12) demonstrates a good structural agreement, with an r.m.s.d. of 1.7Å. Despite the canonical PtdIns binding PH domains clustering together, Kindlin-1 is somewhat distant even though superimposition of pleckstrin, PKB/Akt and PEPP1 demonstrate very similar structures (Figure 4.12). Kindlin-1 is clearly evolutionarily related to these types of modules but its reluctance to cluster with other PtdIns binders suggests a distinct, but as yet unidentified function. Intriguingly, Kindlins -1, -2 and -3 are not only isolated, but distant from each other. Unfortunately, evolutionary structural neighbours are limited and have as yet undefined functions. Overall, this phylogenetic analysis suggests diversity among protein binding PH modules whereas PtdInsP binding modules are more structurally conserved and that a non-canonical PtdInsP binding modality is a rare adaptation, but requires a conserved structural arrangement.

The electrostatic surface potential of the Kindlin-1 PH domain crystal structure (Figure 4.13) and of the structure with modelled β1-β2 loop compared to a non-canonical

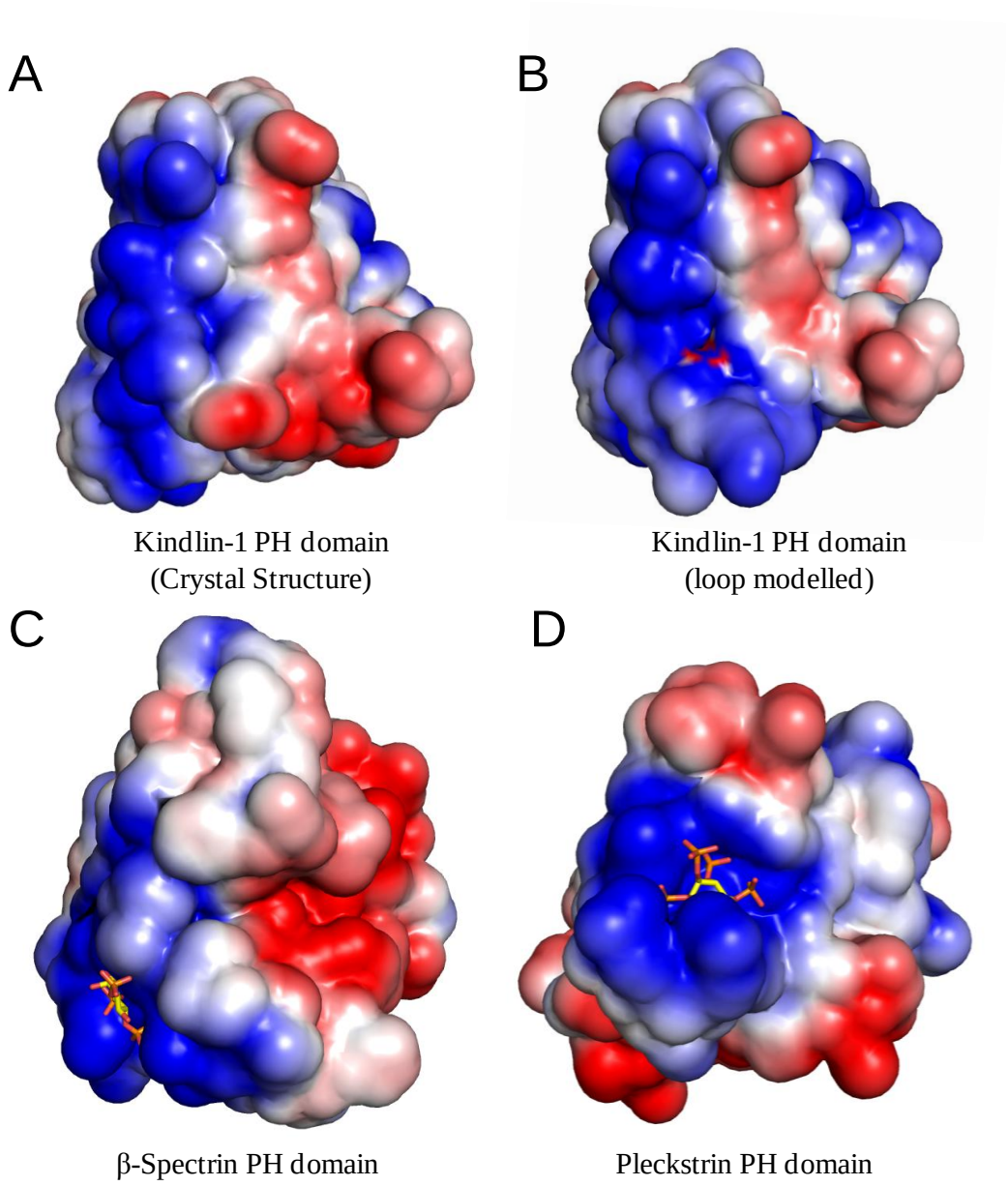


Figure 4.13 Surface representation and electrostatic potential of (A) the Kindlin-1 PH domain crystal structure; (B) Kindlin-1 PH domain complete with modelled disordered loop (see Materials and Methods); (C) β-spectrin crystal structure (PDB code: 1BTN), a non-canonical PtdInsP binding module, bound to D-Myo-inositol 1,4,5-triphsophate rendered in sticks; and (D) Pleckstrin C-terminal PH domain crystal structure (PDB code: 2I5F), a canonical PtdInsP binder, together with D-myo-inositol 1,2,3,5,6-pentakisphosphate bound rendered in sticks. The PH domains, in all cases, are viewed looking down the β-barrel opening and are in identical orientations. Electrostatic calculations were performed using ABPS in PyMol (http://www.pymol.org). The potential is shaded from -2.5*kT/e* (red) to +2.5*kT/e* (blue).

PtdIns binding PH domain (β-spectrin [Figure 4.13]) and a canonical PtdIns binding PH domain (Pleckstrin [Figure 4.13]), reveal a surface charge distribution reminiscent of a non-canonical binding modality, with the positively charged surface away from the β-barrel opening (Figure 4.13). However, despite possessing similar charged surface potentials, the residues that are responsible for co-ordinating PtdIns phosphate ligands at this non-canonical site, as in β-spectrin are not present at equivalent positions in the Kindlin-1 PH structure (Figure 4.14).

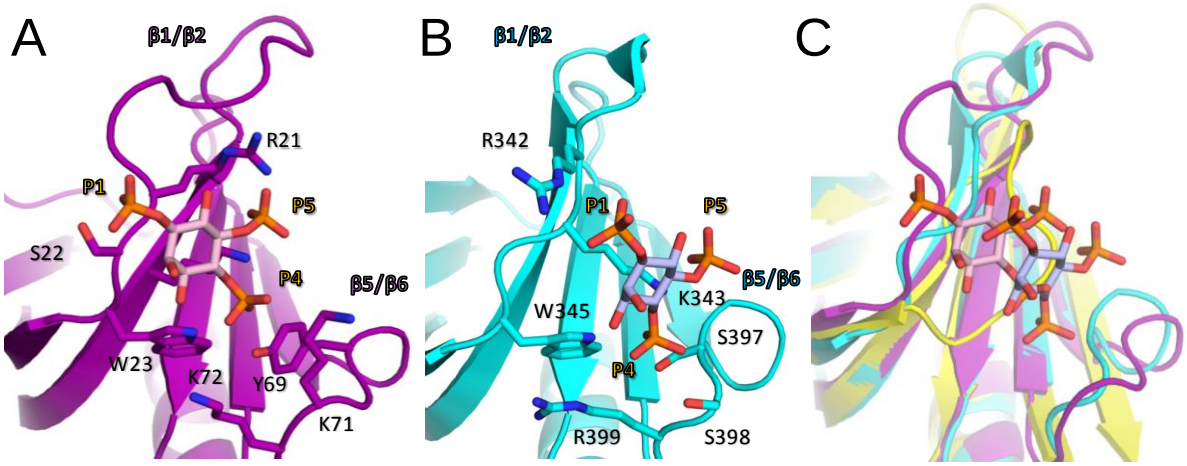


Figure 4.14. Molecular detail of non-canonical PtdIns binding modalities. (A) β-Spectrin PH domain crystal structure (violet) bound to inositol 1,4,5-trisphosphate (IP_3) (PDB: 1BTN) together with side chains of residues (labelled) co-ordinating the phosphates of the ligand. (B) ArhGAP9 PH domain crystal structure (cyan) bound to inositol 1,4,5-trisphosphate (IP_3) (PDB: 2P0D) together with side chains of residues (labelled) that co-ordinate the ligand phosphates. Both β-spectrin and ArhGAP9 bind to PtdInsP ligand in a non-canonical manner (see Figure 4.3), which binds to the outside of the N-terminal half of the β-barrel. Specificity is conferred by residues from the β1-β2 loop and β5-β6 loop (labelled). Each structure was superimposed along with the Kindlin-1 PH domain, with modelled β1-β2 loop using SHP (Stuart *et al.* 1979) (C) and shows a similar binding site for the PtdIns ligand at a non-canonical site. The Kindlin-1 PH domain does not have a β5-β6 loop that large enough to participate in the co-ordination of phosphates nor residues on the β-barrel surface capable of ligation (not shown). The modelled β1-β2 loop is also clearly occluding the binding site. Figures were rendered in PyMol (http://www.pymol.org).

4.2.5. *In vitro* Lipid Binding and Molecular Dynamics (MD) Simulations Reveals That Ligand Binding is Dictated by a Salt-Bridge

Pleckstrin homology domains are well known to function as membrane targeting modules (Lemmon, 2008). However, structural analysis of the Kindlin-1 PH domain suggested a lack of conserved residues that mediate lipid binding. To assess the function of the Kindlin-1 PH domain as a PtdInsP binding module we investigated the function in two capacities, [1] *in vitro* lipid 'dot' blot protein overlay assays using commercially available reagents to identify specificity (if any) for lipid ligands; and [2] molecular dynamics (MD) simulations and molecular docking to determine ligand specificity and a potential binding site and possible binding modality. The MD simulation work was performed by C. N. Lumb and Prof. M. S. P. Sansom (Department of Biochemistry, University of Oxford, UK) as part of a collaboration. It has been reported elsewhere that PtdInsP binding is mediated by side chains in the variable loops, particularly variable loop 1, or the β1-β2 loop (DiNitto and Lambright, 2006), exemplified in (Jackson *et al.*, 2007). Therefore, prior to molecular dynamics simulations the missing β1-β2 loop (Pro-Lys-Lys-Leu-Met-Leu-Lys) in the crystal structure was modeled using the program *MODELLER* (Sali and Blundell, 1993) (see Materials and Methods). After modeling these residues we conducted MD simulations of the Kindlin-1 PH domain. The r.m.s.d. and root-mean-square fluctuation (r.m.s.f.) of the Cα atoms over the course of the simulation trajectories are shown in Figure 4.15. The Cα r.m.s.d. indicated that the protein experienced little conformational drift over a period of 100 ns. The Cα r.m.s.f. suggested that the three loops crowning the open end of the β-barrel (VL1, VL2 and VL3) were flexible and, in agreement with the purported flexibility based on weak electron density in the crystal structure, the β1/β2 loop (VL1) appeared to be the most flexible (Figure 4.15).

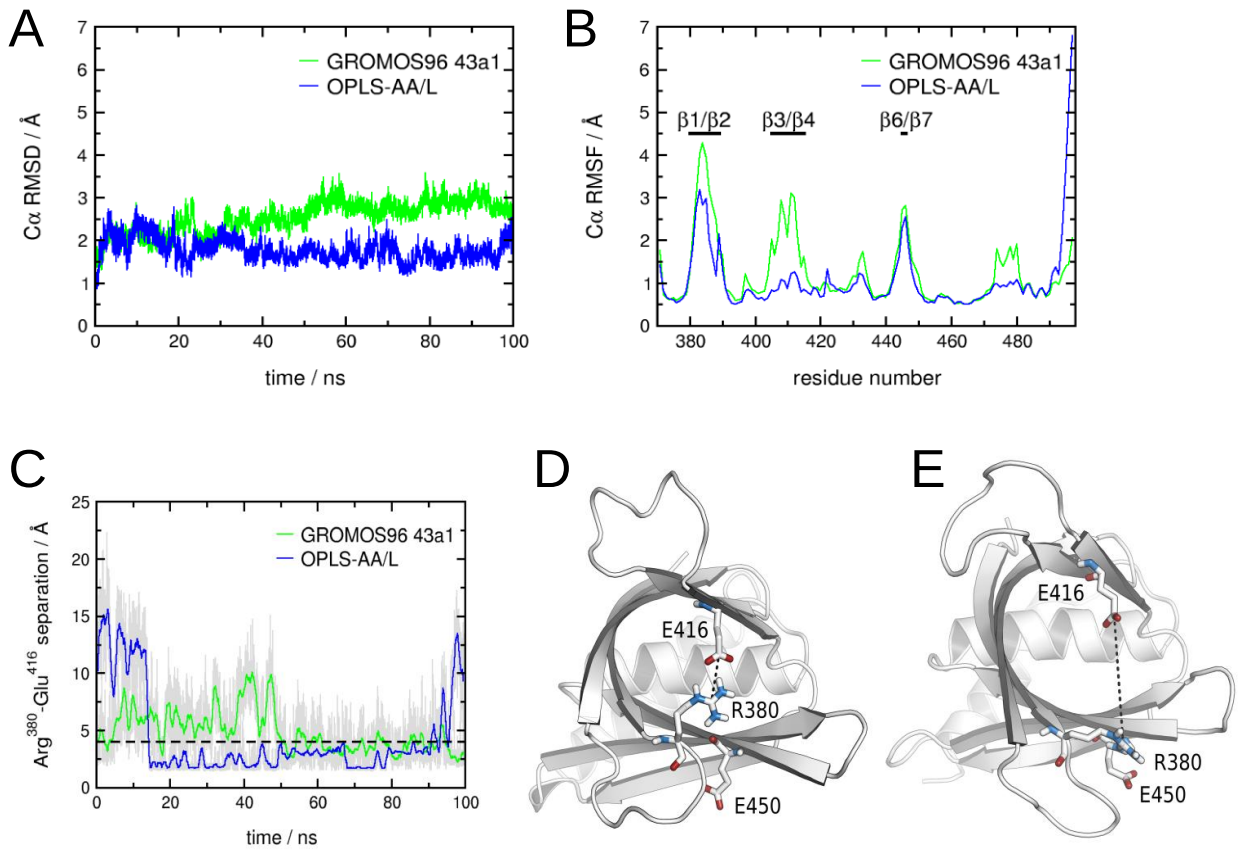


Figure 4.15 – Conformational dynamics of the Kindlin-1 PH domain during the MD simulations. (A) Cα r.m.s.d. calculated as a function of time over the MD simulations using two forcefield protocols (GROMOS96 43a1 forcefield, green; and the OPLS-AA/L forcefield, blue). (B) r.m.s.f. of the protein Cα atoms calculated for each residue over the course of the MD simulations. The so-called variable loops are considered the most flexible. The β1-β2 loop exhibits the greatest conformational fluctuations over the time-course suggesting it to be the most flexible. (C) Separation between the charged groups of Arg^{380} and Glu^{416}, which together form a salt bridge in the crystal structure that appears to partially occlude the putative lipid binding cavity. Running averages taken every 100 ps show the transient breaking of the salt bridge over the course using two forcefield protocols (GROMOS96 43a1 simulation, green; and the OPLS-AA/L simulation, blue). The dashed line at 4 Å indicates the separation at which the salt bridge is considered to be formed/broken, following the published definition (Kumar *et al.* 2002). (D). Simulation snapshot of the occluded conformation used for docking, with the Arg^{380}-Glu^{416} salt bridge intact. (E) Simulation snapshot of the open conformation used for docking, with the Arg^{380}-Glu^{416} salt bridge now broken and replaced by an alternative Arg^{380}-Glu^{450} salt bridge, leaving the binding pocket exposed. Figure contents kindly provided by C. N. Lumb (Department of Biochemistry, University of Oxford).

In the crystal structure the open end of the β-barrel that is crowned by variable loops, a known PtdInsP binding site in other PH domains (DiNitto and Lambright, 2006), is traversed by a salt-bridge between Arg^{380} and Glu^{416}, potentially blocking any PtdInsP binding. Interestingly, although the Arg^{380}- Glu^{416} salt-bridge was typically preserved during the simulations, we also observed a transient switching to an alternative conformation. This involved the formation of a different salt bridge, Arg^{380}- Glu^{450}, which left the binding pocket exposed (Figure 4.15). In the canonical PH domain the ligand would be expected to bind in the central binding pocket that is the open end of the β-barrel. Examples have been reported where PH domains bind to PtdInsP in a non-canonical modality, with the binding site away from the β-barrel opening (Ceccarelli *et al.*, 2007, Hyvonen *et al.*, 1995, Terawaki *et al.*, 2010). Therefore in an effort to identify any putative lipid binding sites, we performed molecular docking of the headgroups of six different PtdIns species to the Kindlin-1 PH domain using AutoDock. A representative conformation of the Kindlin-1 PH domain was extracted from the MD simulations in which the Arg^{380}-Glu^{416} salt bridge remained intact. However, when molecular docking was performed using this partially occluded state the results were generally poor. Clustering the ligand binding conformations resulted in a large array of small, spatially separated clusters and little consensus emerged as to the location of any possible binding site (Figure 4.16). Thus, in the absence of a predicted binding site, we turned to the alternative open conformation we had observed in our MD simulations. Molecular docking using this open conformation extracted from the simulations led to improved results. Clustering of these results produced a smaller number of larger clusters, and most of the docking runs placed the ligand in or near the central cavity (Figure 4.16). Docking using the open conformation rather than the occluded conformation approximately halved the number of conformational clusters and multi-member clusters. This higher level of

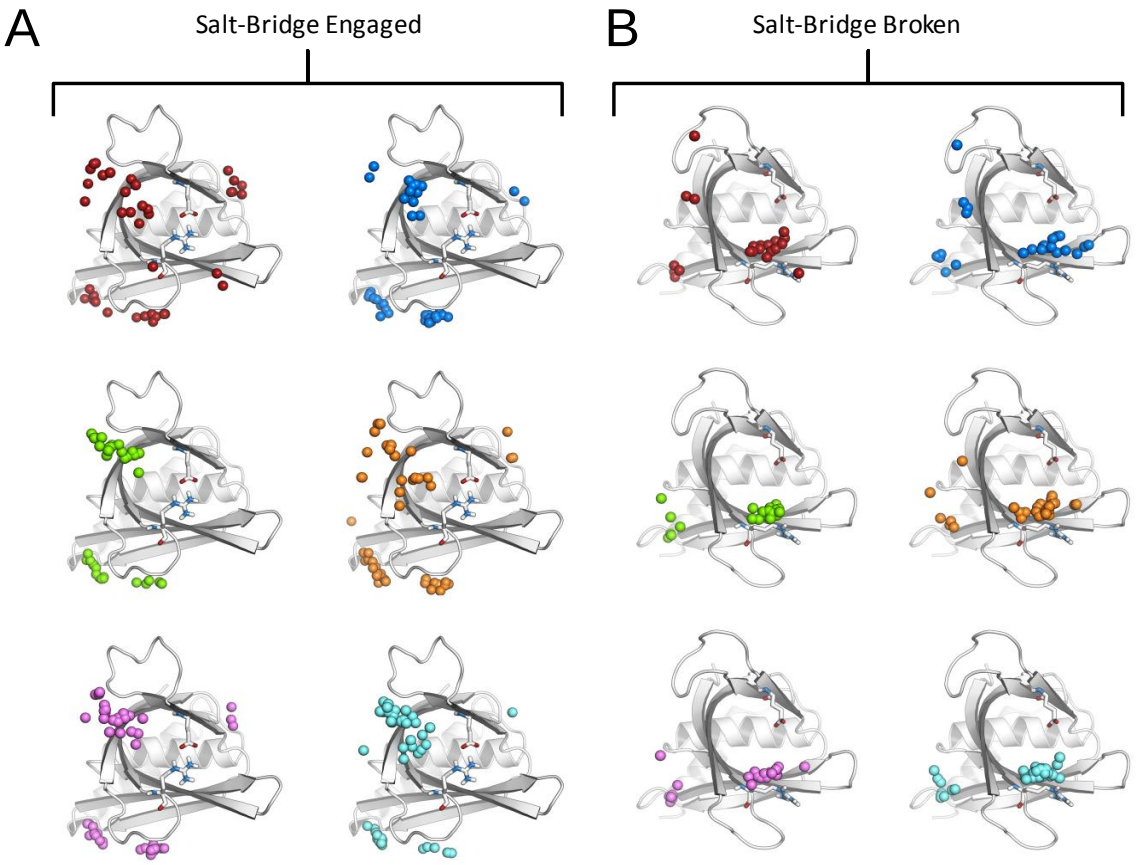


Figure 4.16 – Molecular Dynamic (MD) simulations and Molecular Docking of the Kindlin-1 PH domain. Molecular docking centres of mass of the ligand clusters when docking to the occluded conformation, based on a cluster tolerance of 1.5 Å r.m.s.d. from 100 docking runs. Working from left to right: Ins(1,3)P_2 (red); Ins(1,4)P_2 (blue); Ins(1,3,4)P_3 (green); Ins(1,3,5)P_3 (orange); Ins(1,4,5)P_3 (purple) and Ins(1,3,4,5)P_4 (cyan). The centres of mass of the ligand clusters when docking to the open conformation are shown in parallel. Ligands are represented in the same sequence as for the occluded docking experiments. Figure contents kindly provided by C. N. Lumb (Department of Biochemistry, University of Oxford).

consensus resulted in a reasonably consistent prediction of a putative binding site for all six PtdIns headgroups, lined by Arg^{380}, Lys^{382}, Lys^{390} and Asn^{449} (Figure 4.16). Performing similar docking runs using another docking program (SwissDock) yielded qualitatively similar results. This effect was even more pronounced when performing docking runs with the Glu^{416}>Ala mutant in the open state. It therefore appears that PtdIns binding is enhanced in the Glu^{416}>Ala mutant for two reasons. First, the breaking of the Arg^{380}-Glu^{416} salt bridge, which otherwise would block the entrance to the binding pocket. Second, the replacement of

the sentry glutamate residue, Glu^{416}, by alanine in this open state, which appears to further increase the accessibility of the binding site for negatively charged lipid headgroups. This stepwise enhancement of PtdIns binding appears to be confirmed by the estimated free energies of binding predicted by AutoDock, which suggest that docking to the Glu^{416}>Ala mutant (open) PH domain is more favourable than docking to the wild type (open), which in turn is an improvement over the wild type (closed) PH domain. In general, this trend is observed for all six PtdIns species, and the more favourable binding energies are accompanied by an improved level of clustering (Figure 4.17).

To complement the MD simulation and molecular docking experiments we performed lipid dot blots (Echelon Biosciences, Salt Lake City, Utah, US) using purified Kindlin-1 PH domain (>95% purity assessed by SDS-PAGE) to evaluate the lipid binding specificity, if any, of this PH domain. The phosphatidylinostol phosphate (PtdInsP) lipid overlay assay revealed that the Kindlin-1 PH domain demonstrated preferential binding to inositol monophosphate PtdIns species [PtdIns(3)P, PtdIns(4)P, PtdIns(5)P], irrespective of phosphate position on the inositol ring (Figure 4.18). Weak signal corresponding to phosphatidylserine binding was also observed. Given the striking difference in docking results between the formed/broken salt bridge conformers, site directed mutagenesis of the isoform-specific invariant glutamate was used to permanently disrupt the salt bridge, with the prediction that this disruption would lead to an improved ligand binding module and extend the species of PtdIns lipids that it recognizes. We considered the removal of the glutamate (Glu^{416}) as this was invariant in Kindlin-1 isoforms only, and the countering arginine (Arg^{380}) was conserved in other PH domains from other Kindlin isoforms. Furthermore, the alternative salt-bridge conformation (Arg^{380}-Glu^{450}) seen during MD simulations may be promoted by removing Glu^{416}. Therefore recombinant Glu^{416}> Ala mutant PH domain was generated by

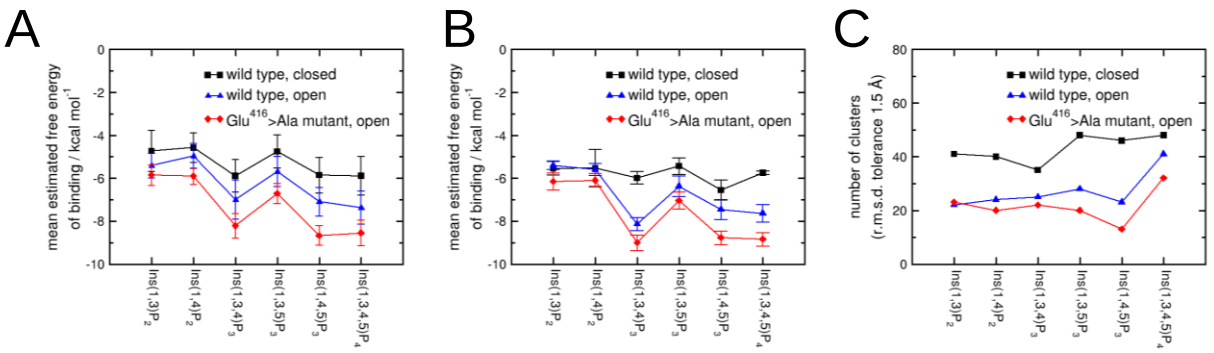
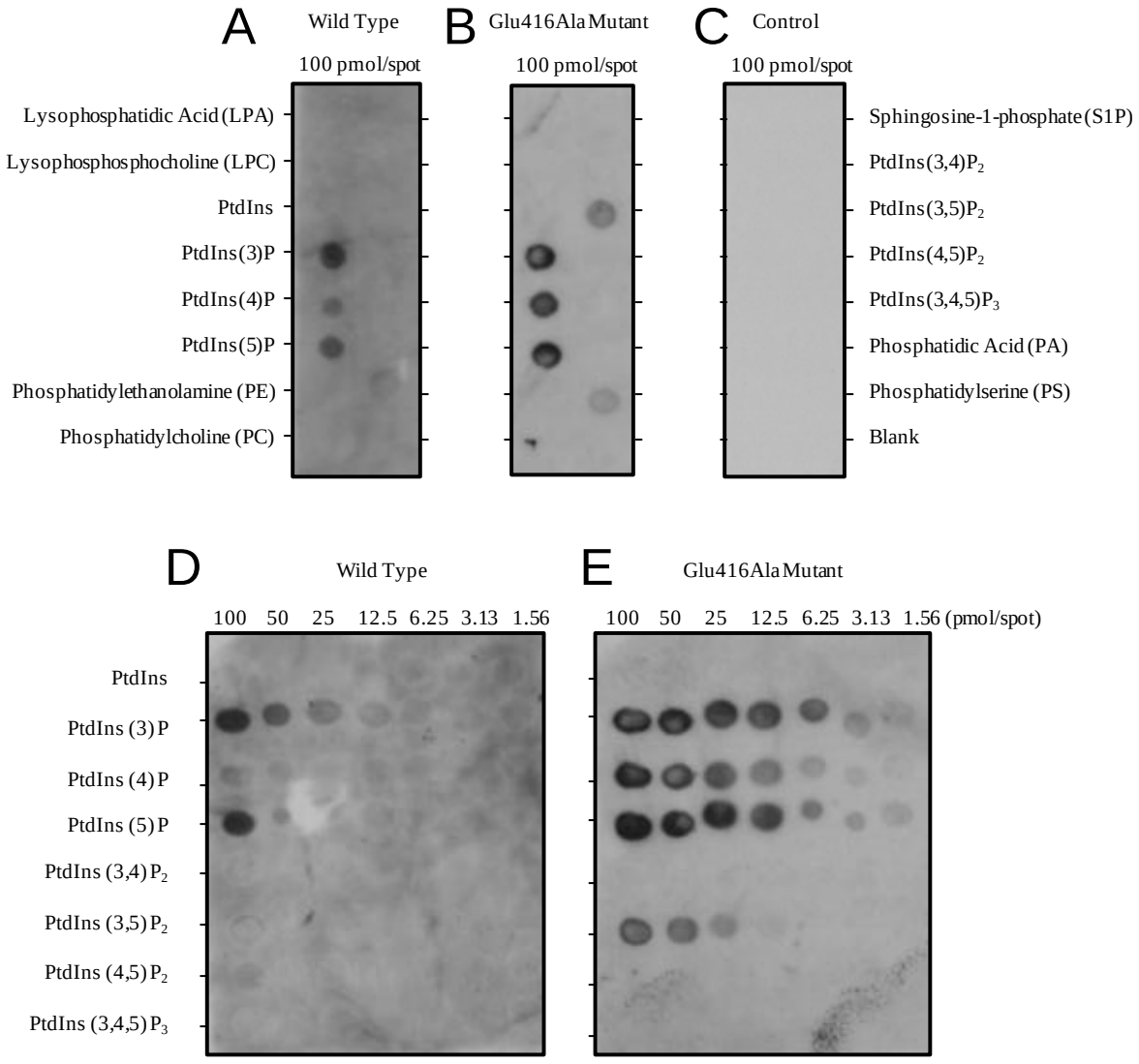


Figure 4.17 – Estimated free energies of binding and number of clusters using AutoDock. (A) Mean estimated free energy of binding for all 100 docking runs using AutoDock for each of the six PtdIns species. Results are shown for docking to the wild type (closed) PH domain (black squares), the wild type (open) PH domain (blue triangles) and the Glu^{416}>Ala mutant (open) PH domain (red diamonds). Error bars are plus or minus one standard deviation from the mean. Lines are guides for the eye. A more negative value is indicative of more favourable binding. (B) Mean estimated free energy of binding for the members of the largest cluster after docking based on a cluster tolerance of 1.5 Å r.m.s.d. Similar trends to those in panel (A) are observed. (C) Number of clusters after 100 docking runs for each case. A lower number of clusters suggest a greater degree of consensus as to the location of the binding site. Figure contents kindly provided by C. N. Lumb (Department of Biochemistry, University of Oxford).

site directed mutagenesis and purified analogously to wild type. The recombinant mutant protein was used in lipid overlay assays, alongside wild type, which revealed a conservation of preferential binding to monophosphate species and phosphatidylserine, analogous to wild type (Figure 4.18). However, the salt-bridge disrupted mutant also displayed additional lipid recognition of PtdIns(3,5) bisphosphate. To test this observation further we incubated purified Kindlin-1 PH domain and the salt-bridge disrupted mutant with membranes with different quantities of PtdInsPs. Again, the Kindlin-1 PH domain interacted with mono-phosphorylated PtdInsP species only and the Glu^{416}>Ala mutant PH domain bound to mono-phosphorylated lipids, as in the 'wild-type' and bound to PtdIns-3,5P_2. Despite equal concentrations of protein used in the PtdInsP arrays, which was performed in parallel, it was obvious that the 'wild-type' Kindlin-1 PH domain bound moderately to the lipids [with signal

Figure 4.18 (Opposite) – Protein Lipid Overlay Assay. (A) Phosphatidylinositol phosphate (PtdInsP) 'Dot-blot' (Echelon Bioscience Inc. Salt Lake City, US) assay for recombinant purified Kindlin-1 PH domain. Kindlin-1 PH domain demonstrates a preference for monophosphorylated inositol moieties over di- and tri-phosphate species. Weak signal can be observed for PtdIns 4 phosphate and very weak signal can be observed corresponding to phosphatidylserine. (B) PtdIns 'dot-blot' assay using a salt-bridge disruption mutant (Glu^{416}>Ala). The mutant Kindlin-1 PH domain demonstrates a conserved mono-phosphorylated inositol preference and has gained an additional PtdIns (3,5) bisphosphate binding capability compared to the wild type. (C) PtdIns 'dot-blot' assay control, using no recombinant protein. (D) Phosphatidylinositol phosphate (PtdIns) serial dilution array using recombinant WT Kindlin-1 PH domain. Kindlin-1 PH domain demonstrates mono-phosphorylated inositol binding, however it has the greatest affinity for 3-phosphate species, with weak signal detected at 6.25pmol compared to weak signal observed corresponding to PtdIns 4-phosphate at 100pmol and signal observed corresponding to PtdIns 5-phosphate at 50pmol. (E) PtdIns serial dilution array using the salt bridge-disruption mutant. Kindlin-1 PH domain mutant demonstrates greater affinity for all mono-phosphorylated PtdInsP species, with equivalent signal observed corresponding to PtdIns 3-phsophate and 5-phsophate, both at 1.56pmol, and slightly weaker signal observed for equivalent concentration for PtdIns 4-phosphate. In all cases the Kindlin-1 PH domain binding was detected by probing the dot-blot membranes with a mouse anti-penta-His antibody, as the recombinant material still possessed the N-terminal His_6 -tag.


A
Wild Type
100 pmol/spot
B
Glu416Ala Mutant
100 pmol/spot
C
Control
100 pmol/spot
Lysophosphatidic Acid (LPA)
Lysophosphosphocholine (LPC)
PtdIns
PtdIns(3)P
PtdIns(4)P
PtdIns(5)P
Phosphatidylethanolamine (PE)
Phosphatidylcholine (PC)
Sphingosine-1-phosphate (S1P)
PtdIns(3,4)P2
PtdIns(3,5)P2
PtdIns(4,5)P2
PtdIns(3,4,5)P3
Phosphatidic Acid (PA)
Phosphatidylserine (PS)
Blank
D
Wild Type
E
Glu416Ala Mutant
100 50 25 12.5 6.25 3.13 1.56
100 50 25 12.5 6.25 3.13 1.56 (pmol/spot)
PtdIns
PtdIns (3) P
PtdIns (4) P
PtdIns (5) P
PtdIns (3,4) P2
PtdIns (3,5) P2
PtdIns (4,5) P2
PtdIns (3,4,5) P3

observed at 6.25pmol quantities for PtdIns(3)P, PtdIns(4)P and PtdIns(5)P] whereas the Glu^{416}>Ala mutant bound strongly to the same lipids [signal observed at 1.56pmols for PtdIns(3)P, PtdIns(4)P and PtdIns(5)P] (Figure 4.18). Moreover, the salt bridge-disrupted mutant exhibited moderate binding to phosphatidylinositol 3, 5-bisphosphate [signal observed at 12.5pmol quantities (Figure 4.18)], whereas the 'wild-type' PH domain demonstrated undetectable binding. This suggests that the liberated arginine from the salt bridge may co-ordinate the additional phosphate in PtdIns (3,5) bisphosphate. Interestingly, the arginine residue involved in the occluding salt-bridge is equivalent to a histidine residue (His^{256}) in the Pleckstrin C-terminal PH domain crystal structure (PDB code: 2I5F) that forms a salt-bridge with the phosphate at the 3 position on the inositol ring of D-*myo*-inositol 1,2,3,5,6-pentakisphosphate (IP_5) (Jackson *et al.*, 2007).

4.2.6. The PH Domain is Important for Integrin Activation and Focal Adhesion Assembly

To investigate the importance of the inserted PH domain in Kindlin-1-mediated integrin activation we performed integrin activation assays in mammalian cells. This work was carried out, as part of a collaboration, by N. N. Brahme and Dr. D. A. Calderwood (Department of Pharmacology, Yale University, USA). We generated a GFP-Kindlin-1 expression construct that lacked the PH domain (Kindlin-1ΔPH) and tested it in a well validated flow cytometric assay for $\alpha_{IIb}\beta_3$ integrin activation. As previously reported (Harburger *et al.*, 2009, Goult *et al.*, 2009b, Ussar *et al.*, 2008), when co-expressed with DsRed-tagged talin head, GFP-Kindlin-1 enhanced activation of $\alpha_{IIb}\beta_3$ stably expressed in CHO cells (Figure 4.19). However, the ability of Kindlin-1ΔPH to activate $\alpha_{IIb}\beta_3$ was

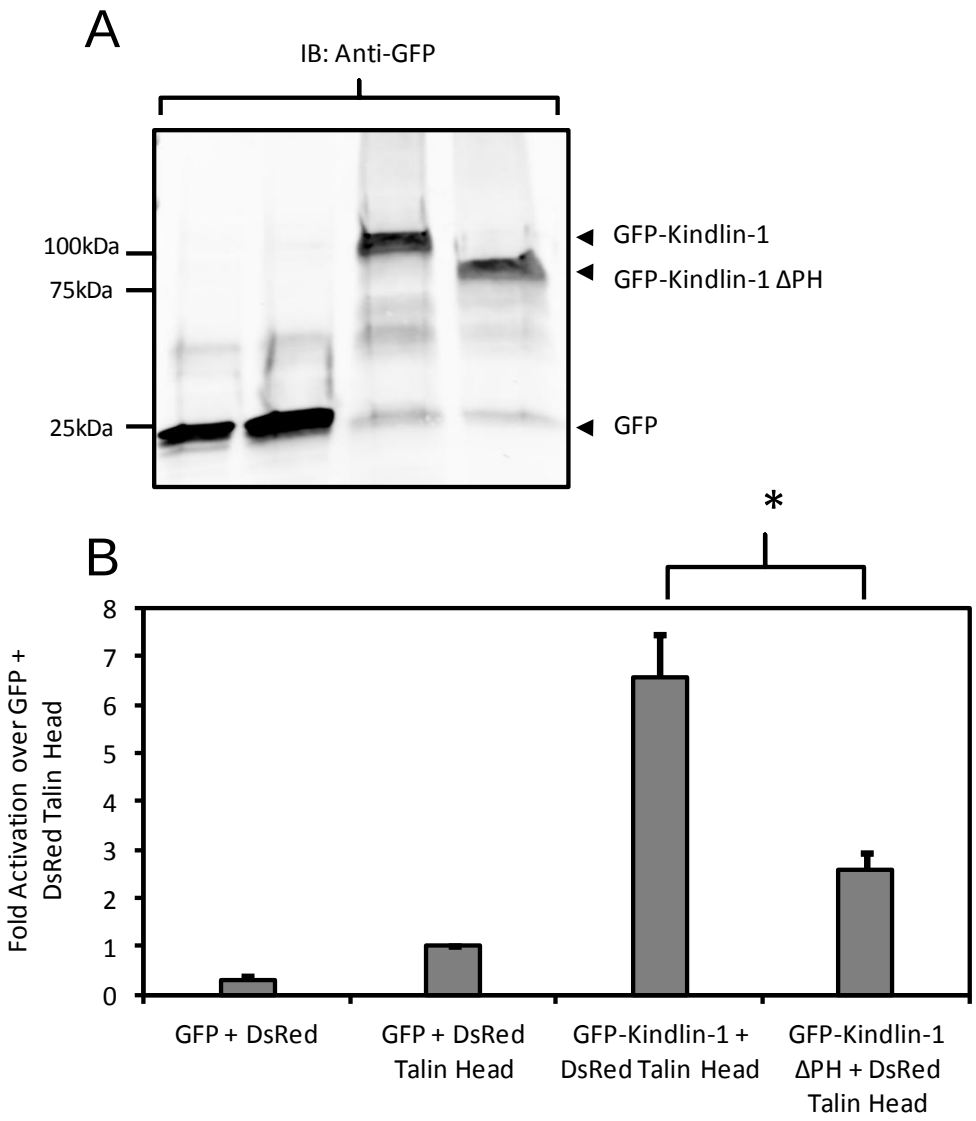


Figure 4.19 – The Kindlin Pleckstrin Homology (PH) domain is important for Kindlin-1-mediated activation of αIIβ3 integrins. DsRed-Talin Head (1-433) and GFP-Kindlin-1 or GFP-Kindlin-1 ΔPH were transiently co-expressed in CHO cells stably expressing αIIbβ3 integrins. Expression of the Kindlin constructs was assessed by immunoblotting whole cell lysate with anti-GFP antibodies (A). Activation indices (normalized for total surface integrin levels) were calculated for cells with similar DsRed and GFP fluorescence intensities and expressed relative to cells expressing GFP + DsRed-Talin Head. Results represent the means ± standard error (n≥3). All conditions are significantly different (p-value < 0.01) from each other. The GFP-Kindlin-1 ΔPH construct resulted in an impaired integrin activation compared to talin- and Kindlin-mediated full-activation (highlighted). Figure contents kindly provided by D. A. Calderwood (Department of Pharmacology, Yale University, U.S).

significantly impaired compared to wild-type Kindlin-1. Kindlin-1ΔPH did nonetheless induce a small, but statistically significant, increase in integrin activation compared to talin head alone (Figure 4.19). In all cases cells were gated to have comparable GFP fluorescence levels and fixed DsRed fluorescence ensuring that the impaired ability of Kindlin-1ΔPH to activate $\alpha_{IIb}\beta_3$ was not related to altered Kindlin or talin head expression levels. Furthermore, integrin activation was normalized to total $\alpha_{IIb}\beta_3$ integrin expression assessed in parallel by antibody staining. Western blotting confirmed that the GFP-Kindlin-1 and GFP-Kindlin-1ΔPH were of the expected size (Fig 5.19). These data show that the Kindlin-1 PH domain is important for Kindlin-1-mediated integrin activation consistent with published data on the Kindlin-2 PH domain (Ma *et al.*, 2008).

4.3. Discussion

The pleckstrin homology domain is the 11th most common module in the human genome (Lemmon *et al.*, 2002), and despite a common scaffold is diverse in functionality. In the work reported here we have shown that the Kindlin PH domains have a C-terminal, additional helix not present in other reported PH module structures. We have performed a phylogenetic analysis based on structural superposition which has defined two distinct clades within the PH domain superfold family. We have made use of molecular dynamics simulations to search for ligand binding pockets on the surface of the protein and we have shown how specific evolutionary modifications to the canonical PH scaffold have resulted in the conferral of limited affinity and restricted specificity of binding to inositol phosphate ligands.

Kindlin-1 PH domain is dimeric in solution: association states are seen in SEC and sedimentation velocity and sedimentation equilibrium AUC show a clear dimeric weight for a single sedimentation coefficient distribution peak, and a weight-averaged dimeric weight overall (Figure 4.6). Protein from peak C in Figure 4.5 (dimer) was used to generate crystals and their X-ray structure was solved by molecular replacement (see Materials and Methods). The crystal structure contains one molecule per asymmetric unit but a dimer does exist across a twofold crystallographic axis with a buried interface of some 700 $Å^2$. However, the PISA web server (http://www.ebi.ac.uk/msd-srv/prot_int/cgi-bin/piserver) considers this to be a crystallisation artefact and this suggests that the dimerisation interface observed in solution is not present in the crystal, therefore an *in vitro* dimer but a monomer *in crystallo*. There are three possible interpretations for the dimerisation of the PH domain we observe. Firstly, this may be a native conformation of the full-length Kindlin-1 protein and indicate that it functions as a dimer, at least prior to integrin activation. Secondly, Kindlin-1 may be a monomer normally but be capable of dimerisation through its PH domain on unmasking of an interaction surface – perhaps due to integrin binding or phosphorylation. Thirdly, the dimerisation of the PH domain we observe may be artefactually mimicking an interaction between the PH domain and F3 domain, both of which are PH superfamily folds. In any case dimerisation is easily an adaptational difference found between homologous proteins in the same species (e.g. the B cell protein B7-1 is a dimer, its homologue B7-2 is a monomer) (Collins, 2002) and whatever the association state of Kindlin-1 this cannot be taken to indicate the association state of the other Kindlins in any way.

The role of the C-terminal helix extension, an integral portion to the Kindlin PH fold, must be considered within the context of the complete Kindlin structure, where the PH domain intersects the FERM F2 subdomain into an N-terminal and a C-terminal portion (Goult *et al.*, 2009b). The most highly conserved region of the structure is found in the α1-α2

loop by which the C-terminal α2 helix is routed off the canonical α1 PH helix prior to it in the sequence (Figure 4.10). Localised sequence conservation is also found within the α2 helix itself, in particular at its interface with α1. The α1-α2 loop conservation (especially the sequence Met-Ala-Asp-Ser-Ser-Tyr) is greater even than that within the core of the domain fold (Figure 4.10).

The phylogenetic tree provides further insight into the evolutionary conservation of features with respect to function within the PH family. In this tree PH domains divide evolutionarily into two groups, which can be rationalised by one clade showing a preference for phosphorylated inositol ligands, and the other for protein ligands (Lemmon, 2008; DiNitto and Lambright, 2006). A divide thus apparently occurred early on in the development of PH domains that resulted in alternative adaptations. While protein-protein interaction adaptation can be expected to involve variation over a substantial fraction of the protein surface, variation in the phosphatidylinositol-binding capability of the proteins can be the result of very limited substitutions. Thus, the protein-interacting PH domains are widely spaced on their phylogenetic branches whereas the PtdIns-binding protein branches are tightly clustered, unsurprising given the adaptation of different regions of the PH domain surface to bind different ligands. Indeed, as shown here a single amino acid change converts the monophosphate-specific Kindlin-1 PH domain into a domain with stereospecific diphosphate binding (Figure 5.18). Also in agreement with this, it seems that acquisition of PtdInsP binding (in these cases of a canonical variety) is relatively trivial as isolated cases occur among the protein-binding clade. Non-canonical PtdInsP binding seems in any case to be a relatively uncommon adaptation. The Kindlin-1 PH domain sits in an isolated cluster of the PtdIns-binding PH domains (Figure 4.12), along with the Kindlin -2 and -3 PH domains and that from oxysterol binding protein (the latter two are unpublished structures from the RCSB PDB and there has not therefore been a structure-function analysis performed for them of the

sort reported here). Specific adaptations present in distinct members of the Kindlin superfamily perhaps allow them to perform different biological functions in line with their specificity in tissue expression (Moser *et al.,* 2009; Ussar *et al.,* 2006). The structure and position of this phylogenetic branch suggests that Kindlin PH domains may even be representative of a grouping distinct from both the protein-binding and PtdInsP binding ones.

Ligand binding assessment by MD suggested the need for an 'occluded → open' transition to occur if PtdIns lipids are to bind to Kindlin-1 PH (Figures, 4.15 - 4.17). The MD simulations showed that, in the wild type protein, the binding pocket is predominantly occluded owing to the presence of the salt bridge (Figure 4.15). Breaking this salt bridge using site-directed mutagenesis was therefore predicted to increase the range of PtdIns ligands that may be accommadated by the Kindlin-1 PH domain. This was confirmed by lipid binding analysis showing generalised monophosphate specificity although testing against different concentrations of inositol phosphate showed a preference for Ins5P > Ins3P > Ins4P, but all with relatively weak binding (Figure 4.18). Subsequent experiments in which the modification of the salt bridge gave additional ligand recognition properties and stronger binding were in excellent accord with the predictions from the simulations. Very recently the NMR structure of Kindlin-2 PH domain was solved (Liu *et al.,* 2011). The specificity of the PtdIns binding pocket found in Kindlin-1 for monophosphates is strikingly at variance with data from Kindlin-2 binding PtdIns3,4,5P_3 (Qu *et al.,* 2011, Liu *et al.,* 2011). A direct comparison of the Kindlin-2 structure with our structure highlights why this is the case: the Arg^{380}-Glu^{146} salt bridge makes the InsP binding pocket much more constricted, drawing its β-sheet wall in by ~6Å – collapsing the β-barrel. Comparisons with the all three Kindlin PH domains by phylogenetic analysis, shows that that Kindlin-3 PH is more similar to the Kindlin-1 PH domain reported here than to the Kindlin-2 PH domain (Figure 4.12). Furthermore, several conformations of side chains in Kindlin-1 further occlude the open end

of the β-barrel thus preventing the recognition of 3-phosphorylated lipid species, a reported function across the Kindlin family (Liu *et al.,* 2011). Interestingly, this is consistent with suggestion made earlier in this chapter, where it was noted that the arginine residue involved in the occluding salt-bridge in Kindlin-1 is equivalent to a histidine residue in Pleckstrin that interacts with the 3-phosphate of its cognate ligand. We therefore suggest that, unlike Kindlin-2, the Kindlin-1 PH domain is unlikely to recognise Phosphoinositide 3-Kinase (PI3K) products, due to steric clashes.

In the end, we judge that the apparent inositol phosphate binding interaction of the Kindlin-1 PH domain is an artefact and that the true ligand is as yet unidentified. PtdIns monophosphates are found on the *trans* side of the Golgi (PtdIns4P) and early endosome and multivesicular body (PtdIns3P) (Kutateladze, 2010) and not at the plasma membrane where Kindlins act. However, it is known that modified lipid headgroups can be good surrogates for phosphotyrosine moieties (Lingwood *et al.,* 2005). What is more, focal adhesions are generally stained and visualized with anti-phosphotyrosine antibodies therefore demonstrating a localised concentration at adhesion assemblies (Patla *et al.,* 2010). These facts together suggest the possibility this protein module in Kindlin-1 actually recognises phosphotyrosines and not PtdIns monophosphates since it would be unlikely to encounter them *in vivo* but would encounter phosphotyrosine moieties.

Overall, our work has provided an atomic description of the Kindlin-1 PH domain structure and its dynamics relating to the protein's function. Our data has suggested that lipid binding found in other Kindlins is not conserved in Kindlin-1 and this is due to an isoform specific salt bridge that prevents lipid recognition.

CHAPTER 5

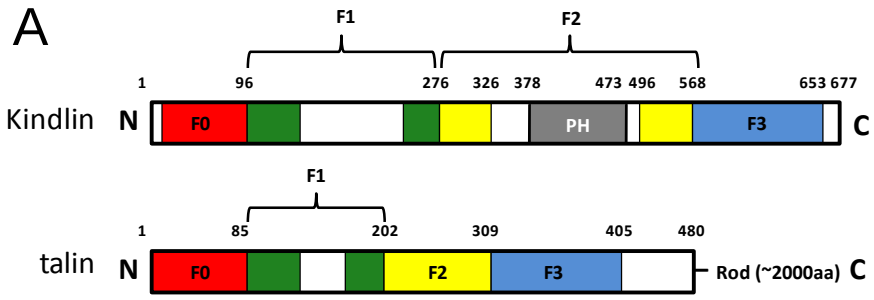
5.1. The Conformation of Kindlin-3 and its Interaction with β-Integrin

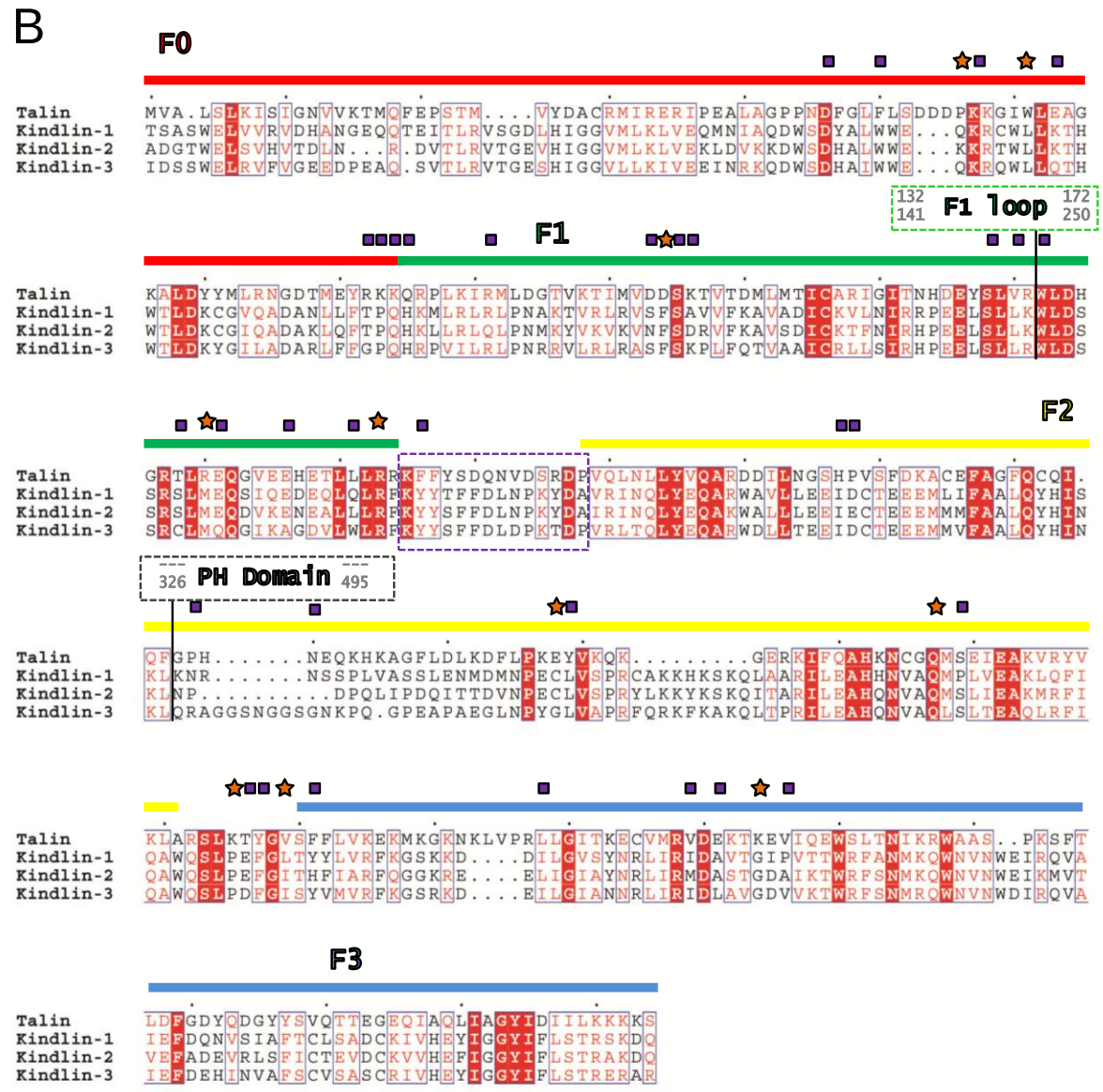
5.1.1. Introduction

Kindlin-3, a 75kDa protein, has been shown to be essential for integrin activation and is therefore critical for haemostasis, immunity and bone metabolism. The Kindlin family of proteins has been, and will continue to be a focus of attention due to their critical role in integrin activation and focal adhesion assembly. An in depth review of the biology of the Kindlin family is provided elsewhere (see Chapter 3 Introduction). The domain structure of the Kindlins and their homology with talin, however, is illustrated in Figure 5.1. Whilst initial work focussed upon the Kindlin family, this study focuses on a single member - the erythropoietically important isoform, Kindlin-3. Kindlin-3 is the smallest member (a 75kDa protein) that has been shown to be essential for integrin activation in clotting, immunity and bone metabolism (Schmidt *et al.,* 2011; Moser *et al.,* 2009; Malanin *et al.,* 2009). Kindlin-3's role in these processes has resulted in extensive cellular and physiological studies. In spite of the surge in these types of studies, there is still an overwhelming lack of structural and biophysical data, which we aim to address.

This chapter and the work therein is part of a collaborative project with Prof. I. D. Campbell and Dr. Anna Fuezery (Department of Biochemistry, University of Oxford, UK) and therefore contributions by collaborators are acknowledged in the text where appropriate. Experimental results and figure material from collaborators is reproduced with permission.

Figure 5.1 (Opposite) – (A) Schematic representation of domains of the Kindlins (Kindlin-1 numbering) compared to talin-1. (B) Sequence alignment of the talin-head (residues 1-400) from the crystal structure (PDB: 3IVF) compared to the FERM domains of mouse Kindlin-1, -2 and -3. The sequence corresponding to the PH domain and the F1 loop regions of the Kindlin sequences were removed prior to sequence alignment. The F1 loop in talin is missing in the crystal structure and therefore is not present in the sequence. The talin sequence was aligned with the Kindlin-1 FERM domain, as in (Goult *et al.* 2010), using ClustalW. The remaining family members were aligned to Kindlin-1 using ClustalW and the comparison between talin and the Kindlin family was adjusted by hand. Espript was used for the final alignment. The domain boundaries are denoted by coloured bars (as in A) above the sequence and the F1 loop and PH position highlighted. Interfacial residues that are buried in the talin structure (PDB: 3IVF) are highlighted (purple squares) together with those that mediate an inter-domain interaction (orange stars).


A
F1
F2
1
96
276
326
378
473
496
568
653
677
Kindlin
N
F0
PH
F3
C
F1
1
85
202
309
405
480
talin
N
F0
F2
F3
Rod (~2000aa)
C


B
F0
Talin
Kindlin-1
Kindlin-2
Kindlin-3
132
141
F1 loop
172
250
F1
F2
326
PH Domain
495
F3

5.1.2. Summary

We sought to biophysically characterise the Kindlins and previously demonstrated that expression of an array of constructs of this family was unmanageable in *E. coli* but subsequently showed good levels of heterologous expression in baculovirus infected *S. frugiperda* cells (see Chapter 3). Here, we describe the first successful optimized expression of recombinant Kindlin-3, using baculovirus-infected Sf9 cells, yielding sufficient quantities to perform the first biophysical characterization of the full-length protein. We therefore purified recombinant Kindlin-3 and investigated the three-dimensional arrangement of the domains within Kindlin-3 at low resolution using synchrotron radiation small-angle X-ray scattering in solution. *Ab initio* envelopes from SAXS together with other shape estimates in solution using additional techniques revealed that Kindlin-3 is elongated and conformationally similar to talin but with the prominent addition of the PH domain. We further characterized interactions between Kindlin and integrins, talin and lipids. We show no detectable direct interaction with talin in solution, with or without a β-integrin cytoplasmic tail, but Kindlin binding of tails was detectable, both by analytical ultracentrifugation and NMR. NMR indicated that Kindlin-3 specifically recognizes the tail's membrane-distal NPxY motif and that an upstream Serine/Threonine cluster is critical for integrin binding. These data support current biological and mutational data on Kindlin-3:β-tail binding and provide novel insights that include the demonstration of an extended conformation for the various Kindlin-3 sub-domains and the provision of direct evidence for sequence-specific determinants of Kindlin-3-integrin tail binding.

5.1.3. Expression Optimisation of Recombinant Kindlin-3

Owing to the versatility of the pOPIN vectors used in the initial HTP study (Berrow *et al.,* 2007), we previously demonstrated that mouse Kindlin-3 could be expressed successfully, along with additional fragments and other Kindlin isoform constructs, using a baculovirus-infected insect cell expression host, but could not be produced using transient mammalian expression systems (see Chapter 3). Given that the C-terminally His-tagged constructs were expressed most successfully, those full-length constructs of the Kindlin family were selected for further expression trials. Therefore, the same C-terminally His_6–tagged full-length Kindlin-3, used in the initial bacterial protein production, was utilised for large-scale protein production in baculovirus-infected insect cells (*Spodoptera frugiperda*, Sf9). We generated recombinant baculovirus by co-transfecting Sf9 cells with pOPINE-*FERMT3* plasmid together with an engineered linearised bacmid (BAC10:KO_{1629}) and harvested the newly formed virions after 6-7 days (see Chapter 2). In this approach, the bacmid (BAC10) is engineered so that an essential viral gene, ORF1629, is inactivated by the insertion of chloramphenicol acetyl transferase resulting in a knock-out bacmid (BAC10:KO_{1629}), and an unviable baculovirus genome (Zhao *et al.,* 2003). Co-transfection of Sf9 cells with linearised BAC10:KO_{1629} and the pOPINE-*FERMT3* vector repairs the inactive ORF1629, through transposition, resulting in a viable genome that has also incorporated the *FERMT3* gene – under the control of the polyhedrin promoter (Zhao *et al.,* 2003). This method of baculovirus generation results in 100% recombinant viruses and essentially dispenses the need for plaque purification (Nettleship *et al.*, 2010; Zhao *et al.,* 2003).

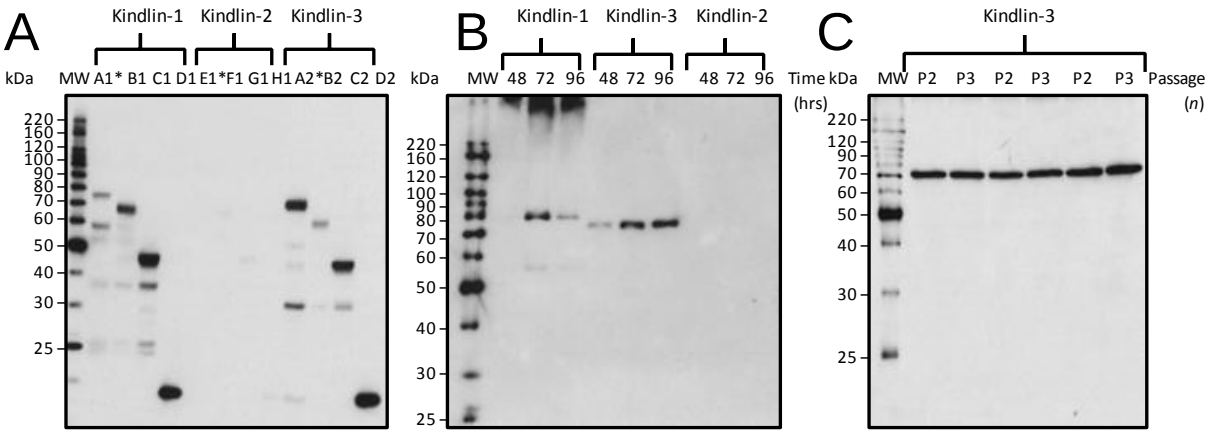


Figure 5.2 –Western Blot Analysis of Kindlin Expression in Sf9 Cells. (A) Western blot analysis of small-scale (2ml) expression of the Kindlin isoforms in baculovirus infected Sf9 cells and of additional truncation mutants. Lanes are labelled according to the construct positions (see Table 3.3). Constructs A1, E1 and A2 (asterisked) are used for further analysis. (B) Western blot analysis of small-scale (2ml) expression of full-length Kindlin in baculovirus infected Sf9 cells. Cells were harvested 48, 72 and 96 hours post-infection. (C) Western blot analysis of Kindlin-3 expression (construct A2, from A) using amplified baculovirus from different passages to determine to what extent rounds of amplification altered protein expression. Cells (2ml) were infected with equal volumes of virus from one of two passages of virus stock. The experiment was performed in triplicate.

Recombinant viruses for the full-length Kindlin isoforms were generated and amplified using standard procedures (Zhoa *et al.*, 2003; Merrington *et al.*, 1997) and the newly formed viruses subsequently used to assess the expression as a function of time after infection. Typically, several variables can influence protein expression in this system and those parameters must be determined empirically for each recombinant virus generated. Therefore expression trial experiments must be carried out in order to determine the optimal conditions for expression, which includes; multiplicity of infection (MOI) for optimal protein expression, expression time course post-infection, optimal number of virus amplifications and cell maintenance conditions. Thus, virus inoculums generated for each of the Kindlins were used to infect small-scale Sf9 cell monolayer cultures (2ml) and the expression of recombinant protein assessed by SDS-PAGE and western blotting (see Material and

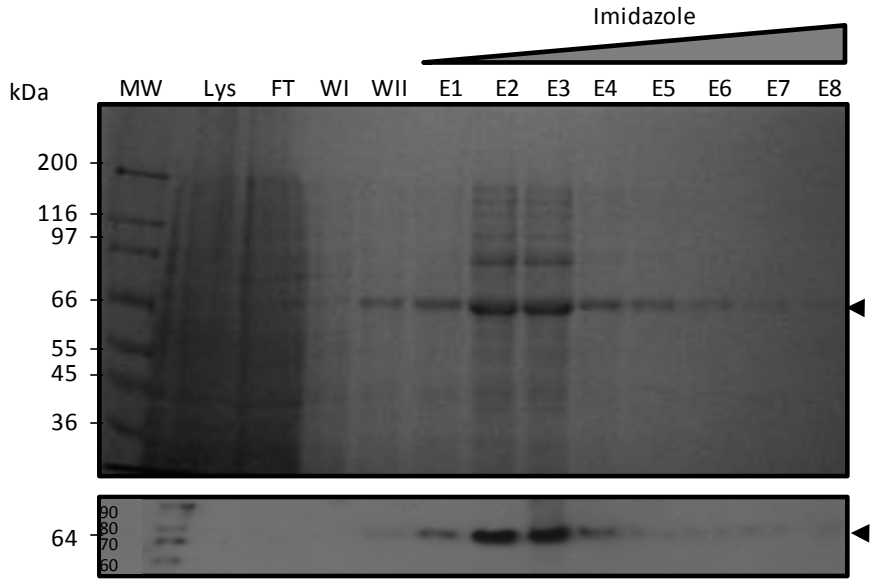
Methods). The time-course expression trial revealed that Kindlin-1 expression was maximal after 72 hours post-infection, subsequently reducing after 96 hours, whereas Kindlin-3 expression was optimal after 96 hours. However, Kindlin-2 expression was not detectable (Figure 5.2). We had previously determined that large scale protein production of recombinant Kindlin-1 using baculovirus could not be efficiently extracted from Sf9 cells, even after several solubilisation optimisation attempts. We also observed improper migration of the 77kDa protein Kindlin-1 during SDS-PAGE analysis of whole cell extract, and accordingly, deemed this as a 'precursory marker' that suggested protein aggregation. As for Kindlin-3, normal electrophoretic migration was observed and this protein was therefore suggested, and later shown (see below), to be expressed in the soluble fraction in Sf9 cells and could therefore be isolated using standard purification techniques. It is now suggested that the unusual nuclear localisation of Kindlin-1 and Kindlin-2, but not Kindlin-3, observed in human cells, possibly by bindng to a nuclear-cytoplasmic shuttling protein such as migfillin (Meves, *et al.,* 2009) may, in part, explain the abnormal observations of Kindlin-1 migration and purification. Thus, Kindlin-3 was the attention of our efforts, given that the production of recombinant protein was perceived to be tractable. What is more, the role of Kindlin-3 in cells of haematopoietic origin was recently highlighted by knock-out mice which suffer from severe bleeding due to inactive platelet integrins, mimicking the human disease Leukocyte Adhesion Deficiency type-III (LAD-III) (Malinin, *et al.,* 2009; Moser, *et al.,* 2009; Moser *et al.,* 2008).

We used baculovirus-infected Sf9 cells in order to express sufficient quantities of Kindlin-3 for biophysical studies and crystallization screening. Standard in-house optimization techniques were implemented in order to maximize the protein expression. Initially, the number of recombinant viral passages and its influence on Kindlin-3 expression was tested. The recombinant virus carrying *KINLDIN-3* (*FERMT3*) was amplified in Sf9 cells

using a virus inoculation dilution of 1:100 and the subsequent virus harvested after 144 hours (6 days) post-infection, as per the standard protocol. The resulting virus solution was used in a second round of amplification, analogous to the first. The resulting virus solutions, denoted as passage 1 (P1) and passage 2 (P2) were inoculated into small-scale Sf9 cell monolayer culture, at a uniform dilution (1:10). The expression of recombinant Kindlin-3 was assessed 96 hours post-infection *via* SDS-PAGE of whole cell extract followed by western blotting. Western blot analysis using an anti-penta-His antibody probe demonstrated comparable protein expression of a 75kDa protein between the viral stocks (Figure 5.2). Furthermore, the inoculation ratio for expression was assessed using the viral stocks and analysed by SDS-PAGE and western blotting. Western blotting revealed that a 1:10 inoculate dilution gave optimal protein expression and was therefore implemented for large scale protein production. Due to the volumetric restraints of a 1:10 dilution the P3 virus stock was used as its amplification could yield virus solutions to the order that we required for litre-scale expression. Therefore, one litre suspension cultures of Sf9 cells were inoculated with amplified P3 viral stocks at a 1:10 dilution and incubated for 96 hours before harvesting the insect cells for protein purification. Standard metal affinity chromatography (IMAC) methods partially purified a ~75kDa protein, assumed to be Kindlin-3 due to an immunoreactivity of the polypeptide with an anti-His antibody probe during western blotting and an appropriate molecular weight estimation. The partially purified Kindlin-3 was further purified by gel filtration chromatography, which gave a symmetrical peak when the elution was monitored using absorbance at 280nm and therefore suggested that Kindlin-3 was predominantly monomeric. SDS-PAGE analysis revealed a homogeneous solution of Kindlin-3. Moreover, we verified that we had purified recombinant murine Kindlin-3 by determining its identity from SDS-PAGE analysis of the IMAC purified protein using in-house fingerprinting (Central Proteomics Facility (CPF), Centre for Cellular and Molecular Physiology (CCMP),

Figure 5.3 (Opposite) – SDS-PAGE Analysis of Kindlin-3 Purification. (A) Representative SDS-PAGE of immobilised metal (nickel) affinity chromatography (IMAC) purification of Kindlin-3 expressed in baculovirus infected Sf9 cells (~2g cell weight). Adsorbed protein was eluted using an imidazole gradient. Western blot analysis (below) was also performed to confirm the presence of the engineered His-tag on the recombinant protein. (B) Identification of the IMAC purified Kindlin-3 protein using in-house fingerprinting (Central Proteomics Facility (CPF), Centre for Cellular and Molecular Physiology (CCMP), Oxford) by Tryptic Digest and Liquid Chromatography-Mass Spectroscopy/Mass Spectroscopy (LC-MS/MS). Peptide coverage of 68% corresponded to Unirpot accession code Q8K1B8. (C) Gel filtration of IMAC purified Kindlin-3 using Superdex S200 (10/30) exhibiting a single symmetrical peak at an elution volume of approximately 14.5ml, which according to a standard curve is equivalent to ~75kDa weight – suggesting that Kindlin-3 is monomeric.

A



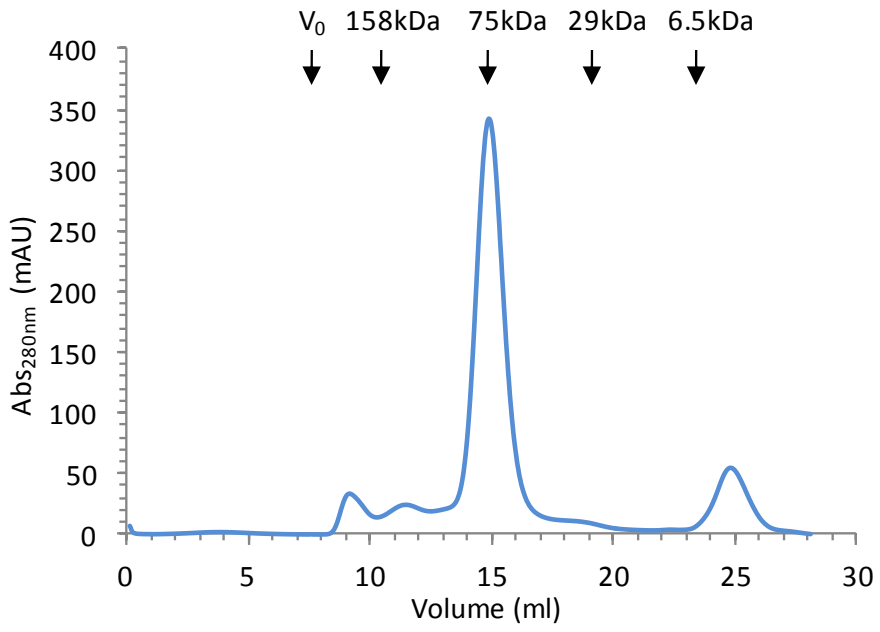
B

Match to: Q8K1B8 Score: 2869

```
  1 MAGMKTASGD YIDSSWELRV FVGEEDPEAQ SVTLRVTGES HIGGVLLKIV
 51 EEINRKQDWS DHAIWWEQKR QWLLQTHWTL DKYGILADAR LFFGPQHRPV
101 ILRLPNRRVL RLRASFSKPL FQTVAAICRL LSIRHPEELS LLRAPEKKEK
151 KKKEKEPEEE VHDLTKVVLA GGVAPTLFRG MPAHFSDSAQ TEACYHMLSR
201 PQPAPDPLLL QRLPRPSSLP DKTQLHSRWL DSSRCLMQQG IKAGDVLWLR
251 FKYYSFFDLD PKTDPVRLTQ LYEQARWDLL TEEIDCTEEE MMVFAALQYH
301 INKLTLSGDV GELASGDLGL DDLDAALNNL EVKLKGSAPS DMLDSLTTIP
351 ELKDHLRIFR PRKLTLKGYR QYWVVFKDTT LSYYKSQDEA PGDPTQQLNL
401 KGCEVVPDVN VSGQKFCIKL LVPSPEGMSE IYLRCQDEQQ YAQWMAACRL
451 ASKGRTMADS SYASEVQAIL AFLSLQRAGG SNGGSGNKPQ GPEAPAEGLN
501 PYGLVAPRFQ RKFKAKQLTP RILEAHQNVA QLSLTEAQLR FIQAWQSLPD
551 FGISYVMVRF KGSRKDEILG IANNRLIRID LAVGDVVKTW RFSNMRQWNV
601 NWDIRQVAIE FDEHINVAFS CVSASCRIVH EYIGGYIFLS TRERARGEEL
651 DEDLFLQLTG GHEAF
```

Sequence Coverage: **69%**

C



Oxford) by tryptic digestion and liquid chromatography-mass spectroscopy/mass spectroscopy (LC-MS/MS), which returned 69% peptide coverage that matched to the Uniprot accession Q8K1B8 with very high mascot scores of approximately 2900, which shows that we purified authentic murine Kindlin-3. The approximate yield of protein was around 1mg per litre of Sf9 cell culture (~2-4g cell pellet), however attrition of material throughout purification meant that the final yield from this initial one litre scale experiment was <1mg. Therefore, multiple litre-cultures were infected in order to acquire sufficient quantities of protein and therefore sufficient virus was required. It was observed that the harvested virus had a limited longevity even at 4°C in the dark, with diminishing protein yields correlating to the lifetime of the stored virus. Invariably, the protein yields obtained from the initial trial were not reproducible and varying amounts of the recombinant Kindlin-3 protein could be purified, as before, but with unsatisfactory purity for further experimentation. Repeating the transfection for baculovirus generation was also attempted together with amplification; however, this did not improve the protein expression nor improve reproducibility.

Consequently, a method was devised in order to obtain reproducible and predictable quantities of recombinant protein in insect cells. Baculovirus-infected Sf9 cell expression protocols used by the Structural Biology department at the Institute of Cancer Research (ICR), London, UK, were adopted and trialled as an alternative method to generate sufficient quantities of recombinant Kindlin-3. It had been suggested that viral amplification is maximal at 72 hours (3 days) post-infection, with a significant reduction in viral titre at 96 hours at a significantly greater cell density and that additional aeration is necessary (http://strubiol.icr.ac.uk/extra/baculovirus/baculo_protocols.html). Previously, virus was harvested 144 hours (6 days) post infection using suspension cultures (at a density of $1x10^6$ cells/ml) shaking at 100 rpm with a culture volume to flask volume ratio of 1:5. Accordingly,

viruses were generated and harvested using the standard transfection methods and were subsequently amplified in suspension culture (at $2x10^6$ cell/ml), ensuring that aeration was adequate *via* a 1:20 culture volume-to-flask volume ratio and increased shaking (110 rpm), and viruses were harvested at 72 hours post infection. These viruses were inoculated into the Sf9 suspension culture at a 1:50 dilution ratio. Due to the imprecision of the plaque assay (Dr. G. Sutton, Division of Structural Biology, University of Oxford, personal communication), it was assumed an expected $1x10^7$ plaque forming units (pfu) were yielded per millilitre of culture when virus was generated and therefore the amplification culture was infected at an estimated multiplicity of infection (eMOI). The harvested virus was assumed to possess an expected viral titre of $2x10^8$ pfu/ml after virus release, based on a conservative estimate of 100pfu/cell during the amplification (Wasilko *et al.*, 2009). The harvested virus was subsequently used in a large scale expression experiment, in which Sf9 suspension culture ($2x10^6$cells/ml) were inoculated with amplified virus at an eMOI equal to 1, or at a dilution of 1:100. Large scale Sf9 cell culture for protein production was performed using a culture volume to flask volume ratio of 1:5 together with shaking at 100 rpm, as per standard protocol. Protein production was achieved after 72 hours, as it was observed that cell density decreased during incubations of greater than 3 days. Given that Kindlin-3 is a cytoplasmic protein and not a secreted protein, it was vital that an adequate cell density was maintained. It was observed that by implementing an improved virus amplification method the expression of Kindlin-3 improved three-fold, yielding between 3-4mg purified protein (see below) per litre of culture (4g cells), and not only required less virus for inoculation but also reduced the number of viral passages. It was also observed that this protocol resulted in reproducible and predictable protein production in Sf9 cells. Therefore, this method allows the expression of recombinant mouse Kindlin-3 in Sf9 cells, yielding a 6-fold improvement on recombinant

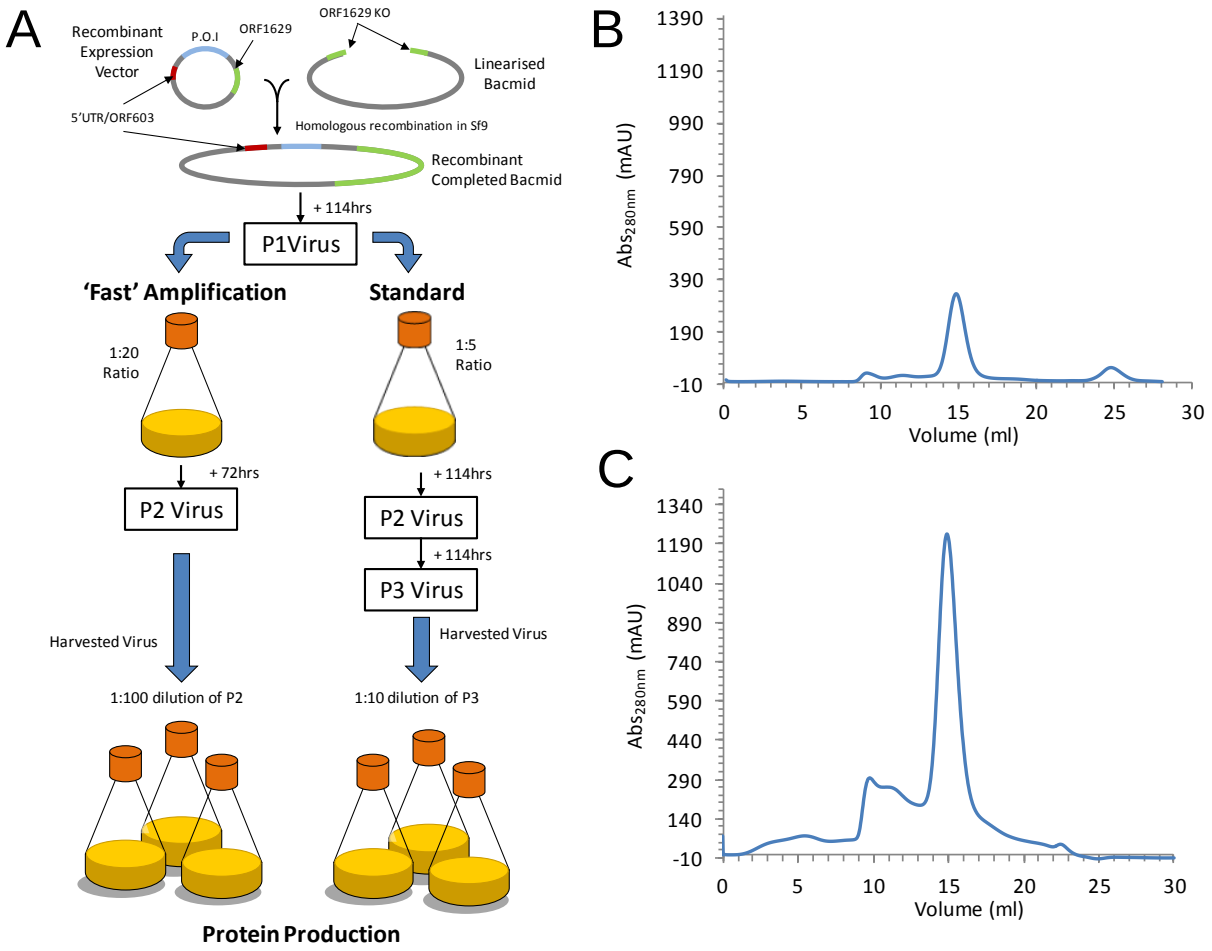


Figure 5.4 – Kindlin-3 Expression Optimisation in Baculovirus-Infected Sf9 Cultures. (A) Schematic comparison of two procedures used in the expression of Kindlin-3. In the standard procedure (right hand side) the baculovirus is amplified in Sf9 suspension cultures that occupy one fifth of the total flask volume (1:5 ratio). The amplified virus is harvested after 114 hours (6 days) and the procedure is repeated using the amplified virus as a new stock until sufficient quantities of virus is generated in order to infect litre-scale Sf9 suspension cultures at a 1:10 dilution. In the "fast" procedure (left hand side) the baculovirus is amplified in Sf9 suspension cultures that occupy one twentieth of the total flask volume (1:20 ratio). Sufficient quantities of virus are generated in a single amplification. Due to the efficient viral amplification litre-scale Sf9 suspension cultures are infected at a 1:100 dilution. (B and C) Size exclusion chromatography (SEC) of IMAC purified Kindlin-3 from three one litre cultures of baculovirus infected Sf9 cells using the standard (B) and the "fast" (C) procedure.

protein expression in *Escherichia coli* and a 3-fold improvement on protein production in insect cell culture, using standard methods implemented in the host Division of Structural Biology, University of Oxford, together with the concomitant reduction in the timescale of protein production in insect cells by approximately 6 days (see Figure 5.4, below).

5.1.4. Purification of Kindlin-3

As described previously, we have demonstrated that Kindlin-3 expression can be improved significantly using an alternative baculovirus amplification protocol. Therefore, large scale baculovirus-infected Sf9 cells, yielding 3-4mg of Kindlin-3 per litre of culture, were cultivated and harvested for large scale purification. Kindlin-3 was purified using IMAC followed by size exclusion chromatography, as before. Whilst there was significantly more Kindlin-3, the purified protein was not of sufficient purity for further experiments, especially crystallisation screening. Therefore an additional chromatographic step was required to yield homogenous protein. It was noted that Kindlin-3 possesses a large number of basic residues (Arginines and Lysines), with a particular stretch of polylysine in the primary sequence. As a result, it was surmised that Kindlin-3 could be further purified by pseudo-affinity chromatography, by making use of an anion exchange step, particularly the polyanionic heparin column, whereby the lysine-rich basic patches on Kindlin-3 would bind tightly to the sulphate groups of heparin. Consequently, Kindlin-3 was buffer exchanged into a low salt buffer (20mM TrisHCl, pH 7.5, 50mM NaCl, 1mM DTT) prior to its application onto a heparin column. Kindlin-3 bound to the heparin column and was subsequently eluted at 0.6M NaCl, when using a linear NaCl gradient, which gave a single symmetrical peak when the absorbance at 280nm was monitored and yielded near-homogeneous protein as assessed by SDS-PAGE. Repeating the purifications, it was observed that Kindlin-3 bound to the heparin sepharose in buffers containing up to 200mM NaCl, which resulted in

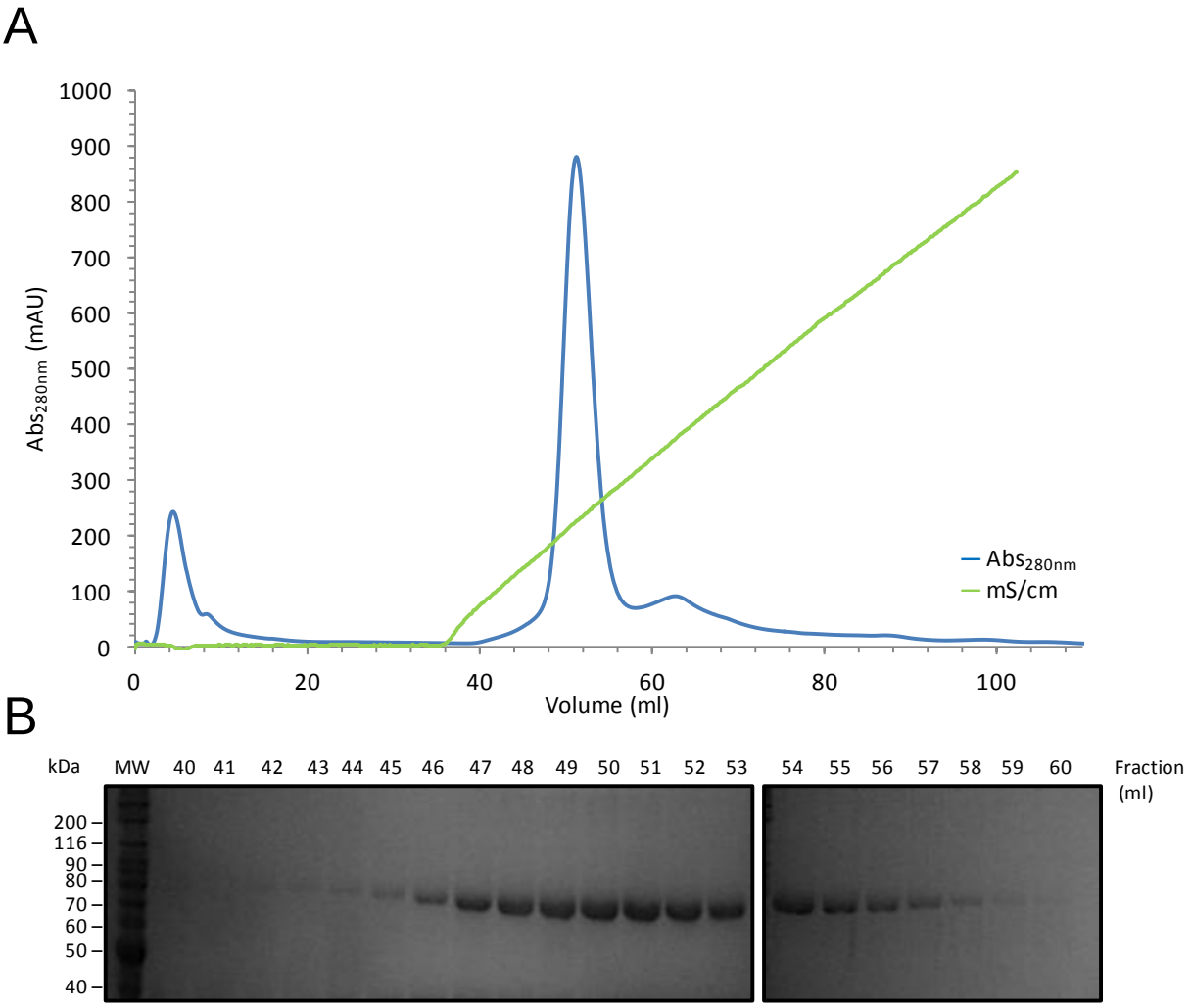


Figure 5.5 –Purification of Kindlin-3 by Heparin Affinity Chromatography. (A) Elution profile observed at 280nm (blue) displaying a single symmetrical peak eluting under a liner sodium chloride gradient (green) using the range 0.05M to 1.0M. (B) SDS-PAGE of the fractionated elution demonstrating the presence of a 75kDa protein, Kindlin-3.

an improvement of purity. Heparin affinity purified Kindlin-3 was finally polished to homogeneity *via* a final gel filtration step that gave a single protein solution, estimated to be >95% pure by SDS-PAGE. Gel filtration (Figure 5.4) using Superdex S200 (10/30) or (16/60) (see Materials and methods) exhibited a single symmetrical peak at an elution volume equivalent to ~75kDa weight – suggesting, as before, that Kindlin-3 is predominantly monomeric. Interestingly, SDS-PAGE analysis of highly concentrated protein (15mg/ml) demonstrated small contamination of two additional polypeptides of equal intensity (Data not

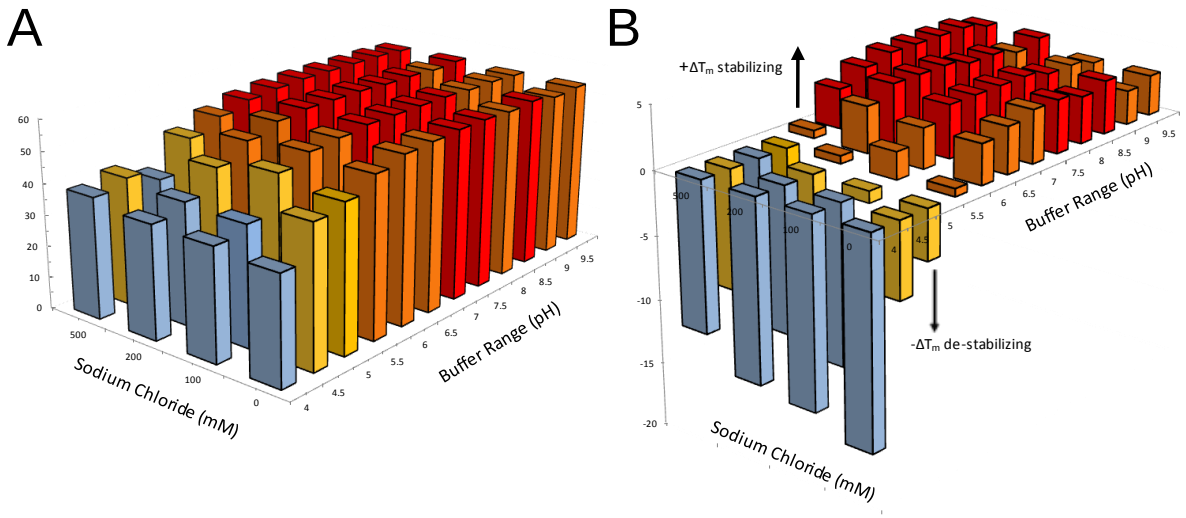


Figure 5.6 – Thermofluor-based Thermal Shift Assay Buffer Screening. Kindlin-3 was diluted into various buffers comprising a two-dimensional screen of pH *versus* sodium chloride concentration. Transition temperatures (temperature mid-points) were observed by fluorescence of the hydrophobically bound dye, Spyro orange (molecular probes) and calculated using the Opticon Monitor software. (A) A 3D histogram of the transition temperatures is plotted together with (B) the change in transition temperature from the average 50°C. For clarity the bars are coloured according to the range of temperatures to which they correspond.

shown). The combined molecular weight of these polypeptides was equal to the full length protein, suggesting that a small fraction of Kindlin-3 was nicked.

In order to assess the optimal stability of the protein in the Kindlin-3 purification buffer and ultimately the ideal buffer for concentration and crystallisation screening, we performed a thermofluor-based thermal shift assay. The protein solution was diluted into buffers that covered a range of pH and sodium chloride concentrations, therefore forming a 2-dimensional screen. The melting of the protein is measured by observing the fluorescence from Spyro orange dye (Molecular Probes) that binds to hydrophobic residues within the folded protein core, over the temperature range 20°C – 95°C (293K – 368K). The temperature mid-point at which the protein unfolds (transition temperature, T_m) was calculated using the *Opticon Monitor* software as described (Sainsbury *et al.*, 2008). Interestingly, the average transition temperature of Kindlin-3 in all conditions was ~50°C and

this was used as a baseline to calculate whether the buffer conditions gave a positive shift in the transition temperature (ΔT_m) (Figure 5.6). Kindlin was observed to be most stable at high sodium chloride concentrations (500mM) within the pH range 7.0 – 9.0, with a consistent transition temperature (T_m) of 55°C. It was also observed that the T_m of Kindlin-3 was approximately 55°C within the pH range 7.0 – 7.5 irrespective of sodium chloride concentration. According to the thermal shift assay Kindlin-3 was least stable at low pH (pH <5.5) irrespective of sodium chloride concentration, with the melt curves at very low pH exhibiting a non-sigmoidal shape. Furthermore, additives were screened in parallel to the buffer screening, which showed no improvement in thermal stability with all T_m calculated less than the average transition temperature.

5.1.5. Biophysical characterization of Purified Kindlin-3

To assess the monodispersity and oligomeric state of the purified Kindlin-3 in solution we used dynamic light scattering (DLS) and analytical ultracentrifugation (AUC). DLS suggested that the protein existed as a single species with a hydrodynamic radius of 3.6 nm (±0.29) (Figure 5.7, overleaf), with a percentage of polydispersity of ~10%, suggesting that the protein solution was monodispersed. Both DLS and size exclusion chromatography (SEC) demonstrate that Kindlin-3 is predominately monomeric in solution, which was further validated by AUC where the model-independent c(s,f/f_0) method of data analysis showed a single major peak with a sedimentation coefficient of 4.3 S (Figure 5.7). A species of 4.3 S with a molecular weight of 75kDa has a predicted hydrodynamic radius of 3.8 nm, in agreement with the DLS results, and a frictional coefficient of 1.46. These data suggest that Kindlin-3 has an extended conformation with an axial ratio of 5.75 when modelled as a

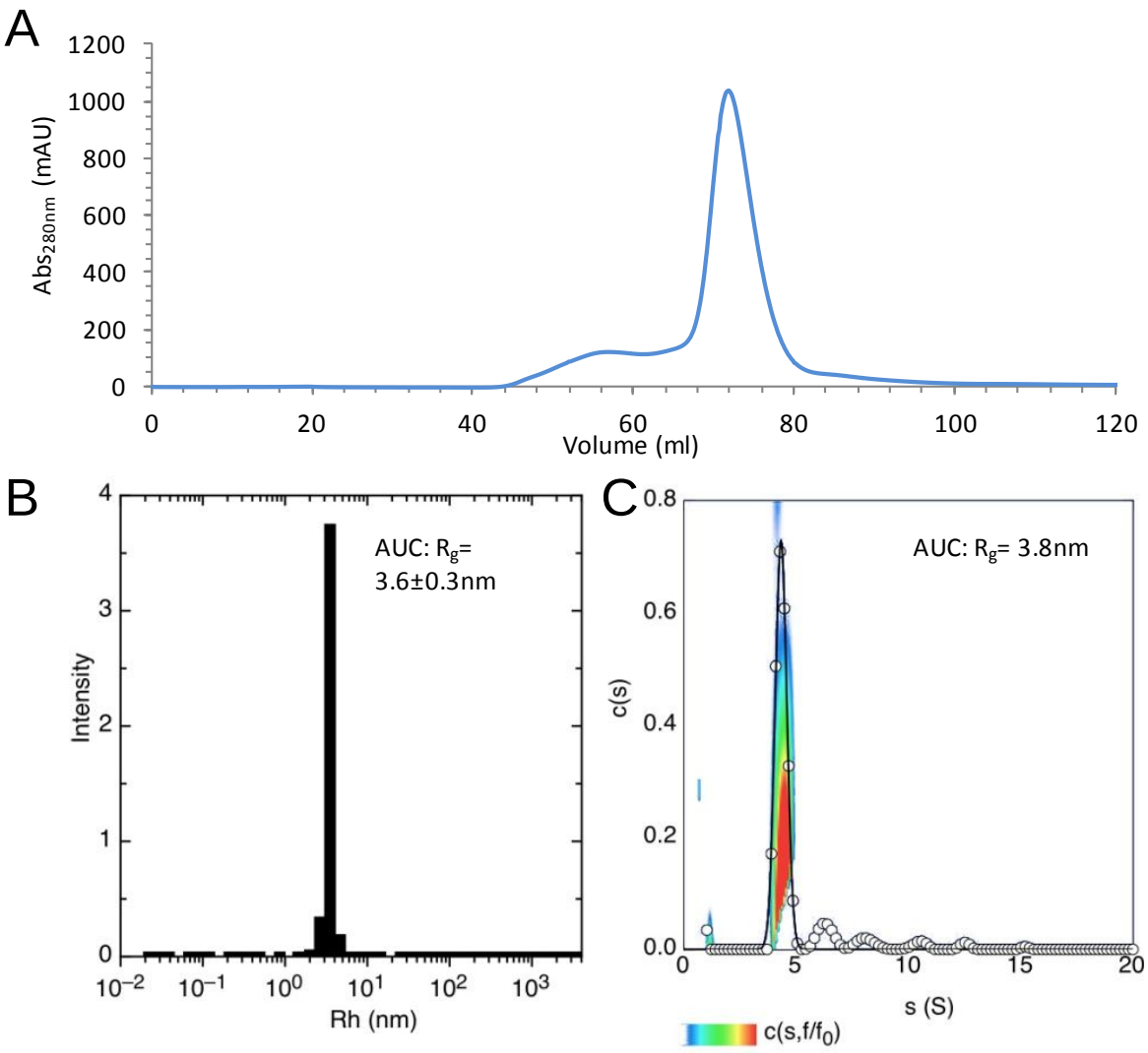


Figure 5.7 – Purification and Preliminary Biophysical Characterisation of Kindlin-3. (A) Representative gel filtration elution profile of purified Kindlin-3 using a Superdex S200 (16/60) (GE Healthcare) in Tris-HCl, pH 7.5, 150mM NaCl and 1mM DTT at 20°C. Based on the elution volume, Kindlin-3 migrates as expected for a 75kDa protein, suggesting a monomer. (B) Dynamic light scattering (DLS) measurements of kindlin-3 in 20mM Tris-HCl, pH 7.5, 150mM NaCl, and 1mM DTT at 22°C. Radius estimated for a 75kDa protein using a spherical model is 3.77nm. (C) Analytical ultracentrifugation (AUC) of homogenous Kindlin-3 solution with c(s,f/f_0) analysis of Kindlin-3. The c(s) component is shown in open circle symbols with the predominant species fitted with a Gaussian, giving Svedberg (s) value = 4.3 +/- 0.01S and variance (σ) = 0.25 +/- 0.01S. The temperature plot shows the distribution of apparent frictional coefficients in the sample.

prolate ellipsoid providing the first evidence that Kindlin-3 is extended in conformation (see Materials and Methods, section 2.8.4).

5.1.6. Crystallisation Screening of Kindlin-3

We have previously demonstrated that milligram quantities of Kindlin-3 can be obtained using an optimised baculovirus expression system and as such performed crystallisation screening of the full-length protein. What is more, biophysical characterisation by SEC, DLS and AUC suggest a monodispersed monomer and the sample therefore seems suitable for crystallisation screening. Kindlin-3 was concentrated prior to crystallisation screening in a buffer of pH 7.5 with 300mM NaCl, as determined earlier by thermal shift analysis. The protein was highly soluble, concentrating to >15mg/ml. However, pre-crystallisation trials and their criteria (i.e. the sitting drops are clear or exhibit light granular precipitate) for assessing optimal protein concentrations for crystallisation screening (Hampton Research, CA) suggested that 9mg/ml met the requirements. Therefore concentrated protein (9mg/ml in 20mM Tris-HCl, pH 7.5, 300mM NaCl, 1mM DTT) was used for crystallization screening with commercially available crystallization reagents and vapour diffusion methods using nanolitre sitting drops (see Materials and Methods). The His_6-tag was not removed prior to crystallization screening, but could be removed *via* treatment with carboxypeptidase A (Sigma) if required. Crystallisation screening failed to yield any crystals, even after several months, at either room temperature or 4°C. Additional crystallisation techniques were employed and attempted in order to obtain crystals including reductive lysine methylation, which has been shown to improve crystallisation success where proteins have failed previously (Walter *et al.*, 2006). Given the polylysine patch within the Kindlin family F1 subdomain used to purify the protein by heparin affinity chromatography, it was reasonable to suggest that this region may hinder crystallisation and that reductive

methylation may circumvent this (Walter *et al.*, 2006). Kindlin-3 was methylated using standard protocols as described in (Walter *et al.*, 2006) and utilised in crystallisation screening as described previously, however using a concentrated methylated protein solution at 8mg/ml.

It was also observed that protein aggregated within the sitting drops in a time-dependent manner and was suggested to be indicative of oxidation of the protein, in particular the oxidation of cysteine residues. There are ten cysteines in the purified full-length protein. It has been shown that the covalent modification of cysteines by iodoacetic acid treatment can aid crystallisation (Soisson *et al.*, 2008). Therefore, protein was incubated with iodoacetic acid for several hours at room temperature in the dark and purified by gel filtration to remove the excess iodoacetic acid, which gave a similar elution peak to the native protein (see Figure 5.7 for SEC profile of the native protein). Interestingly the elution volume of carboxymethylated-Kindlin-3, was shifted by ~10ml (sephadex S200 16/60) compared to the unmodified protein. The reason for this is unknown. Again, this modified protein was concentrated and utilised in standard crystallisation screening, at a concentration of 5mg/ml as determined by pre-crystallisation trial analysis (Hampton Research, CA).

Finally, it was shown that for Kindlin-1 the F1 subdomain possesses a ~100 amino acid loop that, in isolation, was considered completely unstructured by NMR (Goult *et al.*, 2010) and is consistent with disorder predictions using the disorder probability prediction program, regional order neural network (RONN) (Yang *et al.*, 2005). The analogous loop in Kindlin-3, which is the shortest of all the Kindlins, was also suggested to be unstructured, with a small region at the centre predicted to possess some order. Therefore, we used non-detergent sulphobetaine-201 (NDSB-201) in order to attempt stabilization of this disordered region of the F1 domain, as NDSBs are known refolding agents. Furthermore, Kindlin-3 was shown to be prone to aggregation events at concentrations of greater than 4mg/ml by small

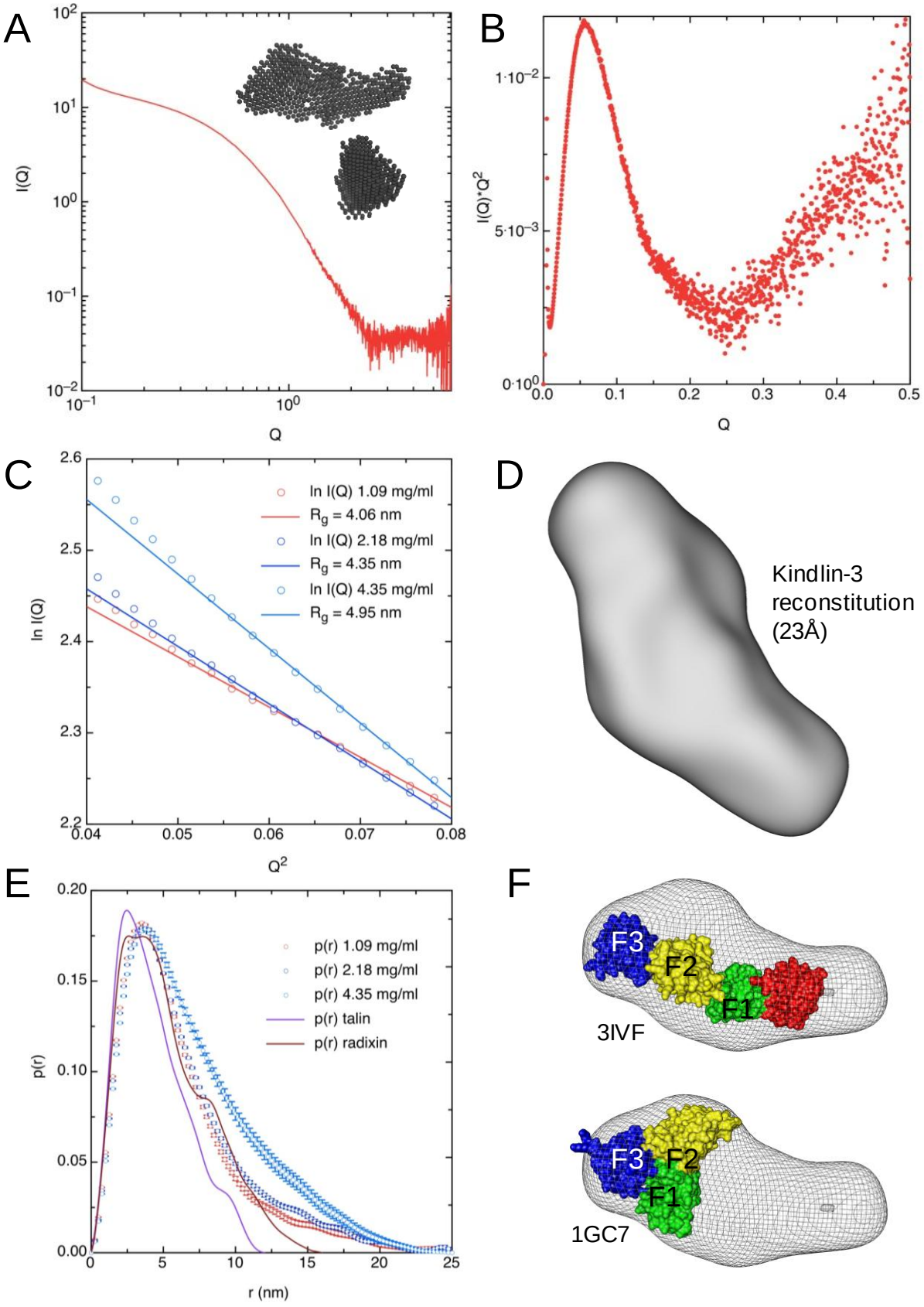
angle X-ray scattering (see below) and despite its solubility overall at even 15 mg/ml. NDSB have been shown to prevent aggregation of proteins and as a result has directly led to the successful crystallisation and structure solution of two Vaccinia virus proteins (Graham *et al.*, 2008). However, the presence of NDSB did not yield any crystals of Kindlin-3. Therefore after extensive crystallisation screening, in solution techniques, such as small angle X-ray scattering (SAXS) and nuclear magnetic resonance (NMR) were employed to gain a structural insight into the protein and its function.

5.1.7. Small-angle X-ray scattering (SAXS) of Kindlin-3 reveals an extended conformation

The Kindlins have yet to be structurally characterised in their entirety, with only the structure of the N-terminal F0 subdomain from Kindlin-1 and Kindlin-2, solved by NMR (Perera *et al.*, 2011; Goult *et al.*, 2010) and the crystal structure of the Kindlin-1 Pleckstrin Homology (PH) domain solved by crystallography and Kindlin-2 and Kindlin-3 PH domains by NMR (see Chapter 4), to date. It has been noted that the difficulty in acquiring milligram quantities has been a major barrier to biochemical characterisation (Harburger *et al.*, 2009) and therefore structural studies (Prof. I. D. Campbell, University of Oxford, personal communication). Given that extensive attempts at crystallising Kindlin-3, with some 4600 crystallisation conditions screened to date without success, small angle X-ray scattering was used to determine the overall solution structure of Kindlin-3 at low resolution.

Accordingly, SAXS data were obtained at four concentrations: 1.09, 2.18, 4.35 and 8.7 mg/ml (see Materials and Methods). Data were processed using PRIMUS (Konarev *et al.*, 2003) and analysed using GNOM (Svergun, *et al.*, 1992). All four plots of I(Q) (e.g. Figure 5.8A) were analyzed using the distance distribution function approach p(r). In this method of analysis, a manually fitted maximum dimension (D_{max}) for the scattering species is used to

Figure 5.8 (Opposite) – Solution Structure of Kindlin-3 by SAXS. (A) The scattering intensities as functions of momentum transfer ($Q = 4\pi \sin(\theta)/\lambda$ where 2θ is the scattering angle and $\lambda = 0.93$Å and is the X-ray wavelength), together with two orthogonal views of a single scattering bead model averaged from 10 calculated *ab initio* models constructed with the DAMAVER package (Volkov *et al.* 2003) inset. (B) A Kratky plot (Kratky, 1951), demonstrating that the protein is partially unfolded or flexible. (C) Guinier analysis used to estimate the radius of gyration (R_g) at low scattering angles for 1.09 (red symbols), 2.18 (blue symbols) and 4.35mg/ml (cyan symbols) samples. (D) 23Å scattering envelope for Kindlin-3 reconstituted as described in Materials and methods. (E) A pair distance distribution plot, p(r) function, for Kindlin-3 at 1.09 (red symbols), 2.18 (blue symbols) and 4.35mg/ml (cyan symbols) with p(r) functions of the crystal structures of talin-1 (PDB: 3IVF) and Radixin (PDB: 1GC7) FERM domains, which are fitted to the reconstituted Kindlin-3 envelope in (F). Radius of gyration is estimated from the slope of the line fitted the data points. (A, C, D, E) were drawn using ProFit software.


A
I(Q)
Q
B
I(Q)*Q²
Q
C
ln I(Q)
Q²
ln I(Q) 1.09 mg/ml
Rg = 4.06 nm
ln I(Q) 2.18 mg/ml
Rg = 4.35 nm
ln I(Q) 4.35 mg/ml
Rg = 4.95 nm
D
Kindlin-3
reconstitution
(23Å)
E
p(r)
r (nm)
p(r) 1.09 mg/ml
p(r) 2.18 mg/ml
p(r) 4.35 mg/ml
p(r) talin
p(r) radixin
F
F3
F2
F1
3IVF
F3
F2
F1
1GC7

calculate a plot of rotationally-averaged scattering vectors found within the particle, via Fourier transformation. As the concentration of the samples increased so did the apparent D_{max}, from 14nm to 20nm, indicating some inter-particle interactions giving rise to cross-scattering vectors (Figure 5.8B). This indicates the presence of sample aggregation and inter-molecular scattering. Guinier analysis also revealed that the radius of gyration calculated for each concentration increased with respect to the protein concentration, consistent with the notion of sample aggregation. The two lowest concentrations however gave very similar p(r) functions (Figure 5.8B), with D_{max} = 14 nm and D_{max} = 16 nm (R_g 4.07 ± 0.01 and 4.51 ± 0.01, respectively) and Guinier plots in which the R_g closely matched that calculated using the p(r) function (Figure 5.8C). These p(r) functions indicate an elongated conformation for Kindlin-3 and were then used in the calculation of *ab initio* shape reconstitution using the program GASBOR (Svergun *et al.*, 2001). *Ab initio* scattering bead models were consistently elongated and therefore, in line with AUC and DLS, a prolate assumption was used in the model calculations. By determining multiple independent ensembles of such models we calculated estimated resolutions for the resulting structures of 23-29Å using the Fourier shell correlation approach used in electron microscopy (see Materials and Methods, section 2.8.2.) (Grigorieff, 2000).Thus ten pairs of independently calculated *ab initio* models from each data set were aligned and correlated in Fourier space and a resolution estimated using a Fourier shell correlation (FSC) criterion of 0.5. Three of the structures – two collected at the lowest concentration and one collected at 2.18 mg/ml – agreed with each other to a resolution of 23Å using an FSC analysis and these were averaged to generate the reconstruction shown in Figure 5.8C. The approximate overall dimensions of this envelope are 130 x 60 x 30 Å. Calculation of a theoretical sedimentation coefficient for it (see Materials and Methods, section 2.8.1) indicates a value of 5.0S, in reasonable agreement with the experimental 4.3S. Additional drag reducing the experimental value is expected to be caused by the disordered

loop within the F1 subdomain. The radius of gyration (R_g) of this model is 3.5 nm, in excellent agreement with the values determined by DLS and AUC (3.6 nm and 3.8 nm, respectively). Kindlin-3 is therefore much more elongated than the clover-leaf conformation of the canonical FERM domain, as in radixin for example, but is more similar to the non-canonical elongated talin head FERM domain (Figure 5.8F) – which is the only known example of a linearly arranged FERM domain. Fitting of the talin head domain crystal structure (PDB: 3IVF) into the Kindlin-3 SAXS envelope suggests that Kindlin-3 possesses features additional to those described by the linear talin FERM domain and is significantly more swollen and slightly longer than the talin crystal structure (Figure 5.8F). This can be accounted for by, and is consistent with, the presence of additional polypeptide sequence (~25kDa) in Kindlin-3 (75kDa), in particular the F1 subdomain and the existence of an additional Pleckstrin Homology (PH) domain not present in talin (50kDa). As described in Figure 5.1, the Kindlin FERM domain is most closely related to that of talin but differs by the possession of a large 107 residue loop within the F1 subdomain and a bipartite F2 subdomain intersected by the PH domain. Given that SAXS yields solution structures averaged over ~10^{15} molecules in the sample volume and the resulting reconstitution is conservatively filtered and is a consensus from 30 *ab initio* models, assessed conservatively to provide structure resolution at 23Å, we therefore treated the reconstituted and filtered density as a truly averaged solution structure. We subsequently fitted atomic models within the envelope to approximate the positions of the subdomains of Kindlin-3 using atomic models derived from crystallography or NMR of talin and Kindlin subdomains. Therefore, models were constructed by fitting the F0-F1 and F2-F3 domain tandems of talin (Elliott *et al.*, 2010), and the PH domain structure from Kindlin-1 (See Chapter 4) into the reconstituted Kindlin-3 envelope using the software VEDA (http://mem.ibs.fr/VEDA/). We then assembled various

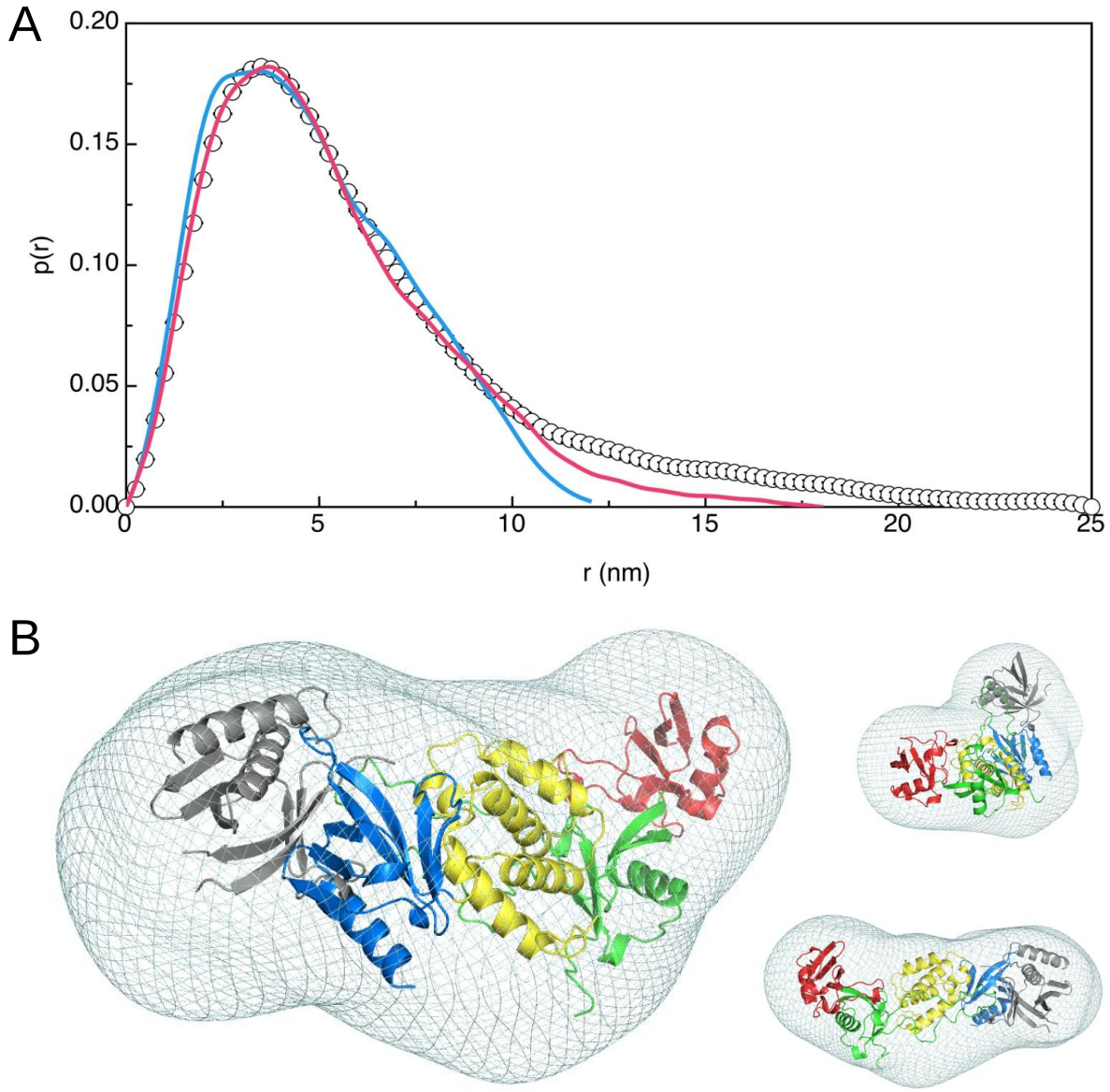


Figure 5.9 – Theoretical Solution Structure of Kindlin-3 by SAXS. (A) Pair distribution function p(r) from experimental scattering data of Kindlin-3 at 1.09mg/ml and a theoretical Kindlin-3 model based on several homologous domains determined by X-ray crystallography or NMR. The addition of the F1 loop to the model clearly improves the theoretical p(r) (pink) over the best model without the F1 loop (blue) when compared the experimentally derived Kindlin-3 p(r) (open circles). (B) Three views of the composite atomic models fitted into the reconstituted Kindlin-3 SAXS density (filtered to 23Å) illustrating reasonable fitting and approximate locations of the domains. The domains are coloured as in Figure 3.1. Additional density may be accounted for by unmodelled polypeptide, especially the N and C termini, and a large proportion of the 109 residue Kindlin F1 loop (the F1 loop shown here belongs to talin). Figure was rendered in Pymol (http://www.pymol.org).

possible alternative arrangements of F0-F1, F2-F3 and PH domains and computed scattering curves and a p(r) function for these models and compared these to the experimentally derived p(r) from Kindlin-3. We found that a talin-like linear arrangement of subdomains F0-F3 with the additional PH domain approximately positioned sitting above the F2-F3 tandem, agreed well with the experimental data. The model could be further improved by the inclusion of an F1 loop. Whilst the Kindlin-1 F1 loop has been shown to be disordered in isolation (Goult *et al.,* 2010) the structure, if any, is unknown. Kratky analysis (Figure 5.8F) suggests a disordered/unfolded region in Kindlin-3, which could correspond to its F1 loop region (Figure 5.1). An NMR structure of the F1 subdomain of talin alone complete with its loop (Goult *et al.,* 2010) does provide a partial model for this modeling study. It is however shorter than the Kindlin-3 loop by ~80 residues. Therefore we assembled a composite model in which the F1 subdomain of the talin crystal structure was replaced by one model from the F1 NMR structure ensemble to indicate a possible loop location in our envelope. The eighth model in the NMR ensemble was used because it was assessed as being the most probable using OLDERADO cluster analysis (http://www.ebi.ac.uk/pdbe-apps/nmr/olderado) (Kelley *et al.,* 1997). Allowing simultaneous rigid-body refinement of the composite F0-F1 tandem, the F2-F3 tandem and the PH domain in VEDA gave a further improved fit over the initial 'best fit', with an improved p(r) similarity to the experimental data (Figure 5.9). This model is therefore a good approximation of the Kindlin-3 structure and in particular accounts better for the long-vector p(r) tail, presumably due to some account being taken of the F1 loop. The fit shown also appears as if it could be improved by the inclusion of additional material between the F0-F1 and F2-F3 tandems, which would be (part of) the unmodeled 80 residues of the Kindlin-3 F1 loop. More detailed modeling is precluded by the unknown structure of the extended Kindlin loop and the likelihood it possesses a degree of disorder. Attempts to

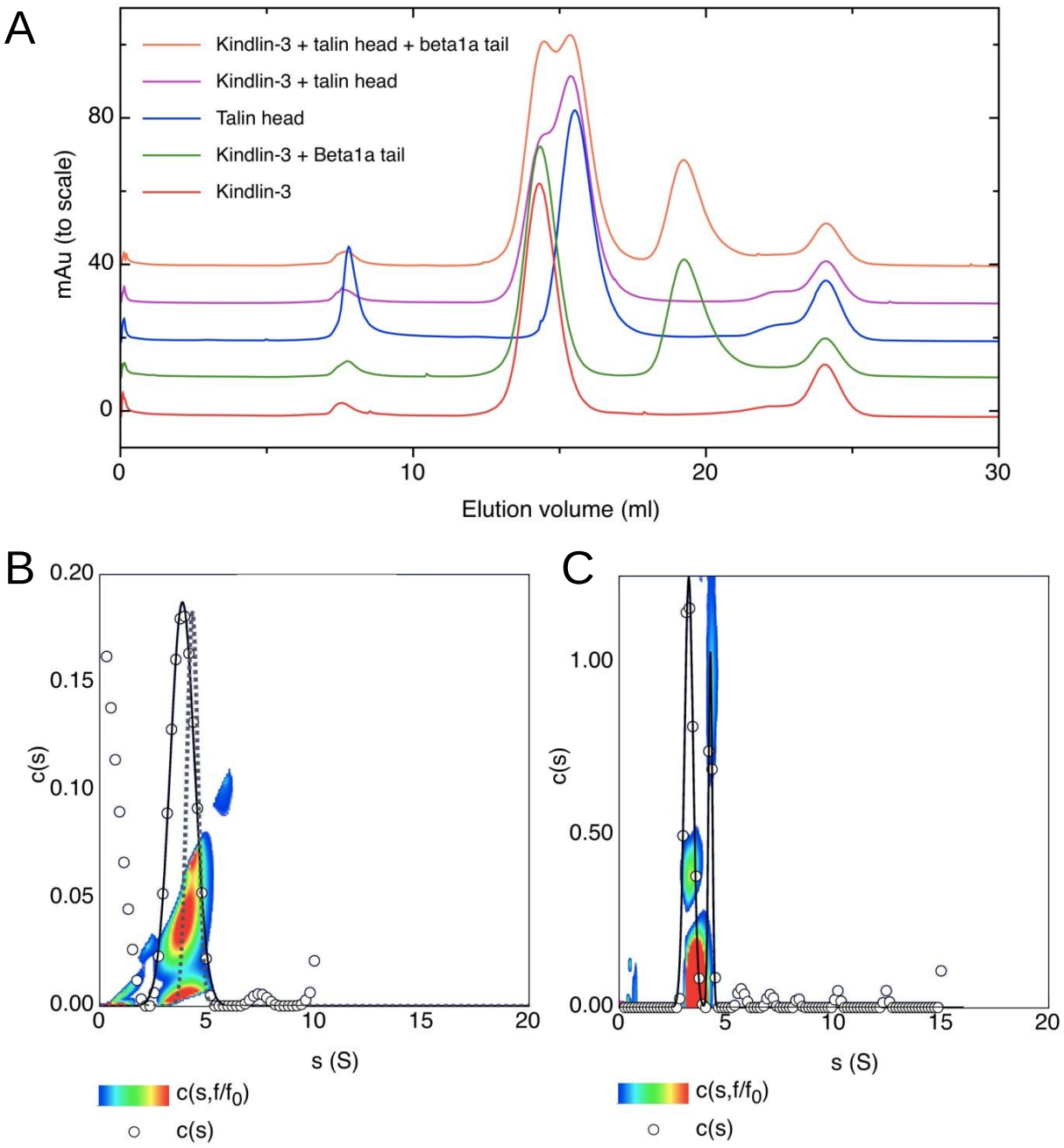
rigid body fit domains into the envelope using other methods, such as SASREF (Petoukhov *et al.,* 2005) did not yield better results.

5.1.8. Interaction partners of Kindlin-3

Heterologous interactions undergone by Kindlin-3 were initially assessed using size-exclusion chromatography (SEC) and AUC. Size exclusion chromatography (Figure 5.10) of Kindlin-3 (~75kDa), talin-head (~50kDa) and integrin β_{1A} cytoplasmic tails (~6kDa) could be resolved individually as separate symmetrical absorbance peaks when observed at 280nm. However, complexes between Kindlin-3 and β_{1A} or Kindlin-3 and talin head domain, or Kindlin-3-talin head-integrin β_{1A} cytoplasmic tails, could not be resolved as all protein mixtures could be separated as individual proteins. The elution volume for the individual proteins; Kindlin-3, talin-head, and β_{1A} peptide were 14.4 ml, 15.5 ml, and 19.2ml respectively. Peaks monitored for the complex mixtures were centred at 14.4 ml and 19.2 ml for Kindlin-3-integrin β_{1A} tail solution, 14.4 ml, 15.4 ml and 19.2 ml for Kindlin-3-talin head-integrin β_{1A} tail mixed solution, and 14.5 ml and 15.4 ml for Kindlin-3-talin head mixed solution. No additional peaks were detected within the predicted elution volume range to suggest the presence of binary or ternary complexes. The elution volumes should have been altered with the binary or ternary complexes. In the case of the relatively small β_{1A} tail (6kDa) a binary complex with Kindlin-3 should reduce the elution volume slightly; however the peaks were centred on the expected elution volume for individual proteins. Significant interactions with talin should, however, have been readily detected.

In the case of AUC (Figures 5.10B and C) a stoichiometric mixture of Kindlin-3 and the integrin β_3 cytoplasmic tail showed a marked difference in behaviour compared to Kindlin-3 alone. The Kindlin-3 peak was down-shifted with respect to its unliganded form and showed a broader range of apparent sedimentation coefficients (Figure 5.10). This indicates a

Figure 5.10 (Opposite) – Protein Interactions of Kindlin-3. (A) Interactions Analytical size exclusion chromatography (SEC) elution profiles of Kindlin-3 (red trace), talin head (blue trace), kindlin-3 with integrin β_{1A} tail (1:5 molar ratio, green trace), kindlin-3 with talin head (1:1 molar ratio, magenta trace) and kindlin-3 with talin head and integrin β_{1A} tail (1:1:5, orange trace). Peaks observed at ~8ml elution volume are in the void volume and are therefore aggregates, and peaks observed at ~24ml elution volume represent buffer components from the resuspended lyophilized talin protein. (B) c(s,f/f_0) analysis of Kindlin-3 with β_3 tail. The c(s) component is shown in open circle symbols with the predominant species fitted with a Gaussian, giving s = 3.8 +/- 0.01S and variance (σ) = 0.57 +/- 0.01S. The temperature plot shows the distribution of apparent frictional coefficients in the sample. The equivalent Gaussian for Kindlin-3 alone is shown superimposed as a grey dashed line to show the broadening in the c(s) as a result of kindlin-tail interactions. (C) As for (B) but with a combination of kindlin-3 and talin head domain. Two species are observed, with s = 3.2 +/- 0.01 S, σ = 0.20 +/- 0.01 S and s = 4.2 +/- 0.01 S and σ = 0.09 +/- 0.01 S.


A
Kindlin-3 + talin head + beta1a tail
Kindlin-3 + talin head
Talin head
Kindlin-3 + Beta1a tail
Kindlin-3
mAu (to scale)
Elution volume (ml)
B
c(s)
s (S)
c(s,f/f0)
c(s)
C
c(s)
s (S)
c(s,f/f0)
c(s)

reduction in its diffusion coefficient i.e. it has become less hydrodynamically ideal in its behaviour. This is clear from both the c(s) and the 2D c(s,f/f_0) plot. Possible causes of this include the adoption of a less compact conformation and rapid exchange with the β_3 tail. A less compact conformation would result from a trailing, bound β_3 tail; the data indicate a frictional coefficient of 1.72 (up from 1.46) and an ellipsoidal ratio of 9.8 (up from 5.75). Whilst AUC data for the interaction of Kindlin-3 with β_3 are shown (Figure 5.10) analogous observations were made for β_{1A} peptide in a separate AUC experiment. The absence of an apparent interaction for the β_{1A} integrin tail assessed by SEC but the presence of an interaction detectable by AUC is in agreement with the data suggesting a rapid exchange between the β tail and the Kindlin and relatively weak binding between them: the purely solution phase of an AUC experiment makes the detection such of rapid and weak interactions possible.

The mixture of Kindlin-3 and talin, by contrast, did not indicate an obvious interaction between these species (Figure 5.10) consistent with SEC observations. Instead, two well-defined peaks were observed in the c(s,f/f_0) spectrum: one with s = 3.2 ± 0.01 S, σ = 0.20 ± 0.01 S and the other with s = 4.2 ± 0.01 S and σ = 0.09 ± 0.01 S. The smaller species is talin head domain, indicating a frictional coefficient of 1.49 and an ellipsoidal ratio of 6.28. The larger species is clearly Kindlin-3, by comparison with Figure 5.7C. These data show that the interaction between Kindlin-3 and talin, if it occurs, is very weak or reliant on the rod domain or the modification of either binding partner, e.g phosphorylation. Note that the frictional coefficient and ellipsoidal ratio for talin demonstrate a linearly arranged protein, consistent with the X-ray crystal structure (Elliot *et al.*, 2011).

Analogously to the studies on the Kindlin-1 PH domain (Chapter 4), we investigated the lipid binding properties of Kindlin-3, making use of lipid dot-blots. Similarly to the Kindlin-1 PH domain (see Chapter 4) we found that Kindlin-3

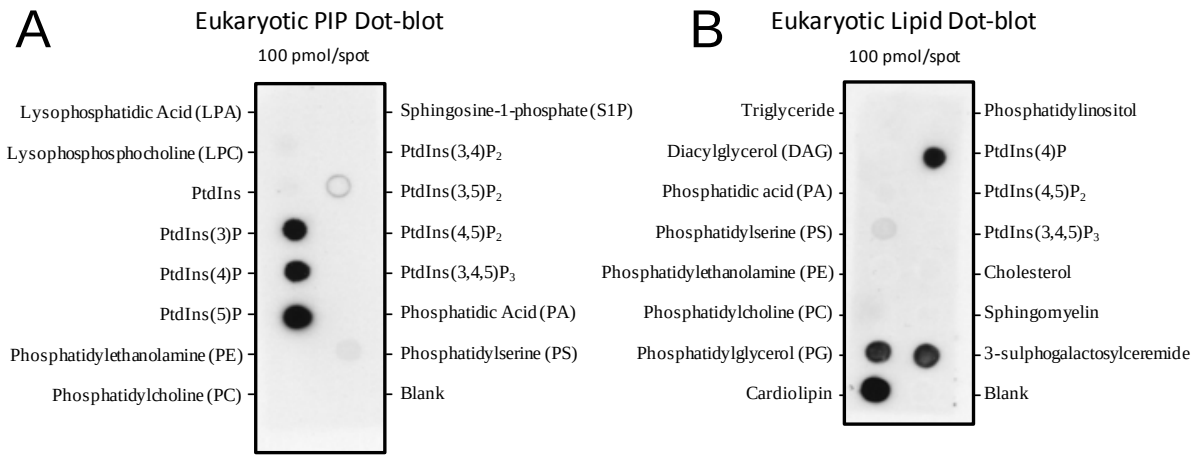


Figure 5.11 – Interaction of Kindlin-3 with eukarytoic lipids and phosphatidylinositol lipids. (A) Phosphatidylinositol phosphate dot-blot for Kindlin-3 probed with anti-pentaHis antibody showing a preference for monophosphorylated inositol moieties. (B) Lipid dot-blot covering a wider range of eukaryotic lipid types, again showing a preference for monophosphorylated inositol moieties over di- and tri-phosphate species but also strong binding to sulfatide lipids as well as cardiolipin and phosphatidyl glycerol.

bound to the monophosphorylated inositol species in a non-stereospecific manner (Figure 5.11) and also bound to phosphatidylglycerol, cardiolipin and a sulfatide. The binding of cardiolipin is not believed to be physiologically relevant as it is present in the inner mitochondrial membrane whereas Kindlin-3 is cytoplasmic and is associated with the plasma membrane. However, the binding of sulfatide is interesting as it has been suggested that the 3-sulfagalatose moiety of this lipid is a mimic of phosphotyrosine (Lingwood *et al.*, 2005) and taken together with the mono-phosphate inositol species that Kindlin-3 also binds, could suggest that Kindlin-3 is capable of binding to phosphorylated-tyrosine proteins. However this hypothesis awaits further investigation.

5.1.9. Interaction of Kindlin-3 with β integrin tails

The Kindlin family of proteins are known to bind to the cytoplasmic tail of the β-subunit of integrins (β-tails) and thus modulate their affinity. Therefore, we assessed the β-

tail binding capability of the recombinant Kindlin-3 by using peptides corresponding to the β1 integrin isoforms; β_{1A} and $\beta_{1D,}$ together with nuclear magnetic resonance (NMR) as part of a collaboration with Dr A. Fuezery and Prof. I. D. Campbell (Department of Biochemistry, University of Oxford, U.K.). Biochemical characterisation has led to several key sequence features of the β-tails that are associated with their function, including determinants that govern Kindlin binding (Figure 5.12). NMR has been previously used in this capacity to investigate the interaction between talin and various β-tails (Anthis, *et al.,* 2009; Garcia-Alvarez, *et al.,* 2003; Goult, *et al.,* 2009a; Oxley, *et al.,* 2008) but has not been used to examine the binding of β-tails by the Kindlins. Critically, these investigations were carried out mostly at pH 6.1. We found that Kindlin-3 is not particularly stable at pH 6.1, as suggested by Thermofluor and observations in solution, so the experiments were therefore carried out at the near physiological, pH 7.0, in which the protein was more stable. The ^{15}N-Heteronuclear Single-Quantum Correlations (HSQCs) of free β_{1A} and at pH 6.1 and 7.0 (Figure 5.12B) were almost identical, making ^{1}H and ^{15}N resonance assignment at pH 7.0 possible with 43 out of 45 non-prolyl residues being correctly assigned for β_{1A}.

The ^{15}N-HSQCs of free β_{1D} at pH 6.1 and 7.0 (Figure 5.12) were significantly different and an additional ^{15}N-HSQC spectrum was therefore recorded at pH 6.4, to make additional assignments. The ^{1}H and ^{15}N resonances of 43 out of 47 non-prolyl residues were thus assigned for β_{1D}. Notably, three of the four remaining unassigned residues were asparagines (Asn^{780}, Asn^{788}, and Asn^{789}). While a distinct or partially overlapping peak could be observed for each residue at pH 6.4, at pH 7.0 these peaks broadened out or, possibly, became much more overlapped with other peaks (Figures 5.12). The broadening of these exposed residues could be due to rapid ^{1}H exchange with solvent at pH 7.0 (Bai *et al.*, 1993); alternatively, it could be the result of conformational averaging on the ms-s time scale.

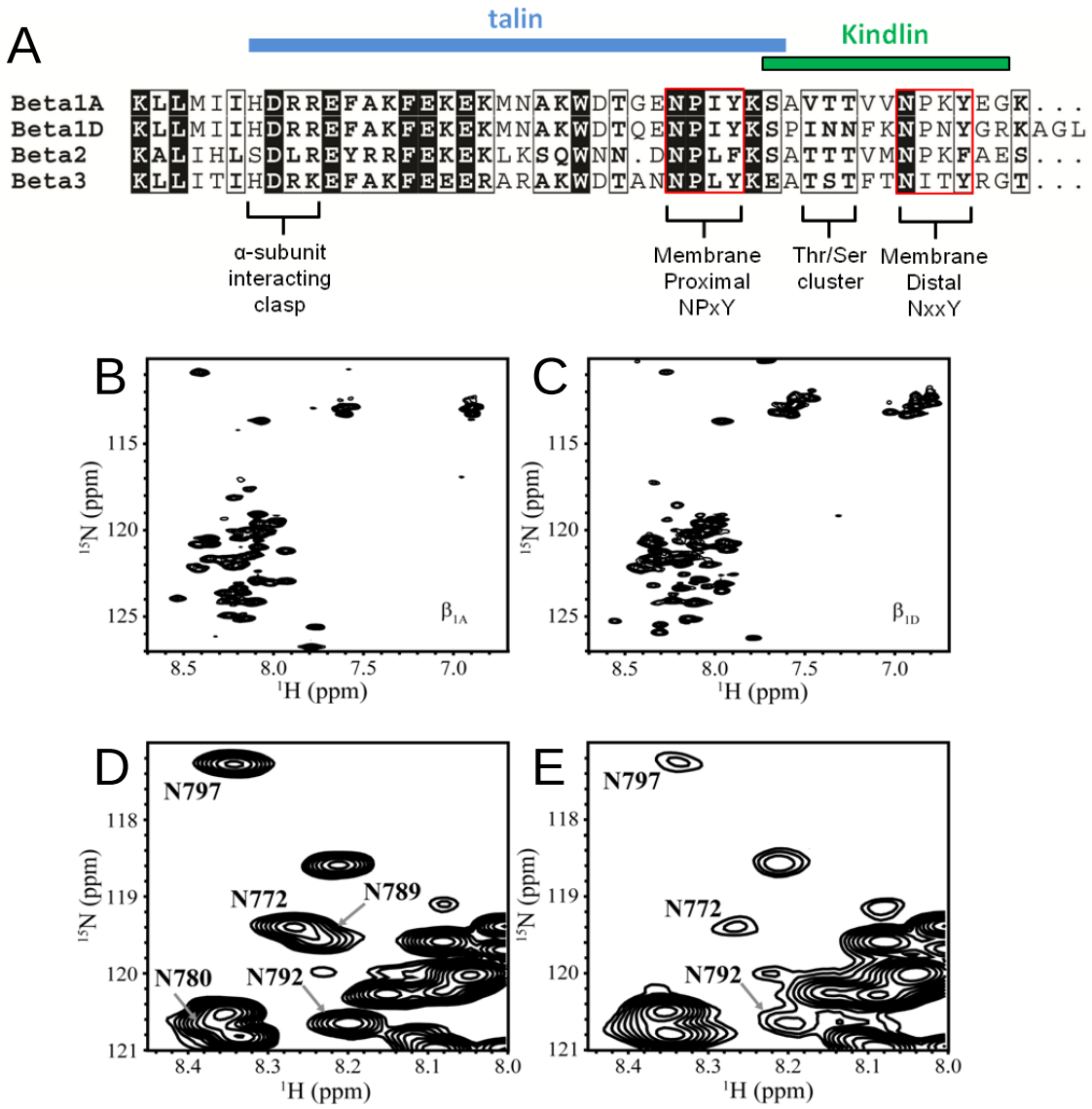
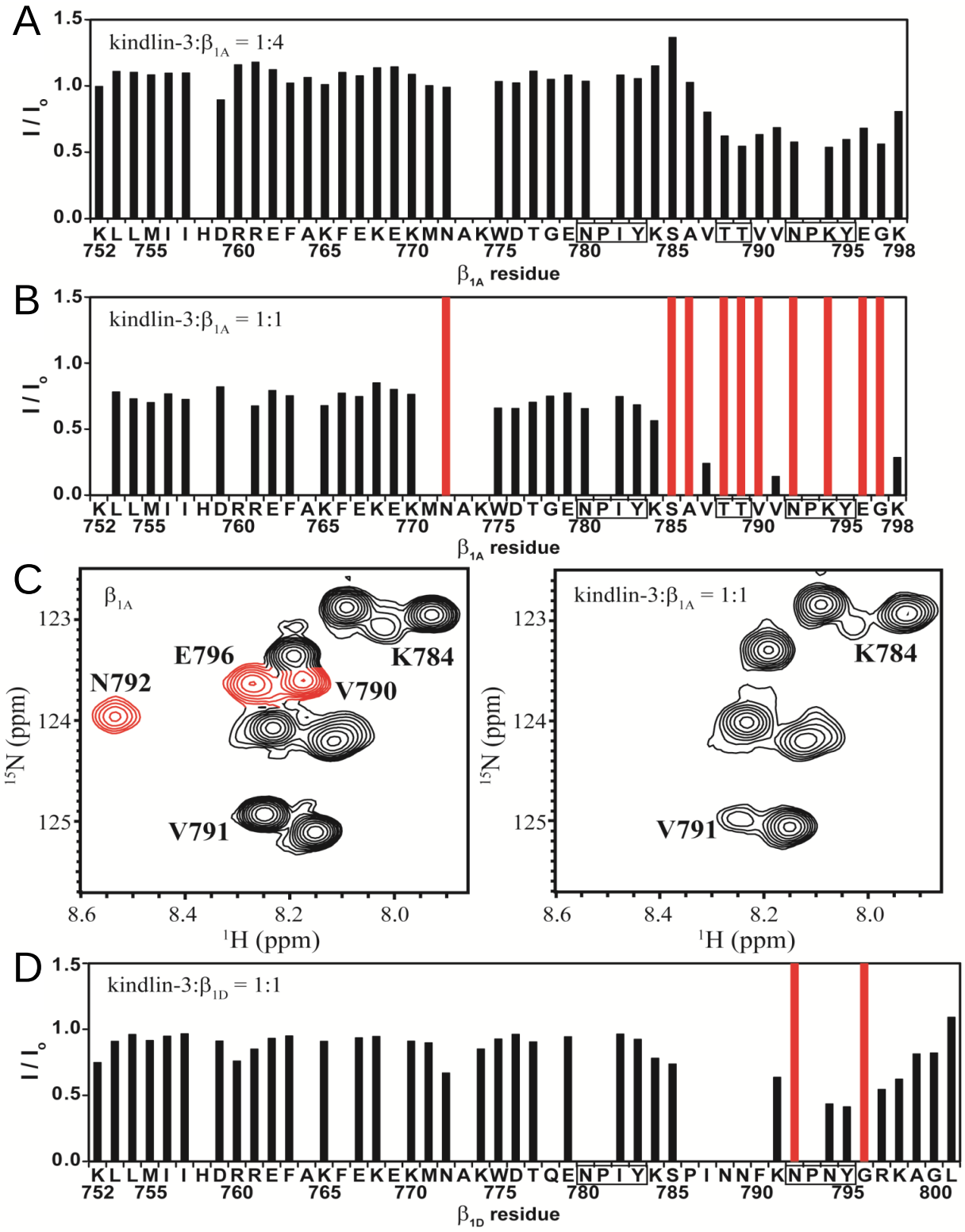


Figure 5.12 – Sequence alignment and ^{15}N-HSQCs of the free β_{1A} and β_{1D} tails. A sequence alignment of four integrin β tails that are known to be bound by Kindlin and talin-1. The conserved NPxY motifs are highlighted in red boxes and the regions of the β-tails bound by talin, according to (Anthis *et al.* 2011), and Kindlin-3, in this study, are denoted by coloured bars. (B) A section of the β_{1A} HSQC. The side chain NH peak of W775 is the only peak that falls outside of the region shown here. (C) A section of the β_{1D} HSQC. The side chain NH peak of W775 is the only peak that falls outside of the region shown here. (D) A section of the β_{1D} HSQC recorded at pH 6.4 and illustrating several Asparagine backbone NH peaks. (E) The same section of the HSQC recorded at pH 7.0. The backbone NH peaks of N780 and N789 have disappeared while those of N772, N792, and N797 have become weaker in intensity. (D-E) Peaks are labeled with their corresponding residue number and one letter amino acid code. Figure (B-E) contents were kindly provided by A. Fuezery and I. D. Campbell, Department of Biochemistry, University of Oxford, U.K.

^{15}N-HSQC spectra of [^{15}N]-labelled β_{1A} tail peptide were recorded in the presence of increasing amounts of Kindlin-3. While no chemical shift changes were observed, a progressive decrease in the intensity of peaks occurred as Kindlin-3 was added due to the size of Kindlin-3 (Figure 5.13). At a 1:4 molar ratio of Kindlin-3 to tail, peak intensities for the first 35 residues remained mostly unchanged whereas peak intensities for the membrane-distal NxxY motif and the preceding Threonine-rich region decreased by ~50% (Figure 5.13). Upon addition of more Kindlin-3 (1:1 Kindlin:tail ratio), the intensity changes became more pronounced (Figure 5.13). While the intensity of the first 33 residues decreased uniformly by ~25%, the peaks of the last 14 residues disappeared completely or decreased in intensity by more than 75%. This pattern of changes indicates that the last twelve β_{1A} residues undergo a significant environmental change, which is consistent with Kindlin-3 binding to this region of the peptide *via* the well known membrane-distal NxxY motif. Similar experiments were carried out by adding Kindlin-3 to the β_{1D} tail. This tail contains the membrane distal NxxY motif but has two asparagines in place of the two threonines on the β_{1A} tail (Figure 5.13). A previous study has demonstrated the importance of this Serine/Threonine rich region for Kindlin-tail interactions (Harburger *et al.*, 2009). Therefore in line with these data we proposed that the interaction between Kindlin-3 and a β-tail isoform lacking this Serine/Threonine rich region (i.e. β_{1A}) would be weaker than for β-tails that contain such residues. Consistent with our assumptions the results indicated this to be the case. At a 1:1 molar ratio of Kindlin to β_{1D}, most peaks of the membrane distal NPxY motif and its immediate vicinity decreased in intensity by only ~50-75% and were still visible in the HSQC (Figure 5.13). The peak intensities of the remaining β_{1D} residues were approximately the same as in the free form of the tail. Taken together with data from the β_{1A}, these results demonstrate an observable and specific change in the chemical environment of the second NPxY motif and are consistent with a specific interaction between Kindlin-3 and

Figure 5.13 (Opposite) - Intensity changes in the ^{15}N-HSQC peaks of β-tails brought about by the presence of Kindlin-3. (A) β_{1A} tails in the presence of Kindlin-3 at a molar ratio of 4:1 (β_{1A}: Kindlin-3) the peak intensities for the membrane distal NxxY motif and the preceding threonine-rich region of β_{1A} decreased by ~50%. (B) β_{1A} tails in the presence Kindlin-3 at a molar ratio of 1:1. The peak intensities of the first 33 residues of β_{1A} decreased uniformly by ~25% while the peak intensities of the last 14 residues disappeared completely (red bars) or decreased in intensity by more than 75%. (C) Close-ups of a region of the β_{1A} HSQC illustrating peak intensity changes brought about by the presence of Kindlin-3. The left panel shows free β_{1A} while the right panel shows β_{1A} in the presence of a molar equivalent of Kindlin-3. Peaks experiencing large reductions in intensity are labeled with their corresponding residue number and one letter amino acid code. In addition, peaks that disappear completely are colored red. (D) β_{1D} in the presence of Kindlin-3 at a molar ratio of 1:1. The peak intensities of most residues in the membrane distal NxxY motif and its immediate vicinity decreased by only ~50-75%. For all three panels, Kindlin-3-induced changes in amide peak intensities of the β tails (I/I_o) are reported as the peak height in the presence of Kindlin-3 (I) divided by the height of the same peak in the absence of Kindlin-3 (I_o). Several peaks corresponding to Asparagine residues are broad and difficult to detect at pH 7.0. Figure contents were kindly provided by A. Fuzery and I. D. Campbell, Department of Biochemistry, University of Oxford, U.K.


A
kindlin-3:β1A = 1:4
I / Io
1.5
1.0
0.5
0.0
K L L M I I H D R R E F A K F E K E K M N A K W D T G E N P I Y K S A V T T V V N P K Y E G K
752 755 760 765 770 775 780 785 790 795 798
β1A residue
B
kindlin-3:β1A = 1:1
I / Io
1.5
1.0
0.5
0.0
K L L M I I H D R R E F A K F E K E K M N A K W D T G E N P I Y K S A V T T V V N P K Y E G K
752 755 760 765 770 775 780 785 790 795 798
β1A residue
C
β1A
123
124
125
15N (ppm)
E796
K784
N792
V790
V791
8.6
8.4
8.2
8.0
1H (ppm)
kindlin-3:β1A = 1:1
123
124
125
15N (ppm)
K784
V791
8.6
8.4
8.2
8.0
1H (ppm)
D
kindlin-3:β1D = 1:1
I / Io
1.5
1.0
0.5
0.0
K L L M I I H D R R E F A K F E K E K M N A K W D T Q E N P I Y K S P I N N F K N P N Y G R K A G L
752 755 760 765 770 775 780 785 790 795 800
β1D residue

the membrane distal NPxY motif of β_{1A} and β_{1D}. Furthermore, these data are also consistent with the previous finding that the Serine/Threonine-rich region upstream of the membrane-distal NPxY motif is important for modulating the strength of the interaction (Harburger *et al.*, 2009) Overall, the NMR experiments described herein strongly suggest that our Kindlin-3 preparations are capable of binding in the expected manner and provide the first directly observed interactions.

5.2. Discussion

The Kindlin family of proteins has been the focus of much attention since their unexpected role as essential co-activators of integrins *in vivo* was discovered. This has triggered much effort to express them recombinantly and resolve their structures. To date little success has been reported but we have here described how the use of a baculovirus system permits large scale expression at levels where structural studies become feasible. While attempts to crystallise Kindlin-3 are ongoing, we have here described how a series of accessory techniques can shed much light on how Kindlins act in the regulation of cell adhesion.

It was previously envisaged that one possible explanation for the difficulties found in expressing recombinant fragments of Kindlins is that their domain arrangement is compact and similar to the cloverleaf arrangement found in most FERM domains (e.g. radixin) rather than the extended FERM domain of talin (Elliott *et al.*, 2010). Through the use of biophysical techniques, we have shown here, however, that Kindlin-3 is most likely an elongated structure (Figure 5.8). A consensus reconstitution of the protein envelope from the SAXS data clearly shows its elongated conformation. Various possible arrangements of the FERM F0-F1 and F2-F3 subdomains were tested by simulating scattering data and comparing them to the experimental findings. The best of the models constructed involved F0-F1 and F2-F3

forming the backbone of the protein, with the PH domain above the F2-F3 interface and the loop assumed to be in the region also occupied by the F0-F1 domain pair (Figure 5.9).

By investigating the ligand-binding properties of Kindlins we have shown that if a direct interaction with the talin head exists, it is very weak or requires some other factor to enhance it, such as the plasma membrane, phosphorylation, or an additional protein. We cannot discount an interaction between the rod domain of talin and Kindlin proteins at the plasma membrane during integrin activation as our investigation made use of the talin head domain alone. We have also found evidence for a rapid binding and release of integrin β tails by Kindlin and that the complex is more elongated than the protein on its own, causing a downshift in sedimentation coefficient. This is in agreement with the binding of the tails by the F3 subdomain at one end of the Kindlin protein.

We have previously shown that the Kindlin-1 PH domain is important for proper integrin activation (see Chapter 4), and this has also been shown for Kindlin-2 (Qu *et al.*, 2011; Liu *et al.*, 2011). We have shown here that Kindlin-3 binds monophosphorylated inositides only (Figure 5.11). This observation is consistent with that observed for the Kindlin-1 PH domain and suggests that this lipid binding property of Kindlin-3 is mediated *via* its PH domain. This observation of lipid preference is, however, puzzling. It has been shown that both talin and vinculin, but not Kindlins, are recruited to focal adhesions at the plasma membrane in a phosphatidylinositol 4,5-phosphate-dependent manner, specifically by the local generation of this lipid by phosphatidylinositol phosphate kinase (PIPK) - γ1 (Legate *et al.*, 2011). Given that full-length Kindlin-3 does not bind to this focal adhesion-associated lipid we suggest that its lipid binding may indicate another ligand, such as phosphotyrosine, rather than inositol phosphate. In the context of focal adhesions (FA), this may suggest, either interactions between Kindlin and a tyrosine phosphorylated component of the FA complex, or that Kindlin interacts with the β tail itself in a manner involving PH and

F3 domains. Our data also show that although Kindlin-2 is known to bind to inositol triphosphates (Liu *et al.*, 2011), Kindlin-3 does not and in this matches Kindlin-1 (see Chapter 4). Thus, the inositol phosphate binding region is not a conserved feature of the Kindlin PH domain but varies between isoforms rather significantly.

An NMR spectroscopy study of Kindlin-β-tail peptide interactions has allowed us to observe and map the residues in the β-integrin cytoplasmic tail that Kindlin-3 interacts with. Thus Kindlin-3 is observed to interact with the second 'membrane-distal' NPxY motif and upstream Serine/Threonine rich region and not the first 'membrane proximal' NPxY motif that talin recognizes. Although previous reports using mutagenesis suggest interactions of the sort demonstrated here, we show, for the first time, a direct interaction with full-length protein at a physiological pH and show that the Serine/Threonine cluster adjacent to the membrane-distal NPxY is important for binding. This is consistent with cellular, biochemical and clinical observations (Harburger and Calderwood, 2009; Moser *et al.*, 2008) that suggest that the Serines and Threonines in this cluster are critical for efficient membrane distal NPxY motif binding by the Kindlins. This NMR study also demonstrates that Kindlin-3 does not significantly interact with the membrane-proximal NPxY – the talin binding site. It is thus possible that Kindlin and talin could bind to the integrin tail at the same time: consulting the crystal structure of talin F2-F3 and a β_1 tail peptide (Anthis *et al.*, 2009) we find that the membrane-distal region containing the second NPxY motif hangs out into the solvent, pointing away from the talin:integrin binding site. This notion is further supported by our AUC data, where the binding of the β-tail peptide at a single site (the second membrane distal NPxY and Serine/Threonine region) would result in the remainder of the polypeptide extending into the solvent. This notion is however supported by the Kindlin-3 sedimentation coefficient downshift observed on both β_{1A} and β_3 tail binding. However, further experimental investigation will clearly be required to resolve this question.

Clearly, the Kindlin family of proteins has been, and will continue to be a focus of attention due to their critical role in integrin activation and focal adhesion assembly. By generating sufficient quantities of recombinant Kindlin-3, using a baculovirus expression system, we have made their first biophysical characterisation, and provided the first structural insight into full-length Kindlin conformation and interaction with integrins, lipids and other proteins.

CHAPTER 6

6.1. Caffeine-Induced Death Suppressor Protein-1 (Cid1) is a Polyuridylyl Polymerase Involved in S-M Checkpoint Control

6.1.1. A Brief History of RNA Uridylation

The first reports of template-independent homopolyribonucleotide addition to RNA were described over 50 years ago, when Edmonds and Abrams (1960) isolated an enzyme from calf thymus that was capable of polymerising adenines from an ATP substrate. The enzyme responsible for the polyadenine synthesis is the poly(A) polymerase (PAP), for which there are crystal structures from yeast (*Saccharomyces cerevisiae*) (Bard *et al.,* 2000) and cow (*Bos tauris*) (Martin *et al.,* 2000). It is now widely accepted that adenines are added to the 3' end of mRNAs in both prokaryotes and eukaryotes and that this poly(A) tail in eukaryotes is important for multiple functions, including promoting translation and transcript stability (see Proudfoot *et al.,* 2002, for instance).

Much like the discovery of adenylated RNAs in the 1960s, evidence for the addition of uridines to the 3' end of RNA in cytoplasmic extracts was observed within the same decade (Wykes and Smellie, 1966) - long before insights into its functional implication in biology. UMP incorporation into RNA was observed even as early as 1957 in rat liver microsome preparations (Canellakis, 1957). Throughout the 1960s, fractions from homogenized rat liver were studied to accrue evidence for cytoplasmic PUPs (Wilkie and Smellie, 1968a; Wilkie and Smellie, 1968b). The microsome fraction of the rat liver homogenate was found to possess strong PUP activity, thus supporting the notion of cytoplasmic PUP enzymes in cells (Wilkie and Smellie, 1968a). Subsequent work suggested the association of cytoplasmic poly(U) polymerases and, interestingly, poly(A) polymerases

with fractions containing cytoplasmic ribosomes (Milchev and Hadjiolov 1978). Studies in plants have also revealed enzymes with PUP activity (Brishammar and Junti, 1975; Zabel *et al.*, 1981). A PUP was identified in tobacco (*Nicotia tobacciana*) leaves and biochemical characterisation revealed a poly(U) activity that was highly specific for UTP (Brishammar and Junti, 1975). Additionally, Zabel and colleagues (1981) identified a 58kDa polypeptide capable of transferring a single UMP residue to the 3'end of RNA, termed a terminal uridine transferase (TUTase), from the membrane fraction of cowpea (*Vigna unguiculata*) leaves.

Terminal uridine transfer to RNA was also observed and suggested to be essential during viral replication. Studies into poliovirus replication suggested a host TUTase was responsible for replication initiation (Andrews *et al.*, 1985, Andrews *et al.*, 1986, Dasgupta *et al.*, 1980). In an *in vitro* poliovirus replication system a 'host factor' was necessary for transcription of the viral genome, together with the viral RNA-dependent RNA polymerase p63 (Andrews *et al.*, 1985). Partial isolation of this 'host factor' was achieved from rabbit reticulocyte lysate, which also contained translation initiation factor eIF2 (Andrews *et al.*, 1985). Further studies using human (HeLa) cells allowed the isolation of a 68kDa cytoplasmic TUTase that adds uridines to the 3' end of the viral RNA (Andrews *et al.*, 1986), although the function of this modification remains controversial (Aphasizhev, 2005).

6.1.2. Known Poly(U) Polymerases

As discussed, the notion of poly(U) polymerases was described over 50 years ago, with the first observation of uridinylated RNA in the 1950s (Canellakis, 1957). To date, there are several characterised PUPs from the mitochondria of the trypanosomal protists (see below) however more recently, substantial progress has been made in ascribing this PUP activity to several cytoplasmic proteins in metazoans. Uridylation of RNA and its associated

enzymes has been brought into the focus of attention, particularly given their new found significance in RNA and miRNA metabolism (Rissland and Norbury, 2009; Heo *et al.,* 2009; Lehrbach *et al.,* 2009; Shen and Goodman, 2004), stem cell biology and cancer (Hagan *et al.,* 2009).

Nearly all of our structural knowledge of terminal uridylyl transferases comes from RNA editing within the mitochondria of the protozoan *Trypanosoma brucei* (Aphasizhev and Aphasizheva, 2007). The processes of RNA editing and the protein complexes involved are well-understood. This process is expertly reviewed in (Aphasizhev and Aphasizheva, 2007) and (Simpson *et al.,* 2003). However, for contextual purposes a brief description of editing is given below. The mitochondrial DNA of kinetoplastid protozoans is a concatenated network of a few maxicircles and thousands of minicircles (Aphasizhev and Aphasizheva, 2007). The maxicircles are transcribed as polycistronic RNAs, encoding ribosomal RNA and protein encoding pre-mRNAs, that are nucleolytically cleaved and polyadenylated and subsequently edited by the insertion/deletion of uridine residues (Aphasizhev and Aphasizheva, 2007; Simpson *et al.,* 2003). The post-transcriptional editing *via* U insertion/deletion acts to correct RNA frameshift, creation of start/stop codons or mRNA coding sequences (Martin and Keller, 2007). In trypanosomes, the RNA editing is directed by guide RNAs (gRNA) that are 5' unmodified, but 3'polyuridinylated (~15nt) RNA molecules. gRNAs anneal with the pre-mRNA and regulate endonucleolytic cleavage at a non-base-paired site of the duplex and subsequently direct U-insertion or deletion (Aphasizhev and Aphasizheva, 2007; Martin and Keller, 2007; Simpson *et al.,* 2003). The isolation of multi-protein complexes involved in U-insertion and deletion identified RET1 (RNA Editing TUTase 1) (Aphasezhev *et al.,* 2002), involved in 3'polyuridine addition to all classes of mitochondrial RNAs (Aphasezhev *et al.,* 2003b). However, gRNAs and rRNAs, receive an ~20nt poly(U) 3'tail, whereas mRNAs receive a heteropolymeric (A/U) 3'tail catalysed by RET1 and the poly(A) polymerase,

KPAP1 (Aphasizhev and Aphasizheva, 2007). The second uridylation enzyme, RET-2 (RNA Editing TUTase 2), is involved in uridine insertion/deletion RNA editing as part of the so-called 20S editosome multiprotein complex (Aphasezhev *et al.,* 2003). Thus RET1 successively incorporates uridine residues to the 3'end of the RNA molecule creating a poly(U) tract, whereas RET2 catalyses the addition of a single uridine molecule to the 3'end of the RNA. Recently, a third TUTase MEAT1 (minor editosome associated TUTase1) has been identified and has also been studied crystallographically (Stagno *et al.,* 2010). Furthermore, several non-mitochondrial TUTases, TUTase3 and TUTase4 have been described in *Trypanosoma brucei* (Stagno *et al.*, 2007b, Aphasizhev *et al.*, 2004). To date, several crystal structures of mitochondrial and cytoplasmic TUTases have been solved thus providing a molecular basis of uridine selectivity, over other nucleotides, and insight into uridine addition (Stagno *et al.,* 2010, Stagno *et al.,* 2007b, Deng *et al.,* 2005) which is discussed in depth in Chapter 7.

The structures of the trypanosome mitochondrial TUTases, RET2 and MEAT1, and the trypanosome cytoplasmic TUTase, TUT4, have all been determined at resolutions better than 2.0Å (Stagno *et al.*, 2010, Stagno *et al.*, 2007a, Stagno *et al.*, 2007b, Deng *et al.,* 2005) and in one case in multiple ligand bound states (Stagno *et al.,* 2007). The overall structure of all enzymes display a bi-domain catalytic module at its core that, as with other members of the polβ-like nulceotidytransferase family, shares structural homology with the catalytic domain and central domain of eukaryotic poly(A) polymerase (PAP). RET2 TUTase possesses an additional domain, denoted as the middle domain (MD), inserted into a loop of the catalytic domain, an organisation previously unseen in the nucleotidyltransferase family (Deng *et al.,* 2005). Although the N-terminal catalytic domain and C-terminal nucleotide recognition domains are structurally conserved in these TUTases the presence of an auxiliary domain usually correlates with divergent biological functions (Ringpis *et al.,* 2010). Indeed,

mutagenesis and biochemical studies of RET2 have suggested that the middle domain interacts with the mRNA:gRNA duplex during RNA editing, but also provides a protein-interaction module for an editosome protein KREPA1 (also known as *Tb*MP81) important for enhancing RET2's enzymatic activity (Ringpis *et al.*, 2010). Given that U insertion/deletion RNA editing is unique to kinetoplastid mitochondria and is not found in other organisms, unlike 3'uridylation which is now considered widespread in metazoans (Martin and Keller, 2007), it would make evolutionary sense that the TUTase RET2 possesses an additional domain specific to its function.

6.1.3. The Discovery of the Cid1 Polymerase

Caffeine induced death suppressor protein-1 (Cid1), from *Schizzosaccharomyces pombe*, is a member of the non-canonical poly(A) polymerase family (Read *et al.*, 2002, Stevenson and Norbury, 2006). Cell biological and biochemical characterisation of the protein surprisingly revealed an unusual polyuridinylating activity, modifying the 3' end of mRNA by the addition of uridine monophosphates, thus creating a homopolymeric tract (Rissland and Norbury, 2009, Rissland and Norbury, 2008, Rissland *et al.*, 2007). Despite the role of Cid1 in 3'modification of mRNA, the protein was first identified during a study into cell-cycle checkpoint control in fission yeast, particularly the S – M checkpoint (Wang *et al.*, 1999). The S – M checkpoint control regulates cell cycle progression into mitosis, with entry into mitosis prevented in event of incomplete genome replication (Rissland and Norbury, 2008). This checkpoint is coupled, as expected, to chromosomal DNA replication initiation. The deletion of the catalytic subunit of yeast DNA polymerase α, involved in initial primer-dependent DNA replication, results in loss of S – M checkpoint function (D'Urso *et al.*, 1995). Interestingly, yeast cells possessing temperature sensitive (*ts*) alleles encoding

subunits of the DNA replication elongation polymerases, Pol ε or Pol δ, activate the S – M checkpoint when shifted to restrictive temperatures (Rissland and Norbury, 2008). Replication stress activates the S – M checkpoint and can be induced exogenously in fission yeast *via* the addition of the ribonucleotide reductase inhibitor, hydroxyurea (HU), reducing the dNTP pool and inhibiting S (Synthesis) phase progression (Rissland and Norbury, 2008).

Activation of the S – M checkpoint, by hydroxyurea for example, can be over-ridden by the addition of caffeine; although a variety of chemical agents that over-ride checkpoint control have been identified (Wang *et al.,* 1999). Thus fission yeast cells are unviable after the concomitant treatment with hydroxyurea (HU) and caffeine, though the cells are mildly sensitive to treatment with each drug alone (Wang *et al.,* 1999). Fission yeast cells that have their S – M checkpoint over-ridden display a "cut" (cell untimely torn) phenotype, that results from septation in the absence of normal chromosomal segregation (Wang *et al.,* 1999, Enoch and Nurse, 1990). The methylxanthine, caffeine inhibits the kinase activity of Rad3 (ATR in humans) (Moser *et al.,* 2000) thus preventing the activation of the signalling kinase cascade as a result of HU-induced replication stress. Rad3 inhibition ultimately results in the dephosphorylation and subsequent activation of Cdc2 allowing the premature entry into mitosis. Over-expression of the upstream kinases Chk1 or Cds1 can rescue the cells from HU and caffeine treatment (Wang *et al.,* 1999). Thus S – M checkpoint control downstream effectors could be identified by their ability to rescue HU and caffeine treatment in fission yeast. Use of such a screening method identified the protein caffeine-induced death suppressor-1 (Cid1) (Wang *et al.,* 1999).

The role of Cid1 in S – M checkpoint control was further investigated using genetic analysis (Wang *et al.,* 2000). Cid1 deletion mutants (*Cid1*Δ) are sensitive to combinatorial treatment with hydroxyurea and caffeine at concentrations ordinarily tolerated by wild-type cells, but display wild-type tolerance to either drug alone (Wang *et al.,* 2000). Furthermore,

Cid1 over-expression in checkpoint *rad* deletion mutants (*Rad3*Δ, *Rad9*Δ and *Rad17*Δ) partially suppresses their HU sensitivity, with similar observations reported for *Rad1*Δ, *Rad26*Δ and *Hus1*Δ (Wang *et al.,* 2000). What is more, Cid1 deletion mutants (*Cid1*Δ), whilst viable and not unusually sensitive to HU treatment, display an accelerated loss of viability in *ts* Pol ε or Pol δ mutants after shifting to the restrictive temperatures resulting from S-phase arrest, suggesting that S – M checkpoint activation is Cid1 dependent (Wang *et al.,* 2000).

The amino acid sequence of Cid1, surprisingly, suggested that this enzyme belongs to the DNA Pol β-like nucleotidyltransferase family due to significant sequence similarity to poly(A) polymerase and the possession of a putative palm domain similar to that found in the rat DNA polymerase β (Wang *et al.,* 2000). A loss-of-function is observed with the mutation of two aspartic acid residues in Cid1 that correspond to catalytic residues in DNA Pol β, suggesting a similar catalytic mechanism and function (Wang *et al.,* 2000). Due to the critical role of Cid1 in S – M checkpoint integrity when DNA Pol ε or Pol δ is inhibited, it was suggested that Cid1 is a DNA polymerase subunit (Wang *et al.,* 2000), although this function was discounted on further investigation (Read *et al.,* 2002). Further characterisation of recombinant Cid1 revealed an *in vitro* ATPase activity that was diminished with the mutation of two of the catalytic aspartates residues previously determined as critical for Cid1-mediated S – M checkpoint function (Read *et al.,* 2002, Wang *et al.,* 2000). Cid1 was found to be constitutively cytoplasmic and recombinant enzyme was shown to possess polyadenylation activity (Read *et al.,* 2002). Thus Cid1 was suggested to play a role in mRNA metabolism in the cytoplasm.

6.1.4. Cid1 is a Poly(U) Polymerase Involved in mRNA Decay

As discussed, Cid1 was demonstrated to be localised to the cytoplasm and shown to possess poly(A) polymerase activity suggesting an involvement in mRNA metabolism and checkpoint control (Read *et al.,* 2002). In addition to the nuclear poly(A) polymerases involved in pre-RNA maturation, cytoplasmic poly(A) polymerases have been described in metazoans, including *C. elegans* (Suh *et al.*, 2006; Wang, Eckmann *et al.*, 2002), *S. pombe* (Saitoh *et al.*, 2002) and mammals (Nakanishi *et al.*, 2006, Kwak *et al.*, 2004). The role of these cytoplasmic poly(A) polymerases is to regulate the length of the mRNA 3'poly(A) tail and therefore transcript stability (Barnard *et al.*, 2004). One example of these modifications in mammals is the polyadenylation of deadenylated mRNAs by GLD-2, resulting in translation activation of latent transcripts in germ line cell development (Nakanishi *et al.*, 2006). Surprisingly, despite bacterially expressed recombinant Cid1 possessing poly(A) polymerase (PAP) activity (Read *et al.*, 2002), further biochemical characterisation of the enzyme revealed that recombinant Cid1 displayed additionally strong poly(U) polymerase (PUP) activity but weak activity with CTP or GTP (Rissland *et al.*, 2007). Moreover, Cid1 displayed selectivity for UTP even in the presence of ATP or all four nucleotides – even when the ATP concentration was in 10-fold molar excess, Cid1 still incorporated UMP (Rissland *et al.,* 2007). Further support that the *in vitro* PUP activity observed was the physiological activity of the enzyme *in vivo* came from the biochemical characterisation of the native enzyme. Thus, tandem affinity purified (TAP) Cid1 from yeast displayed robust PUP activity and weak poly(C) polymerase (PCP) and poly(G) polymerase (PGP) activity, consistent with that observed for the recombinant enzyme (Rissland *et al.,* 2007). Interestingly, TAP-Cid1 was isolated as a Cid1 complex that, as determined by mass spectrometry, demonstrated the presence of poly(A) binding protein, Pab1, as well as

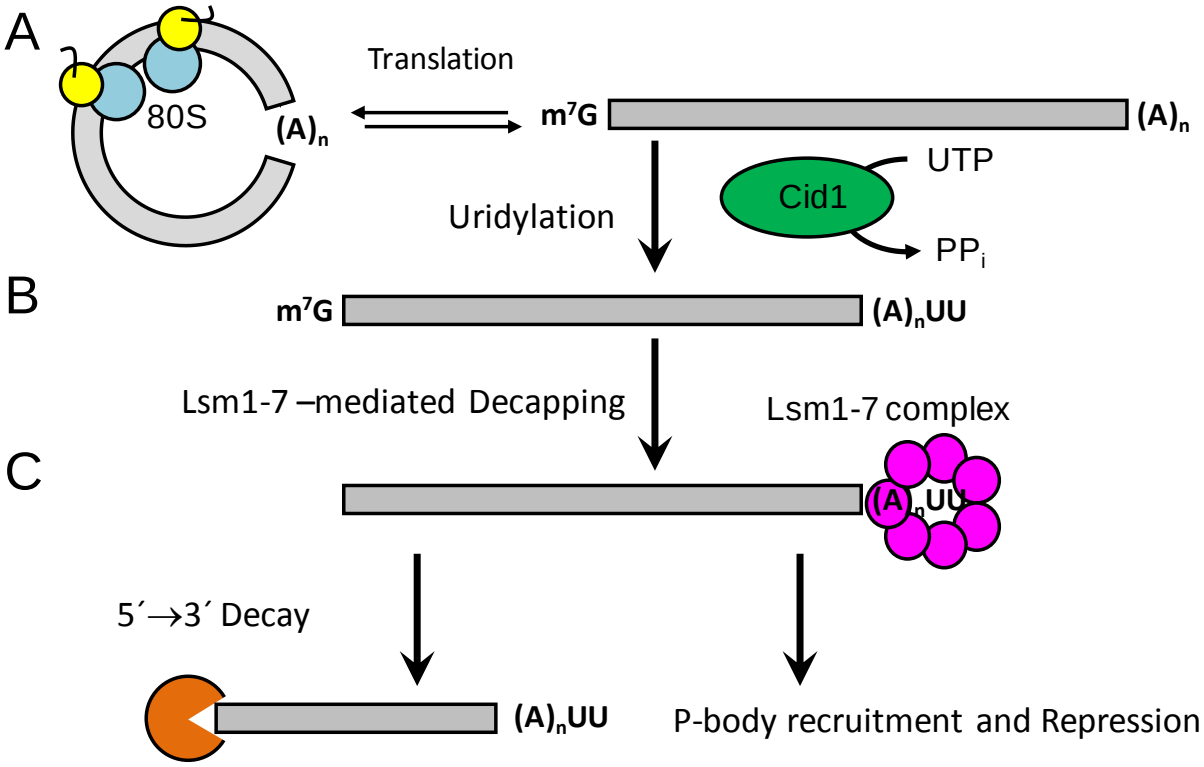


Figure 6.1 – Model of Uridylation–mediated mRNA Turnover in *S. pombe.* (A) Mature messenger RNA (mRNA), which possesses a 5' m^7G (7-methylguanosine) cap and a 3' poly(A) tail, is competent to be translated by the 80S ribosome. mRNAs can undergo deadenylation independent decapping followed by 5'→3' exonucleolytic decay, or deadenylation followed by Lsm1-7-mediated decapping and degradation (Rissland and Norbury, 2009). (B) Capped transcripts are uridylated in a Cid1-dependent manner, which catalyses the addition of UMP, from UTP, to the 3' end of mRNA and releases inorganic pyrophosphate (PP_i). (C) Uridylation of mRNA recruits Lsm1-7 complex and promotes Lsm1-7-mediated decapping and as a result 5'→3' exonucleolytic decay. Alternatively, instead on decapping, Lsm1-7 association localises these mRNAs to Processing bodies (P-bodies) where they are silenced (O.S. Rissland, unpublished)

additional RNA binding components (Rissland and Norbury, 2007 unpublished observations). Isolated Cid1 from *S. pombe*, unlike bacterially expressed recombinant Cid1, displays weak PAP activity *in vitro* with the PAP activity outcompeted by the enzyme's PUP activity at physiological ratios of nucleotides (Rissland *et al.*, 2007). Final proof that Cid1 was a physiological PUP *in vivo* came from direct sequencing of mRNA during S-phase

arrest that demonstrated Cid1-dependent post-transcriptional 3' uridylation of polyadenylated actin mRNA (Rissland *et al.,* 2007). Capturing of the 5' and 3' ends of mRNA required the technique circularized rapid amplification of cDNA ends (cRACE) (Rissland and Norbury, 2009, Rissland *et al.,* 2007). Briefly, the mature adenylated and uridylated mRNAs are treated with Shrimp Alkaline Phosphatase (SAP), which inhibits RNA ligase dependent circularisation of uncapped transcripts undergoing 5'-3' decay. The capped transcripts are treated with Tabacco Acid Pyrophosphatase (TAP) which allows their RNA ligase circularisation. The circularised RNA is subject to sequencing after reverse transcription. Further studies into Cid1-dependent uridylation of adenylated mRNAs were carried out using cRACE by Rissland and Norbury (2009), whereby the investigators revealed that Cid1-dependent uridylation of polyadenylated mRNAs stimulated their decapping *via* the recruitment of the RNA degradation Lsm1-7 complex. Human cells possess seven Cid-1 family polymerases with proposed functions analogous to their fission yeast orthologues, with the Cid1-like human enzyme ZCCHC6 (also known as Hs2) demonstrated to possess Cid1 PUP activity (Rissland *et al.,* 2007). Therefore, uridylation-mediated mRNA decay is suggested to be a more widespread and conserved mechanism for cytoplasmic RNA metabolism (Rissland and Norbury, 2009).

6.2. Expression, Purification, Crystallisation and X-ray Diffraction Analysis of Cid1

6.2.1. Summary

Recently, the *Schizosaccharomyces pombe* cytoplasmic polymerase Cid1, a protein involved in S – M checkpoint control, was described as an efficient and robust 3'-

polyuridylation enzyme (Rissland *et al.,* 2007) that regulates RNA decay through a novel exonucleolytic pathway of RNA metabolism (Rissland and Norbury, 2009). The *S. pombe* polyuridylation enzyme, Cid1, was expressed and purified as a GST-fusion protein using a bacterial expression system. Unexpectedly, the enzyme displayed nucleic acid binding capacity as determined by the co-purification of RNAse-A and DNAse-1 sensitive nucleic acids. The contaminating nucleic acids were removed using chromatographic techniques in order to obtain homogenous protein solution for structural studies. The purified recombinant protein appeared to be monomeric and displayed efficient polyuridylation activity *in vitro.* The enzyme was crystallised in two space groups (C2 and $P2_12_12_1$) in highly similar conditions, in the presence of a 5-mer of DNA and its substrate UTP, with magnesium chloride.

6.2.2. Expression and Purification Optimisation of tCid1

The Caffeine induced death suppressor protein-1 (Cid1) was cloned into a pGEX-6P vector, which encodes a glutathione S-transferase (GST) fusion tag, and was a kind gift from O. Rissland, Sir William Dunn School of Pathology, University of Oxford, UK. The fusion protein was expressed as described in (Rissland *et al.,* 2007) using the *E. coli* strain BL21 (DE3) pLysS. Recombinant protein had been previously shown to be susceptible to degradation (O. Rissland, personal communication) and therefore the expression construct was a truncation mutant lacking the first 31 amino acids and as such is denoted as tCid1.

Bacterially expressed recombinant polymerase tCid1 was purified from the soluble fraction of lysate using methods described (Rissland *et al.,* 2007). However, whilst this method allowed sufficient quantities for the biochemical characterisation of this enzyme it was insufficient for structural studies. Therefore the expression and purification were

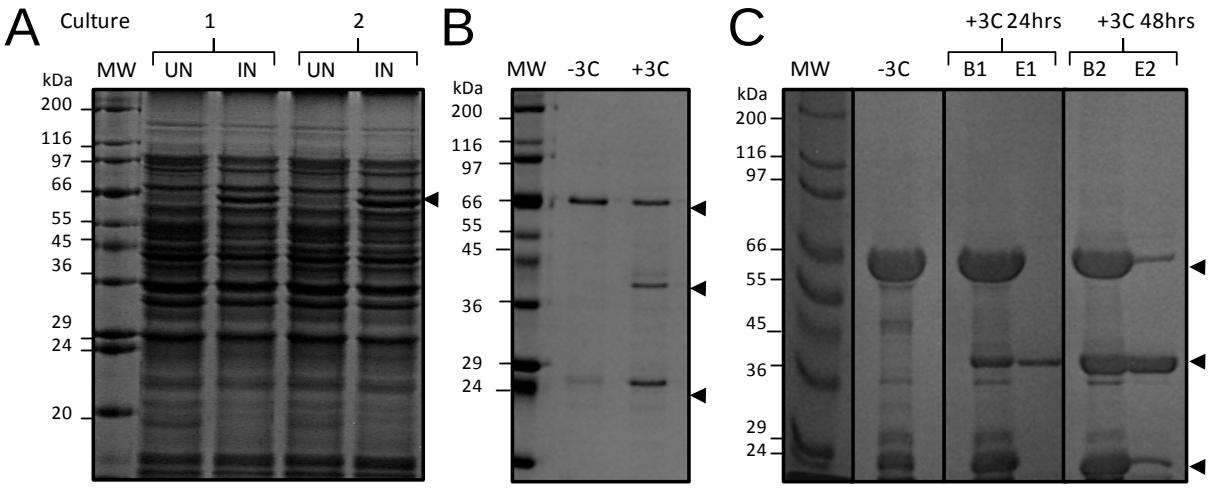


Figure 6.2 – SDS-PAGE Analysis of Expression of tCid1 in BL21 (DE3) pLysS and Purification (A) Representative expression of an ~66kDa protein (tCid1-GST fusion protein, arrowed) in the presence (lane labelled IN) or absence (lane labelled UN) of IPTG (1mM final) of two colonies. (B) Initial purification of tCid1-GST fusion protein from 2 litres of *E. coli* using glutathione sepharose (GE Healthcare) affinity chromatography and subsequent cleavage of GST by 3C protease digestion, yielding tCid1 (42kDa), GST (26kDa) with fusion protein remaining (66kDa) (arrowed). (C) Composite SDS-PAGE of purified tCid1-GST fusion protein bound to glutathione sepharose from 2 Litres of *E .coli* after incubation of the lysate supernatent with glutathione sepharose overnight at 4°C. On-column cleavage of the fusion protein was performed using 3C protease at 4°C. Samples of material bound to beads (labelled, B) were taken along with the corresponding eluant (labelled, E) to assess the extent of cleavage and quantity of liberated enzyme (arrowed) and contamination by GST (arrowed) at 24 hours (lanes labelled B1 and E1) and 48 hours (lanes labelled B2 and E2).

optimised for maximal yield using two modifications. Firstly, *E. coli* carrying the pGEX-6P-*tCid1* plasmid was cultivated in the rich media, Terrific Broth (TB), so that the total cell weight was increased and therefore the total available protein (Figure 6.2). The expression levels were comparable to those found in *E. coli* cultivated in LB. Secondly, the soluble fraction of bacterial cell lysis was incubated with glutathione sepharose for >12 hours at 4°C, as the protocol described (Rissland *et al.,* 2007) only allowed the purification of a fraction of

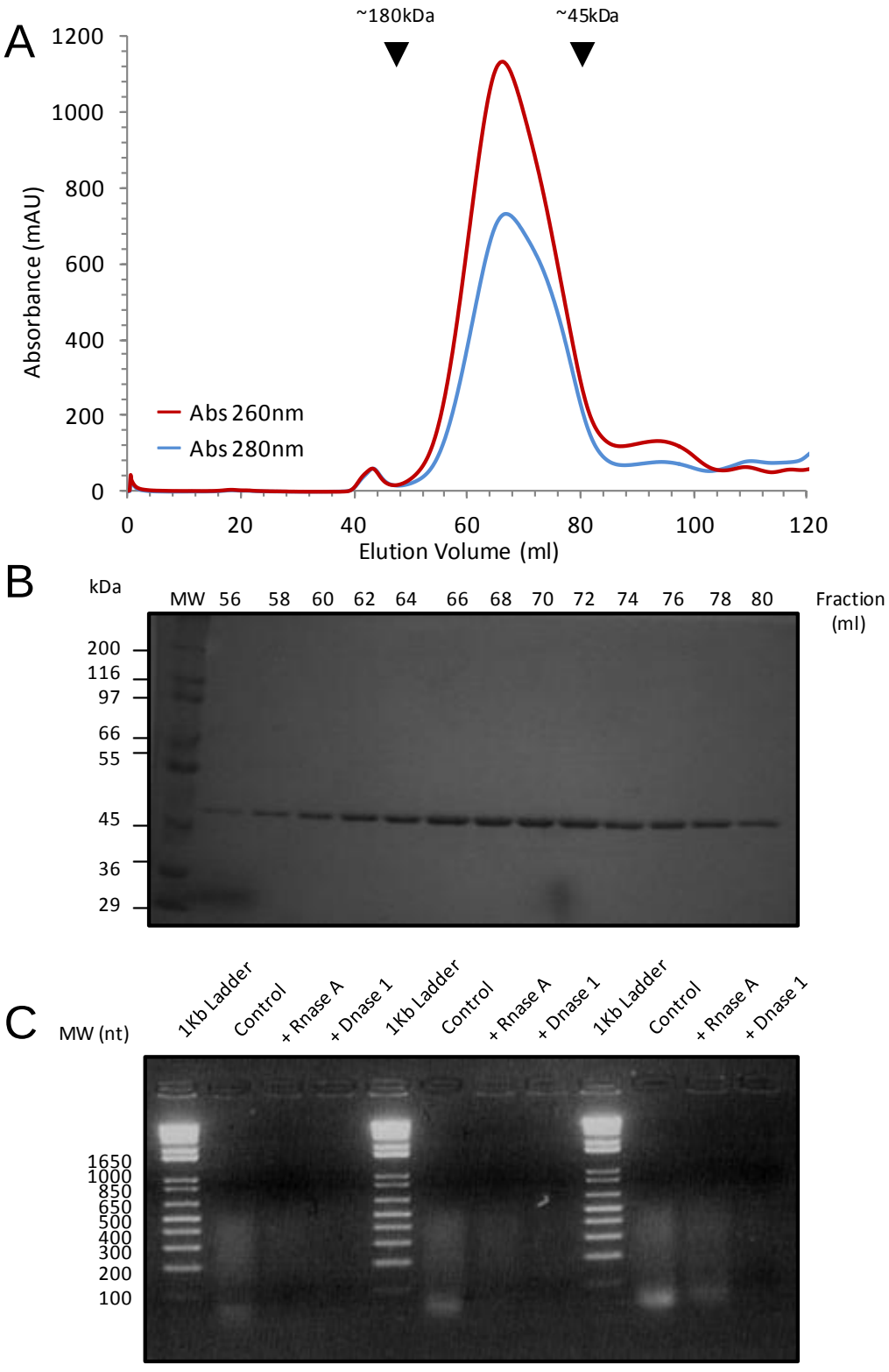
total available protein (Figure 6.2). Our early work with the enzyme suggested that the rate of this fusion protein binding to glutathione sepharose was slow, as recombinant fusion protein did not adsorb when applied to a pre-packed glutathione sepharose column (GE Healthcare) using an FPLC – even using a minimal flow rate (<1ml/min for a 5ml column). Using these modifications milligram quantities of reasonably pure fusion protein could be obtained and was deemed sufficient for further purification and structural studies.

With a reproducible method for obtaining purified fusion protein at sufficient quantities, the removal of the fusion partner, GST, was considered. The removal of the GST-tag was suggested to be pre-requisite for crystallisation, although some have reported the successful crystallisation of fusion protein (Smyth *et al.,* 2003). It was noted that the GST-tCid1 fusion protein possessed an engineered cleavage site between the enzyme and the fusion partner that is recognised and cleaved by the commercially available 3C protease. Therefore, human rhinovirus 3C protease digestion could liberate the tCid1 from its GST fusion tag. It is well known that GST dimerises (Fabrini *et al.,* 2009) and a GST dimer (50kDa) would be equivalent in molecular weight to the enzyme (~45kDa) and therefore could be co-purified when using size exclusion chromatography. It was rationalised that *in situ* cleavage of the adsorbed fusion protein, using the GST-tagged 3C protease, PreScission protease (GE Healthcare), would allow the successful liberation of the enzyme alone and the GST fusion partner and the contaminating protease would remain adsorbed to the affinity matrix. This proved successful and yielded milligrams of near-homogenous (>90% homogenous as assessed by SDS-PAGE) protein that could be further purified to homogeneity using size exclusion chromatography.

The polymerase tCid1 could be purified to homogeneity using size exclusion chromatography, thus allowing the removal of protein aggregates and providing an estimation

of the oligomeric state of the enzyme – an observation that had not previously been reported. As the oligomeric state was unknown, we applied protein onto a Sephadex S200 16/60 (GE Healthcare), which can resolve proteins of ≤450 kDa, using an FPLC (GE Healthcare). The protein migration was monitored spectroscopically using a wavelength of 280nm. To our surprise, the elution profile demonstrated a broad, asymmetrical, peak of absorbance that suggested multiple conformers and oligomeric states as the elution volume range was 55ml to 85ml which corresponds to a molecular weight range of 180kDa - 43kDa (Figure 6.3). SDS-PAGE analysis of the fractionated elution, corresponding to the single peak, revealed a single protein of ~45kDa protein, consistent with tCid1. We posited that the broad molecular weight range observed using size exclusion chromatography could be attributed to the contamination of the protein with a mixture of nucleic acids, given that the enzyme modifies the 3'end of RNA. We set out to provide evidence for the binding of nucleic acids by tCid1 given that no RNA binding domain could be detected using the enzyme's primary sequence. We first repeated the size exclusion chromatography and monitored the elution using two wavelengths at 260nm and 280nm. We quickly discovered that the absorbance at 260nm exceeded that at 280nm, suggesting the co-purification of nucleic acid (Figure 6.3). This was confirmed when the purified enzyme was subjected to agarose gel electrophoresis, which revealed a 'smear' of nucleic acids using SYBER® safe DNA gel stain (InVitrogen), suggesting that a heterogeneous mixture of nucleic acids was bound to the protein. To identify the nucleic acid contamination the purified enzyme was treated with RNase A (1µg) or DNase 1 (1µg) in the presence of $MgCl_2$ and the solution incubated at room temperature for 1 hour. The solutions were analysed by agarose gel electrophoresis and the nucleic acids visualised as before, which revealed that the contaminating nucleic acids were both DNase 1 and RNase A sensitive, suggesting not only a heterogeneous population of nucleic acids of differing lengths but also contamination of two different molecules. It was therefore posited that tCid1 was

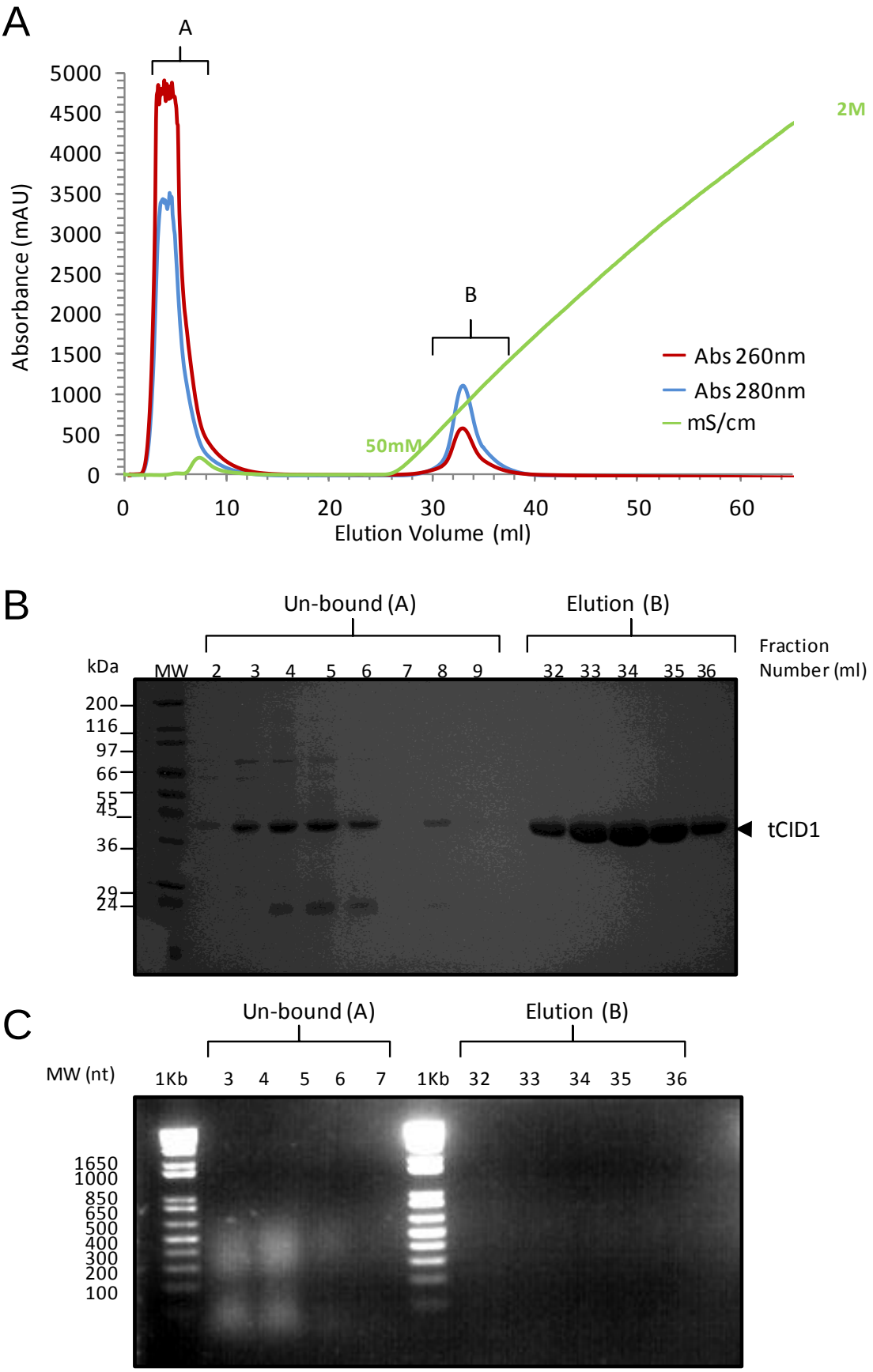
Figure 6.3 (Opposite) **-** Purification of tCid1 with Contaminating Nucleic Acids. (A) Size exclusion chromatography (SEC) of tCid1 (without GST fused) using a Sephadex S200 16/60 (GE Healthcare). A single large asymmetrical peak encompassing a substantial elution volume corresponding to the molecular weight (MW) range ~180kDa – ~45kDa is observed. Monitoring the elution profile using 280nm and 260nm wavelength suggests significant nucleic acid contamination. (B) SDS-PAGE analysis of the protein composition from SEC peak, suggesting that a single protein is present. (C) Agarose electrophoresis of SEC purified tCid1 alone (Control) or incubated at 37°C for 1 hour in the presence of RNAseA (1μg) or DNaseI (1 μg) both in the presence of 1mM $MgCl_2$. A smear of material is visible in the control lanes and is slightly sensitive to RNaseA treatment and very sensitive to DNaseI treatment.


A
~180kDa
~45kDa
Absorbance (mAU)
Abs 260nm
Abs 280nm
Elution Volume (ml)
B
kDa
MW 56 58 60 62 64 66 68 70 72 74 76 78 80
Fraction (ml)
200
116
97
66
55
45
36
29
C
MW (nt)
1Kb Ladder
Control
+ Rnase A
+ Dnase 1
1650
1000
850
650
500
400
300
200
100

capable of general nucleic acid binding. However, the modality of such binding by tCid1 and its biological significance was studied in depth and is discussed later.

The discovery that tCid1 possessed a general nucleic acid binding capacity was unexpected, and posed a significant problem for the crystallisation of the enzyme. Therefore several strategies were employed to remove the contamination. It had previously been demonstrated that the contaminating nucleic acids were sensitive to enzymatic digestion with RNase A and DNase 1. Therefore several litres of *E. coli* were cultured and the fusion protein and subsequently the tCid1 polymerase were purified as described previously. However, the method was modified by the addition of both RNAse A and DNase 1 with $MgCl_2$ (1mM final concentration) during the overnight incubation of the cell lysate soluble fraction with glutathione sepharose. Initially the cleaved and purified enzyme was assessed spectroscopically for the presence of nucleic acids based on the apparent 260nm:280nm absorbance ratio. However, it was considered that the enzyme's substrate, uridine triphosphate (UTP) could have bound during purification and would therefore contribute to the 260nm absorbance owing to the pyrimidine base. Therefore, the removal of contaminating nucleic acids was assessed by size exclusion chromatography and the elution monitored using two wavelengths (260nm and 280nm). Using a sephadex S200 (16/60) gel filtration column revealed an elution profile very similar to that seen previously, that is a broad peak where the 260nm absorbance exceeds the 280nm absorbance profile. However, at an elution volume close to that expected for a 45kDa protein, the corresponding fraction possessed a 260nm:280nm absorbance ratio of ~0.6. This was further resolved using size exclusion chromatography with a sephadex S75 (16/60) gel filtration column (GE Healthcare) as it possesses a narrower separation range. The enzyme was re-concentrated and applied to the column and the elution profile monitored. Interestingly, two separate peaks were observed, an initial broad peak with a 260nm:280nm absorbance ratio equal to 1.7

Figure 6.4 (Opposite) **-** Purification of tCid1 and Removal of Contaminating Nucleic Acids. (A) Heparin affinity chromatography of tCid1 (without GST fused) contaminated with nucleic acids. The nucleic acids bound to tCid1 enzyme were displaced by the immobilised highly-sulphated glycosaminoglycan carbohydrate, heparin leading to the unbound fraction that absorbs strongly at 260nm and eluted early (labelled A). The protein that bound to the immobilised heparin was eluted using a linear sodium chloride gradient from 50mM to 2M. A single symmetrical peak eluted as a result of the sodium chloride gradient and clearly displays a significant absorbance at 280nm and a greatly reduced 260nm.280 absorbance ratio. (B) SDS-PAGE analysis of the protein composition from peaks A and B, demonstrated a limited quantity of protein present in the unbound fractions (peak A), despite its strong absorbance profile, and a single ~45kDa protein is present in peak B. (C) Agarose electrophoresis of protein fractions from peak A and peak B after heparin affinity chromatography. The unbound fraction demonstrates the presence of nucleic acids and the eluted protein demonstrates an absence of nucleic acids.


A
A
B
2M
50mM
Absorbance (mAU)
Elution Volume (ml)
Abs 260nm
Abs 280nm
mS/cm
B
Un-bound (A)
Elution (B)
Fraction Number (ml)
kDa
MW
tCID1
C
Un-bound (A)
Elution (B)
MW (nt)
1Kb

and a second sharp, apparently symmetrical peak with a 260nm:280nm absorbance ratio of 0.6. Both peaks possessed a single ~45kDa protein, as assessed by SDS-PAGE. Although this method could separate the tCid1 polymerase from populations that are contaminated with nucleic acids it represented a highly inefficient method, with the majority of protein still contaminated. Furthermore, the ideal 260nm:280nm absorbance ratio for protein-only solution should equal 0.5. Therefore a more robust and reproducible method was considered.

An adequate DNA/RNA mimic was sought that could out-compete the bound nucleic acids to yield a homogenous protein solution suitable for crystallisation and for the robust and reproducible removal of nucleic acid contamination. The evidence thus far has suggested that the general nucleic acid binding capability is sequence non-specific and is likely to be mediated through a charge-charge mediated interaction between the nucleic acid phosphodiester backbone and the enzyme. The highly-sulphated glycosaminoglycan carbohydrate heparin was considered a suitable mimetic of the nucleic acid phosphate (deoxy) ribose backbone. Therefore, the commercially available immobilised heparin agarose was used for chromatographic purification of the enzyme.

The purified tCid1 polymerase bound to nucleic acids was demonstrated to bind to heparin agarose in low salt (50mM NaCl) conditions using a batch method. Consequently, protein was purified as described earlier and was buffer exchanged into a low salt HEPES buffered saline (20mM HEPES, 50mM NaCl, pH 7.0) using successive rounds of concentration and dilution using a concentration filter device. The protein solution was concentrated and applied to a heparin column (5ml GE Healthcare) using an FPLC and the elution profile monitored using two wavelengths (260nm and 280nm) as described previously. Predictably, a significant absorbance peak was observed directly after the application of the protein solution to the heparin column and possessed an absorbance at

260nm that far exceeded the absorbance at 280nm, suggesting unbound material and displaced nucleic acids. The protein that was bound to the heparin column was subsequently eluted using a linear NaCl gradient from 50mM to 2M. A single symmetrical peak was observed at a NaCl concentration range of 0.5M – 0.6M. Despite the linear gradient reaching the maximum NaCl concentration of 2M only a single peak was observed (Figure 6.4). The absorbance profile of the eluted protein revealed a 260nm:280nm ratio indicative of pure protein, therefore illustrating the removal of the contaminating nucleic acids. Unexpectedly, the absorbance of the unbound fraction was greater than the uppermost limit of the scale and demonstrated a substantial absorbance at 280nm that could be construed as unbound protein. However, SDS-PAGE analysis of the fractionated elution corresponding to unbound/displaced material and the eluted protein revealed only minimal protein present in the unbound fraction and highly pure protein eluted under a NaCl gradient, despite the absorbance of the unbound fraction greatly exceeding that of the eluted protein (Figure 6.4). Moreover, agarose gel electrophoresis of the fractionated elution revealed the presence of nucleic acids in the unbound/displaced fraction and no such contamination in the eluted protein (Figure 6.4).

The polymerase was purified to homogeneity by size exclusion chromatography subsequent to the removal of the nucleic acid contamination by heparin anion exchange chromatography. This final chromatographic step demonstrated that the protein possessed a 260nm:280nm ratio of 0.5 and is monomeric, indicated by the single symmetrical 260nm/280nm absorbance peak observed during chromatography using a sephadex S200 (16/60) gel filtration column (GE Healthcare) and its relative elution volume. SDS-PAGE analysis of the fractionated eluant revealed a homogenous protein solution suggested to be

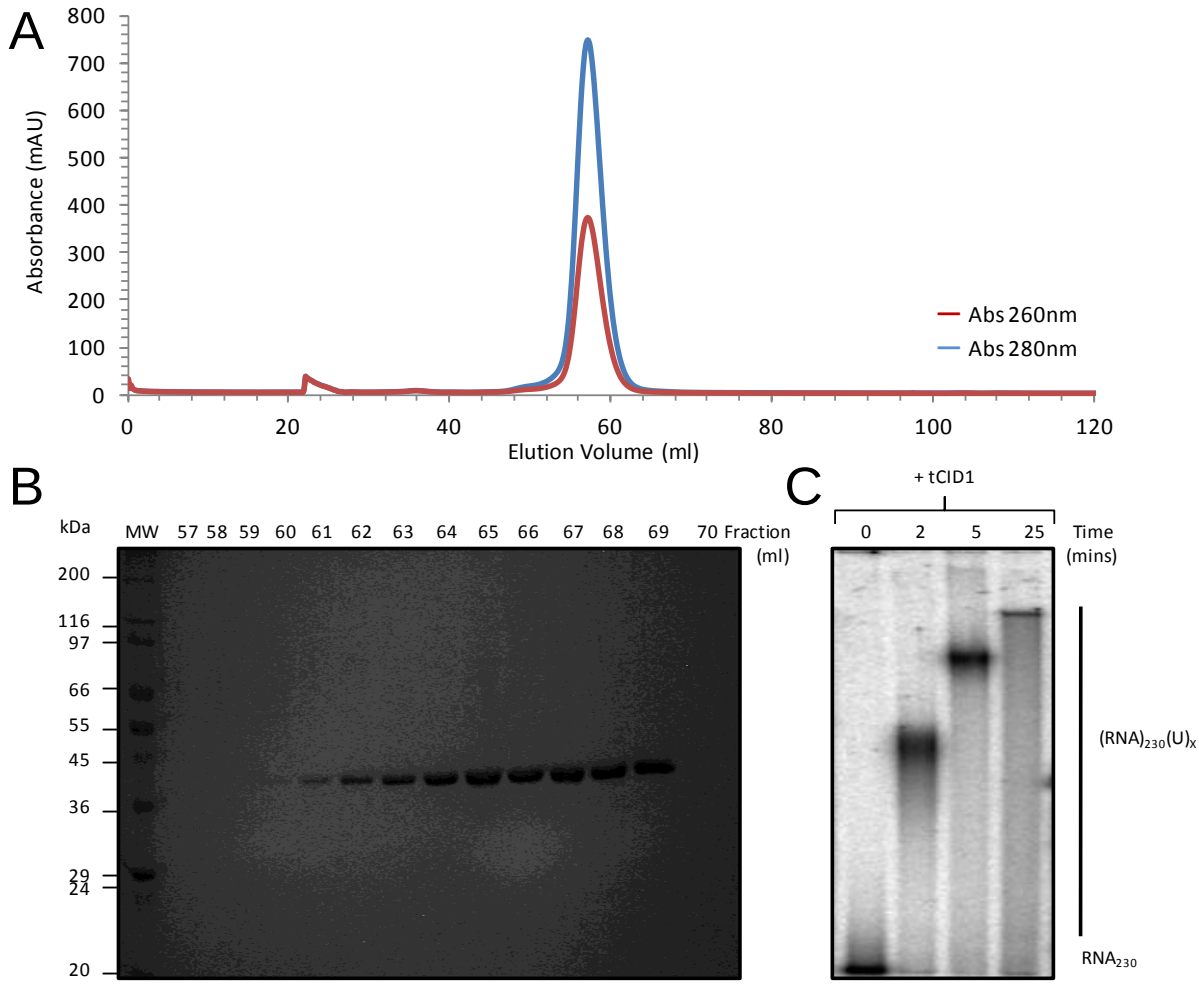


Figure 6.5- Purification of Homogenous tCid1. (A) Size exclusion chromatography (SEC) of tCid1 (without GST fused or contaminating nucleic acids). A single symmetrical peak with a 260nm/280nm ratio of 0.5 was observed when monitoring the elution at 280nm and 260nm. (B) SDS-PAGE analysis of the protein composition SEC demonstrating a single ~45kDa protein present (tCid1). (C) Acryl-8M Urea gel demonstrating uridylation activity of the SEC purified tCid1. Template RNA of 230nt (RNA_{230}) that includes a 30nt 3' poly(A) tail was incubated with tCid1 (1.5 pmol) at 30° C in the presence of 500 µM UTP and 7mM magnesium. Samples were taken at time 0 and after 2, 5 and 25 minutes. Panel C was kindly provided by S. Fleurdepine (Sir William Dunn School of Pathology, University of Oxford, UK).

>95% pure (Figure 6.5). The fractions corresponding to the absorbance peak from the elution profile were pooled and the protein concentrated to 11mg/ml. The concentration of the protein was assessed spectroscopically using an exctinction co-efficient ε=36330

(alternatively expressed as, Abs_{280nm} = 0.869 at 1mg/ml). The protein, once concentrated, was used immediately for crystallisation trials and a fraction was tested for its polyuridylyl polymerase activity (Figure 6.5).

6.2.3. Crystallisation Optimisation of tCid1

As described earlier, purified homogenous active enzyme (Figure 6.5) was concentrated and utilized in crystallization screening using commercially available crystallization reagents and vapour diffusion methods. Therefore, 100nl sitting drops of protein solution (10mg/ml in 20mM HEPES, pH 7.0, 200mM NaCl, 1mM DTT) were dispensed using a nanodrop crystallization robot (Cartesian Technologies) and diluted with an equal volume of crystallization solution. The sitting drops were equilibrated against 100µl of crystallization reagent at 16˚C (289K) and were duplicated for equilibration at 4°C (277K). During initial crystallisation trials, using tCid1 and UTP (1mM final concentration), thin plate-like hexagonal crystals appeared in Hampton Crystal Screen Cryo reagent number 39 (Hampton Research, CA) [1.7 % (v/v) Polyethylene Glycol 400, 15% (v/v) Glycerol, 1.7 M Ammonium Sulphate, 85mM HEPES, pH 7.5] at 4.0°C (277K). These crystals were unsuitable for X-ray diffraction data collection or structure determination. However, synchrotron radiation X-ray diffraction analysis revealed that these crystals were indeed protein. Further crystallisation optimisation was attempted using this initial ‘hit’ as a crystallisation lead. This condition was further optimized using a 3-way serial dilution protocol, with the crystallization condition [1.7 % (v/v) polyethylene glycol 400, 15% (v/v) glycerol, 1.7 M ammonium sulphate, 85mM HEPES, pH 7.5] ranging from 100% - 79% (diluted with sterile water) in one direction of the crystallization plate experiment together with three different protein: crystallization solution drop ratios i.e. 1:1 (protein : reservoir),

2:1 and 1:2, in another direction. However, this optimisation method was unsuccessful in producing protein crystals. Expanding on the serial dilution method, a fine dilution series protocol was performed using 2% dilution steps of the crystallisation solution in one direction of the crystallisation plate in conjunction with streak-seeding using crystals from the original crystallisation 'hit'. Therefore, nanolitre sitting drops were dispensed as described and small crystalline fragments, harvested from the original crystals using a cat's whisker, were transferred to the new crystallisation nanodrop. A crude dilution of crystalline seeds was achieved by the subsequent 'streaking' of the cat's whisker through identical nanoliter sitting drops. This was repeated until all crystallisation solution dilutions were screened. Thus, the crystallisation plate experiment represents a 2-dimensional optimisation with a dilution series in one direction and a seeding dilution in the second direction. Small plate-like crystals appeared in close proximity to the path of the cat's whisker and some single crystals were observed. X-ray analysis of these crystals using synchrotron radiation revealed that these crystals diffracted very weakly at a limited resolution (~6Å). Further optimisations could not yield better crystals.

It was proposed that the weak diffraction could be due to the inherent flexibility of the enzyme and by replacing the UTP with a non-hydrolysable ligand may capture a rigid 'transitional' state leading to stronger diffraction. Accordingly, additional crystallisation screening trials were performed using protein bound to the non-hydrolysable UTP analogues, where the α-β phosphodiester is replaced with a methylene group (UMPcPP) or an amine (UMPnPP). Crystallisation screening in the presence of these UTP analogues resulted in several salt crystals, as determined by in-house X-ray diffraction analysis.

Previously, we had determined that tCid1 was capable of binding to DNA and RNA (Figure 6.3); therefore including either nucleic acid in the crystallisation screening may aid the formation of single crystals suitable for structure determination. As DNA is inexpensive

and commercially synthesised and more resistant to degradation in solution it was reasoned that DNA would represent a suitable natural ligand mimetic, given the interaction between tCID1 and its natural ligand, RNA, was assumed to be mediated via the phosphodiester backbone and charged regions on the surface of the enzyme. Furthermore, given the previously determined poly(U) polymerase activity with a poly(A) primer and poly(A) polymerase activity with a poly(U) primer (Rissland *et al.*, 2007), one could not rule out the contribution Watson-Crick base pairing in UTP selection, much like that of the trypanosomal TUTase TUT4 (Stagno, *et al.*, 2007a). Therefore, a short 5-mer of DNA was synthesised and purified (SourceBio) with the sequence 5'AAAAA 3'. Additionally, the enzyme's interaction with endogenous *E. coli* nucleic acids could be outcompeted at low sodium chloride concentrations (50mM NaCl) using heparin affinity chromatography (see Materials and Methods and Figure 6.4), suggesting that efficient binding may occur under low salt conditions. As the interaction was assumed to be sequence non-specific it was reasoned that mixing the DNA and the enzyme in low sodium chloride conditions would promote the formation of a stable complex. Accordingly, homogenous tCID1 solution was buffer exchanged into a low salt crystallisation buffer (20mM HEPES, pH 7.0, 50mM NaCl, 1mM DTT) using a series of dilution and concentration steps. Subsequently, the 5-mer of DNA (5'AAAAA 3') was mixed with the homogenous protein at a ratio of 1.2:1, DNA: protein. Additionally, UTP (1mM final concentration) and $MgCl_2$ (1mM final concentration) were also included. Given that the activity assay was performed at 37°C, suggesting the enzyme was stable at that temperature, the concentrated tCid1:DNA:UTP•Mg solution was subsequently heated to 37°C for 30 minutes to allow the complex to form. Surprisingly some material precipitated during the incubation, which was removed by centrifugation. The precipitated protein was thought to be incorrectly/partially folded enzyme and therefore represents a purification step. The remaining protein in solution was assumed to be a ligand

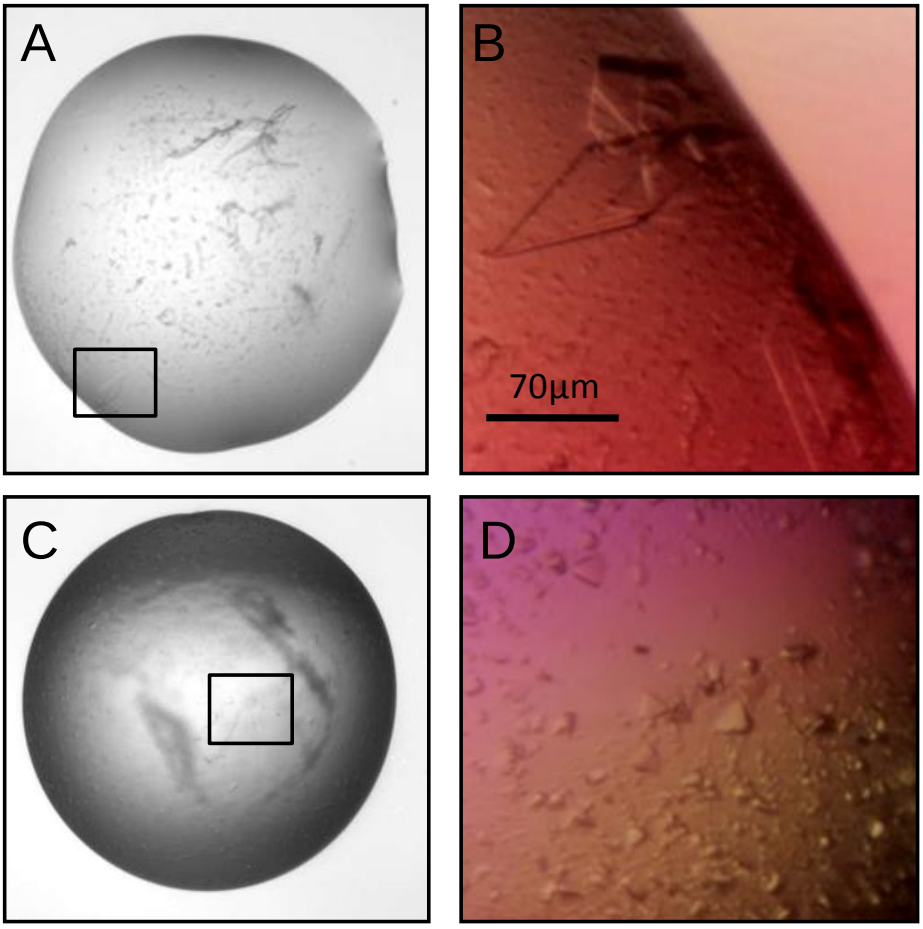


Figure 6.6 - Crystallisation of tCid1. Images of nanolitre sitting drop vapour diffusion crystallisation experiments. Purified tCid1 (in 20mM HEPES, pH 7.0, 50mM NaCl, 0.5mM TCEP) was incubated with a 5-mer of DNA (5'AAAAA3'), UTP and $MgCl_2$ (at a molar ratio of 1:1.2:8) at 37°C for 30 minutes prior to crystallisation screening. (A) Crystallisation of tCid1 in Hampton Crystal Screen Cryo (Hampton Research, CA, USA) reagent number 9. Thin multiple plate-like crystals of tCid1 appeared after 5 days and grew to ~100µm in the longest dimension after 2 weeks (magnified in B from box in A). (C) A shower of crystals of tCid1 grown in Hampton Crystal Screen (Hampton Research, CA, USA) reagent number 46 and appeared after 24 hours (magnified in D from box in C). All crystals were grown at 298K.

bound complex and was used for crystallisation screening using the nanolitre sitting drop vapour diffusion method. The apparent concentration of the protein sample was artificially increased by the addition of DNA and UTP, both of which absorbed strongly at 260nm but also contribute to an absorbance background at 280nm. Therefore the protein concentration

could not be accurately determined, but the presence of the enzyme was determined by SDS-PAGE analysis. Therefore, the absorbance at 280nm, not the protein concentration, is reported prior to crystallisation (See Materials and Methods).

A shower of small plate-like triangular crystals appeared in Hampton Crystal Screen I reagent number 46 (Hampton Research, CA) [18% (w/v) PEG 8000, 200mM calcium acetate, 100mM sodium cacodylate, pH 6.5] and Hampton Crystal Screen Cryo reagent number 46 [20% (v/v) glycerol, 14.4% (w/v) PEG 8000, 160mM calcium acetate, 80mM sodium cacodylate, pH 6.5] both at 16.0°C, within 24 hours. Analysis of these crystals by synchrotron radiation demonstrated no appreciable X-ray diffraction. Within the same crystallization screening experiment larger multiple plate-like crystals appeared in Hampton Crystal Screen Cryo reagent number 9 [15% (v/v) glycerol, 25.5% (w/v) PEG4000, 170mM ammonium acetate, 85mM tri-sodium citrate, pH 5.6] at room temperature within 5 days and grew to an optimal size (~140µm x ~70µm) within 2 weeks. X-ray diffraction, using synchrotron radiation demonstrated reasonable diffraction properties, with reflections observed to 4Å. Crystallization optimization was performed using the Crystal Screen Cryo reagent number 9 (CSC9) and an additive screen (Hampton Research, CA). Crystals of similar morphology grew over the course of 2-5 weeks in conditions including the volatile compounds; dichloromethane (0.05% [v/v]), 2,5-hexandiol (8% [v/v]), and ethyl acetate (1% [v/v]). Further crystallization optimization was performed using a micro-seeding protocol in conjunction with the nanolitre sitting drop vapour diffusion method. This protocol differs from the previously described streak-seeding using a cat's whisker as it is automated using the nanodrop crystallization robot (Cartesian Technologies). The protocol used was developed in the OPPF protein crystallization laboratory and is described in detail elsewhere (Walter *et al.,* 2008), but briefly, a single crystal is harvested and transferred to a solution,

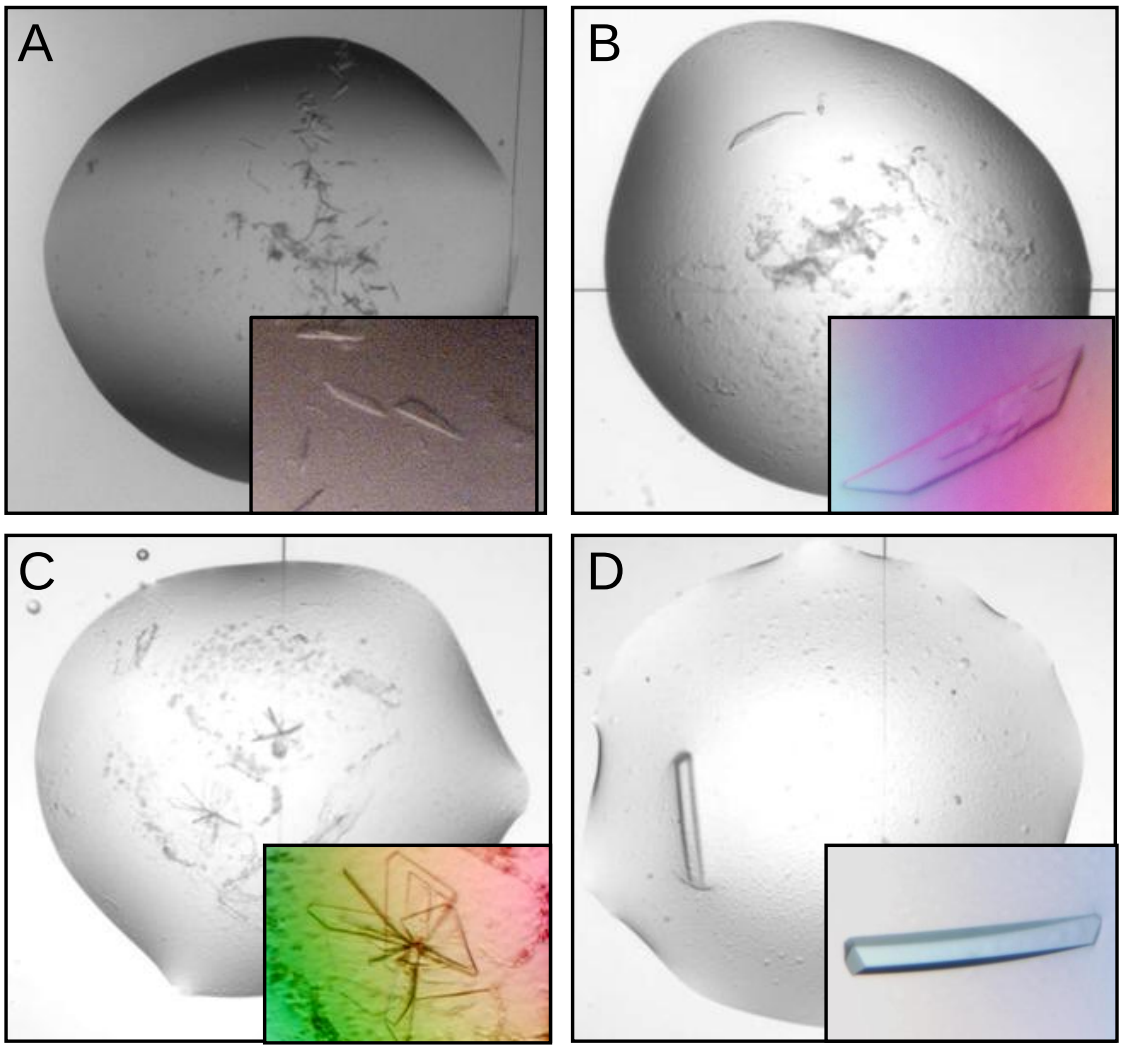


Figure 6.7 - Examples of Crystallisation Optimisation of tCid1 Using Several Crystallisation Optimisation Techniques. Images of nanolitre sitting drop vapour diffusion crystallisation experiments along with magnified crystals (inset) to show crystal morphology. (A) tCid1 crystal optimisation using steak seeding into a sitting drop using an early crystallisation hit (15% (v/v) glycerol, 1.7% (v/v) PEG400, 1.7M ammonium sulphate, 85mM HEPES, pH 7.5) at 277K. (B) Microseeding optimisation of tCid1 crystals using Crystal Screen Cryo reagent 9 (CSC9; Hampton Research, CA, USA) described by (Walter *et al.* 2008) at 298K using, in this example, a seed stock dilution of 1:100000. (C) Optimisation of crystallisation condition CSC9 (Hampton Research, CA, USA) by supplementing the sitting drop with small molecules from the Additive Screen (Hampton Research, CA, USA). Crystals shown were grown in the presence of 1% (v/v) ethyl acetate at 298K. (D) Combinatorial optimisation using microseeding and additives of crystallisation condition CSC9, yielding crystals of different morphology, which also belong to a different space group. Crystals grew to greater than 400µ in the longest dimension and appeared after 6 weeks at 298K.

which is comprised of crystallization mother liquor and dilute recombinant protein (tCid1). The crystal was subsequently shattered using a 'seed bead' (Hampton Research, CA) and vigorous agitation. The micro seed suspension, or seed stock, was serially diluted into a micro-seed solution, which contains mother liquor and diluted recombinant protein, to generate a dilution series. A 'no-seed' control was also created for comparative purposes. This dilution series of micro-seed suspensions was dispensed on top of freshly prepared 100nl sitting drops, thus forming a micro-seed dilution series in one direction of the crystallization plate experiment.

Analogous to the 'streak-seeding' protocol implemented and described earlier, the crystallization solution for the sitting drops and reservoir was a stepwise dilution (100% - 67%) with sterile water to form a dilution series in a second direction, perpendicular to the micro-seed dilution series. The sitting drops were equilibrated against 100µl of the stepwise diluted crystallization reservoir (CSC9, Hampton Research, CA) at room temperature. Crystals of similar morphology, that is multiple thin plates, grew in conditions with the greatest concentration of micro-seeds (1:10000 dilution of seed stock). A single plate-like crystal grew where the seed concentration was an order of magnitude less (1:100000 dilution of seed stock).

Further crystallization optimizations were also performed using a combination of the crystallization optimization techniques that had previously been shown to improve crystallization reproducibility or morphology. Thus micro-seeding optimization was used in conjunction with additive screening using a diluted crystallization solution as the basis for the concurrent optimizations. Unexpectedly, two conditions that contained taurine or L-proline as additives, yielded very large single crystals (~400µm in the longest dimension) different in morphology to the previous crystals obtained in similar conditions (Figure 6.7) that grew after 6 weeks at 25°C.

Despite the protein crystallizing in the presence of UTP, $MgCl_2$ and DNA, the solved structure (see Chapter 7) represented an unbound (apo) state. Therefore, in order to obtain a UTP bound structure the crystals were soaked with UTP prior to data collection. As the thin plate-like crystals were so numerous (similar to those in Figure 6.7D), crystals of this morphology were used in UTP soaking experiments. Therefore the nanolitre sitting drops containing these crystals were supplemented with mother liquor containing UTP, yielding a final concentration of 10mM, and the crystals incubated at room temperature for 0.5 – 1 hour. The soaked crystals were harvested and frozen for data collection as previously described.

6.2.4. X-ray Diffraction Analysis of tCid1

The multiple plate-like crystals that grew in many optimized conditions could be separated into single-plate crystals, or pried in such a way as to produce a fragment with a singular region suitable for data collection. Thus all crystals that grew from the optimizations and those subsequently soaked in a solution containing 10mM UTP, were treated in this way and were frozen directly in liquid nitrogen prior to synchrotron X-ray radiation exposure. The crystallization condition (CSC9) used for optimization contains glycerol; therefore no cryo-protectant was required for freezing. The large single crystals were cut into 2-3 fragments before freezing. Although the crystallization condition contains glycerol, the concentration present was not sufficient for adequate cryo-protection. Therefore, the crystals were transferred and soaked in mother liquor supplemented with 25% (v/v) glycerol before freezing in liquid nitrogen.

X-ray diffraction data (see Figure 6.8 for diffraction images) from the two types of native tCid1 crystals together with the UTP soaked crystals were collected under cryo-temperature (~100K) conditions using single wavelength synchrotron radiation at Diamond Light Source (Didcot, UK) and the European Synchrotron Radiation Facility (ESRF). Due to

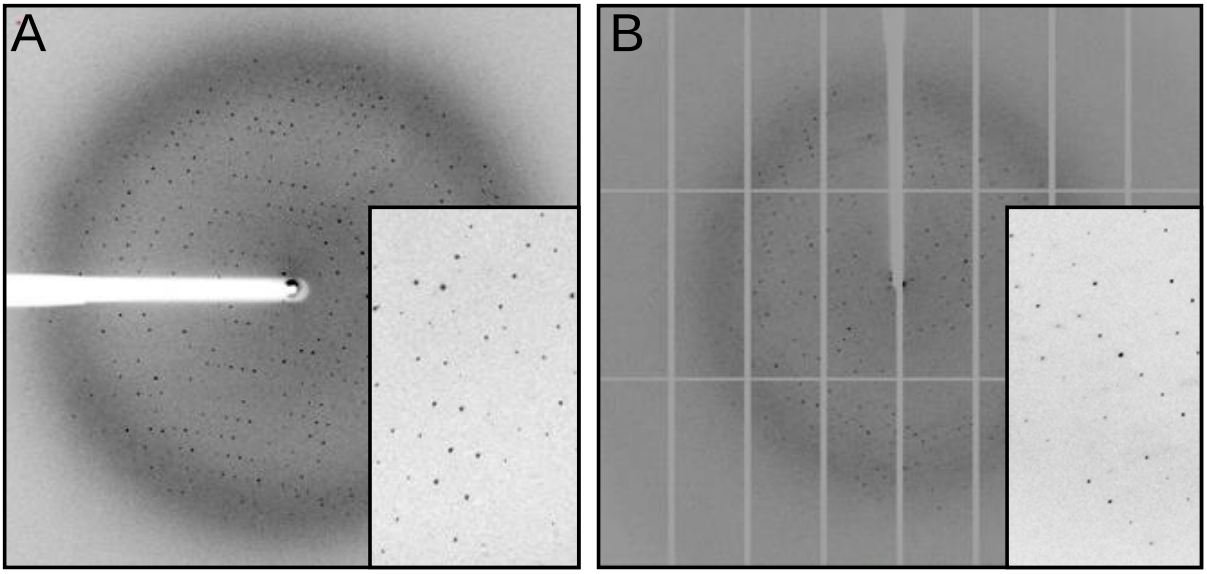


Figure 6.8 – Representative diffraction of tCid1 Crystals. (A) Single diffraction image from crystal form I soaked in UTP. (B) Single diffraction image from crystal form II. Diffraction data was collected on a MarCCD and a Pilatus 6M detector, respectively at a single wavelength.

the marked difference in size of the crystals it was appropriate to collect diffraction data using two different beamlines. In the case of the thin-plate like crystals the specialised microfocus beamlines I24, at Diamond light source, and ID23-1, at the ESRF, were used for data collection of native and UTP-soaked crystals, respectively. Due to the size of the crystals a beamsize of 10μm x 10μm, at a single wavelength, was used in both cases. In the case of the large rectangular rod-like crystals the standard macromolecular crystallography beamline I04 was used for data collection at a single wavelength (λ=0.9773Å). Diffraction data for all crystals were collected using a Pilatus 6M detector, at I24, and a MarCCD detector, at ID23-1, with 0.2° oscillations in all cases.

X-ray diffraction data from each frozen thin plate-like crystal were tested for optimal diffraction prior to data collection. All crystals of this type demonstrated diffraction, with reflections typically observed to a resolution range of 3Å – 4Å. It was also observed that these types of crystals were very sensitive to radiation damage, demonstrated by the loss of diffraction resolution after prolonged exposure to synchrotron radiation. However, the soaked

TABLE 6.1 – Data Collection Statistics

	Crystal Form I (Native)	Crystal Form I (UTP soak)	Crystal Form II (Native)
Crystallisation Condition	15% (v/v) Glycerol, 25.5% (w/v) PEG 4000, 170mM Ammonium Acetate, 85mM tri-Sodium Citrate, pH 5.6	15% (v/v) Glycerol, 25.5% (w/v) PEG 4000, 170mM Ammonium Acetate, 85mM tri-Sodium Citrate, pH 5.6	1% Taurine (w/v), 15% (v/v) Glycerol, 25.5% (w/v) PEG 4000, 170mM Ammonium Acetate, 85mM tri-Sodium Citrate, pH 5.6
Beamline	Diamond I24	Diamond I24	Diamond I04
Wavelength (Å)	0.973	0.973	0.9173
Oscillation (°)	0.2	0.2	0.2
X-ray Detector	Pilatus 6M	Pilatus 6M	Pilatus 6M
Temperature (K)	100	100	100
Unit cell (Å, °)	a = 164.4, b = 78.0, c = 152.5, α =γ = 90, β =109.5	a = 164.2, b = 77.9, c= 151.2, α = γ = 90, β =109.2	a = 81.2, b = 101.2, c=113.6, α = β = γ = 90
Space Group	C2	C2	$P2_12_12_1$
No. Of Molecules in the ASU	4	4	2
Resolution (Å)	64.24–3.20 (3.22–3.20)[a]	77.72 – 3.02 (3.10-3.02)	56.8 – 2.60 (2.74 – 2.60)
Observed Reflections	130475	151755	530572
Unique Reflections	30013 (9703)	35591 (2736)	29430 (4231)
Completeness (%)	99.5 (97.7)	99.5(99.5)	99.6 (99.7)
Redundancy	4.1 (4.1)	3.9 (3.9)	18.0 (18.4)
$\langle I/\sigma I \rangle$	8.9 (2.4)	12.8 (3.2)	24.7 (6.2)
R_{merge}[b]	0.138 (0.712)	0.093 (0.479)	0.073 (0.481)
$R_{p.i.m.}$[c]	0.078(0.402)	0.055 (0.279)	0.018 (0.011)

[a] Values in parentheses account for the highest resolution shell,

[b] $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| \,/\, \sum_{hkl}\sum_i I_i(hkl)$

[c] $R_{p.i.m.}$ (precision-indicating merging R-value) $= \sum_{hkl}[1/N-1]^{\frac{1}{2}}\sum_i |I_i(hkl) - \overline{I(hkl)}| / \sum_{hkl}\sum_i I_i(hkl)$

crystals demonstrated diffraction, on average, to a resolution of ~2.9 – 3.0 Å and were markedly less sensitive to synchrotron radiation but high resolution reflections still diminished as a result of extensive illumination during data collection. Therefore in both cases single complete data sets to intermediate resolutions (2.9Å - 3.5Å) could not be collected using a simple single sweep data collection strategy. Thus, data collection was performed by collecting several overlapping wedges of data from adjacent positions within a single mounted crystal. Initial indexing of diffraction images was performed by *MOSFLM* (Leslie and Powell, 2007), which identified that the crystals belong to the C-centred monoclinic space group C2 and possess the average unit cell dimensions; a=165Å, b=78Å, c=152Å (α=γ=90°, β=109°). The data were initially processed with *MOSFLM* (Leslie and Powell, 2007) and *SCALA* (Evans, 2006) or XDS (Kabsch, 2010) and SCALA (Evans, 2006) implemented in Xia2 (Winter, 2009). The subsequent merging statistics were used to

determine which diffraction images to exclude in each sweep in order to create an optimised subset of data to be merged. Optimal datasets were comprised of a total of ≥90° of diffraction data from three to four positions on the crystal. These data from each sweep were combined, merged and scaled in Xia2 (Winter, 2009) to give a final working native dataset at a resolution of 3.15Å and a UTP soaked dataset at a resolution of 2.9Å – the statistics of which are summarised in Table 6.1. The crystals contained four molecules in the asymmetric unit as determined by Matthews' coefficient (Matthews, 1968) with a suggested solvent of content of 49% using a protein molecular weight of 41kDa.

X-ray diffraction data from single frozen crystals derived from the large crystal species demonstrated strong diffraction, with reflections typically observed at a resolution of 2.5Å. *In situ* indexing and data reduction, performed by Xia2 (Winter, 2009), identified that the crystals belonged to the primitive orthorhombic space group $P2_12_12_1$, and possessed the average unit cell dimensions; a=81Å, b=101Å, c=114Å (α= β= γ= 90°). Several complete datasets to 2.5Å were collected from a single frozen crystal by collecting diffraction data from adjacent positions within a single mounted crystal. Each dataset was indexed and integrated in *MOSFLM* (Leslie and Powell, 2007), or XDS implemented in Xia2, and merged and scaled using the CCP4 program *SCALA* (Evans, 2006). Autoindexing solutions in *MOSFLM* also suggested that the crystal belonged to the primitive orthorhombic space group $P2_12_12_1$ consistent with Xia2, and was confirmed as the most probable using the CCP4 program *POINTLESS*. Owing to the data collection strategy and that each dataset was collected from the same crystal and the difference in unit cell parameters between the datasets was less than 0.3Å in any direction, the datasets were combined, remerged and rescaled together using *SCALA* (Evans, 2006), with a resolution cut-off of 2.6Å. The combined reflection file was converted into structure factors using the CCP4 program TRUNCATE (French and Wilson, 1978). This yielded a complete dataset to 2.6Å with a greater number of

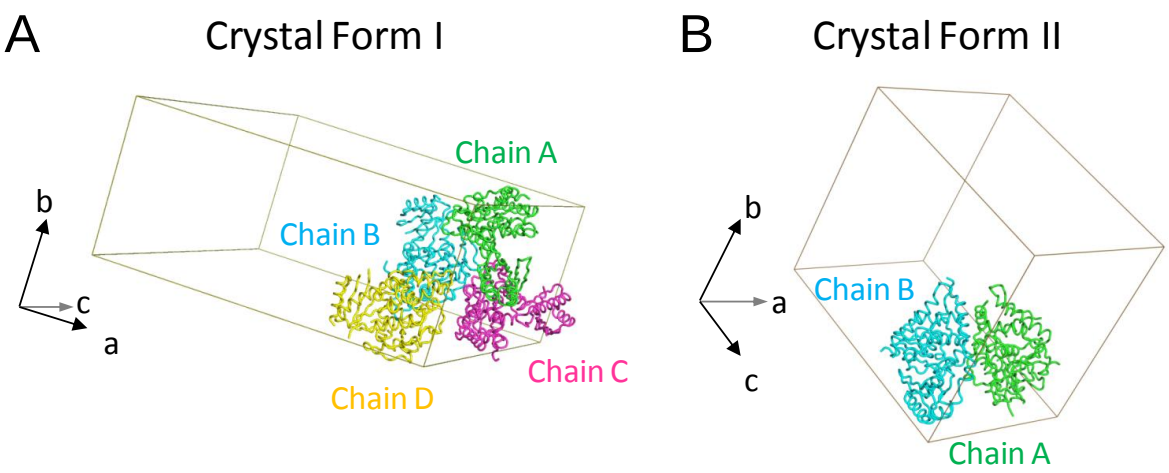


Figure 6.9 – tCid1 Crystallisation in Two Crystal Forms. (A) Unit cell of crystal form I, which belongs to space group C2 and possesses dimensions of a=165Å, b=78Å, c=152Å (α=γ=90°, β=109°) together with the four molecules (chain A to D) per asymmetric unit, shown as tubes. (B) Unit cell of crystal form II, which belongs to space group $P2_12_12_1$ with dimensions a=81Å, b=101Å, c=114Å (α= β= γ= 90°) together with the two molecules in the asymmetric unit, shown as tubes.

observations and thus a greatly increased redundancy – the statistics of which are summarised in Table 6.1. The unit cell of this crystal form was suggested to possess two molecules per asymmetric unit as determined by Matthews' coefficient (Matthews, 1968) implemented in *PHENIX* Xtriage (Adams *et al.*, 2010). During data collection, it was observed that some diffraction images demonstrated obvious anisotropy. Therefore, the resulting structure factors were submitted to the anisotropy server (Strong *et al.*, 2006), which confirmed an intensity falloff in a directionally dependent manner. The data were not processed using an ellipsoidal truncation procedureas directional resolution cut-offs in a*, b* and c*, to define the limits of the ellipse, were 1/2.6Å in all directions. Therefore anisotropic scale factors were applied by the server together with an isotropic B factor of -17.32Å^2 parrallel to c* to correct for the anisotropic scaling (Strong *et al.*, 2006). Usually the reciprocal resolution limits along a*, b* and c* for the ellipsoidal truncation are markedly different for strongly anisotropic data due to the directionally dependent intensities (Strong *et al.*, 2006). Therefore analysis of the

intensity falloff *versus* resolution of the three principle directions suggested that a higher resolution dataset should have been collected as intensity falloff for the a* and b* direction were less severe. Unfortunately, due to the rarity of this crystal form further higher resolution data collection was not possible.

The crystal structures were subsequently solved by molecular replacement (see Materials and Methods, section 2.6.4) and the resulting structures are discussed in depth in the following chapter, Chapter 7.

CHAPTER 7

7.1. The Structural Basis for the Activity of the Cytoplasmic RNA Terminal Uridylyl Transferase, Cid1

7.1.1. A Brief Introduction to Cid1-like 3'Uridylation Enzymes in Humans

As described in Chapter 6, the discovery of Cid1 in cell cycle regulation led to its biochemical characterisation and to definition of the enzyme's uridylation activity (Rissland *et al.*, 2007). Furthermore, the decay of mRNAs was shown to be Cid1-dependent with 3' uridinylation of cytoplasmic polyadenylated mRNAs targeted for degradation (Rissland and Norbury, 2009). The post-transcriptional addition of uridines to cytoplasmic RNA 3' ends has recently been implicated in several key aspects of eukaryotic RNA biology, including mRNA turnover and regulation of the biogenesis and activity of micro-RNAs (miRNAs). For example, cytoplasmic uridylation targets the precursor of *let-7* tumour suppressor miRNAs for destruction. Additional roles for post-transcriptional 3'uridinylation include; U6 snRNA 3' end processing (Trippe *et al.*, 1998), cell cycle dependent histone mRNA metabolism (Schmidt *et al*, 2011), and miRNA-directed RNA decay (Shen and Goodman, 2004).

The metazoan terminal uridylyltransferase (TUT) ZCCHC11/TUT4 was originally identified due to its sequence homology with the yeast terminal uridylyltransferase, Cid1 (Stevenson and Norbury, 2006). Whilst Cid1 is a 46 kDa protein containing two recognisable sequence motifs: a nucleotidyl transferase motif common to all members of the DNA polymerase β (Polβ) superfamily and a poly(A) polymerase (PAP)-associated motif, by contrast, ZCCHC11 is much larger (185 kDa), containing a duplication of both motifs found in Cid1. Of the two duplicated polymerase β-like nulceotidyltransferase domains the

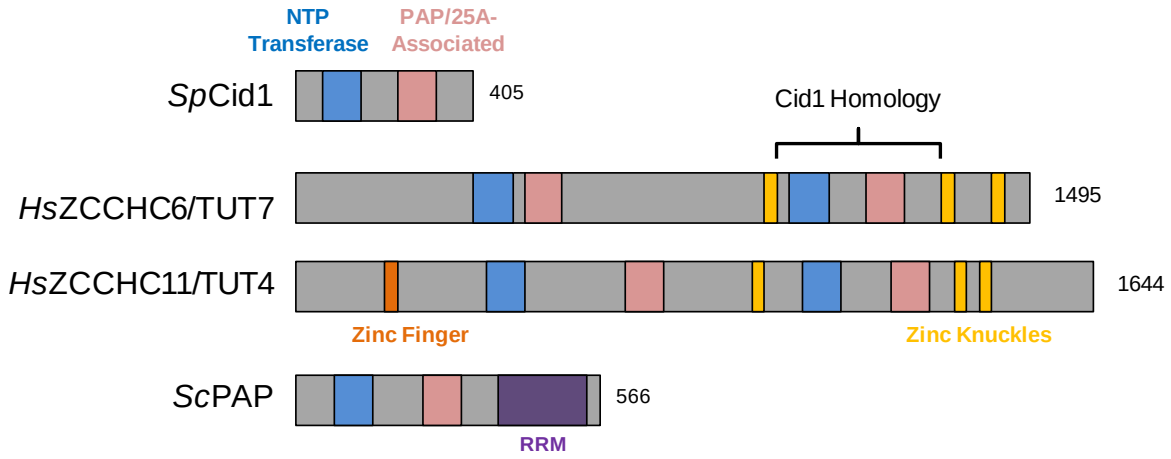


Figure 7.1 – A Schematic of Cid1 and Metazoan TUTases. The *S. pombe* Terminal Uridyltransferase (TUTase) Cid1 (*Sp*Cid1) possesses two recognisable motifs; the catalytic NTP transferase motif (blue) and the Poly(A) polymerase/ 2'-5'Oligoadenylate Synthetase-associated motif (pink). In the metazoan TUTases, ZCCHC6 (TUT7) and ZCCHC11 (TUT6), these motifs are duplicated with the C-terminal pair of domains most similar to Cid1. Poly(A) polymerase from yeast (*S. cerevisiae*) is also shown for comparison.

N-terminal domain is catalytically inactive, whereas the more C-terminally situated domain is able to transfer UMP to the 3'end of RNA substrates (Jones *et al.*, 2009; Hagan *et al.*, 2009; Rissland *et al.*, 2007). Furthermore, three CCHC-type zinc knuckle motifs are also present in ZCCHC11, which has provided a basis for its name – zinc-finger, CCHC domain-containing (Figure 7.1). ZCCHC11 catalyses the 3' uridylation of cytoplasmic RNAs, which include, miRNA precursors (pre-miRNAs), mature miRNAs and replication-dependent histone mRNAs. Whereas ZCCHC11 interacts with the RNA-binding protein Lin28A (Hagan *et al.*, 2009) in order to associate stably with pre-miRNA substrates, no equivalent interaction has been described for Cid1, which in monomeric recombinant form efficiently transfers UMP residues from UTP to RNA substrates *in vitro*. An additional metazoan cytoplasmic TUT, ZCCHC6/TUT7, orthologous to Cid-1 and implicated in miRNA modification (Wyman *et al.*, 2011) was originally characterised as an uridylation enzyme alongside Cid1 in the same study (Rissland *et al.*, 2007). However, ZCCHC11 has taken centre stage, with numerous studies

demonstrating its importance in stem cell biology and implications in cancer (Schmidt *et al.*, 2011; Heo *et al.*, 2009; Jones *et al.*, 2009; Hagan *et al*. 2009; Lehrbach *et al.*, 2009). Many RNAs that ZCCHC11 uridylates are involved in proliferation pathways. For example, ZCCHC11 uridylates miRNA-26a, which allows the increased expression of miRNA-26 targets, which includes many cell cycle control genes (Blahna *et al.*, 2011). Additional RNAs that are uridylated by ZCCHC11 are the precursor miRNAs of the let-7 family members, inhibiting the generation of mature let-7 miRNAs (Heo *et al.*, 2009; Hagan *et al.*, 2009), the targets of which include the cyclins. Further interest to ZCCHC11 and its family members has been added from studies of the enzyme's N-terminal regions, in which these regions contribute to cell cycle progression in the absence of uridylation activity through an as yet undermined mechanism.

To date, there are no structural data on these enzymes but sequence alignments have shown that they contain a Cid1-like region (Stevenson and Norbury, 2006), which is illustrated in Figure 7.1. Interestingly, in a simple CLUSTALW alignment Cid1 is immediately aligned with the catalytically active nulceotidyltransferase domain. Therefore, understanding the structure of Cid1 may provide clues as to the relatedness of the activity of the ZCCHCs and whether Cid1 serves as a suitable surrogate for adequately describing the ZCCHC catalytic domains, while being substantially more straightforward than the direct study of the metazoan enzyme.

7.1.2. Summary

The *Schizosaccharomyces pombe* uridylation enzyme, caffeine induced death suppressor-1 (Cid1), was previously crystallised in two crystal forms and in a UTP bound form by soaking the crystals with UTP (Chapter 6). Here the X-ray crystal structures of the

Schizosaccharomyces pombe cytoplasmic RNA terminal U-transferase Cid1 in two Apo conformers and in complex with UTP are reported. The architecture of the active site of Cid1 identifies a single histidine residue critcal for the UTP selectivity of the enzyme. We also observe a dramatic conformational change and remodelling of the active site on transition between two conformations suggesting a mechanism for the clearance of uridylated product. Cid1 lacks a canonical RNA recognition motif, but efficiently binds RNA substrates with high affinity via three basic patches of amino acid side chains on the surface that are non-contiguous in the primary sequence. This is likely to be a novel but widespread method employed by many proteins for efficient RNA binding. This work presents the first structure of a cytoplasmic RNA uridylyl transferase and opens the way to novel avenues in anti-cancer drug design.

7.1.3. Structure Determination of tCid1

As described in Chapter 6, a functional but slightly truncated variant of Cid1 (residues 33-405; termed tCid1) was crystallised in two crystal forms (space groups C2 and $P2_12_12_1$). A UTP-bound structure was later obtained by soaking the crystals of crystal form I in UTP. The apo structure was initially determined by molecular replacement and refined to 3.2Å, with the UTP-bound structure determined using the Apo structure refined to 3.0 Å (see Materials and Methods). The structure of the second crystal form was determined, again by molecular replacement, using a different search model to that used in the initial MR calculation (see Materials and Methods).

The refined structures demonstrate good agreement with the crystallographic data, with good R-factors and expected peptide geometry (bond angles, lengths and Ramachandran analysis). A table of structure refinement statistics is given in Table 7.1. The asymmetric unit

of crystal form I (space group C2) possesses four molecules (chains A to D), which superimposed with an average r.m.s.d. value of 0.25 Å over ~238 Cα atoms. The asymmetric unit of crystal form II, on the other hand, possesses two molecules per asymmetric unit (chain A and B). Of the two molecules, Chain B displays an unusual domain rearrangement, which is discussed in detail later, and therefore the molecules are not isomorphous. Importantly, Chain A is similar to the structures in crystal form I, with average r.m.s.d. values of 0.55Å. The oligomers in the crystal do not seem physiologically relevant, given that tCid1 was purified as a monomeric enzyme, except where nucleic acid is present, and that the monomer to monomer contacts in the asymmetric unit are not conserved between the crystal forms.

During the determination of the UTP-bound structure it was obvious that a single UTP molecule was bound to the enzyme as clear unbiased electron density was present in the omit map (F_c-F_o) directly after molecular placement (Figure 7.2). Interestingly, in two of the four molecules in the asymmetric unit electron density corresponding to a disordered loop in the C-terminal domain that was absent in either apo crystal structure, became apparent in the UTP-bound structure. The subsequent building and refinement revealed a two-stranded anti-parallel β-sheet disordered in the absence of UTP.

7.1.4. The Overall Structure of tCid1

The structure of tCid1 contains two globular domains; a C-terminal and an N-terminal domain, separated by a deep cleft approximately ~15Å X ~32Å X ~18Å (Figure 7.2). In the tCid1-UTP complex structure the nucleotide is bound in the cleft, analogous to other terminal uridine transferases from *T. brucei* and poly(A) polymerase (Stagno *et al.,* 2007a, Deng *et al.,* 2005, Bard *et al.,* 2000). The UTP structure also reveals that residues 310 – 324 in the C-terminal domain form a structured protrusion that looks similar to a ‘trap door’ in the

orientation given in Figure 7.2., which extends and contacts the N-terminal domain. This feature, denoted as the trap door, is only found in Cid1 and is not present in other TUTase structures. Furthermore, sequence alignments have also shown its absence in the metazoan orthologues ZCCHC11 and ZCCHC6 (Figure 7.3).

The structures encompass residues 41 – 380 of Cid1 according to the UniProt Accession code O13833 and are comprised of eleven α-helices, designated A through K, and seven β-stands, which are numbered accordingly (Figure 7.2). The enzyme's two globular domains share sequence and structural similarities with other polymerases. Accordingly, each domain and its structural similarities will be discussed in turn, below.

The N-terminal domain (NTD) is comprised of a mixed 5-stranded twisted β-sheet (β1 – β5) and four α-helices (αA to αD) and is considered a two-layered α-β sandwich, which adopts a DNA β-polymerase palm domain topology (CATH). The twist of the beta sheet is imposed from the two α-helices (αB and αD) in the palm domain fold that stack against one face of the β-sheet. The NTD of Cid1 is topologically homologous to the palm domain of DNA polymerase β (Sawaya *et al.,* 1994) in which strands β1 – β4 are anti-parallel and strand β5 is inserted into the middle of an otherwise antiparallel β-sheet between β2 and β4 and therefore runs parallel to the β2 strand and antiparallel to β4 (Figure 7.2). This topology is consistent with other palm domains of the DNA polβ-like nucleotidytransferase family members, for example the eukaryotic poly(A) polymerase (PAP; Bard *et al.,* 2000, Martin *et al.,* 2000), CCA-adding enzymes (Li *et al.,* 2002), terminal uridine transferases (TUT; Stagno *et al.,* 2007a) and RNA editing enzymes (Deng *et al.,* 2005) from *Trypanosoma brucei.* Consistent with other characterised members the polβ-like family nucleotidyltransferases and biochemical characterisations (Rissland *et al.,* 2007) the palm domain (or NTD) is considered

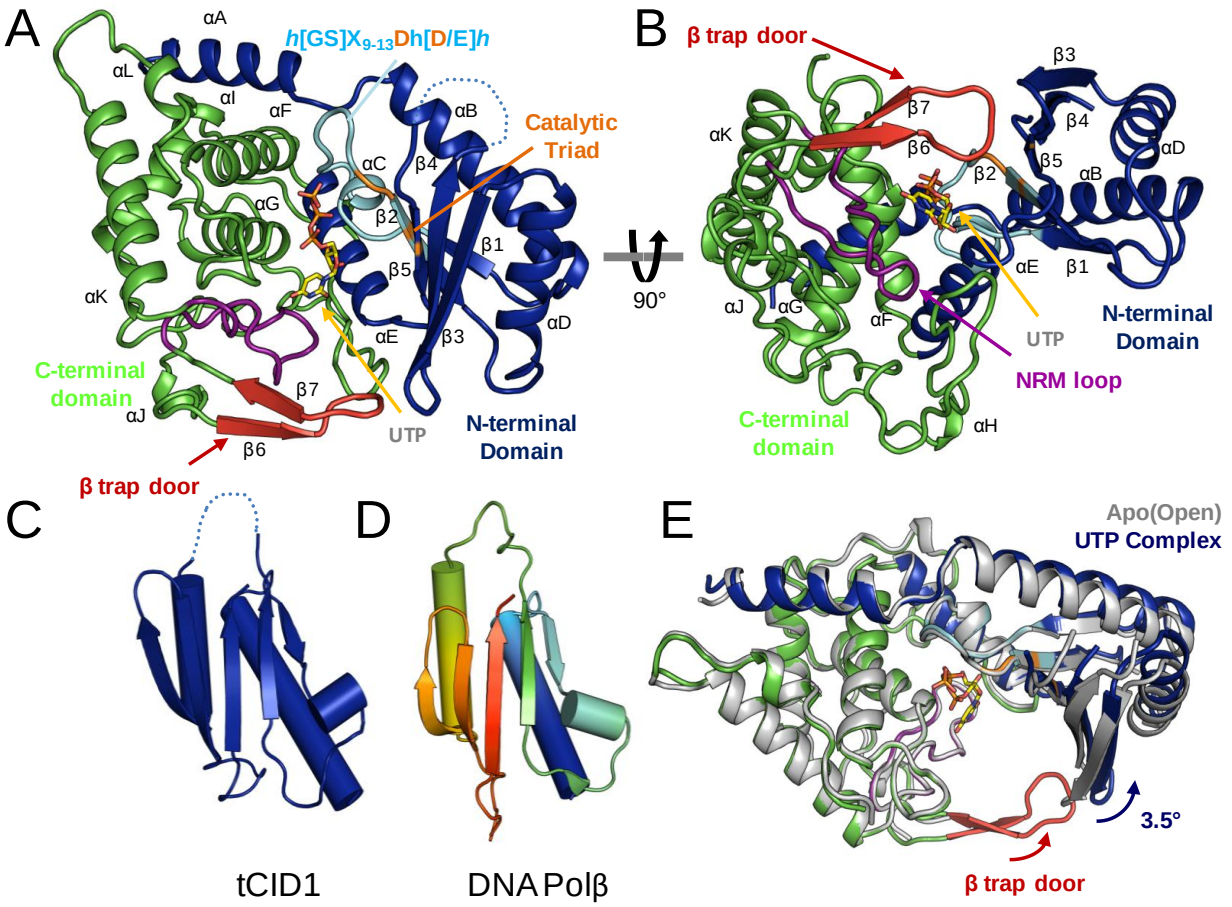


Figure 7.2 – The Structure of Cid1. (A) Overall structure of tCid1 displaying Cid1 as a bi-domain enzyme with an N-terminal domain (blue) and a C-terminal domain (green) bordering a deep central catalytic cleft where UTP is bound in an *anti*-conformation. A Cid1-specific two stranded β-sheet denoted as a trap-door (red) extends from the CTD to the NTD only in the UTP bound form. The secondary structure elements are labelled along with the consensus catalytic sequence *h*[GS]$X_{9\text{-}13}$Dh[D/E]*h*, (cyan) where *h* represents a hydrophobic residue and X represents any amino acid. The NTD-side of the catalytic cleft houses three Aspartate residues essential for catalytic activity (orange). (b) Orthogonal view of the tCid1 crystal structure and labelled as in (A). (C) The N-terminal domain of tCid1 (rendered blue) together with the palm domain of DNA polymerase β (PDB: 1BPB, coloured as rainbow), illustrating the topology and structural conservation of the catalytic domain of the polβ nucleotidyltransferase family. Secondary structure elements are labelled as for (A) for Cid1 and according to and (Sawaya *et al.*, 1994) for DNA polβ. (D) Comparison between Apo (grey) and UTP-bound forms (coloured as in A) showing an slight domain motion and the deployment of the trap door in a UTP-bound state only.

Table 7.1. - Data Collection and Refinement Statistics

	tCid1 Apo (I)	tCid1 UTP Bound	tCid1 Apo(II)
Data Collection			
Space Group	C2	C2	$P2_12_12_1$
Cell Dimensions			
a, *b*, *c* (Å)	164.4, 78.0, 152.5	164.7, 78.0, 151.7	81.2, 101.2, 113.6
α, *β*, *γ* (°)	90, 90, 109.5	90, 90, 109.3	90, 90,90
Resolution (Å)	64.24–3.20 (3.28–3.20)*	77.72–3.02 (3.10-3.02)	56.8 – 2.6 (2.74 – 2.6)
R_{merge} [a] (%)	13.8 (71.2)	9.3 (47.9)	7.3 (48.1)
R_{pim} [b] (%)	7.8(40.2)	5.5 (27.9)	1.8 (11.0)
$I/\sigma I$	8.9 (2.4)	12.8 (3.2)	24.7 (6.2)
Completeness (%)	99.5 (97.7)	99.4 (99.5)	99.6 (99.7)
Redundancy	4.1 (4.1)	3.7 (3.9)	18.0 (18.4)
Refinement			
Resolution (Å)	64.27 – 3.20 (3.31 – 3.20)	28.72 – 3.02 (3.11-3.02)	56.78 – 2.60 (2.69-2.60)
No. Reflections	30013	35591	29430
R_{work}/R_{free}	17.88/21.30	17.64/20.11	20.28/24.76
No. atoms			
Protein	10262	10339	5180
Ligand/ion	26	133	95
Water	10	17	79
B factors (Å^2)			
Protein	73.28	61.20	74.25
Ligand	67.00	55.44	107.25
Water	42.95	42.88	63.13
R.m.s deviations			
Bond lengths	0.009	0.009	0.009
Bond angles	0.93	0.98	0.98
Ramachandran			
Favoured (%)	98.6	98.1	96.2
Allowed (%)	1.4	1.9	3.8

* Values in parentheses account for the highest resolution shell,

[a] $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$

[b] $R_{p.i.m.}$ (precision-indicating merging R-value) $= \sum_{hkl}[1/N-1]^{\frac{1}{2}}\sum_i \left|I_i(hkl) - \overline{I(hkl)}\right| / \sum_{hkl}\sum_i I_i(hkl)$

the catalytic domain, housing the three invariant catalytic aspartates. A short conserved helix (αC) is located in a loop between strands β1 and β2 that forms the base of the cleft and contacts the CTD *via* hydrophobic interactions. Importantly the β1-β2 loop is part of the conserved hallmarking catalytic sequence *h*[GS]X_{9-13}D*h*[D/E]*h*, where *h* is a hydrophobic residue, X is any amino acid and the Apsartate/Glutamate residues are catalytic. The catalytic aspartates in Cid1 are located on strands β2 (Asp^{101} and Asp^{103}) and β5 (Asp^{160}), all of which have previously been shown to be essential for catalysis (Rissland *et al.,* 2007) and, like other enzymes is thought to co-ordinate a single magnesium ion. Interestingly, the most N-terminal helix that belongs to this domain is distant from the core NTD fold and packs onto the CTD *via* extensive hydrophobic interactions.

The C-terminal domain (CTD) is predominantly an α-helical domain (α5 – α9) that shares structural homology with the central domain of poly(A) polymerase (PAP) (Bard *et al.,* 2000) and C-terminal domains of the Trypanosomal TUTases (Stagno *et al.,* 2007b). Sequence comparisons of the C-terminal domains of enzymes in the polβ-like polymerase class, including; interferon-induced oligo(A) synthetase (OAS), poly(A) polymerases, Trf4-family polymerases, CCA-adding enzymes and trypanosomal uridylyltransferases, revealed a conserved region with a corresponding structural relatedness with one another in addition to the conserved catalytic domain (Rogozin *et al.,* 2008). This so-called polβ-C domain (PβCD) is structurally conserved, possessing four helices and is most closely related to the regulatory nucleotide-binding ATP-cone found in ribonucleotide reductase (Rogozin *et al.,* 2008). Interestingly, the primary sequence of *S. pombe* Cid1 was included in alignments performed by Rogozin and colleagues (2008) and therefore domain boundaries for the CTD, or PβCD, of Cid1 were suggested. The C-terminal half of the Cid1 crystal structure possesses a four

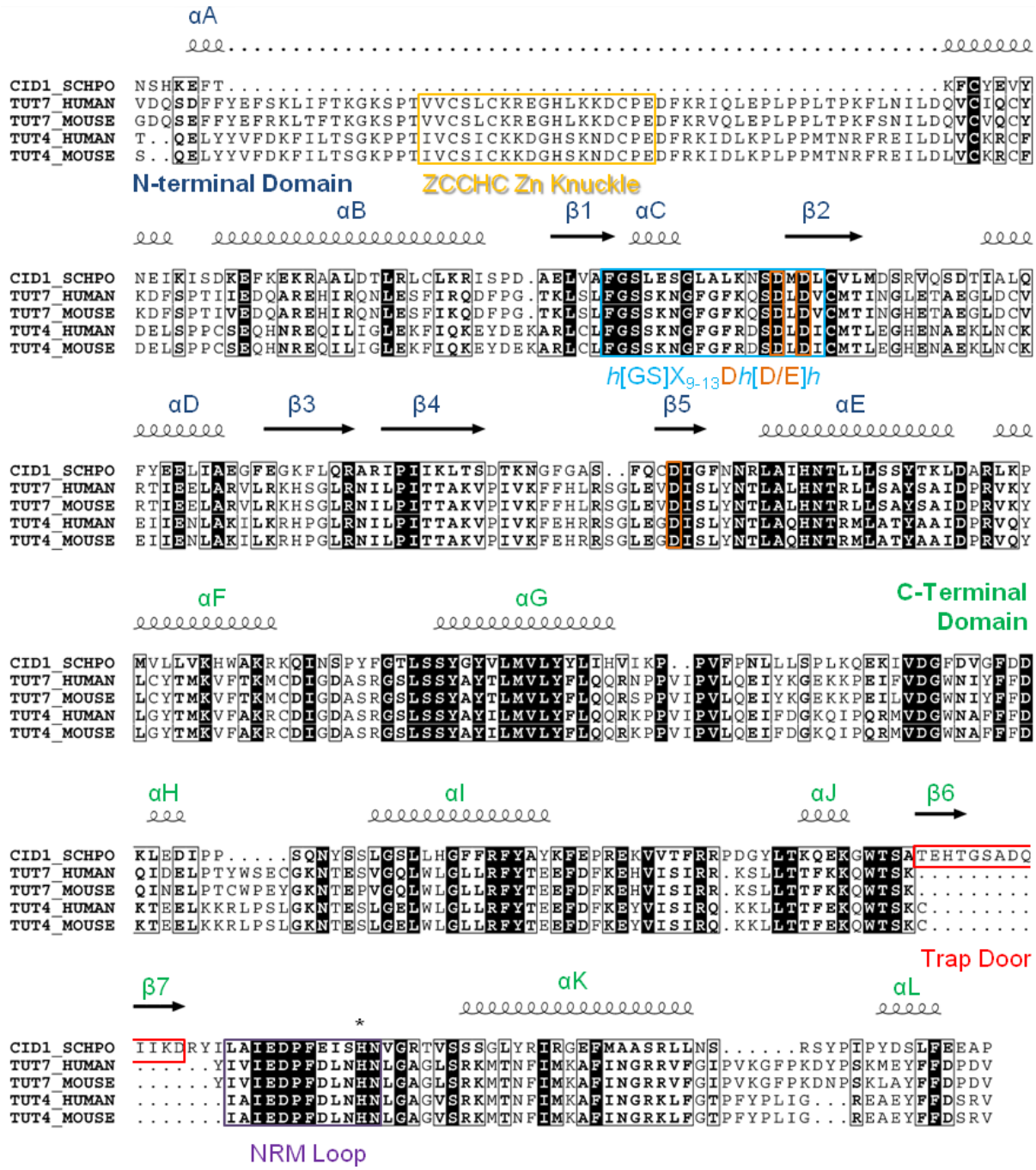


Figure 7.3 – Sequence Alignment Between Cid1 and the ZCCHCs. The Cid1 sequence observed in the crystal structure is aligned to the Cid1 homology region of the ZCCHC6/TUT7 sequence from human (*H. sapien*) and mouse (*M. musculus*) and the ZCCHC11/TUT4 sequences from the same species. Sequences were aligned in pairs in CLUTALW and then aligned together manually. The final alignment together with the secondary structure elements of the Cid1 crystal structure was prepared using Espript (Gouet *et al.*, 1999). Invariant residues are highlighted as white letters on a black background, conserved residues are boxed in black and variable residues are rendered black. The secondary structure elements are labelled as in Figure 7.2 and important features of Cid1 highlighted in boxes and labelled.

helix bundle, reminiscent of an ATP-cone domain, which is consistent with predictions made by Rogozin and colleagues. Accordingly, this four helix bundle in tCid1 structure is considered analogous to the PβCD and forms the core of Cid1's C-terminal domain.

The interface between the two domains is limited due to the large, solvent accessible, active site cleft. The 'trap door' structure from the CTD and the β1-β2 loop from the NTD form interdomain contacts. Additionally, helices α1 and α2 of the NTD traverse both domains and form a 'spine' to the overall structure - with the α1 helix packed against the surface of the CTD *via* hydrophobic residues and the α2 helix being integral to the fold of the NTD. Due to the limited interface between the domains there are some minor movements of the N-terminal domain with respect to the C-terminal domain between the four molecules in crystal form I.

Sequence comparisons between Cid1 and the ZCCHC11 and 6 suggest that the catalytically active portion of the ZCCHCs is homologous to the two domains present in Cid1 (Figure 7.3). During initial investigation of the ZCCHCs for structural studies (not this work), it was noted that the disorder probability plot of Cid1 calculated by the disorder prediction program RONN (Yang *et al.*, 2005) was reminiscent of the disorder probability plot corresponding to the catalytic portion of the ZCCHCs predicted by the same program. Therefore the enzymes were directly compared using this disorder probability distribution and a CLUSTALW sequence alignment. In Figure 7.4 the constructed sequence *versus* probability of disorder plot for each enzyme is shown, where they have been manually aligned based upon the sequence similarity from CLUSTALW sequence alignment. From the sequence aligned disorder probability plots, it is clear that the pattern of disorder probability of Cid1 matches a region of the ZCCHCs disorder plot corresponding to the catalytic portion. This, along with functional similarities proposed here and elsewhere (Rissland *et al.*, 2007) suggests a probable structural homology between the catalytic portion of ZCCHCs and Cid1.

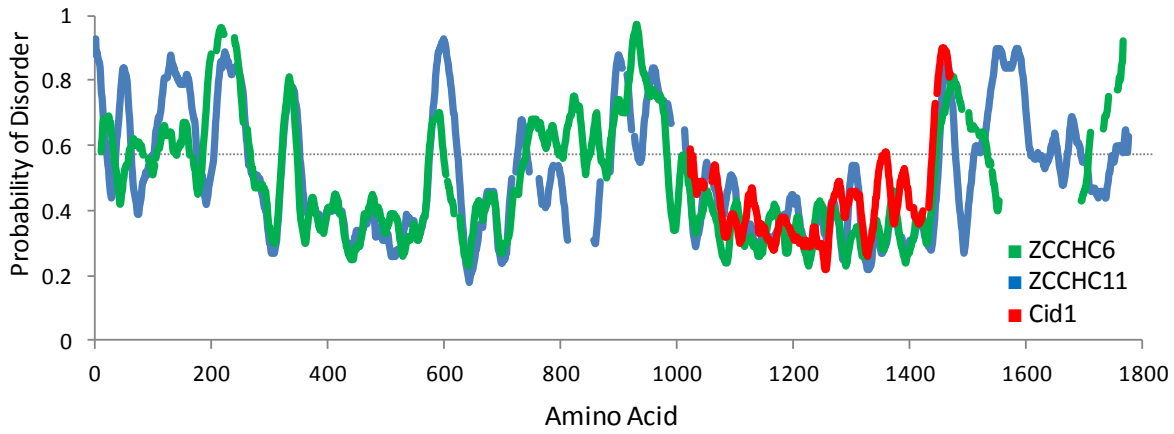


Figure 7.4 – Disorder Probability Plot of Cid1 and the human ZCCHC TUTases. Disorder probability was predicted based upon the amino acid sequence of Cid1 (Uniprot accession O13833), human ZCCHC11 (Uniprot accession Q5TAX3) and human ZCCHC6 (Uniprot accession Q96KX5) using the regional order neural network (RONN) (Yang *et al.* 2005). Probabilities of disorder were predicted an a single amino acid basis and therefore the amino acid sequences were aligned manually based upon a CLUSTALW alignment along with their corresponding disorder probabilities. Gaps were introduced, as for CLUTALW alignment and the disorder predictions plotted against the total number of amino acids and gaps.

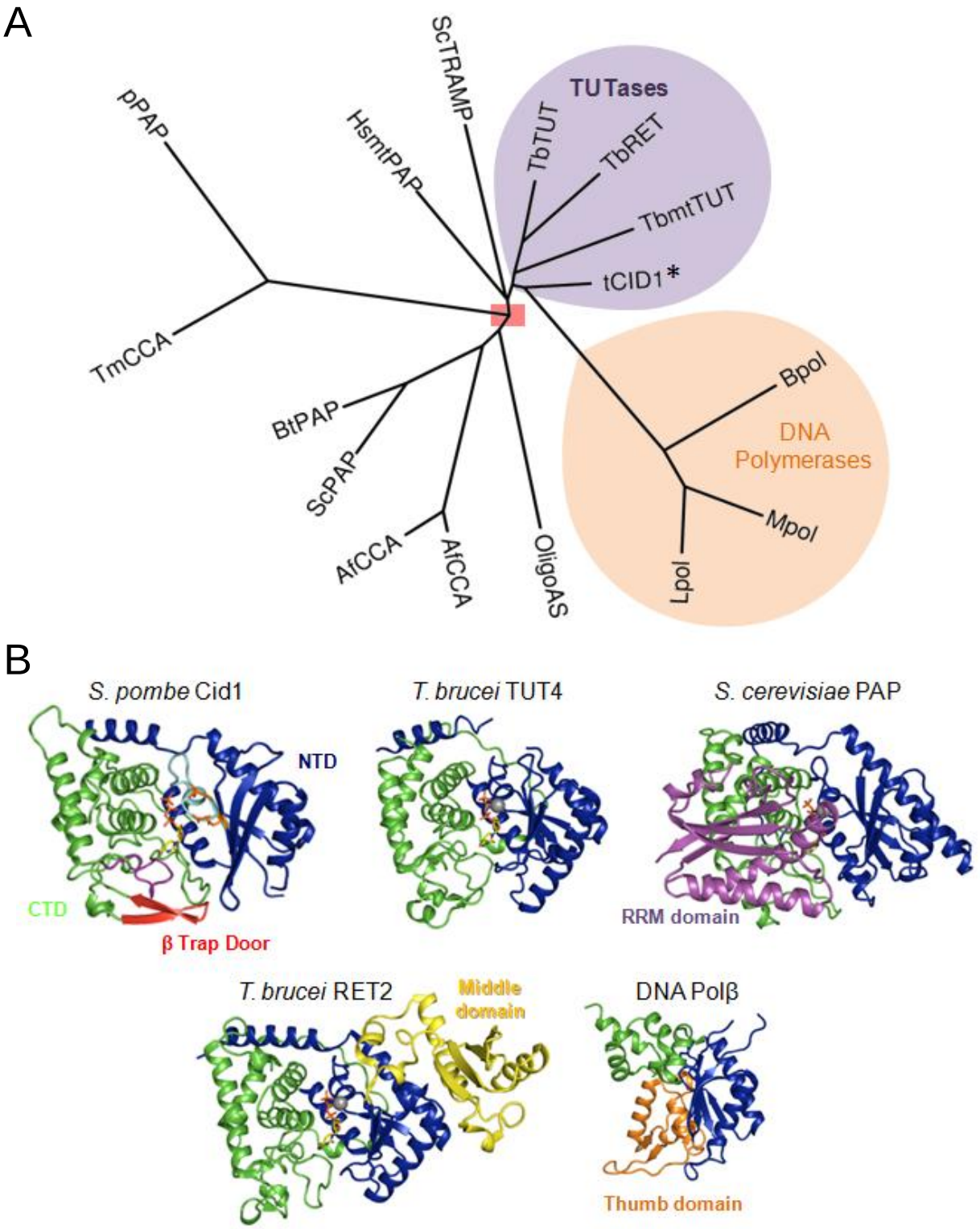
7.1.5. Structural Comparison of Cid1 and Other Polymerase β Family Nucleotidyltransferases

A structural similarity search using the Cid1 structure in DALI (Holm and Rosenstrom, 2010) indicated that Cid1 was most similar to other known TUTases from *T. brucei*. However despite reasonably high Z-scores, these enzymes possessed less than 25% sequence identity. The top three structural homologues, according to DALI are the Trypanosomal uridylation enzymes, minor editosome-associated TUTase1 (MEAT1) (PDB code; 3HJ1, Z-score = 29.4, r.m.s.d. = 2.6Å, 20% identity), RNA editing complex, RET2 (PDB code; 2B56, Z-score = 25.6, r.m.s.d. = 2.6Å, 24% identity), and TUT4 (PDB code; 2Q0E, Z-score = 25.5, r.m.s.d. = 2.6Å, 25% identity). Following the expected TUTases is the

yeast non-canonical poly(A) polymerase Trf4 (PDB code; 3NYB, Z-score = 24.1, r.m.s.d. = 3.2Å, 23% identity) and the yeast canonical poly(A) polymerase (PDB code; 3C66, Z-score = 23.2, r.m.s.d. = 2.8Å, 17% identity).

Given that Cid1 represents the first cytoplasmic terminal uridylyltransferase (TUTase) structure we performed a structural phylogenetic analysis (Riffel *et al.*, 2002) with several known TUTases, including Cid1, together with CCA-adding enzymes, canonical nuclear poly(A) polymerases and non-canonical cytoplasmic poly(A) polymerases and several DNA polymerases all belonging to the DNA polymerase β nucleotidyltransferase family. Therefore an all-pairs, pairwise structural alignment was performed in SHP (Stuart *et al.*, 1979) and by using the probability of equivalence between the pairwise alignments a phylogenetic tree was constructed to assess the structurally-based evolutionary position of Cid1 with respect to other family members. In this analysis, it was found that the terminal uridylyltransferases were suggested to be closely related, forming a tight cluster and clearly stemming from a common ancestor. Of the TUTase clade, the *T. brucei* minor editosome-associated TUTase 1 (*Tb*MEAT1) is a structural neighbour of Cid1, consistent with DALI-based comparisons and in spite of MEAT1 possessing a helical bridge domain (Stagno *et al.*, 2010) that is absent in Cid1. Interestingly, the *S. cerevisiae* non-canonical poly(A) polymerase Trf4 seemed to be positioned near to the TUT clade along with the human mitochondrial non canonical poly(A) polymerase PAPD1 (Figure 7.5). The DNA polymerases β, μ and λ, unsurprisingly, cluster together, but are more related to Cid1 than any other TUTase. The larger and multidomain canonical poly(A) polymerases and tRNA CCA-adding enzymes are positioned away from the TUTases. This is possibly due to the predominantly larger structurally conserved domains, i.e. the central domain and palm domain of poly(A) polymerase, which are structurally homologous to Cid1 CTD and NTD, respectively. Cid1 is clearly the closest to the root of the phylogenetic tree and therefore is suggested to be closest to the last common

Figure 7.5 (Opposite) – Structural Phylogenetic Comparison of DNA Polymerase β Nucleotidyltransferases. (A) Structural phylogenetic tree constructed using an all pairs-pairwise structural alignment in SHP (Stuart *et al.* 1979) to generate an evolutionary distance matrix. The names of the proteins and their PDB codes corresponding to labels of each branch are given as follows: tCID1, caffeine induced death suppressor protein 1 (this work); TbmtTUT, *T. brucei* minor editosome-associated TUTases 1 (MEAT1), PDB 3HJ1; TbRET2, *T. brucei* RET2, PDB 2B4V; TbTUT *T. brucei* TUTase 4, PDB 2Q0F; ScTRAMP, *S. cerevisiae* Trf4, PDB 2NYB; hsmtPAP, human mitochondrial poly(A) polymerase PAPD1, PDB 3PQ1; pPAP *E. coli* poly(A) polymerase, PDB 3AQK; TmCCA, *B.stearothermophilus* CCA adding enzyme, PDB code 1MIV; ScPAP, *S. cerevisiae* poly(A) polymerase, PDB 2HHP; BtPAP, *Bos taurus* poly(A) polymerase, PDB 1Q78; AfCCA, *A. Fulgidus* CCA adding enzyme PDB 1SZ1 and 1R8C; OligoAS, 2'-5'Oligoadenylate Synthetase, PDB code 1PX5; Lpol, human DNA polymerase λ, PDB code, 1XSL; Mpol, mouse DNA polymerase μ , PDB code, 2IHM; Bpol, rat DNA polymerase β, PDB code, 1BPB. Cid1 belongs to the TUTase clade (purple) but sits between this clade and the DNA polymerase clade (orange). Cid1 is closest to the root of the tree (red) suggesting that it is closest to the last common ancestor of this nucleotidyltransferase family. (B - F) Structural array of the DNA polβ nucleotidyltransferase family, aligned in SHP, demonstrating a structural conservation of the C-terminal domain (green) and N-terminal domain (blue) found in Cid1. Structures are labelled as in the phylogenetic tree.


A
ScTRAMP
pPAP
HsmtPAP
TbTUT
TbRET
TbmtTUT
tCID1*
TUTases
TmCCA
BtPAP
ScPAP
AfCCA
AfCCA
OligoAS
Bpol
DNA
Polymerases
Mpol
Lpol
B
S. pombe Cid1
NTD
CTD
β Trap Door
T. brucei TUT4
S. cerevisiae PAP
RRM domain
T. brucei RET2
Middle
domain
DNA Polβ
Thumb domain

ancestor of the TUTases.

7.1.6. The Architecture of the Active Site: a comparison with other TUTases

As discussed in Chapter 6, the majority of our structural knowledge of 3' RNA uridinylation comes from crystallographic studies of mitochondrial terminal uridine transferase enzymes in trypanosomatids. Thus UTP bound states of these TUTases has revealed the molecular basis for residues that confer selectivity of uridine over other pyrimidines and purines (Ringpis *et al.*, 2010, Stagno *et al.*, 2010, Stagno *et al.*, 2007a, Stagno *et al.*, 2007b, Deng *et al.*, 2005). Owing to Cid1 in complex with UTP being the first structure of a eukaryotic and cytoplasmic TUTase involved in mRNA decay, a comparative analysis between the modality of UTP binding between the trypanosmatid TUTases can be established. Unlike template-dependent RNA polymerases where the selectivity of the enzyme for its nucleotide substrate is governed by Watson-Crick base pairing, the TUTase nucleotidyltransferases are template independent and therefore rely on the active site to confer specificity for UTP. High resolution X-ray crystallographic analysis of the Trypanosomal TUTases, *Tb*RET2 (Deng *et al.*, 2005) and *Tb*TUT4 (Stagno *et al.*, 2007b), where *Tb* refers to *Trypanosoma brucei*, has provided the mechanistic detail that underpins the TUTases' UTP selectivity. In all TUTases solved to date, including Cid1, the UTP molecule is bound between the two N-terminal and C-terminal domains (Figures 7.5 and 7.6). Superimpositions of the active site of Cid1 with that of *Tb*RET2 and *Tb*TUT4 have revealed, as expected, structurally conserved catalytic aspartates, suggesting a conserved catalytic mechanism, but also a closely related UTP binding architecture to the Trypanosomal TUTases despite low sequence similarities between the enzymes.

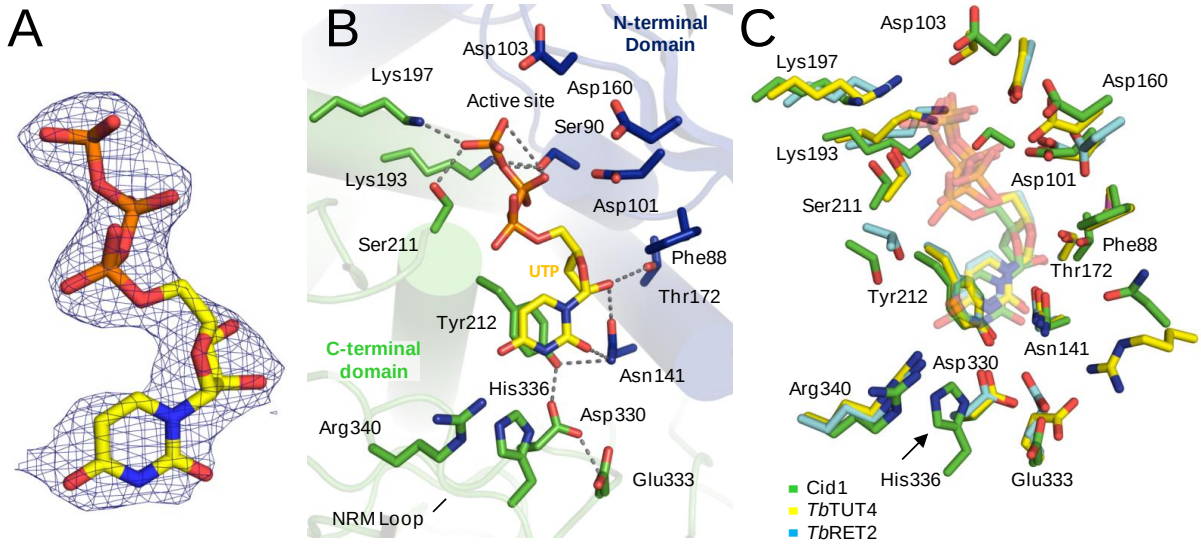


Figure 7.6 – Architecture of the Cid1 Active Site. (A) Unbiased electron density ($F_o - F_c$) map contoured at 2 σ corresponding to a single Uridine Triphosphate (UTP) molecule in the active site. A single UTP molecule model is rigid-body fitted into the electron density for clarity. (B) Detail of the active site of Cid1 showing residues that interact with the UTP molecule. The secondary structure is shown faintly for orientation purposes and is coloured according to the domains as in Figure 7.2. (C) Active site comparison between Cid1 and the high resolution *T. brucei* TUTase structures, TUT4 and RET2, showing a grossly similar modality of UTP binding. Residues that interact with UTP are coloured green for Cid1, yellow for TUT4 and blue for RET2. The residues shown are labelled for Cid1 only. A Cid1 specific Histidine^{336} is shown (arrowed) that is in close proximity (~4.4Å) to the N3 amine of the Uridine base.

The nucleotide in Cid1 is bound into a deep active site cleft at the centre of the enzyme. The uridine base adopts an *anti*-conformation deep in the binding pocket, with each domain forming a side of the cleft and interacting with the UTP molecule. In all three enzymes UTP binding is conferred *via* π-stacking interactions between the pyrimidine ring of uracil and the aromatic ring of an invariant Tyrosine residue (Tyr^{212} in Cid1) (Figure 7.6). What is more, the co-planarity of Tyr^{211} and the uracil base is maintained in all three enzymes by the simultaneous hydrogen bonding of the 2' hydroxyl group of the Tyrosine side chain and a carbonyl oxygen atom (O2 atom) of the uracil pyrimidine ring by the amine of an

Asparagine residue (Asn^{141} in Cid1), which is invariant in *Tb*RET2 and *Tb*TUT4. Furthermore, the carbonyl of Asn^{141} also contacts the 2' hydroxyl group of the ribose moiety. Thus the interaction between Asn^{141}, the pyrimidine carbonyl and Tyr^{212} hydroxyl group allows the maintenance of the ring stacking interactions between the uracil base and the aromatic ring of the Tyr^{212} residue but also, through the interaction of Asn^{141} allows the concomitant selection of ribose sugars found in RNA (Figure 7.6). Due to the resolution of the tCid1•UTP crystal structure it is not absolutely clear whether the uracil base of the bound UTP molecule is in an *anti*-conformation, as is the UTP in the high resolution crystal structures of Trypanosomal TUTases (Stagno *et al.*, 2007b; Deng *et al.*, 2005). However, based upon the conservation of a Uridine interacting Asparagine residue in all three structures together with guidance from the electron density omit map during refinement it became clear that the Uridine base adopts an *anti*-conformation and that the conformation of Asn^{141} must allow contacts with the carbonyl of the pyrimidine ring, hydroxyl of Tyr^{212} and 2' hydroxyl group of the sugar (Figure 7.6). Recently Kondo and Westhof (2011) classified protein-nucleotide base pseudo-base-pairing using structures of protein-nucleotide complexes and categorised this type of interaction as a sugar-edge pseudo pair. Furthermore, the ribose moiety is further discriminated by additional interactions between the 2' hydroxyl group of the ribose sugar, which is in a C3'-*endo* conformation, with the side chain hydroxyl group of a Threonine residue in Cid1 (Thr^{172}) but a conserved Serine in both Trypanosomal enzymes (Figure 7.6). Interestingly, both Cid1 and *Tb*TUT4, but not *Tb*RET2, possess a Phenylalanine residue (Phe^{88} in Cid1) in close proximity to the ribose of UTP (Figure 7.6) that are structural homologues of Phe^{100} of poly(A) polymerase (Martin *et al.*, 2000) and Tyr^{37} in kanamycin nucleotidyltransferase (KanNt) (Pedersen *et al.*, 1995). In the mammalian poly(A) polymerase, Phe^{100} is within van de Waals contact with the ribose sugar moiety and its mutation into an aspartate in PAP has a detrimental effect on k_{cat} (Martin *et al.*, 1999). A

mutation of Phe^{88} in Cid1 to an aspartate, analogous to (Martin *et al.*, 1999), results in an enzyme with no activity (O.S. Rissland, unpublished results). The tri-phosphates are co-ordinated by direct hydrogen bonding with Serine and Lysine residues (Ser^{90}, Ser^{211}, Lys^{193}, Lys^{197}) that are invariant amongst the polβ nulceotidyltransferase family. Importantly, the highly conserved polβ nulceotidyltransferase catalytic domain possesses three invariant aspartates in close proximity to the α and β phosphate groups of the UTP molecule. These aspartates, like other enzymes, are though to co-ordinate metal ions.

A structural comparison between the both Apo structures and the UTP-bound enzyme reveals that the enzyme only displays small movements of the N-terminal domain, with the UTP molecule forcing the cleft to widen, slightly, in its presence. A close inspection of the UTP-binding amino acid side chains also reveals that the majority of the residues move slightly, but do not deviate far in their UTP-binding positions. A single residue side chain, Asparagine^{141} displays an alternative conformation (rotamer) in the absence of UTP. In both Apo structures, Asparagine^{141} interacts with Aspartate^{330}, whereas in the UTP bound structure Asn^{141} is responsible for the simultaneous interaction between Asn^{171}, the pyrimidine carbonyl and ribose 2' OH at the sugar-base edge and Tyr^{212}. Given the importance of the Uridine sugar-edge interaction with Asn^{141} for the maintenance of planarity of base and Tyr^{212} it could be suggested that Cid1 employs an induced fit mechanism for UTP binding.

7.1.7. Structural and Biochemical Identification of a Uridine Selection Mechanism

As discussed previously, polβ nucleotidyltransferases possess a conserved structure within the C-terminal domain, which is structurally homologous to the regulatory ATP-cone

module found in ribonucleotide reductase. This ATP-cone structure at the core of the ribonucleotidyltransferases is proposed to possess the so-called nucleotide recognition motif within an extended loop (Martin and Keller, 2007). This nucleotide recognition motif contacts the Adenine base in poly(A) polymerase (Martin and Keller, 2007; Martin *et al.*, 2000) and the Uridine base in the *T. brucei* terminal uridylyl transferases (Stagno *et al.*, 2007b, Deng *et al.*, 2005) and also in our Cid1 crystal structure. A superposition of the Cid1 crystal structure bound to UTP with poly(A) polymerase bound to ATP (PDB: 1Q78), *T. brucei* TUTase RET2 (PDB: 2B56) and *T. brucei* TUTase TUT4 (PDB: 2IKF), bound to UTP, demonstrates a structural homology between the 'Fingers' (C-terminal domain in TUTases) and 'Palm' domains (N-terminal domain in TUTases) and also a structurally conserved nucleotide binding site. Furthermore, the loop that contacts the nucleotide base, possessing the so-called nucleotide recognition motif (NRM), in these four cases is also structurally conserved, with this loop in Cid1 between α-helices αK and αI (Figure 7.2). This NRM loop is part of an extended unstructured region that stacks onto the surface of the C-terminal domain through conserved hydrophobic interactions. The NRM portion sits on top of the ATP-cone-like module and extends towards, and contacts, the N-terminal domain where it is held in place by a single conserved hydrophobic interaction with the N-terminal domain. Although this loop is unstructured it is well ordered with average main chain B-factors in the Cid1-UTP crystal structure of ~44Å^2 compared to an overall mean B-factor of 57Å^2. This structurally homologous loop is also conserved in other non-canonical poly(A) polymerase crystal structures, for example *S. cerevisiae* Trf4 (PAP2), but sequence comparisons show that the sequences of the canonical poly(A) polymerases are well conserved whereas the non-canonical poly(A) polymerases are more divergent but are closely related to one another (Figure 7.7). Martin and Keller (2007) have suggested two distinct nucleotide recognition motifs (NRM) within the loop of the ATP-cone are characteristic of

the canonical and non-canonical ribonucleotidytransferases. Given that the non-canonical poly(A) polymerases include both professional poly(A) and poly(U) polymerases this structurally conserved NRM loop can be divided further based on UTP or ATP selectivity.

The high resolution *T. brucei* TUTase crystal structures have allowed a detailed examination of uracil selectivity over other pyrimadines. In RET2 and TUT4, the uridine base is recognised by a series of invariant amino acid side chains together with a network of water molecules. The selectivity of uridine over other pyrimidines is determined by acidic residues Asp^{297}or Asp^{421} and Glu^{300}or Glu^{424} (where amino acid numbering refers to TUT4 and RET2, respectively) that cluster around the N3 amine group of the uracil base (Figure 7.7). The aspartate residues contact the uracil amine *via* a co-ordinated water molecule (denoted Wat1 in both enzymes). In addition, the very high resolution crystal structure of mitochondrial minor-editosome associated TUTase1 (MEAT1) from *T. brucei* also reveals a conserved aspartate, Asp^{335} that also contacts the N3 amine group of the Uridine pyrimidine ring *via* a co-ordinated water molecule (Figure 7.7).

In RET2 the hydrogen bonding pattern between the water and the carboxyl oxygens of the invariant acidic residues is such that the oxygen atom of the water can only accept hydrogen bond donors, in this case, the hydrogen from the endocyclic (N3) amine of the pyrimidine ring. This co-ordination of the water (Wat1) by both glutamate and aspartate confers the absolute specificity of the enzyme for uracil, as a cytosine pyrimidine ring nitrogen atom is a hydrogen bond acceptor and would therefore be repelled by the lone pair of electrons of the oxygen of the co-ordinated conserved water. Simplistically, the highly co-ordinated water serves as a hydrogen atom detection mechanism for the N3 amine in the pyrimidine ring.

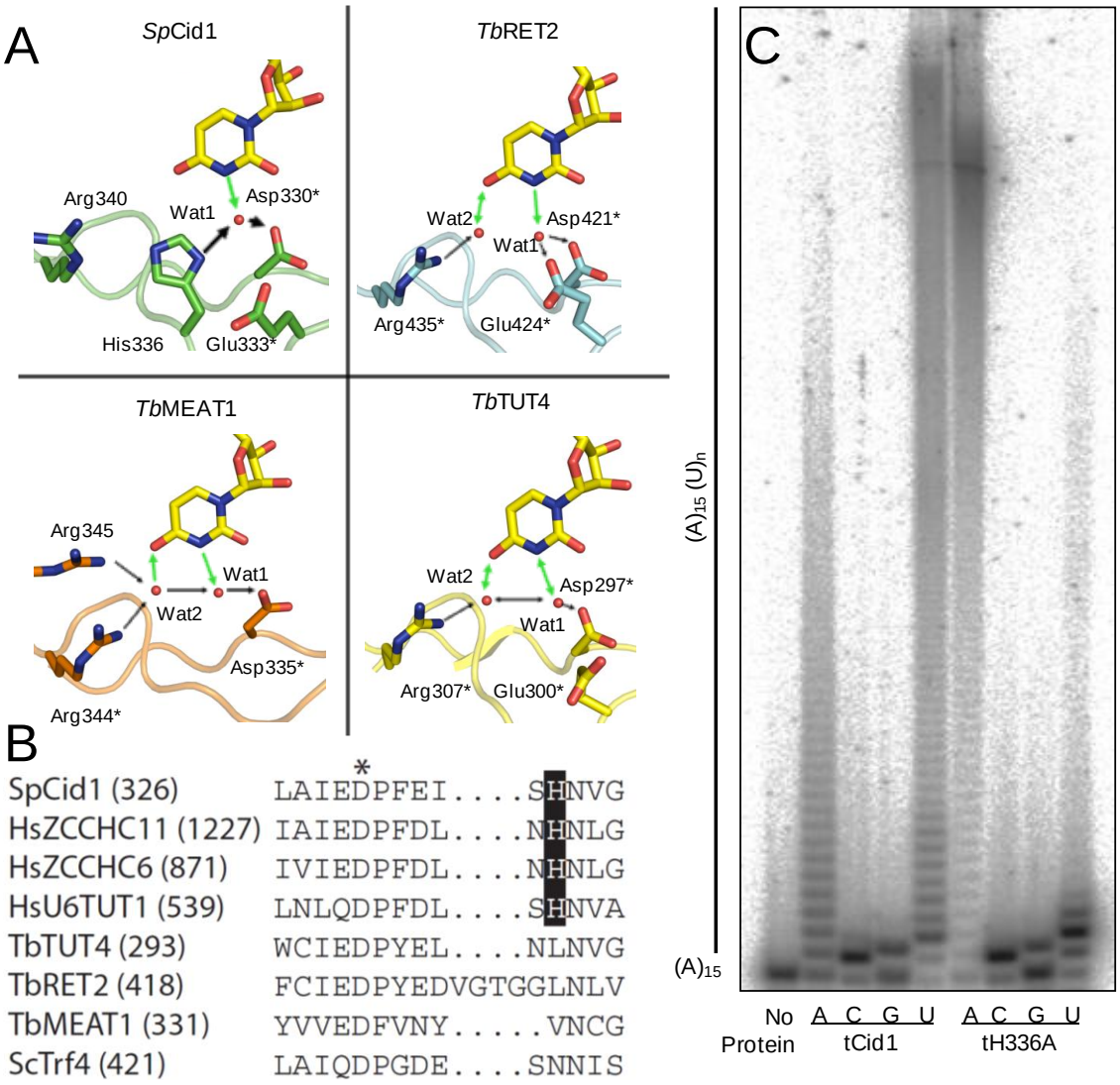


Figure 7.7 – A Uracil-selectivity Mechanism for Cid1. (A) Comparison of Uracil selection mechanisms found in Cid1 and the high resolution crystal structures of TUT4 (PDB: 2IKF), RET2 (PDB: 1B51) and MEAT1 (PDB: 3HJ1) from *T. brucei*. Highly co-ordinated waters that form hydrogen bonds with the Uridine base from the *T. brucei* structures are shown as red spheres. In Cid1 a proposed water is modelled (from 3HJ1) based on structural superposition. Hydrogen bonds involved in achieving U-specificity are shown as green arrows. Invariant residues shared by the trypanosomal TUTases and Cid1 are asterisked. (B) Sequence alignment of NRM loops from the polβ nulceotidyltransferase family. Histidine^{336} is highlighted as white on a black background along with the conserved Histidines present in the metazoan enzymes. An invariant Aspartate found in all the enzymes is asterisked. (C) Substitution of His^{336} to Ala converts the enzyme's native polyuridylation activity into polyadenylation activity *in vitro*. Cid1 polymerase activity assays were performed in the presence of each nucleotide as described in Rissland *et al.* (2007) using 1pmol of purified enzyme. Panel (C) was kindly provided by S. Fleudepine and C. Norbury, Sir William Dunn School of Pathology, University of Oxford, UK.

In the *T. brucei* enzymes, TUT4 and MEAT1, a network of hydrogen bonds between two co-ordinated water molecules achieves Uridine specificity (Figure 7.7). The second water molecule in both cases is co-ordinated by Arginine (Arg^{344} Arg^{345}, in MEAT1 or Arg^{307} in TUT4) residues, which allows the second water molecule (Wat2) to contact Wat1 and the carbonyl oxygen atom of the Uracil pyrimidine ring (Figure 7.7). Thus two co-ordinated water molecules in MEAT1 and TUT4 confer selectivity by contacting the Watson-Crick edge of the Uridine base. RET2 also contacts the carbonyl of the Uridine base *via* a second water molecule co-ordinated by an equivalent Arginine (Arg^{435}). An active site comparison between UTP bound Cid1 and the *T brucei* TUTase crystal structures, together with a sequence comparison, demonstrates a conservation of both the invariant Asparatate and Glutamate residues (Asp^{330} and Glu^{333} in Cid1) that are responsible for the pyrimidine amine 'detection' and a conservation of the second invariant Arginine (Arg^{340} in Cid1) in uracil selectivity found in the *T. brucei* TUTase (Figure 7.7). Additionally, a Cid1 specific Histidine (His^{336}) residue is situated within the NRM loop and is positionally equivalent to a Leucine (Leu303 or Leu431) residue in the *T. brucei* TUTases and comes into close proximity with the Uridine base (Figure 7.7). Given the intermediate resolution of the tCid1 crystal structure, potential structural waters responsible for uracil selection were not observed. Therefore, a biochemical analysis was undertaken in collaboration with Dr S. Fleurdepine and Dr C.J. Norbury (Sir William Dunn School of Pathology, University of Oxford) to determine the contribution of residues in the proposed uracil selection mechanism of Cid1. Site-directed mutagenesis of the conserved Aspartate and Glutamate residues in TAP-purified Cid1 (D330A and E333A), equivalent to those found in the *T. brucei* enzymes RET2 and TUT4 and suggested to comprise a universal U-selection mechanism in TUTases (Stagno *et al.,* 2007a), surprisingly, did not affect the poly(U) polymerase activity (O. Rissland, unpublished observations). Both mutant enzymes possess wild-type processivity and UTP incorporation

onto an A_{15} primer. Additionally, these mutations did not alter the very weak/absent poly(C) or poly(G) activity. This suggests that Cid1 employs a different mechanism of U-selection that differs from the Trypanosomal TUTases, despite possessing analogous residues. A structural comparison of Cid1 with the trypanosomal TUTases led to the identification of a Cid1-specific Histidine (His^{336}) residue that is positioned in close proximity (~4.4Å) to the endocyclic amine of the Uridine base, a uracil moiety often utilised by *T. brucei* TUTases in nucleotide discrimination (Figure 7.7). Substitution of His^{336} for Alanine resulted in a recombinant enzyme with the complete loss of poly(U) activity. Unexpectedly, together with the loss of poly(U) activity, the enzyme was transformed into a proficient poly(A) polymerase (Figure 7.7). Interestingly, poly(C) and poly(G) activity was absent/very weak, analogous to wild-type Cid1. These data, along with the structure, confirms that $Histidine^{336}$ is responsible for the selectivity of pyrimidines over purines but more importantly Uridine over Adenosine. What is more, this Histidine residue is also invariantly conserved in the mammalian TUTases (Figure 7.7), suggesting a mechanism distinct from trypanosomes, which is conserved in *S. pombe* and mammals. Given the striking effect of the Histidine substitution on poly(U) activity of Cid1 it was still unclear how this interaction may be mediated given that the distance between the His^{336} and the uracil, particularly the endocyclic amine moiety, is too great for direct hydrogen bonding. Therefore it was proposed that a single water molecule may be co-ordinated by the N3 group of the uracil and the amine of the imidazole side chain of histidine, effectively bridging the gap. A superposition of the active sites of Cid1 with TUT4, RET2 and MEAT1 all in complex with UTP reiterated that uracil selection is mediated by highly co-ordinated waters (Wat1 and Wat2) in Typanosomes, but demonstrated a conserved region where the waters from each enzyme clearly cluster. We therefore modelled the water molecules equivalent to Wat1 from each structure onto the Cid1 active site and measured the distances between the His^{336} and the Uridine base. This

suggested distances of 2.7Å – 3.1Å between the water molecule Wat1 and the amine of the imidazole of His^{336} and distances of 1.9Å - 2.5Å between Wat1 and the endocyclic amine of the pyrimidine ring – both of which are within the range of putative hydrogen bonding distances. Waters equivalent to Wat1 are also in close proximity (~3.0Å) with the invariant Apsartate (Asp^{330}) residue, which in all structures of *T. brucei* TUTases thus far participates in the co-ordination of Wat1. However, given that the substitution of $Aspartate^{330}$ to Alanine in Cid1 has no marked effect of the enzyme's Uridine-selectivity, it could be suggested that this residue plays a minor role in Uracil discrimination.

Using sequence comparisons between several TUTases from metazoans, namely the ZCCHC11 and 6 from human and mouse (Figure 7.7), in particular the region corresponding to the NRM loop, reveals that the U-selection dependent Histidine residue in Cid1 is invariant in the other known TUTases. What is more, the sequence conservation between all residues in the NRM loop is high therefore suggesting this mechanism of uracil identification is conserved in *S. pombe* and mammals yet absent from trypanosomal TUTs. This is important for two reasons, (1) providing more evidence for the relatedness of Cid1 and its metazoan orthologues, and perhaps a model for U-selectivity, and (2) suggesting a mechanism distinct from *T. brucei* allowing anti-trypanosomal compounds to be explored for the treatment of trypanosomiasis (African sleeping sickness) without a affecting the host TUTases.

7.1.8. tCID1 Efficiently Binds to RNA Which is Important for TUT Activity

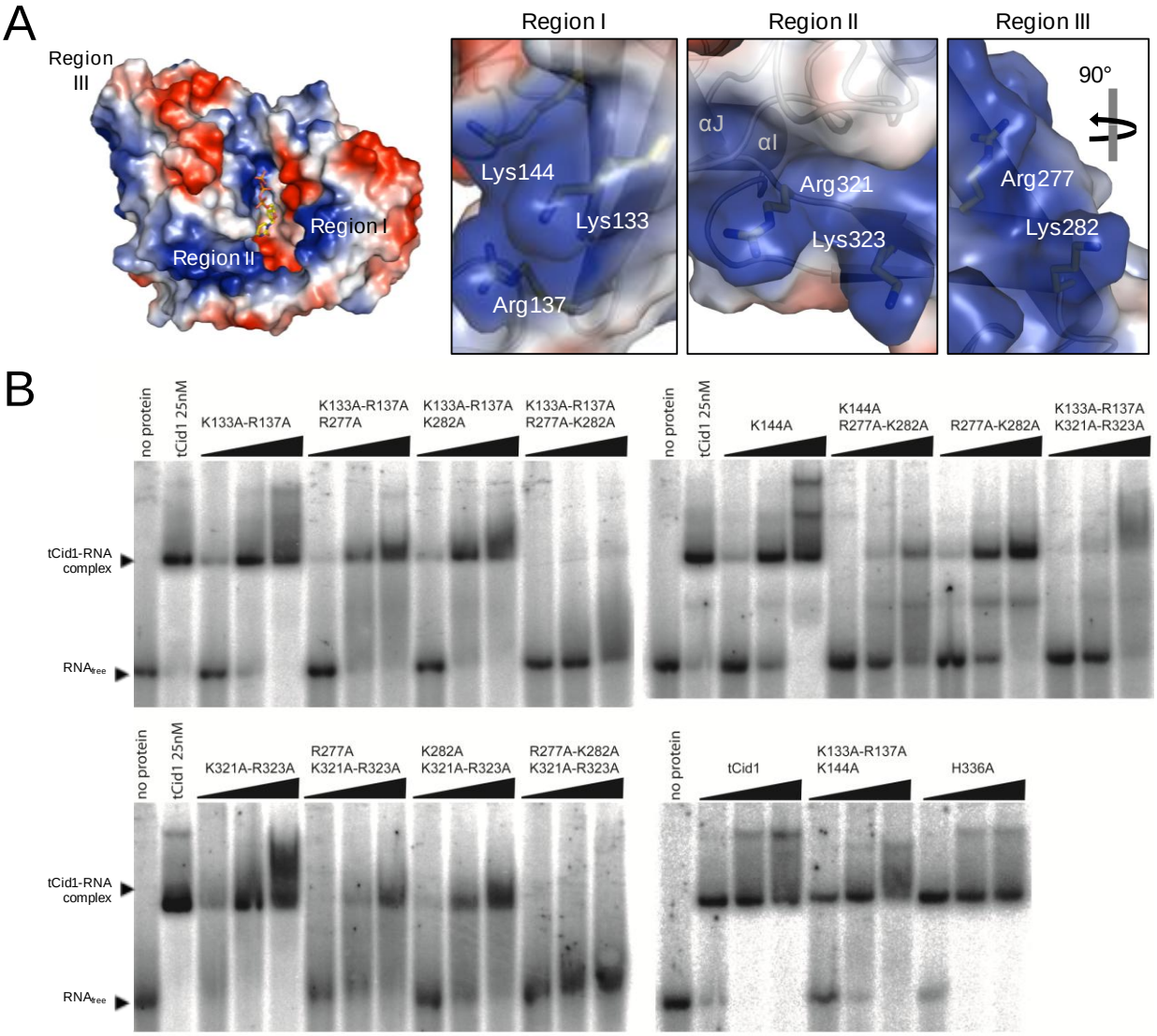
During the purification of tCid1, we observed the contamination of purified protein with RNaseA and DNaseI sensitive *E. coli* nucleic acids (see Chapter 6). Unlike other members of the Polβ superfamily such as the non-canonical PAP Trf4, Cid1 has not been found in a stable multi-protein complex with RNA-binding proteins and, unlike canonical

PAPs, Cid1 lacks an RNA recognition motif. Putative evidence, in the form of a crude electrophoresis mobility shift assay using purified protein and DNA molecular weight standards (Sigma) clearly demonstrated that Cid1 was capable of binding to DNA. Further investigation, as a part of a collaboration with Dr. Sophie Fleurdepine and Dr. Chris Norbury (Sir William Dunn School of Pathology, University of Oxford, U.K.) futher provided evidence that Cid1 was capable of binding to ssRNA, ssDNA, dsDNA, in a sequence independent manner, but was unable to bind to highly structured RNAs such as tRNA (S. Fleurdepine, unpublished observations). It was surmised that Cid1 must therefore contain a cryptic RNA binding domain or site(s), which on the basis of the nucleic acids bound is likely to depend, in part, on sequence-independent charge interactions. This was clarified when the structure of Cid1 was determined, which revealed several prominent basic patches on the surface of the enzyme. Of the basic patches on the surface of the enzyme, three striking basic regions (denoted as region I, II and III) were of particular interest. These regions are primarily comprised of Arginine and Lysine clusters. The residues that contribute to each basic region are as follows; Region I, which sits on the edge of the β-sheet of the palm domain, is composed of Lys^{133}, Arg^{137} and Lys^{144}; Region II, which lies close to the so called 'trap door', is composed of Arg^{321}, Lys^{323} with His^{336} and Arg^{340} in close proximity; Region III, situated on the far side of the C-terminal domain, is composed of Arg^{321} and Lys^{323} with Arg^{359} and Arg^{364} in close proximity. The residues that comprise these basic regions are dispersed throughout the primary sequence but cluster together in the three-dimensional structure, thereby eluding any prediction. Basic regions I and II are in close proximity to one another and to the active site cleft either side of the catalytic cleft entrance. Region III is located on the surface of the C-terminal domain some distance from the active site, but together with Region II forms a perceptible basic groove that is proposed to be the path of the RNA. In order to investigate the biological relevance of such interactions a biochemical

analysis was undertaken in collaboration with the laboratory of Dr. C. Norbury, Sir William Dunn School of Pathology, University of Oxford, UK.

The RNA binding capacity of tCid1 was proved using electrophoretic mobility shift assays, in which the presence of a protein: RNA complex retards the mobility of the RNA molecule during gel electrophoresis. Compared to free RNA the presence of purified tCid1 markedly altered the migration of the RNA molecule at all concentrations used (see Figure 7.8). Interestingly, with increased protein concentration an additional higher molecular weight band became apparent possibly arising from multiple tCid1 molecules binding to a single RNA molecule. Furthermore, early studies revealed that tCid1 is capable of binding to ssRNA and not structured RNAs such as tRNA (S. Fleurdepine and C. Norbury, unpublished observations). The principal residues that contribute to each region were substituted for alanines, analogous to surface entropy reduction (SER) mutations which do not disturb protein solubility (Derewenda *et al.*, 2004). Therefore two or three residues that contribute significantly to the electrostatic characteristics of each region were substituted for Alanines and the RNA binding capacity of the resulting mutant protein analysed using EMSA. Initially, Lysine or Arginine substitutions were made for each region and each tested in turn. Substitutions in regions I and II demonstrated RNA binding that was slightly impaired observed by a reduction in proportion of RNA shifted at low concentrations of mutant compared to tCid1. Substitutions in region III impaired RNA binding further, with a relatively greater proportion of free RNA at low concentrations (Figure 7.8). In all cases, impaired RNA binding was surmounted by increasing the protein concentration. Combination mutagenesis of residues from different patches revealed that substitutions of patches with additional residues from other patches impaired the tCid1-dependent mobility shift of the RNA substrate and in most cases could be overcome by increasing the protein concentration (Figure 7.8). However, the RNA binding capacity of tCid1 could be ablated at all protein

Figure 7.8 (Opposite) – tCid1 Efficiently Binds to RNA through a Series of Basic Surface Patches. (A) Electrostatic surface representation of Cid1-UTP crystal structure. Basic patches are evident on the enzyme's surface and are denoted as region I, II and III. The molecular detail of each patch is shown in panels to the right. In order to see the detail of region III the structure is rotated by 90°. (B) Electrophoresis mobility shift assay (EMSA) demonstrating that 'native' tCid1 is capable of binding to RNA (40 nucleotides) as shown by the shift in RNA compared to the 'no protein' control. Single, double, triple and quadruple surface mutants were generated, based upon the principal residues to each basic patch, and tested in EMSA to determine the contribution of each residue and patch to the overall RNA binding capacity of Cid1. The most notable are the quadruple mutants which demonstrate severely impaired RNA binding at any protein concentrations. The polyadenylation mutant His^{336}>Ala mutant was also included for completeness. RNA binding was assessed at increasing concentrations of protein (4nM – 125nM). The free RNA is highlighted (arrowed) along with RNA-tCid1 complexes. Panel (B) was kindly provided by S. Fleudepine and C. Norbury, Sir William Dunn School of Pathology, University of Oxford, UK.


A
Region III
Region I
Region II
Region I
Region II
Region III
Lys144
Lys133
Arg137
αJ
αI
Arg321
Lys323
90°
Arg277
Lys282
B
no protein
tCid1 25nM
K133A-R137A
K133A-R137A R277A
K133A-R137A K282A
K133A-R137A R277A-K282A
K144A
K144A R277A-K282A
R277A-K282A
K133A-R137A K321A-R323A
K321A-R323A
R277A K321A-R323A
K282A K321A-R323A
R277A-K282A K321A-R323A
tCid1
K133A-R137A K144A
H336A
tCid1-RNA complex
RNAfree

concentrations tested by the simultaneous substitution of amino acids comprising both regions II and III. Simultaneous substitutions of regions I and II and regions I and III severely impaired RNA binding but did not fully remove the enzyme's RNA binding capacity (Figure 7.8), thus suggesting that all regions contribute to the binding of RNA but their contribution toward the enzyme's capability is not equal. Interestingly, distance measurements of the putative RNA-binding surface of the structure suggest it could be capable of accommodating an RNA molecule of ~13nt, in line with EMSA experiments using labeled oligonucleotides of different lengths (S. Fleurdepine and C. Norbury, personal communication).

We further characterized the RNA binding capacity of tCid1 and its mutant derivatives by surface plasmon resonance (SPR). Wild type tCid1 exhibited tight RNA binding with surprisingly high on- and off-rate constants, which complicated estimation of K_D from kinetic measurements (estimated K_Ds are listed in Table 7.2). Steady-state SPR analyses gave higher apparent K_D values (Table 7.2) of approximately 100 nM for tCid1, 200 nM for mutant I, 450 nM for mutant II and 410 nM for mutant III. Scatchard plots for tCid1 and mutants I and II, were non-linear and concave, consistent with heterogeneous binding at more than one sub-site, whereas Scatchard analysis for the region III mutant was almost linear, indicating that for this mutant binding was dependent on a single site (S. Fleurdepine and C. Norbury, unpublished observations). We therefore propose that RNA binding by Cid1 involves two sites, the first depending on regions I and II and the second depending on region III.

Table 7.2. – Dissociation Constants (K_D) of Cid1:RNA by SPR

Mutant	K_D (nM) Steady State Analysis	K_D (nM) Kinetic Analysis
tCid1 (Native)	99.4	49.3
Region I Mutant (K133A – K137A)	198.9	86.0
Region II Mutant (K321A – R323A)	447.0	208.1
Region III Mutant (R277A – K282A)	414.5	199.0

Given that Cid1 uridylates a variety of mRNAs in *S. pombe* cells, for instance actin mRNA, targeting these for Lsm1-7-mediated mRNA decay or transcriptional silencing (Rissland and Norbury, 2008), the RNA binding capability must be relevant to enzymatic activity. Therefore *in vitro* poly(U) polymerase activity assays were performed using the RNA-binding-defective mutant enzymes. Interestingly, biochemical analysis revealed that impairment of RNA binding also impaired catalytic activity. Surface mutant protein in which RNA binding capacity is severely impaired displayed absent enzymatic activity in the uridylation activity assay (Figure 7.9). What is more, those mutants in which RNA binding was impaired slightly (i.e. impaired RNA binding at low protein concentrations) only displayed reduced activity, incorporating only a single Uridine residue to the RNA substrate. Overall these results demonstrate a relationship between efficient RNA binding and catalysis, particularly impaired processivity related to impaired RNA binding.

A structural comparison between tCid1 and the *T. brucei* TUTases MEAT1 and TUT4, both of which lack any RNA binding domain, demonstrated that the distribution of positively charged residues in Cid1 that confer the RNA binding capacity is unique. Neither MEAT1 nor TUT4 display the surface properties required to bind to RNA. The TUTases RET2 from *T. brucei* possesses an inserted domain that has been shown to be important in RNA binding (Ringpis *et al.,* 2010). The Poly(A) polymerase possess an RRM domain that binds to RNA and is important for catalytic activity (Balbo and Bohm, 2007) and other non-canonical poly(A) polymerases such as Trf4 from *S. cerevisiae* require Air2, a zinc knuckle containing protein for substrate recognition (Hamill *et al.,* 2010).

These data therefore suggest that Cid1 has a two-stage binding mechanism, in which its general RNA binding ability is succeeded by specific binding of the 3' end alongside UTP, which is dependent, as we have shown, on His^{336}. The bound RNA could be threaded along the basic groove on the surface of Cid1 and fed into the active site cleft prior to uridylation.

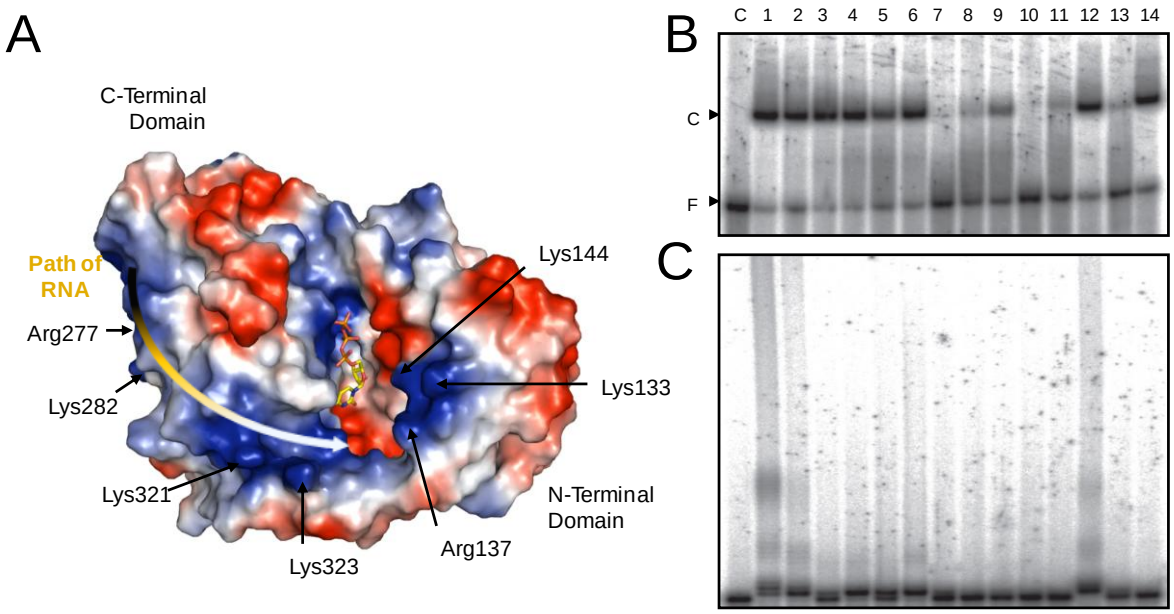


Figure 7.9 – RNA Binding is Important for Uridylation Activity. (A) Electrostatic surface representation of Cid1-UTP crystal structure with the basic residues mutated in this study labelled for reference purposes. (B) EMSA and (C) polyuridylation test of native and RNA binding impaired tCid1 mutants. Lane 1 corresponds to 25 nM of native tCid1, the remaining lanes correspond to the following mutants of tCid1: K133A-R137A (2), K321A-R323A (3), R277A-K282A (4), K133A-R137A-R277A (5), K133A-R137A-K282A (6), K133A-R137A-R277AK282A (7), R277A-K321-R323A (8), K282A-K321A-R323A (9), R277A-K282A-K321AR323A (10), K133A-R137A-K321A-R323A (11), K144A (12), K144A-R277A-K282A (13) and the catalytically 'dead' mutant D101A-D103A (14). Mutants unable to bind RNA display a lack of enzymatic activity, whereas those mutants that have slightly impaired RNA binding capability display reduced processivity and therefore transfer a single UMP residue to the RNA probe. A 'no protein' control lane (labelled C) is also included. The free RNA and protein bound RNA is indicated (arrowed) and labelled. Panel (B) and (C) were kindly provided by S. Fleudepine and C. Norbury, Sir William Dunn School of Pathology, University of Oxford, UK.

Furthermore, the processive nature of recombinant tCid1 *in vitro* poly(U) activity can be explained by its tight and rapid RNA association/dissociation. In this scenario, Cid1 is capable of incorporating 1 - 2 Uridine residues, its physiological function (S. Fleurdepine, unpublished observations, Rissland *et al.*, 2007), onto the 3'end of RNA but rapidly

dissociates from its RNA substrate and binds to and uridylates other RNAs, irrespective of their uridylation state.

7.1.9. Cid1 Displays an Unusual Domain Twist and Active Site Remodelling

We determined the crystal structure of Cid1 in a second crystal form (crystal form II) by molecular replacement. Although this crystal form possessed two molecules per asymmetric unit, it was clear that these were not isomorphous. Surprisingly, a dramatic conformational difference is observed between the two molecules in this crystal's asymmetric unit, whereby the catalytic N-terminal domain is rotated with respect to the C-terminal domain and the catalytic cleft has become remodelled as a result (Figure 7.10). Given the severe rotation and movement we refer to this conformation as the 'twisted conformer'.

The movement of the N-terminal domain is centred on the residues 163 – 166, which comprises the loop that connects the αE helix, which is integrally associated with the C-terminal domain, to strand β5 of the DNA polβ topology portion of the N-terminal domain (see Figure 7.10). There is clear and unambiguous density for this loop and the secondary structure elements surrounding it. In the twisted conformer the αE-β5 loop turns sharply with the rotation of the Asn^{165} and Arg^{166} peptide bond and in doing so forms a partial 3_{10} helical turn with αE. The β5 strand, and thus the remainder of the domain, is translated by ~8.8Å (measured using Phe^{163} from the native conformer and the twisted conformer). Globally, residues 57 – 165 of the N-terminal domain are rotated about Asn^{166} by 42° as determined by DYNDOM (Hayward and Berendsen, 1998), which results in the twisting of the N-terminal domain and the closure of one side of the active site. This rotational motion also perturbs the so-called spinal helix αB, whilst helix αA of the NTD remains fixed to surface of the CTD.

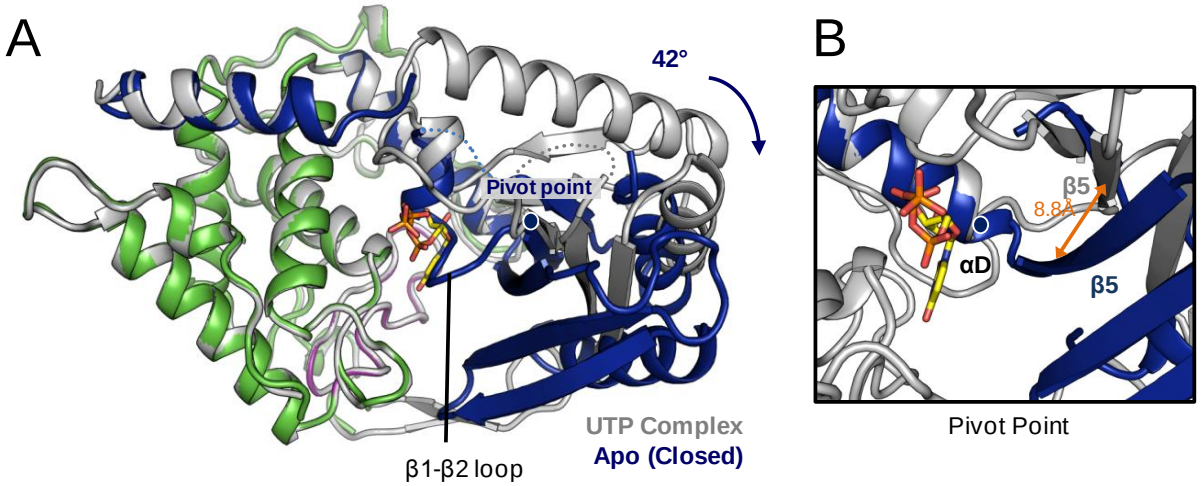


Figure 7.10 – Structure of the Cid1 Twisted Conformer. A structural comparison between the native conformation (grey) that is capable of binding to UTP and the so-called twisted conformer (coloured by domain as in Figure 7.2). The domain rotates about residue Asparagine[166] (pivot point). The β1 – β2 loop containing the catalytic consensus sequence is also remodelled and occupies the cleft, thus excluding UTP. This mechanism is suggested to eject uridylated products. (B) Close-up of the pivot point showing that a rotation about Asn[166] results in an 8.8Å translation of the strand β5 along with the remainder of the domain.

We observed weak electron density for residues 61 – 71, which corresponds to the N-terminal half of the helix αB and the αA-αB interconnecting loop, and suggests that some of the secondary structure has dissolved as a result of such domain motion.

The catalytic motif, which has the consensus sequence *h*[GS]$X_{9\text{-}13}$D*h*[D/E]*h* (described earlier) is remodelled as a result of the domain re-arrangement. It is clear that the β2 strand alters its course at the catalytic residues. Furthermore, the helical turn denoted αC within the β1-β2 loop is also not visible in the electron density. Ambiguous electron density has precluded the exact positioning of amino acids, but the use of a polyalanine segment has suggested that this loop region fills the catalytic cleft and therefore excludes uridylated product (see Figure 7.10). Careful model building and refinement of this region is ongoing to complete the structure of the twisted conformer.

7.3. Discussion

Overall these data suggest a mechanism of uracil discrimination in tCid1 that is dependent on a single Histidine residue. The uracil selectivity Histidine^{336} residue is invariant in the mammalian TUTases, providing further evidence for the functional and structural homology between ZCCHC6/11 and Cid1 (Rissland *et al.*, 2007). The striking conversion of the native uridyltransferase activity, by a Histidine^{336} substitution, to an enzyme that added adenine residues to RNA (poly(A) polymerase activity) suggests that acquisition of His336 during evolution may have been sufficient to convert an ancestral PAP into the common ancestor of Cid1 and its human orthologues. Given the uracil selection mechanism employed by the *T. brucei* TUTases *via* a network of water molecules, it is likely that a single water molecule is co-ordinated by the imidazole amine of His^{336} and the N3 amine of the Uridine base in Cid1 (Figure 7.7). However a high resolution structure will be necessary to confirm this.

We had previously demonstrated that Cid1 was capable of binding to RNA (see Chapter 6). Through a series of structure-guided mutagenesis and biochemical analysis, the single stranded RNA substrates of tCid1 traverse the surface of the enzyme *via* a series of basic patches. Whilst other DNA polβ-nucleotidyltransferases possess RNA binding modules, we propose that the method employed by Cid1 is a novel and widespread mechanism. The RNA-class specificity of the human orthologues of Cid1 presumably reflects their acquisition of additional structural elements (e.g. zinc knuckles) as well as their interaction with substrate-specific RNA-binding proteins such as Lin28A. Sequence comparisons between Cid1 and the ZCCHCs suggest that the RNA binding sites of Cid1 are not well conserved in metazoans, with residues in region I in Cid1 conserved in the ZCCHCs. Overall, this makes evolutionary sense given that pre-miRNAs are structured molecules and a negatively charged groove is not appropriate for such substrates, hence the necessity for other interaction

partners and the use of zinc knuckles. Recently the structure of Lin28 was solved in complex with a range of Pre-let-7 RNA family members (Nam *et al.,* 2011) and demonstrates the use of the two CCHC zinc knuckles in Lin28 for its interaction with the structured pre-miRNA. However, a 'reciprocity' between the RNA and Lin28 is also noted by the authors, whereby the interaction of the CCHC zinc knuckles induces a specific conformation to the RNA backbone and is proposed to then influence downstream signalling (Nam *et al.,* 2011).

Importantly, structural comparisons with other TUTases lacking known RNA binding domains demonstrates that the surface properties of Cid1 that are critical for RNA binding are unique to Cid1. This is not surprising for MEAT1, which exists, as its name suggests, in a stable multi-protein complex that is likely to include RNA binding proteins. The RNA binding capacity of Cid1 is consistent with an ancient enzyme as suggested by the phylogenetic analysis performed. Cid1 could therefore be considered as a minimal TUTase. It is tempting to suggest that in metazoans the last common ancestor of Cid1 evolved additional RNA substrate recognition elements that eventually, alongside the evolution of Drosha and the miRNA biogenesis pathways, resulted in the large and decorated ZCCHCs that exist today.

Inhibition of ZCCHC11 expression was recently shown to suppress cancer-related phenotypes including invasion and metastasis in human cell lines and xenografts (Piskounova, 2011). Furthermore, the identification of His^{336}-dependent nucleotide selectivity in Cid1 may provide a rationale for the identification of selective small molecule inhibitors of trypanosomal TUTs, which lack this feature. The insights afforded by this study into the basis of catalysis and RNA binding by Cid1 therefore open up the possibility of using this structural information to target TUT enzymes therapeutically.

CHAPTER 8

Conclusions, Future Directions and Final Remarks

Structural biology can offer invaluable insights into the fine detail of many biological processes. By using structural biology techniques, the challenges and successes in the course of research have resulted in the investigation and analysis of both cell adhesion and division. Structures have been determined for the cell adhesion proteins Kindlin-3 and the PH domain of Kindlin-1 but also the RNA terminal uridyltransferase, Cid1 in two crystal forms and in complex with its substrate. This concluding chapter will summarise the important findings and outline additional and future work.

8.1. Summary of Results and Future Directions

8.1.1. High-throughput Methods for the Study of the Kindlins

Given the difficulty in obtaining recombinant protein for functional studies (Harburger *et al.,* 2009), and initial unsuccessful cloning strategies (Prof. I.D. Campbell personal communication) a small-scale structural genomics approach was employed to clone, express and crystallise the Kindlin proteins, or domains thereof. We made use of high-throughput (HTP) technologies designed and implemented in the Oxford Protein Production Facility (OPPF) for this study, the conclusions of which are discussed below.

The OPPF HTP pipeline is well established with the automation of many procedures including, cloning, small scale expression screening in *E. coli* and crystallisation screening. Large scale protein production is still largely labour intensive but this has been reduced significantly by the use of autoinduction bacterial growth media (Studier, 2004), dispensing

with the need to manually assess the culture optical density and to supplement the culture medium with IPTG.

We designed and generated 48 mouse Kindlin-1 constructs in a first-pass HTP screening approach. It was found that many constructs, despite being soluble, were difficult to obtain during large scale protein production exercises. However, successful over-expression in *E. coli* was achieved for a subset of constructs that encompass the Pleckstrin Homology domain only – with the single largest PH domain construct successfully crystallising. In light of a report on the solution structure of the Kindlin-1 F0 domain and the redefinition of the domain boundaries we subsequently embarked upon a second pass HTP screen. This second screen was modified from the first by using genes of each isoform and also by using different fusion tags (GST and MBP) that may improve the solubility and yield of protein. To this end we designed a format that utilised the redundancies in the ligation independent cloning process thus saving resources. We therefore generated 48 new constructs from only 24 PCR products using half the quantity of InFusion cloning kits. The use of fusion partners maltose-binding protein (MBP) and glutathione S-transferase (GST) did not improve yields. Furthermore, after large scale production, several insoluble constructs were made artificially soluble by the presence of the large fusion partners, further complicating results. This prompted us to seek additional routes in obtaining the highly sought after recombinant Kindlin proteins. Whilst the OPPF pipeline is relatively straight forward and highly parallelised resulting in many purified and crystallised targets, as for the Kindlin-1 PH domain, in cases where *E. coli* is not suitable, automated systems for screening in additional hosts is under developed. Therefore we designed and performed a manual parallel screen using a restricted panel of constructs in mammalian and insect cell hosts. This revealed that baculovirus-infected Sf9 cells represented the best host for the expression of recombinant Kindlin constructs. Although the number of soluble constructs was limited and proteolysis

was prominent, this was a significant step forward in obtaining recombinant Kindlin and fragments for further study. The use of baculovirus-infected Sf9 cells led to the expression and purification of milligram quantities of full-length Kindlin-3, previously unobtainable from *E. coli* and several lead constructs for further investigation.

8.1.2. Future Directions for the Study of the Kindlins.

Functional work, in collaboration with Dr. David Calderwood, Yale University, USA, using full length protein for assessing the Kindlin binding partners is currently under investigation. We are making use of N-terminally GST-tagged full-length Kindlin (1, 2 and 3) constructs (not included in the original work) which have been shown to express in milligram quantities (data not shown). The susceptibility to proteolysis is noted for these constructs. However this could be circumvented by the addition of a second fusion tag at the C-terminal end, and requires a simple site-directed mutagenesis step by PCR, to remove the stop codon at the C-terminus allowing an encoded His_6-tag to be incorporated into the protein. This strategy is currently under investigation.

8.1.3. Structural Characterisation of the Kindlin-1 PH Domain

We expressed, crystallised and determined the structure of the Kindlin-1 PH domain – the construct for which, was generated from HTP cloning experiments. We solved the structure by molecular replacement and the model was refined to 2.1Å. The structure reveals a PH domain superfamily fold, with the addition of a C-terminal helix that, on detailed inspection, is clearly integral to the fold of this PH domain. Despite structural comparisons with other well documented PH domains it was not immediately clear what the function of

this module is. Therefore we embarked upon a multi-disciplinary approach to assess its function. A structural phylogenetic tree was constructed, given that many PH domains have diverse functions despite a common fold and this method had not previously been employed in comparing PH domain function. An all pairs-pairwise comparison of 30 PH domains was performed. This provided a useful insight into this domain superfamily with several similarly functioning, phosphoinositide binding, PH domains clustering together. This made evolutionary sense given that relatively minor changes in the PH domain scaffold would result in altered ligand specificity. The most sparsely positioned were the protein-interacting modules, whereby larger changes in sequence and structure would result in specific protein binding interfaces. A molecular dynamics (MD) simulation approach in collaboration with Prof. Mark Sansom's laboratory (Department of Biochemistry, University of Oxford) allowed the assessment of ligand binding by the Kindlin-1 PH domain *in silico* and revealed the occlusion of the putative ligand binding site in the Kindlin-1 binding site by a salt bridge. In simulations this salt bridge was transiently broken, forming a new salt bridge elsewhere, 25% of the time and this resulted in an open conformation. Further *in silico* ligand docking revealed that the open conformation (salt bridge broken) was more able to bind to potential ligands. Ligand binding was assessed *in vitro* making use of lipid-dot blots. It was quickly apparent that the Kindlin-1 PH domain bound a limited number of lipids, namely the mono-phosphorylated inositides – all of which were unexpected and potentially not physiologically relevant given the location of these lipids in the cell. *In vitro* lipid dot-blot analysis with a mutant PH domain ($Glu^{416}Ala$), with a permanently disrupted salt bridge, resulted in a protein that was able to recognise an additional ligand – a di-phosphorylated inositide, $PtdIns(3,5)P_2$, in line with MD simulation predictions. Lastly a cellular approach was employed in collaboration with Dr. David Calderwood's laboratory (Department of Pharmacology, Yale University, USA). Deletion of the PH domain from the full length protein Kindlin-1 in cells

resulted in impaired integrin activation. Furthermore, focal adhesion assembly was also impaired, thus revealing an important role for the PH domain in Kindlin-mediated integrin activation.

8.1.4. Future Directions for the Kindlin-1 PH Domain

Overall, this work suggests a role for the Kindlin-1 PH domain in integrin activation, but the exact mechanism by which the PH domain functions is still undetermined. Using a multidisciplinary approach in the domain's characterisation we revealed that this PH domain may not function as expected, because a salt-bridge, only revealed by determining the structure, occludes the canonical phosphoinositide ligand binding site. In a recent report demonstrating Kindlin-2 PH domain binding to the headgroup of the PI3K product phosphatidylinositol (3,4,5) trisphsophate (Liu *et al.*, 2011) the authors suggested this to be a universal function of the Kindlins, our data reveals this to be an untrue assumption. Therefore work in ascertaining the exact ligand will be necessary along with additional structural work to determine complex structures. Putatively, the preferred mono-phosphorylated inositol ligands that the 'native' Kindlin-1 PH domain recognises may be mimics of phosphotyrosine. Therefore, a simple further experiment would be to generate Kindlin-1 PH domain crystals and to perform soaking experiments with phosphotyrosine for structure determination of a complex.

8.1.5. Biophysical Characterisation of Kindlin-3

Kindlin-3 is the blood cell specific isoform of the Kindlin family and has received great interest due to its essential role in clotting (Moser *et al.*, 2009). We improved upon the previous HTP Kindlin cloning and expression work in baculovirus-infected Sf9 cells focussing on a single full-length C-terminally His-tagged Kindlin-3 construct. We optimised the standard recombinant baculovirus expression method for virus amplification and protein expression by aerating the Sf9 cell culture during viral amplification. This resulted in a three-fold improvement in total yield of recombinant Kindlin-3, over standard procedures, and concomitantly reduced the time-scale of baculovirus-based expression work by more than a week. Furthermore, we were able to purify recombinant Kindlin-3 to homogeneity by use of a heparin affinity chromatography – a method that can now be applied to other isoforms. Whilst extensive crystallisation screening proved unsuccessful, we obtained a low resolution solution structure of Kindlin-3 by small angle X-ray scattering (SAXS). SAXS envelopes obtained from the scattering data revealed that Kindlin-3 was elongated. Additional techniques such as dynamic light scattering (DLS) and analytical ultracentrifugation (AUC) also independently confirmed an elongated structure. This supported notions that the Kindlins possesses a linearly arranged FERM domain, analogous to its homologue talin – the only example of a linear FERM domain containing protein. Furthermore, the Kindlin-3 SAXS envelope displayed additional density that could be accounted for by the additional PH domain that Kindlins possess but talin does not. Using atomic models of talin obtained by X-ray crystallography and NMR, together with the PH domain structure we had previously solved we produced an approximate model for Kindlin-3 by fitting and positionally refining these models within the envelope's density.

Using additional methods, AUC and NMR, we confirmed the binding of integrin β_{1A} cytoplasmic tails by Kindlin-3. Whilst NMR spectroscopy, in collaboration with Prof. I.D. Campbell's laboratory (Department of Biochemistry, University of Oxford), was used previously for the study of talin subdomains and β-integrin cytoplasmic tails it had not been performed using full length Kindlin-3. By using diminishing intensity of the isotopically labelled β-integrin peptide instead of chemical shifts we demonstrated that full length Kindlin-3 was capable of recognising the second 'membrane distal' NPxY motif in the sequence. We also observed the interaction with a Serine/Threonine rich region upstream of the NPxY. When Kindlin-3 was presented with a spliced variant of the β_1 tail, namely β_{1D}, which lacked the Serine/Threonine rich region, the binding of the NPxY motif by Kindlin-3 was less pronounced. This is wholly consistent with cellular studies that show that the Serine/Threonine rich region is important for Kindlin binding (Harburger *et al.*, 2009). What is more, these NMR experiments were performed at physiological pH, unlike talin experiments (pH 6.1) further adding to their relevance.

8.1.6. Future Directions for the Study of Kindlin-3

We have already provided recombinant Kindlin-3 protein to the laboratory of David Calderwood, Yale University, USA, for functional studies. We are continuing to try to obtain crystal structures of Kindlin-3, but this is likely to require some construct re-engineering. Given the difficulty in obtaining truncated Kindlin-3 mutant protein, a simple strategy would be to use small quantities of proteases to digest the Kindlin-3 into fragments and separate these fragments by size exclusion chromatography. The purified fragments would then be subject to mass spectroscopy and if sufficient quantities were available crystallisation screening. This experimentally led fragment identification may prove a more successful route

to determining more structures of the Kindlins as predicted domains has yielded little success. Using size exclusion chromatography, we were unable to identify any interactions between Kindlin and talin or β-integrin cytoplasmic tails. The most pertinent follow-up work that may address this, is to generate F3 subdomain containing fragments of Kindlin-3, probably directed by proteolysis sensitive sites, and to perform NMR spectroscopy titrations using isotopically labelled integrin-tails. This will provide clear chemical shifts and confirm observations with full-length Kindlin-3 and will allow further experiments with the talin F3 subdomain to ascertain whether talin and Kindlin bind concomitantly to their corresponding NPxY motifs or sequentially. Perhaps more importantly, these experiments will determine how Kindlins modulate integrin activation – a process that is still very much unclear.

8.1.7. The Crystal structure of tCid1

We expressed and purified recombinant Cid1, revealing that it possessed the capacity to bind to nucleic acids. The removal of contaminating *E. coli* nucleic acids was essential for the crystallisation of the enzyme. We crystallised a slightly truncated, but catalytically active, mutant of Cid1 (tCid1) lacking the first 33 amino acids. Crystals were obtained only after the protein was heated to 37°C prior to crystallisation screening and also required low sodium chloride concentrations (50mM) and the presence of DNA and UTP – although neither of these molecules was present after the structure was determined. Crystals could be optimised by reducing the dielectric constant of the crystallisation experiment *via* the addition of volatile organic compounds. We obtained optimal diffraction data from crystals by making use of the highly specialised microfocus beamline I24 at Diamond Light Source (Didcot, UK) and collecting several overlapping wedges of data to yield a complete dataset. We also managed to acquire a UTP-bound complex by soaking the crystals in mother liquor

supplemented with UTP and collecting a dataset analogously to the native crystals. Additionally we crystallised Cid1 in a second crystal form using micro-seeding and additive optimisation methods in tandem.

The structure of Cid1 was solved by molecular replacement and refined to a resolution of 3.2Å and 3.0Å for the Apo and UTP bound forms, respectively. The second crystal form was solved by molecular replacement and the structure was refined to 2.6Å resolution. Cid1 is clearly related to other polβ nucleotidyltransferases, including poly(A) polymerase and the trypanosomal terminal Uridine transferases, with phylogenetic analysis placing this enzyme closest to the last common ancestor of such enzymes. However an in-depth structural analysis revealed a UTP binding mechanism that was largely conserved in Cid1 and the other TUTases but identified a single histidine that was present in Cid1 alone. In collaboration with Dr. Chris Norbury's laboratory, we investigated this observation further. We demonstrated that the substitution of the proposed Uridine selectivity histidine to an alanine converted the uridine transferase activity of Cid1 to a poly(A) activity *in vitro*. Furthermore, sequence comparisons with mammalian TUTases ZCCHC11 and 6, focussing, in particular, on regions corresponding to the portion of Cid1 that contacts the uracil, revealed a conservation of this important histidine residue. Disorder prediction in RONN and alignment of the Cid1 disorder plot against that of ZCCHC11 and 6 revealed similar characteristics between Cid1 and a portion of the ZCCHCs. This, along with function work and sequence comparisons, suggests that Cid1 is structurally similar to the ZCCHCs. In addition, with the Norbury laboratory, we investigated the biological significance of the nucleic acid binding capacity of Cid1 observed during its purification. Structural and biochemical analysis revealed several basic patches on the surface of the enzyme were important for high affinity ssRNA binding. Furthermore, Cid1 RNA binding capability is also important for the enzyme's activity.

8.1.8. Future Directions for the Study of Cid1

The most obvious and most straightforward is to obtain a Cid1-RNA complex crystal structure and given the relatively high affinity binding (nM range), this should be attainable. It was noted after solving the structure that a large number of crystal contacts were mediated by hydrophobic contacts and main chain hydrogen bonding, with some salt bridges. Perhaps by making use of the basic patch mutants, analogous to surface entropy reduction, different crystal contacts may be favoured and therefore a higher resolution structure may become attainable. Additionally, several inhibitors of this enzyme have been developed in other laboratories, and obtaining a complex for rational drug design will necessitate a high resolution structure.

Given the similarity between the catalytic domains of the ZCCHCs and Cid1, both in sequence similarity, disorder prediction and activity, our structure provides a solid basis for investigating the structure of the human orthologues. This will undoubtedly be invaluable given the role of the ZCCHCs in miRNA biogenesis and cancer. Therefore, drug screening studies in Cid1 and complex structures thereof, may provide a useful foundation for screening in ZCCHCs, with structural data to follow.

8.2. Final Remarks

Accepted manuscripts and published research that has been described in this Thesis are referenced in the appendix along with the copies of the manuscripts. Any structure depositions to the PDB as part of publication are also listed in the appendix.

We have been fortunate enough to have recieved a Cancer Research UK funded project grant based upon the Cid1 work in this thesis. Therefore we will continue to work on

Cid1, in particular the drug screening efforts, together with obtaining structural information on its human orthologues.

CHAPTER 9

9.1. References

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., *et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta crystallographica *66*, 213-221.

Albeck, S., Alzari, P., Andreini, C., Banci, L., Berry, I.M., Bertini, I., Cambillau, C., Canard, B., Carter, L., Cohen, S.X., *et al.* (2006). SPINE bioinformatics and data-management aspects of high-throughput structural biology. Acta crystallographica *62*, 1184-1195.

Altamirano, M.M., Woolfson, A., Donda, A., Shamshiev, A., Briseno-Roa, L., Foster, N.W., Veprintsev, D.B., De Libero, G., Fersht, A.R., and Milstein, C. (2001). Ligand-independent assembly of recombinant human CD1 by using oxidative refolding chromatography. Proceedings of the National Academy of Sciences of the United States of America *98*, 3288-3293.

Andrews, N.C., and Baltimore, D. (1986). Purification of a terminal uridylyltransferase that acts as host factor in the in vitro poliovirus replicase reaction. Proceedings of the National Academy of Sciences of the United States of America *83*, 221-225.

Andrews, N.C., Levin, D., and Baltimore, D. (1985). Poliovirus replicase stimulation by terminal uridylyl transferase. The Journal of biological chemistry *260*, 7628-7635.

Anthis, N.J., and Campbell, I.D. (2011). The tail of integrin activation. Trends Biochem Sci *36*, 191-198.

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

Aphasizhev, R. (2005). RNA uridylyltransferases. Cellular and molecular life sciences : CMLS *62*, 2194-2203.

Aphasizhev, R., and Aphasizheva, I. (2007). RNA editing uridylyltransferases of trypanosomatids. Methods in enzymology *424*, 55-73.

Aphasizhev, R., Aphasizheva, I., Nelson, R.E., Gao, G., Simpson, A.M., Kang, X., Falick, A.M., Sbicego, S., and Simpson, L. (2003a). Isolation of a U-insertion/deletion editing complex from Leishmania tarentolae mitochondria. The EMBO journal *22*, 913-924.

Aphasizhev, R., Aphasizheva, I., and Simpson, L. (2003b). A tale of two TUTases. Proceedings of the National Academy of Sciences of the United States of America *100*, 10617-10622.

Aphasizhev, R., Aphasizheva, I., and Simpson, L. (2004). Multiple terminal uridylyltransferases of trypanosomes. FEBS letters *572*, 15-18.

Aphasizhev, R., Sbicego, S., Peris, M., Jang, S.H., Aphasizheva, I., Simpson, A.M., Rivlin, A., and Simpson, L. (2002). Trypanosome mitochondrial 3' terminal uridylyl transferase (TUTase): the key enzyme in U-insertion/deletion RNA editing. Cell *108*, 637-648.

Aricescu, A.R., Assenberg, R., Bill, R.M., Busso, D., Chang, V.T., Davis, S.J., Dubrovsky, A., Gustafsson, L., Hedfalk, K., Heinemann, U., *et al.* (2006). Eukaryotic expression: developments for structural proteomics. Acta crystallographica *62*, 1114-1124.

Aricescu, A.R., Siebold, C., Choudhuri, K., Chang, V.T., Lu, W., Davis, S.J., van der Merwe, P.A., and Jones, E.Y. (2007). Structure of a tyrosine phosphatase adhesive interaction reveals a spacer-clamp mechanism. Science (New York, NY) *317*, 1217-1220.

Arita, K., Wessagowit, V., Inamadar, A.C., Palit, A., Fassihi, H., Lai-Cheong, J.E., Pourreyron, C., South, A.P., and McGrath, J.A. (2007). Unusual molecular findings in Kindler syndrome. The British journal of dermatology *157*, 1252-1256.

Baca, A.M., and Hol, W.G. (2000). Overcoming codon bias: a method for high-level overexpression of Plasmodium and other AT-rich parasite genes in Escherichia coli. International journal for parasitology *30*, 113-118.

Bai, Y., Milne, J.S., Mayne, L., and Englander, S.W. (1993). Primary structure effects on peptide group hydrogen exchange. Proteins *17*, 75-86.

Baraldi, E., Djinovic Carugo, K., Hyvonen, M., Surdo, P.L., Riley, A.M., Potter, B.V., O'Brien, R., Ladbury, J.E., and Saraste, M. (1999). Structure of the PH domain from Bruton's tyrosine kinase in complex with inositol 1,3,4,5-tetrakisphosphate. Structure *7*, 449-460.

Bard, J., Zhelkovsky, A.M., Helmling, S., Earnest, T.N., Moore, C.L., and Bohm, A. (2000). Structure of yeast poly(A) polymerase alone and in complex with 3'-dATP. Science (New York, NY *289*, 1346-1349.

Barnard, D.C., Ryan, K., Manley, J.L., and Richter, J.D. (2004). Symplekin and xGLD-2 are required for CPEB-mediated cytoplasmic polyadenylation. Cell *119*, 641-651.

Begley, M.J., Taylor, G.S., Brock, M.A., Ghosh, P., Woods, V.L., and Dixon, J.E. (2006). Molecular basis for substrate recognition by MTMR2, a myotubularin family phosphoinositide phosphatase. Proceedings of the National Academy of Sciences of the United States of America *103*, 927-932.

Berenedsen, H.J.C., Postma, J.P.M., Van Gunsteren, W.F., DiNola, A., and Haak, J.R. (1984). Molecular-Dynamics with Coupling to an External Bath. J Chem Phys *81*, 3684-3690.

Berrow, N.S., Alderton, D., Sainsbury, S., Nettleship, J., Assenberg, R., Rahman, N., Stuart, D.I., and Owens, R.J. (2007). A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. Nucleic acids research *35*, e45.

Bialkowska, K., Ma, Y.Q., Bledzka, K., Sossey-Alaoui, K., Izem, L., Zhang, X., Malinin, N., Qin, J., Byzova, T., and Plow, E.F. (2010). The integrin co-activator Kindlin-3 is expressed

and functional in a non-hematopoietic cell, the endothelial cell. The Journal of biological chemistry *285*, 18640-18649.

Bird, L.E. (2011). High throughput construction and small scale expression screening of multi-tag vectors in Escherichia coli. Methods *55*, 29-37.

Blahna, M.T., Jones, M.R., Quinton, L.J., Matsuura, K.Y., and Mizgerd, J.P. (2011). Terminal uridyltransferase enzyme Zcchc11 promotes cell proliferation independent of its uridyltransferase activity. The Journal of biological chemistry *286*, 42381-42389.

Blomberg, N., Baraldi, E., Nilges, M., and Saraste, M. (1999). The PH superfold: a structural scaffold for multiple functions. Trends Biochem Sci *24*, 441-445.

Bohnsack, M.T., Czaplinski, K., and Gorlich, D. (2004). Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. RNA *10*, 185-191.

Bottomley, M.J., Salim, K., and Panayotou, G. (1998). Phospholipid-binding protein domains. Biochimica et biophysica acta *1436*, 165-183.

Boyd, R.S., Adam, P.J., Patel, S., Loader, J.A., Berry, J., Redpath, N.T., Poyser, H.R., Fletcher, G.C., Burgess, N.A., Stamps, A.C., *et al.* (2003). Proteomic analysis of the cell-surface membrane in chronic lymphocytic leukemia: identification of two novel proteins, BCNP1 and MIG2B. Leukemia *17*, 1605-1612.

Brishammar, S., and Juntti, N. (1975). A poly(U) polymerase in tobacco leaves. Biochimica et biophysica acta *383*, 351-358.

Bueno, M.J., and Malumbres, M. (2011). MicroRNAs and the cell cycle. Biochimica et biophysica acta *1812*, 592-601.

Butcher, D.T., Alliston, T., Weaver, V.M. (2009). A tense situation: forcing tumour progression. Nat Rev Cancer *9*, 108-122

Campbell, I.D., and Humphries, M.J. (2011). Integrin structure, activation, and interactions. Cold Spring Harb Perspect Biol *3*.

Canellakis, E.S. (1957). Incorporation of radioactive uridine-5'-monophosphate into ribonucleic acid by soluble mammalian enzymes. Biochimica et biophysica acta *23*, 217-218.

Carpten, J.D., Faber, A.L., Horn, C., Donoho, G.P., Briggs, S.L., Robbins, C.M., Hostetter, G., Boguslawski, S., Moses, T.Y., Savage, S., *et al.* (2007). A transforming mutation in the pleckstrin homology domain of AKT1 in cancer. Nature *448*, 439-444.

CCP4 (1994). The CCP4 suite: programs for protein crystallography. Acta crystallographica *50*, 760-763.

Ceccarelli, D.F., Blasutig, I.M., Goudreault, M., Li, Z., Ruston, J., Pawson, T., and Sicheri, F. (2007). Non-canonical interaction of phosphoinositides with pleckstrin homology domains of Tiam1 and ArhGAP9. The Journal of biological chemistry *282*, 13864-13874.

Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta crystallographica *66*, 12-21.

Chivukula, R.R., and Mendell, J.T. (2008). Circular reasoning: microRNAs and cell-cycle control. Trends Biochem Sci *33*, 474-481.

Cifuentes, M.E., Honkanen, L., and Rebecchi, M.J. (1993). Proteolytic fragments of phosphoinositide-specific phospholipase C-delta 1. Catalytic and membrane binding properties. The Journal of biological chemistry *268*, 11586-11593.

Cimprich, K.A., and Cortez, D. (2008). ATR: an essential regulator of genome integrity. Nature reviews *9*, 616-627.

Collins, A.V., Brodie, D.W., Gilbert, R.J., Iaboni, A., Manso-Sancho, R., Walse, B., Stuart, D.I., van der Merwe, P.A., and Davis, S.J. (2002). The interaction properties of costimulatory molecules revisited. Immunity *17*, 201-210.

Cortez, D., Guntuku, S., Qin, J., and Elledge, S.J. (2001). ATR and ATRIP: partners in checkpoint signaling. Science (New York, NY *294*, 1713-1716.

Cronin, T.C., DiNitto, J.P., Czech, M.P., and Lambright, D.G. (2004). Structural determinants of phosphoinositide selectivity in splice variants of Grp1 family PH domains. The EMBO journal *23*, 3711-3720.

Currie, R.A., Walker, K.S., Gray, A., Deak, M., Casamayor, A., Downes, C.P., Cohen, P., Alessi, D.R., Lucocq, J. (1999). Role of phosphatidylinositol 3,4,5-trisphosphate in regulating the activity and localization of 3-phosphoinositide-dependent protein kinase-1. The Biochemical journal *337*, 575–583

D'Urso, G., Grallert, B., and Nurse, P. (1995). DNA polymerase alpha, a component of the replication initiation complex, is essential for the checkpoint coupling S phase to mitosis in fission yeast. Journal of cell science *108 (Pt 9)*, 3109-3118.

Darden, T., York, D., and Pedersen, L. (1993). Particle mesh Ewald. An N log(N) method for Ewald sums in large systems. J Chem Phys *98*, 10089-10092.

Dasgupta, A., Zabel, P., and Baltimore, D. (1980). Dependence of the activity of the poliovirus replicase on the host cell protein. Cell *19*, 423-429.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR *6*, 277-293.

Deng, J., Ernst, N.L., Turley, S., Stuart, K.D., and Hol, W.G. (2005). Structural basis for UTP specificity of RNA editing TUTases from Trypanosoma brucei. The EMBO journal *24*, 4007-4017.

Derewenda, U., Mateja, A., Devedjiev, Y., Routzahn, K.M., Evdokimov, A.G., Derewenda, Z.S., and Waugh, D.S. (2004). The structure of Yersinia pestis V-antigen, an essential virulence factor and mediator of immunity against plague. Structure *12*, 301-306.

DiNitto, J.P., and Lambright, D.G. (2006). Membrane and juxtamembrane targeting by PH and PTB domains. Biochimica et biophysica acta *1761*, 850-867.

Dowler, S., Currie, R.A., Campbell, D.G., Deak, M., Kular, G., Downes, C.P., and Alessi, D.R. (2000). Identification of pleckstrin-homology-domain-containing proteins with novel phosphoinositide-binding specificities. The Biochemical journal *351*, 19-31.

Dowling, J.J., Gibbs, E., Russell, M., Goldman, D., Minarcik, J., Golden, J.A., and Feldman, E.L. (2008a). Kindlin-2 is an essential component of intercalated discs and is required for vertebrate cardiac structure and function. Circulation research *102*, 423-431.

Dowling, J.J., Vreede, A.P., Kim, S., Golden, J., and Feldman, E.L. (2008b). Kindlin-2 is required for myocyte elongation and is essential for myogenesis. BMC cell biology *9*, 36.

Dumas, J.J., Merithew, E., Sudharshan, E., Rajamani, D., Hayes, S., Lawe, D., Corvera, S., and Lambright, D.G. (2001). Multivalent endosome targeting by homodimeric EEA1. Molecular cell *8*, 947-958.

Elliott, P.R., Goult, B.T., Kopp, P.M., Bate, N., Grossmann, J.G., Roberts, G.C., Critchley, D.R., and Barsukov, I.L. (2010). The Structure of the talin head reveals a novel extended conformation of the FERM domain. Structure *18*, 1289-1299.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta crystallographica *60*, 2126-2132.

Essen, L.O., Perisic, O., Cheung, R., Katan, M., and Williams, R.L. (1996). Crystal structure of a mammalian phosphoinositide-specific phospholipase C delta. Nature *380*, 595-602.

Evans, P. (2006). Scaling and assessment of data quality. Acta crystallographica *62*, 72-82.

Evrard, G.X., Langer, G.G., Perrakis, A., and Lamzin, V.S. (2007). Assessment of automatic ligand building in ARP/wARP. Acta crystallographica *63*, 108-117.

Fabrini, R., De Luca, A., Stella, L., Mei, G., Orioni, B., Ciccone, S., Federici, G., Lo Bello, M., and Ricci, G. (2009). Monomer-dimer equilibrium in glutathione transferases: a critical re-examination. Biochemistry *48*, 10473-10482.

Feig, M., Karanicolas, J., and Brooks, C.L., 3rd (2004). MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. J Mol Graph Model *22*, 377-395.

Felsenstein, J. (1997). An alternating least squares approach to inferring phylogenies from pairwise distances. Syst Biol *46*, 101-111.

Ferguson, K.M., Kavran, J.M., Sankaran, V.G., Fournier, E., Isakoff, S.J., Skolnik, E.Y., and Lemmon, M.A. (2000). Structural basis for discrimination of 3-phosphoinositides by pleckstrin homology domains. Molecular cell *6*, 373-384.

Ferguson, K.M., Lemmon, M.A., Schlessinger, J., and Sigler, P.B. (1994). Crystal structure at 2.2 A resolution of the pleckstrin homology domain from human dynamin. Cell *79*, 199-209.

Ferguson, K.M., Lemmon, M.A., Schlessinger, J., and Sigler, P.B. (1995). Structure of the high affinity complex of inositol trisphosphate with a phospholipase C pleckstrin homology domain. Cell *83*, 1037-1046.

Franke, T.F., Yang, S.I., Chan, T.O., Datta, K., Kazlauskas, A., Morrison, D.K., Kaplan, D.R., and Tsichlis, P.N. (1995). The protein kinase encoded by the Akt proto-oncogene is a target of the PDGF-activated phosphatidylinositol 3-kinase. Cell *81*, 727-736.

Frech, M., Andjelkovic, M., Ingley, E., Reddy, K.K., Falck, J.R., and Hemmings, B.A. (1997). High affinity binding of inositol phosphates and phosphoinositides to the pleckstrin homology domain of RAC/protein kinase B and their influence on kinase activity. The Journal of biological chemistry *272*, 8474-8481.

French, S., and Wilson, K. (1978). On the treatment of negative intensity observations. Acta Crystallogr A *34*, 517-525.

Garcia-Alvarez, B., de Pereda, J.M., Calderwood, D.A., Ulmer, T.S., Critchley, D., Campbell, I.D., Ginsberg, M.H., and Liddington, R.C. (2003). Structural determinants of integrin recognition by talin. Molecular cell *11*, 49-58.

Garcia de la Torre, J., del Rio, G., and Ortega, A. (2007). Improved calculation of rotational diffusion and intrinsic viscosity of bead models for macromolecules and nanoparticles. J Phys Chem B *111*, 955-961.

Geiger, B., Spatz, J.P., and Bershadsky, A.D. (2009). Environmental sensing through focal adhesions. Nature reviews *10*, 21-33.

Geiger, B., and Yamada, K.M. (2011). Molecular architecture and function of matrix adhesions. Cold Spring Harb Perspect Biol *3*.

Giancotti, F.G., and Ruoslahti, E. (1999). Integrin signaling. Science (New York, NY *285*, 1028-1032.

Goddard, T.D., and Kneller, D.G. (2000). SPARKY 3 (University of California San Francisco).

Goult, B.T., Bate, N., Anthis, N.J., Wegener, K.L., Gingras, A.R., Patel, B., Barsukov, I.L., Campbell, I.D., Roberts, G.C., and Critchley, D.R. (2009a). The structure of an interdomain complex that regulates talin activity. The Journal of biological chemistry *284*, 15097-15106.

Goult, B.T., Bouaouina, M., Elliott, P.R., Bate, N., Patel, B., Gingras, A.R., Grossmann, J.G., Roberts, G.C., Calderwood, D.A., Critchley, D.R., *et al.* (2010). Structure of a double

ubiquitin-like domain in the talin head: a role in integrin activation. The EMBO journal *29*, 1069-1080.

Goult, B.T., Bouaouina, M., Harburger, D.S., Bate, N., Patel, B., Anthis, N.J., Campbell, I.D., Calderwood, D.A., Barsukov, I.G., Roberts, G.C., *et al.* (2009b). The structure of the N-terminus of kindlin-1: a domain important for AlphaIIbBeta3 integrin activation. Journal of molecular biology *394*, 944-956.

Graham, S.C., Bahar, M.W., Cooray, S., Chen, R.A., Whalen, D.M., Abrescia, N.G., Alderton, D., Owens, R.J., Stuart, D.I., Smith, G.L., *et al.* (2008). Vaccinia virus proteins A52 and B14 Share a Bcl-2-like fold but have evolved to inhibit NF-kappaB rather than apoptosis. PLoS pathogens *4*, e1000128.

Graslund, S., Nordlund, P., Weigelt, J., Hallberg, B.M., Bray, J., Gileadi, O., Knapp, S., Oppermann, U., Arrowsmith, C., Hui, R., *et al.* (2008). Protein production and purification. Nat Methods *5*, 135-146.

Grigorieff, N. (2000). Resolution measurement in structures derived from single particles. Acta crystallographica *56*, 1270-1277.

Grimes, J.M., Burroughs, J.N., Gouet, P., Diprose, J.M., Malby, R., Zientara, S., Mertens, P.P., and Stuart, D.I. (1998). The atomic structure of the bluetongue virus core. Nature *395*, 470-478.

Grosdidier, A., Zoete, V., and Michielin, O. (2011a). Fast docking using the CHARMM force field with EADock DSS. J Comput Chem.

Grosdidier, A., Zoete, V., and Michielin, O. (2011b). SwissDock, a protein-small molecule docking web service based on EADock DSS. Nucleic acids research *39*, W270-277.

Hagan, J.P., Piskounova, E., and Gregory, R.I. (2009). Lin28 recruits the TUTase Zcchc11 to inhibit let-7 maturation in mouse embryonic stem cells. Nat Struct Mol Biol *16*, 1021-1025.

Hamill, S., Wolin, S.L., and Reinisch, K.M. (2010). Structure and function of the polymerase core of TRAMP, a RNA surveillance complex. Proceedings of the National Academy of Sciences of the United States of America *107*, 15045-15050.

Hammond, C., (2007). The Basics of Crystallography and Diffraction: 12 (International Union of Crystallography Texts on Crystallography). *Oxford University Press.* Oxford.

Harburger, D.S., Bouaouina, M., and Calderwood, D.A. (2009). Kindlin-1 and -2 directly bind the C-terminal region of beta integrin cytoplasmic tails and exert integrin-specific activation effects. The Journal of biological chemistry *284*, 11485-11497.

Harburger, D.S., and Calderwood, D.A. (2009). Integrin signalling at a glance. Journal of cell science *122*, 159-163.

Harlan, J.E., Hajduk, P.J., Yoon, H.S., and Fesik, S.W. (1994). Pleckstrin homology domains bind to phosphatidylinositol-4,5-bisphosphate. Nature *371*, 168-170.

Has, C., Herz, C., Zimina, E., Qu, H.Y., He, Y., Zhang, Z.G., Wen, T.T., Gache, Y., Aumailley, M., and Bruckner-Tuderman, L. (2009). Kindlin-1 Is required for RhoGTPase-mediated lamellipodia formation in keratinocytes. The American journal of pathology *175*, 1442-1452.

Haslam, R.J., Koide, H.B., and Hemmings, B.A. (1993). Pleckstrin domain homology. Nature *363*, 309-310.

Hayward, S., and Berendsen, H.J. (1998). Systematic analysis of domain motions in proteins from conformational change: new results on citrate synthase and T4 lysozyme. Proteins *30*, 144-154.

Heo, I., Joo, C., Kim, Y.K., Ha, M., Yoon, M.J., Cho, J., Yeom, K.H., Han, J., and Kim, V.N. (2009). TUT4 in concert with Lin28 suppresses microRNA biogenesis through pre-microRNA uridylation. Cell *138*, 696-708.

Hermans, J., Berenedsen, H.J.C., Van Gunsteren, W.F., and Postma, J.P.M. (1984). A Consistent Empirical Potential for Water-Protein Interactions. Biopolymers *23*, 1513-1518.

Hess, B., Becker, H., Berenedsen, H.J.C., and Fraaije, J.G.E.M. (1997). LINCS: a linear constraint solver for molecular simulations. J Comput Chem *18*, 1463-1472.

Hess, B., Kutzner, C., von der Spoel, D., and Lindahl, E. (2008). GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. J Chem Theor Comput *4*, 435-447.

Hirose, K., Kadowaki, S., Tanabe, M., Takeshima, H., and Iino, M. (1999). Spatiotemporal dynamics of inositol 1,4,5-trisphosphate that underlies complex Ca2+ mobilization patterns. Science (New York, NY) *284*, 1527-1530.

Holm, L., and Rosenstrom, P. (2010). Dali server: conservation mapping in 3D. Nucleic acids research *38*, W545-549.

Howlett, G.J., Minton, A.P., and Rivas, G. (2006). Analytical ultracentrifugation for the study of protein association and assembly. Current opinion in chemical biology *10*, 430-436.

Huey, R., Morris, G.M., Olson, A.J., and Goodsell, D.S. (2007). A Semi-empirical Free Energy Force Field with Charge-Based Desolvation. J Comput Chem *28*, 1145-1152.

Hynes, R.O. (2002). Integrins: bidirectional, allosteric signaling machines. Cell *110*, 673-687.

Hyvonen, M., Macias, M.J., Nilges, M., Oschkinat, H., Saraste, M., and Wilmanns, M. (1995). Structure of the binding site for inositol phosphates in a PH domain. The EMBO journal *14*, 4676-4685.

Jackson, S.G., Zhang, Y., Haslam, R.J., and Junop, M.S. (2007). Structural analysis of the carboxy terminal PH domain of pleckstrin bound to D-myo-inositol 1,2,3,5,6-pentakisphosphate. BMC Struct Biol *7*, 80.

Jobard, F., Bouadjar, B., Caux, F., Hadj-Rabia, S., Has, C., Matsuda, F., Weissenbach, J., Lathrop, M., Prud'homme, J.F., and Fischer, J. (2003). Identification of mutations in a new gene encoding a FERM family protein with a pleckstrin homology domain in Kindler syndrome. Human molecular genetics *12*, 925-935.

Johnson, C.D., Esquela-Kerscher, A., Stefani, G., Byrom, M., Kelnar, K., Ovcharenko, D., Wilson, M., Wang, X., Shelton, J., Shingara, J., *et al.* (2007). The let-7 microRNA represses cell proliferation pathways in human cells. Cancer research *67*, 7713-7722.

Johnson, S.M., Grosshans, H., Shingara, J., Byrom, M., Jarvis, R., Cheng, A., Labourier, E., Reinert, K.L., Brown, D., and Slack, F.J. (2005). RAS is regulated by the let-7 microRNA family. Cell *120*, 635-647.

Jones, D.T., Taylor, W.R., and Thornton, J.M. (1992). A new approach to protein fold recognition. Nature *358*, 86-89.

Jones, M.R., Quinton, L.J., Blahna, M.T., Neilson, J.R., Fu, S., Ivanov, A.R., Wolf, D.A., and Mizgerd, J.P. (2009). Zcchc11-dependent uridylation of microRNA directs cytokine expression. Nature cell biology *11*, 1157-1163.

Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., and Klein, M.L. (1983). Comparison of simple potential functions for simulating liquid water J Chem Phys *79*, 926-934.

Kabsch, W. (2010). Xds. Acta crystallographica *66*, 125-132.

Kaminski, G.A., Friesner, R.A., Tirado-Rives, J., and Jorgensen, W.L. (2001). Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. J Phys Chem B *105*, 6474-6487.

Kato, K., Shiozawa, T., Mitsushita, J., Toda, A., Horiuchi, A., Nikaido, T., Fujii, S., and Konishi, I. (2004). Expression of the mitogen-inducible gene-2 (mig-2) is elevated in human uterine leiomyomas but not in leiomyosarcomas. Human pathology *35*, 55-60.

Kavran, J.M., Klein, D.E., Lee, A., Falasca, M., Isakoff, S.J., Skolnik, E.Y., and Lemmon, M.A. (1998). Specificity and promiscuity in phosphoinositide binding by pleckstrin homology domains. The Journal of biological chemistry *273*, 30497-30508.

Kawate, T., Michel, J.C., Birdsong, W.T., and Gouaux, E. (2009). Crystal structure of the ATP-gated P2X(4) ion channel in the closed state. Nature *460*, 592-598.

Kelley, L.A., and Sutcliffe, M.J. (1997). OLDERADO: on-line database of ensemble representatives and domains. On Line Database of Ensemble Representatives And DOmains. Protein Sci *6*, 2628-2630.

Kern, J.S., Herz, C., Haan, E., Moore, D., Nottelmann, S., von Lilien, T., Greiner, P., Schmitt-Graeff, A., Opitz, O.G., Bruckner-Tuderman, L., *et al.* (2007). Chronic colitis due to an epithelial barrier defect: the role of kindlin-1 isoforms. The Journal of pathology *213*, 462-470.

Kim, C.U., Lew, W., Williams, M.A., Liu, H., Zhang, L., Swaminathan, S., Bischofberger, N., Chen, M.S., Mendel, D.B., Tai, C.Y., *et al.* (1997). Influenza neuraminidase inhibitors possessing a novel hydrophobic interaction in the enzyme active site: design, synthesis, and structural analysis of carbocyclic sialic acid analogues with potent anti-influenza activity. J Am Chem Soc *119*, 681-690.

Kim, H.H., Kuwano, Y., Srikantan, S., Lee, E.K., Martindale, J.L., and Gorospe, M. (2009). HuR recruits let-7/RISC to repress c-Myc expression. Genes & development *23*, 1743-1748.

Kindler, T. (1954). Congenital poikiloderma with traumatic bulla formation and progressive cutaneous atrophy. The British journal of dermatology *66*, 104-111.

Klarlund, J.K., Guilherme, A., Holik, J.J., Virbasius, J.V., Chawla, A., and Czech, M.P. (1997). Signaling by phosphoinositide-3,4,5-trisphosphate through proteins containing pleckstrin and Sec7 homology domains. Science (New York, NY *275*, 1927-1930.

Klarlund, J.K., Tsiaras, W., Holik, J.J., Chawla, A., and Czech, M.P. (2000). Distinct polyphosphoinositide binding selectivities for pleckstrin homology domains of GRP1-like proteins based on diglycine versus triglycine motifs. The Journal of biological chemistry *275*, 32816-32821.

Klein, E.A., Yin, L., Kothapalli, D., Castagnino, P., Byfield, F.J., Xu, T., Levental, I., Hawthorne, E., Janmey, P.A., and Assoian, R.K. (2009). Cell-cycle control by physiological matrix elasticity and in vivo tissue stiffening. Curr Biol *19*, 1511-1518.

Kleywegt, G.J. (2007). Crystallographic refinement of ligand complexes. Acta crystallographica *63*, 94-100.

Kloeker, S., Major, M.B., Calderwood, D.A., Ginsberg, M.H., Jones, D.A., and Beckerle, M.C. (2004). The Kindler syndrome protein is regulated by transforming growth factor-beta and involved in integrin-mediated adhesion. The Journal of biological chemistry *279*, 6824-6833.

Knoll, R., Postel, R., Wang, J., Kratzner, R., Hennecke, G., Vacaru, A.M., Vakeel, P., Schubert, C., Murthy, K., Rana, B.K., *et al.* (2007). Laminin-alpha4 and integrin-linked kinase mutations cause human cardiomyopathy via simultaneous defects in cardiomyocytes and endothelial cells. Circulation *116*, 515-525.

Koch, M.H., Vachette, P., and Svergun, D.I. (2003). Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. Quarterly reviews of biophysics *36*, 147-227.

Konarev, P., Volkov, V.V., Sokolova, A.V., Koch, M.H., and Svergun, D.I. (2003). PRIMUS - a Windows-PC based system for small-angle scattering data analysis. J Appl Crystallogr *36*, 1277-1282.

Kondo, J., and Westhof, E. (2011). Classification of pseudo pairs between nucleotide bases and amino acids by analysis of nucleotide-protein complexes. Nucleic acids research *39*, 8628-8637.

Kratky, O., Porod, G., and Kahovec, L. (1951). Einige neuerungen in der technik und auswertung von Roentgen-kleinwinkelmessungen. Z Elektrochem *55*, 53-59.

Kruger, M., Moser, M., Ussar, S., Thievessen, I., Luber, C.A., Forner, F., Schmidt, S., Zanivan, S., Fassler, R., and Mann, M. (2008). SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function. Cell *134*, 353-364.

Kumagai, A., Lee, J., Yoo, H.Y., and Dunphy, W.G. (2006). TopBP1 activates the ATR-ATRIP complex. Cell *124*, 943-955.

Kwak, J.E., Wang, L., Ballantyne, S., Kimble, J., and Wickens, M. (2004). Mammalian GLD-2 homologs are poly(A) polymerases. Proceedings of the National Academy of Sciences of the United States of America *101*, 4407-4412.

Langer, G., Cohen, S.X., Lamzin, V.S., and Perrakis, A. (2008). Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. Nature protocols *3*, 1171-1179.

Larjava, H., Plow, E.F., and Wu, C. (2008). Kindlins: essential regulators of integrin signalling and cell-matrix adhesion. EMBO reports *9*, 1203-1208.

Lau, T.L., Kim, C., Ginsberg, M.H., and Ulmer, T.S. (2009). The structure of the integrin alphaIIbbeta3 transmembrane complex explains integrin transmembrane signalling. The EMBO journal *28*, 1351-1361.

Lebowitz, J., Lewis, M.S., and Schuck, P. (2002). Modern analytical ultracentrifugation in protein science: a tutorial review. Protein Sci *11*, 2067-2079.

Legate, K.R., and Fassler, R. (2009). Mechanisms that regulate adaptor binding to beta-integrin cytoplasmic tails. Journal of cell science *122*, 187-198.

Legate, K.R., Takahashi, S., Bonakdar, N., Fabry, B., Boettiger, D., Zent, R., and Fassler, R. (2011). Integrin adhesion and force coupling are independently regulated by localized PtdIns(4,5)(2) synthesis. The EMBO journal.

Lehrbach, N.J., Armisen, J., Lightfoot, H.L., Murfitt, K.J., Bugaut, A., Balasubramanian, S., and Miska, E.A. (2009). LIN-28 and the poly(U) polymerase PUP-2 regulate let-7 microRNA processing in Caenorhabditis elegans. Nat Struct Mol Biol *16*, 1016-1020.

Lemmon, M.A. (2005). Pleckstrin homology domains: two halves make a hole? Cell *120*, 574-576.

Lemmon, M.A. (2008). Membrane recognition by phospholipid-binding domains. Nature reviews *9*, 99-111.

Lemmon, M.A., and Ferguson, K.M. (2000). Signal-dependent membrane targeting by pleckstrin homology (PH) domains. The Biochemical journal *350 Pt 1*, 1-18.

Lemmon, M.A., and Ferguson, K.M. (2001). Molecular determinants in pleckstrin homology domains that allow specific recognition of phosphoinositides. Biochemical Society transactions *29*, 377-384.

Lemmon, M.A., Ferguson, K.M., and Abrams, C.S. (2002). Pleckstrin homology domains and the cytoskeleton. FEBS letters *513*, 71-76.

Leslie, A.G.W., and Powell, H.R. (2007). Processing diffraction data with MOSFLM. In Evolving methods for macromolecular crystallography, R.J. Read, and J.L. Sussman, eds. (Dordrecht: Springer), pp. 41-51.

Li, C., Xu, L., Wolan, D.W., Wilson, I.A., and Olson, A.J. (2004). Virtual screening of human 5-aminoimidazole-4-carboxamide ribonucleotide transformylase against the NCI diversity set by use of AutoDock to identify novel nonfolate inhibitors. J Med Chem *47*, 6681-6690.

Li, F., Xiong, Y., Wang, J., Cho, H.D., Tomita, K., Weiner, A.M., and Steitz, T.A. (2002). Crystal structures of the Bacillus stearothermophilus CCA-adding enzyme and its complexes with ATP or CTP. Cell *111*, 815-824.

Lindvall, J.M., Blomberg, K.E., Valiaho, J., Vargas, L., Heinonen, J.E., Berglof, A., Mohamed, A.J., Nore, B.F., Vihinen, M., and Smith, C.I. (2005). Bruton's tyrosine kinase: cell biology, sequence conservation, mutation spectrum, siRNA modifications, and expression profiling. Immunological reviews *203*, 200-215.

Lingwood, C., Mylvaganam, M., Minhas, F., Binnington, B., Branch, D.R., and Pomes, R. (2005). The sulfogalactose moiety of sulfoglycosphingolipids serves as a mimic of tyrosine phosphate in many recognition processes. Prediction and demonstration of Src homology 2 domain/sulfogalactose binding. The Journal of biological chemistry *280*, 12542-12547.

Liu, J., Fukuda, K., Xu, Z., Ma, Y.Q., Hirbawi, J., Mao, X., Wu, C., Plow, E.F., and Qin, J. (2011). Structural basis of phosphoinositide binding to Kindlin-2 pleckstrin homology domain in regulating integrin activation. The Journal of biological chemistry.

Lodowski, D.T., Pitcher, J.A., Capel, W.D., Lefkowitz, R.J., and Tesmer, J.J. (2003). Keeping G proteins at bay: a complex between G protein-coupled receptor kinase 2 and Gbetagamma. Science (New York, NY *300*, 1256-1262.

Luo, B.H., Carman, C.V., and Springer, T.A. (2007). Structural basis of integrin regulation and signaling. Annual review of immunology *25*, 619-647.

Ma, Y.Q., Qin, J., Wu, C., and Plow, E.F. (2008). Kindlin-2 (Mig-2): a co-activator of beta3 integrins. The Journal of cell biology *181*, 439-446.

Macias, M.J., Musacchio, A., Ponstingl, H., Nilges, M., Saraste, M., and Oschkinat, H. (1994). Structure of the pleckstrin homology domain from beta-spectrin. Nature *369*, 675-677.

Mackinnon, A.C., Qadota, H., Norman, K.R., Moerman, D.G., and Williams, B.D. (2002). C. elegans PAT-4/ILK functions as an adaptor protein within integrin adhesion complexes. Curr Biol *12*, 787-797.

Malumbres, M., and Barbacid, M. (2001). To cycle or not to cycle: a critical decision in cancer. Nature reviews Cancer *1*, 222-231.

Malumbres, M., and Barbacid, M. (2005). Mammalian cyclin-dependent kinases. Trends Biochem Sci *30*, 630-641.

Martin, G., Jeno, P., and Keller, W. (1999). Mapping of ATP binding regions in poly(A) polymerases by photoaffinity labeling and by mutational analysis identifies a domain conserved in many nucleotidyltransferases. Protein Sci *8*, 2380-2391.

Martin, G., and Keller, W. (2007). RNA-specific ribonucleotidyl transferases. Rna *13*, 1834-1849.

Martin, G., Keller, W., and Doublie, S. (2000). Crystal structure of mammalian poly(A) polymerase in complex with an analog of ATP. The EMBO journal *19*, 4193-4203.

Matthews, B.W. (1968). Solvent content of protein crystals. Journal of molecular biology *33*, 491-497.

Mayer, B.J., Ren, R., Clark, K.L., and Baltimore, D. (1993). A putative modular domain present in diverse signaling proteins. Cell *73*, 629-630.

Mayo, C.J., Diprose, J.M., Walter, T.S., Berry, I.M., Wilson, J., Owens, R.J., Jones, E.Y., Harlos, K., Stuart, D.I., and Esnouf, R.M. (2005). Benefits of automated crystallization plate tracking, imaging, and analysis. Structure *13*, 175-182.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J Appl Crystallogr *40*, 658-674.

Merrington, C.L., Bailey, M.J., and Possee, R.D. (1997). Manipulation of baculovirus vectors. Mol Biotechnol *8*, 283-297.

Messerschmidt, A. (2007). X-Ray Crystallography of Biomacromolecules: A Practical Guide. *Wiley VCH*.

Meves, A., Stremmel, C., Gottschalk, K., and Fassler, R. (2009). The Kindlin protein family: new members to the club of focal adhesion proteins. Trends in cell biology *19*, 504-513.

Milburn, C.C., Deak, M., Kelly, S.M., Price, N.C., Alessi, D.R., and Van Aalten, D.M. (2003). Binding of phosphatidylinositol 3,4,5-trisphosphate to the pleckstrin homology domain of protein kinase B induces a conformational change. The Biochemical journal *375*, 531-538.

Milchev, G.I., and Hadjiolov, A.A. (1978). Association of poly(A) and poly(U) polymerases with cytoplasmic ribosomes. European journal of biochemistry / FEBS *84*, 113-121.

Miller, G.J., Wilson, M.P., Majerus, P.W., and Hurley, J.H. (2005). Specificity determinants in inositol polyphosphate synthesis: crystal structure of inositol 1,3,4-trisphosphate 5/6-kinase. Molecular cell *18*, 201-212.

Montanez, E., Ussar, S., Schifferer, M., Bosl, M., Zent, R., Moser, M., and Fassler, R. (2008). Kindlin-2 controls bidirectional signaling of integrins. Genes & development *22*, 1325-1330.

Morgan, S.E., Lovly, C., Pandita, T.K., Shiloh, Y., and Kastan, M.B. (1997). Fragments of ATM which have dominant-negative or complementing activity. Molecular and cellular biology *17*, 2020-2029.

Morris, G.M., Goodsell, D.S., Halliday, R.A., Huey, R., Hart, W.E., Belew, R.K., and Olson, A.J. (1998). Automated Docking Using a Lamarckian Genetic Algorithm and and Empirical Binding Free Energy Function. J Comput Chem *19*, 1639-1662.

Moser, B.A., Brondello, J.M., Baber-Furnari, B., and Russell, P. (2000). Mechanism of caffeine-induced checkpoint override in fission yeast. Molecular and cellular biology *20*, 4288-4294.

Moser, M., Legate, K.R., Zent, R., and Fassler, R. (2009). The tail of integrins, talin, and kindlins. Science (New York, NY *324*, 895-899.

Moser, M., Nieswandt, B., Ussar, S., Pozgajova, M., and Fassler, R. (2008). Kindlin-3 is essential for integrin activation and platelet aggregation. Nature medicine *14*, 325-330.

Munger, J.S., and Sheppard, D. (2011). Cross talk among TGF-beta signaling pathways, integrins, and the extracellular matrix. Cold Spring Harb Perspect Biol *3*, a005017.

Musacchio, A., Gibson, T., Rice, P., Thompson, J., and Saraste, M. (1993). The PH domain: a common piece in the structural patchwork of signalling proteins. Trends Biochem Sci *18*, 343-348.

Nakanishi, T., Kubota, H., Ishibashi, N., Kumagai, S., Watanabe, H., Yamashita, M., Kashiwabara, S., Miyado, K., and Baba, T. (2006). Possible role of mouse poly(A) polymerase mGLD-2 during oocyte maturation. Developmental biology *289*, 115-126.

Nakayama, K.I., and Nakayama, K. (2006). Ubiquitin ligases: cell-cycle control and cancer. Nature reviews Cancer *6*, 369-381.

Nam, Y., Chen, C., Gregory, R.I., Chou, J.J., and Sliz, P. (2011). Molecular basis for interaction of let-7 microRNAs with Lin28. Cell *147*, 1080-1091.

Nettleship, J.E., Assenberg, R., Diprose, J.M., Rahman-Huq, N., and Owens, R.J. (2010). Recent advances in the production of proteins in insect and mammalian cells for structural biology. Journal of structural biology *172*, 55-65.

Norbury, C., and Nurse, P. (1992). Animal cell cycles and their control. Annual review of biochemistry *61*, 441-470.

Nose, K., and Klein, M.L. (1983). Constant pressure molecular dynamics for molecular systems. Mol Phys *50*, 1055-1076.

O'Toole, T.E., Katagiri, Y., Faull, R.J., Peter, K., Tamura, R., Quaranta, V., Loftus, J.C., Shattil, S.J., and Ginsberg, M.H. (1994). Integrin cytoplasmic domains mediate inside-out signal transduction. The Journal of cell biology *124*, 1047-1059.

Oxley, C.L., Anthis, N.J., Lowe, E.D., Vakonakis, I., Campbell, I.D., and Wegener, K.L. (2008). An integrin phosphorylation switch: the effect of beta3 integrin tail phosphorylation on Dok1 and talin binding. The Journal of biological chemistry *283*, 5420-5426.

Painter, J., and Merritt, E.A. (2006). Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta crystallographica *62*, 439-450.

Pardee, A.B. (1974). A restriction point for control of normal animal cell proliferation. Proceedings of the National Academy of Sciences of the United States of America *71*, 1286-1290.

Parrinello, M., and Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method J Appl Phys *52*, 7182-7190.

Paterson, H.F., Savopoulos, J.W., Perisic, O., Cheung, R., Ellis, M.V., Williams, R.L., and Katan, M. (1995). Phospholipase C delta 1 requires a pleckstrin homology domain for interaction with the plasma membrane. The Biochemical journal *312 (Pt 3)*, 661-666.

Patla, I., Volberg, T., Elad, N., Hirschfeld-Warneken, V., Grashoff, C., Fassler, R., Spatz, J.P., Geiger, B., and Medalia, O. (2010). Dissecting the molecular architecture of integrin adhesion sites by cryo-electron tomography. Nature cell biology *12*, 909-915.

Pedersen, L.C., Benning, M.M., and Holden, H.M. (1995). Structural investigation of the antibiotic and ATP-binding sites in kanamycin nucleotidyltransferase. Biochemistry *34*, 13305-13311.

Perera, H.D., Ma, Y.Q., Yang, J., Hirbawi, J., Plow, E.F., Qin, J. (2011). Membrane Binding of the N-Terminal Ubiquitin-Like Domain of kindlin-2 Is Crucial for Its Regulation of Integrin Activation. Structure *19*, 1664-1671

Petoukhov, M.V., and Svergun, D.I. (2005). Global rigid body modeling of macromolecular complexes against small-angle scattering data. Biophysical journal *89*, 1237-1250.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem *25*, 1605-1612.

Piskounova, E., Polytarchou, C., Thornton, J.E., Lapierre, R.J., Pothoulakis, C., Hagan, J.P., Iliopoulos, D., and Gregory, R.I. (2011). Lin28A and Lin28B Inhibit let-7 MicroRNA Biogenesis by Distinct Mechanisms. Cell *147*, 1066-1079.

Prehoda, K.E., Lee, D.J., and Lim, W.A. (1999). Structure of the enabled/VASP homology 1 domain-peptide complex: a key component in the spatial control of actin assembly. Cell *97*, 471-480.

Proudfoot, N., and O'Sullivan, J. (2002). Polyadenylation: a tail of two complexes. Curr Biol *12*, R855-857.

Putnam, C.D., Hammel, M., Hura, G.L., and Tainer, J.A. (2007). X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. Quarterly reviews of biophysics *40*, 191-285.

Qu, H., Tu, Y., Shi, X., Larjava, H., Saleem, M.A., Shattil, S.J., Fukuda, K., Qin, J., Kretzler, M., and Wu, C. (2011). Kindlin-2 regulates podocyte adhesion and fibronectin matrix deposition through interactions with phosphoinositides and integrins. Journal of cell science *124*, 879-891.

Read, R.L., and Norbury, C.J. (2002). Roles for cytoplasmic polyadenylation in cell cycle regulation. J Cell Biochem *87*, 258-265.

Rebecchi, M.J., and Scarlata, S. (1998). Pleckstrin homology domains: a common fold with diverse functions. Annual review of biophysics and biomolecular structure *27*, 503-528.

Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R., and Ruvkun, G. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature *403*, 901-906.

Riffel, N., Harlos, K., Iourin, O., Rao, Z., Kingsman, A., Stuart, D., and Fry, E. (2002). Atomic resolution structure of Moloney murine leukemia virus matrix protein and its relationship to other retroviral matrix proteins. Structure *10*, 1627-1636.

Ringpis, G.E., Aphasizheva, I., Wang, X., Huang, L., Lathrop, R.H., Hatfield, G.W., and Aphasizhev, R. (2010a). Mechanism of U insertion RNA editing in trypanosome mitochondria: the bimodal TUTase activity of the core complex. Journal of molecular biology *399*, 680-695.

Ringpis, G.E., Stagno, J., and Aphasizhev, R. (2010b). Mechanism of U-insertion RNA editing in trypanosome mitochondria: characterization of RET2 functional domains by mutational analysis. Journal of molecular biology *399*, 696-706.

Rissland, O.S., Mikulasova, A., and Norbury, C.J. (2007). Efficient RNA polyuridylation by noncanonical poly(A) polymerases. Molecular and cellular biology *27*, 3612-3624.

Rissland, O.S., and Norbury, C.J. (2008). The Cid1 poly(U) polymerase. Biochimica et biophysica acta *1779*, 286-294.

Rissland, O.S., and Norbury, C.J. (2009). Decapping is preceded by 3' uridylation in a novel pathway of bulk mRNA turnover. Nat Struct Mol Biol *16*, 616-623.

Roessle, M.W., Klaering, R., Ristau, U., Robrahn, B., Jahn, D., Gehrmann, T., Konarev, P., Round, A., Fiedler, S., Hermes, C., *et al.* (2007). Upgrade of the small-angle x-ray scattering beamline x33 at the european molecular biology laboratory, hamburg. J Appl Crystallogr *40*, s190-s194.

Rogalski, T.M., Mullen, G.P., Gilbert, M.M., Williams, B.D., and Moerman, D.G. (2000). The UNC-112 gene in Caenorhabditis elegans encodes a novel component of cell-matrix adhesion structures required for integrin localization in the muscle cell membrane. The Journal of cell biology *150*, 253-264.

Rogozin, I.B., Makarova, K.S., Pavlov, Y.I., and Koonin, E.V. (2008). A highly conserved family of inactivated archaeal B family DNA polymerases. Biology direct *3*, 32.

Rossman, K.L., and Sondek, J. (2005). Larger than Dbl: new structural insights into RhoA activation. Trends Biochem Sci *30*, 163-165.

Rossman, K.L., Worthylake, D.K., Snyder, J.T., Siderovski, D.P., Campbell, S.L., and Sondek, J. (2002). A crystallographic view of interactions between Dbs and Cdc42: PH domain-assisted guanine nucleotide exchange. The EMBO journal *21*, 1315-1326.

Roush, S., and Slack, F.J. (2008). The let-7 family of microRNAs. Trends in cell biology *18*, 505-516.

Rupp, B., (2011). Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology. *Garland Science*.

Sainsbury, S., Ren, J., Saunders, N.J., Stuart, D.I., and Owens, R.J. (2008). Crystallization and preliminary X-ray analysis of CrgA, a LysR-type transcriptional regulator from pathogenic Neisseria meningitidis MC58. Acta Crystallogr Sect F Struct Biol Cryst Commun *64*, 797-801.

Saitoh, S., Chabes, A., McDonald, W.H., Thelander, L., Yates, J.R., and Russell, P. (2002). Cid13 is a cytoplasmic poly(A) polymerase that regulates ribonucleotide reductase mRNA. Cell *109*, 563-573.

Sali, A., and Blundell, T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. Journal of molecular biology *234*, 779-815.

Sambrook, J., Russel, D., (2000) Molecular Cloning (3-volume set): A Laboratory Manual. *Cold Spring Harbor Laboratory Press*. U.S.

Sands, D.E., (1994) Introduction to Crystallography. *Dover Publications.*

Schmeing, T.M., and Ramakrishnan, V. (2009). What recent ribosome structures have revealed about the mechanism of translation. Nature *461*, 1234-1242.

Schmidt, M.-J., West, S., and Norbury, C. (2011a). The human cytoplasmic RNA terminal U-transferase ZCCHC11 targets histone mRNAs for degradation. RNA *in press*.

Schmidt, S., Nakchbandi, I., Ruppert, R., Kawelke, N., Hess, M.W., Pfaller, K., Jurdic, P., Fassler, R., and Moser, M. (2011b). Kindlin-3-mediated signaling from multiple integrin classes is required for osteoclast-mediated bone resorption. The Journal of cell biology *192*, 883-897.

Schuttelkopf, A.W., and van Aalten, D.M. (2004). PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. Acta crystallographica *60*, 1355-1363.

Shaffer, P.L., Goehring, A., Shankaranarayanan, A., and Gouaux, E. (2009). Structure and mechanism of a Na+-independent amino acid transporter. Science (New York, NY *325*, 1010-1014.

Shaw, G. (1993). Identification of novel pleckstrin homology (PH) domains provides a hypothesis for PH domain function. Biochemical and biophysical research communications *195*, 1145-1151.

Shen, B., and Goodman, H.M. (2004). Uridine addition after microRNA-directed cleavage. Science (New York, NY *306*, 997.

Sherr, C.J., and Roberts, J.M. (1999). CDK inhibitors: positive and negative regulators of G1-phase progression. Genes & development *13*, 1501-1512.

Sherwood, D., Cooper, J.B., (2010). Crystals, X-rays and Proteins: Comprehensive Protein Crystallography. *Oxford University Press*. Oxford.

Shi, X., Ma, Y.Q., Tu, Y., Chen, K., Wu, S., Fukuda, K., Qin, J., Plow, E.F., and Wu, C. (2007). The MIG-2/integrin interaction strengthens cell-matrix adhesion and modulates cell motility. The Journal of biological chemistry *282*, 20455-20466.

Siegel, D.H., Ashton, G.H., Penagos, H.G., Lee, J.V., Feiler, H.S., Wilhelmsen, K.C., South, A.P., Smith, F.J., Prescott, A.R., Wessagowit, V., *et al.* (2003). Loss of kindlin-1, a human homolog of the Caenorhabditis elegans actin-extracellular-matrix linker protein UNC-112, causes Kindler syndrome. Am J Hum Genet *73*, 174-187.

Simpson, L., Sbicego, S., and Aphasizhev, R. (2003). Uridine insertion/deletion RNA editing in trypanosome mitochondria: a complex business. RNA *9*, 265-276.

Smyth, D.R., Mrozkiewicz, M.K., McGrath, W.J., Listwan, P., and Kobe, B. (2003). Crystal structures of fusion proteins with large-affinity tags. Protein Sci *12*, 1313-1322.

Snyder, J.T., Worthylake, D.K., Rossman, K.L., Betts, L., Pruitt, W.M., Siderovski, D.P., Der, C.J., and Sondek, J. (2002). Structural basis for the selective activation of Rho GTPases by Dbl exchange factors. Nat Struct Biol *9*, 468-475.

Sobolevsky, A.I., Rosconi, M.P., and Gouaux, E. (2009). X-ray structure, symmetry and mechanism of an AMPA-subtype glutamate receptor. Nature *462*, 745-756.

Soisson, S.M., Nimnual, A.S., Uy, M., Bar-Sagi, D., and Kuriyan, J. (1998). Crystal structure of the Dbl and pleckstrin homology domains from the human Son of sevenless protein. Cell *95*, 259-268.

Soisson, S.M., Parthasarathy, G., Adams, A.D., Sahoo, S., Sitlani, A., Sparrow, C., Cui, J., and Becker, J.W. (2008). Identification of a potent synthetic FXR agonist with an unexpected mode of binding and activation. Proceedings of the National Academy of Sciences of the United States of America *105*, 5337-5342.

Stagno, J., Aphasizheva, I., Aphasizhev, R., and Luecke, H. (2007a). Dual role of the RNA substrate in selectivity and catalysis by terminal uridylyl transferases. Proceedings of the National Academy of Sciences of the United States of America *104*, 14634-14639.

Stagno, J., Aphasizheva, I., Bruystens, J., Luecke, H., and Aphasizhev, R. (2010). Structure of the mitochondrial editosome-like complex associated TUTase 1 reveals divergent mechanisms of UTP selection and domain organization. Journal of molecular biology *399*, 464-475.

Stagno, J., Aphasizheva, I., Rosengarth, A., Luecke, H., and Aphasizhev, R. (2007b). UTP-bound and Apo structures of a minimal RNA uridylyltransferase. Journal of molecular biology *366*, 882-899.

Stein, N.J. (2008). CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. J Appl Cryst *41*, 641-643.

Stevenson, A.L., and Norbury, C.J. (2006). The Cid1 family of non-canonical poly(A) polymerases. Yeast *23*, 991-1000.

Strong, M., Sawaya, M.R., Wang, S., Phillips, M., Cascio, D., and Eisenberg, D. (2006). Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis. Proceedings of the National Academy of Sciences of the United States of America *103*, 8060-8065.

Stuart, D.I., Levine, M., Muirhead, H., and Stammers, D.K. (1979). Crystal structure of cat muscle pyruvate kinase at a resolution of 2.6 A. Journal of molecular biology *134*, 109-142.

Studier, F.W. (2005). Protein production by auto-induction in high density shaking cultures. Protein Expr Purif *41*, 207-234.

Suh, N., Jedamzik, B., Eckmann, C.R., Wickens, M., and Kimble, J. (2006). The GLD-2 poly(A) polymerase activates gld-1 mRNA in the Caenorhabditis elegans germ line. Proceedings of the National Academy of Sciences of the United States of America *103*, 15108-15112.

Svensson, L., Howarth, K., McDowall, A., Patzak, I., Evans, R., Ussar, S., Moser, M., Metin, A., Fried, M., Tomlinson, I., *et al.* (2009). Leukocyte adhesion deficiency-III is caused by mutations in KINDLIN3 affecting integrin activation. Nature medicine *15*, 306-312.

Svergun, D.I. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J Appl Crystallogr *25*, 495-503.

Svergun, D.I., Petoukhov, M.V., and Koch, M.H. (2001). Determination of domain structure of proteins from X-ray solution scattering. Biophysical journal *80*, 2946-2953.

Tadokoro, S., Shattil, S.J., Eto, K., Tai, V., Liddington, R.C., de Pereda, J.M., Ginsberg, M.H., and Calderwood, D.A. (2003). Talin binding to integrin beta tails: a final common step in integrin activation. Science (New York, NY *302*, 103-106.

Tamkun, J.W., DeSimone, D.W., Fonda, D., Patel, R.S., Buck, C., Horwitz, A.F., and Hynes, R.O. (1986). Structure of integrin, a glycoprotein involved in the transmembrane linkage between fibronectin and actin. Cell *46*, 271-282.

Terawaki, S., Kitano, K., Mori, T., Zhai, Y., Higuchi, Y., Itoh, N., Watanabe, T., Kaibuchi, K., and Hakoshima, T. (2010). The PHCCEx domain of Tiam1/2 is a novel protein- and membrane-binding module. The EMBO journal *29*, 236-250.

Thomas, C.C., Deak, M., Alessi, D.R., and van Aalten, D.M. (2002). High-resolution structure of the pleckstrin homology domain of protein kinase b/akt bound to phosphatidylinositol (3,4,5)-trisphosphate. Curr Biol *12*, 1256-1262.

Thomas, C.C., Dowler, S., Deak, M., Alessi, D.R., and van Aalten, D.M. (2001). Crystal structure of the phosphatidylinositol 3,4-bisphosphate-binding pleckstrin homology (PH) domain of tandem PH-domain-containing protein 1 (TAPP1): molecular basis of lipid specificity. The Biochemical journal *358*, 287-294.

Touhara, K., Inglese, J., Pitcher, J.A., Shaw, G., and Lefkowitz, R.J. (1994). Binding of G protein beta gamma-subunits to pleckstrin homology domains. The Journal of biological chemistry *269*, 10217-10220.

Trippe, R., Guschina, E., Hossbach, M., Urlaub, H., Luhrmann, R., and Benecke, B.J. (2006). Identification, cloning, and functional analysis of the human U6 snRNA-specific terminal uridylyl transferase. Rna *12*, 1494-1504.

Tsujishita, Y., Guo, S., Stolz, L.E., York, J.D., and Hurley, J.H. (2001). Specificity determinants in phosphoinositide dephosphorylation: crystal structure of an archetypal inositol polyphosphate 5-phosphatase. Cell *105*, 379-389.

Tu, Y., Wu, S., Shi, X., Chen, K., and Wu, C. (2003). Migfilin and Mig-2 link focal adhesions to filamin and the actin cytoskeleton and function in cell shape modulation. Cell *113*, 37-47.

Ussar, S., Moser, M., Widmaier, M., Rognoni, E., Harrer, C., Genzel-Boroviczeny, O., and Fassler, R. (2008). Loss of Kindlin-1 causes skin atrophy and lethal neonatal intestinal epithelial dysfunction. PLoS genetics *4*, e1000289.

Ussar, S., Wang, H.V., Linder, S., Fassler, R., and Moser, M. (2006). The Kindlins: subcellular localization and expression during murine development. Experimental cell research *312*, 3142-3151.

van Gunsteren, W.F., Billeter, S.R., Eising, A.A., Hunenberger, P.H., Kruger, P., Mark, A.E., Scott, W., and Tironi, I.G. (1996). Biomolecular Simulation: The GROMOS96 Manual and User Guide. (Groningen, The Netherlands: Biomos & Hochschulverlag AG an der ETH Zurich).

van Rossum, D.B., Patterson, R.L., Sharma, S., Barrow, R.K., Kornberg, M., Gill, D.L., and Snyder, S.H. (2005). Phospholipase Cgamma1 controls surface expression of TRPC3 through an intermolecular PH domain. Nature *434*, 99-104.

Varghese, J.N., McKimm-Breschkin, J.L., Caldwell, J.B., Kortt, A.A., and Colman, P.M. (1992). The structure of the complex between influenza virus neuraminidase and sialic acid, the viral receptor. Proteins *14*, 327-332.

Vetter, I.R., Arndt, A., Kutay, U., Gorlich, D., and Wittinghofer, A. (1999). Structural view of the Ran-Importin beta interaction at 2.3 A resolution. Cell *97*, 635-646.

Volkov, V.V., and Svergun, D.I. (2003). Uniqueness of ab initio shape determination in small-angle scattering. J Appl Crystallogr *36*, 860-864.

von Itzstein, M., Wu, W.Y., Kok, G.B., Pegg, M.S., Dyason, J.C., Jin, B., Van Phan, T., Smythe, M.L., White, H.F., Oliver, S.W., *et al.* (1993). Rational design of potent sialidase-based inhibitors of influenza virus replication. Nature *363*, 418-423.

Walter, T.S., Diprose, J.M., Mayo, C.J., Siebold, C., Pickford, M.G., Carter, L., Sutton, G.C., Berrow, N.S., Brown, J., Berry, I.M., *et al.* (2005). A procedure for setting up high-throughput nanolitre crystallization experiments. Crystallization workflow for initial screening, automated storage, imaging and optimization. Acta crystallographica *61*, 651-657.

Walter, T.S., Mancini, E.J., Kadlec, J., Graham, S.C., Assenberg, R., Ren, J., Sainsbury, S., Owens, R.J., Stuart, D.I., Grimes, J.M., *et al.* (2008). Semi-automated microseeding of nanolitre crystallization experiments. Acta Crystallogr Sect F Struct Biol Cryst Commun *64*, 14-18.

Walter, T.S., Meier, C., Assenberg, R., Au, K.F., Ren, J., Verma, A., Nettleship, J.E., Owens, R.J., Stuart, D.I., and Grimes, J.M. (2006). Lysine methylation as a routine rescue strategy for protein crystallization. Structure *14*, 1617-1622.

Wang, L., Eckmann, C.R., Kadyk, L.C., Wickens, M., and Kimble, J. (2002). A regulatory cytoplasmic poly(A) polymerase in Caenorhabditis elegans. Nature *419*, 312-316.

Wang, S.W., Norbury, C., Harris, A.L., and Toda, T. (1999a). Caffeine can override the S-M checkpoint in fission yeast. Journal of cell science *112 (Pt 6)*, 927-937.

Wang, S.W., Toda, T., MacCallum, R., Harris, A.L., and Norbury, C. (2000). Cid1, a fission yeast protein required for S-M checkpoint control when DNA polymerase delta or epsilon is inactivated. Molecular and cellular biology *20*, 3234-3244.

Wang, T., Pentyala, S., Rebecchi, M.J., and Scarlata, S. (1999b). Differential association of the pleckstrin homology domains of phospholipases C-beta 1, C-beta 2, and C-delta 1 with lipid bilayers and the beta gamma subunits of heterotrimeric G proteins. Biochemistry *38*, 1517-1524.

Wasilko, D.J., Lee, S.E., Stutzman-Engwall, K.J., Reitz, B.A., Emmons, T.L., Mathis, K.J., Bienkowski, M.J., Tomasselli, A.G., and Fischer, H.D. (2009). The titerless infected-cells

preservation and scale-up (TIPS) method for large-scale production of NO-sensitive human soluble guanylate cyclase (sGC) from insect cells infected with recombinant baculovirus. Protein Expr Purif *65*, 122-132.

Wegener, K.L., Partridge, A.W., Han, J., Pickford, A.R., Liddington, R.C., Ginsberg, M.H., and Campbell, I.D. (2007). Structural basis of integrin activation by talin. Cell *128*, 171-182.

Weinstein, E.J., Bourner, M., Head, R., Zakeri, H., Bauer, C., and Mazzarella, R. (2003). URP1: a member of a novel family of PH and FERM domain-containing membrane-associated proteins is significantly over-expressed in lung and colon carcinomas. Biochimica et biophysica acta *1637*, 207-216.

Wen, W., Liu, W., Yan, J., and Zhang, M. (2008). Structure basis and unconventional lipid membrane binding properties of the PH-C1 tandem of rho kinases. The Journal of biological chemistry *283*, 26263-26273.

Wen, W., Yan, J., and Zhang, M. (2006). Structural characterization of the split pleckstrin homology domain in phospholipase C-gamma1 and its interaction with TRPC3. The Journal of biological chemistry *281*, 12060-12068.

Wick, M., Burger, C., Brusselbach, S., Lucibello, F.C., and Muller, R. (1994). A novel member of human tissue inhibitor of metalloproteinases (TIMP) gene family is regulated during G1 progression, mitogenic stimulation, differentiation, and senescence. The Journal of biological chemistry *269*, 18953-18960.

Wilkie, N.M., and Smellie, R.M. (1968a). Chain extension of ribonucleic acid by enzymes from rat liver cytoplasm. The Biochemical journal *109*, 485-494.

Wilkie, N.M., and Smellie, R.M. (1968b). Polyribonucleotide synthesis by subfractions of microsomes from rat liver. The Biochemical journal *109*, 229-238.

Winter, G. (2010). xia2: an expert system for macromolecular crystallography data reduction. J Appl Cryst *43*, 186-190.

Worthylake, D.K., Rossman, K.L., and Sondek, J. (2000). Crystal structure of Rac1 in complex with the guanine nucleotide exchange region of Tiam1. Nature *408*, 682-688.

Worthylake, D.K., Rossman, K.L., and Sondek, J. (2004). Crystal structure of the DH/PH fragment of Dbs without bound GTPase. Structure *12*, 1078-1086.

Wykes, J.R., and Smellie, R.M. (1966). The synthesis of polyribonucleotides by cytoplasmic enzymes. The Biochemical journal *99*, 347-355.

Wyman, S.K., Knouf, E.C., Parkin, R.K., Fritz, B.R., Lin, D.W., Dennis, L.M., Krouse, M.A., Webster, P.J., and Tewari, M. (2011). Post-transcriptional generation of miRNA variants by multiple nucleotidyl transferases contributes to miRNA transcriptome complexity. Genome research *21*, 1450-1461.

Xiao, T., Takagi, J., Coller, B.S., Wang, J.H., and Springer, T.A. (2004). Structural basis for allostery in integrins and binding to fibrinogen-mimetic therapeutics. Nature *432*, 59-67.

Xie, C., Zhu, J., Chen, X., Mi, L., Nishida, N., and Springer, T.A. (2010). Structure of an integrin with an alphaI domain, complement receptor type 4. The EMBO journal *29*, 666-679.

Xiong, J.P., Stehle, T., Diefenbach, B., Zhang, R., Dunker, R., Scott, D.L., Joachimiak, A., Goodman, S.L., and Arnaout, M.A. (2001). Crystal structure of the extracellular segment of integrin alpha Vbeta3. Science (New York, NY *294*, 339-345.

Xiong, J.P., Stehle, T., Goodman, S.L., and Arnaout, M.A. (2003). Integrins, cations and ligands: making the connection. Journal of thrombosis and haemostasis : JTH *1*, 1642-1654.

Xiong, J.P., Stehle, T., Zhang, R., Joachimiak, A., Frech, M., Goodman, S.L., and Arnaout, M.A. (2002). Crystal structure of the extracellular segment of integrin alpha Vbeta3 in complex with an Arg-Gly-Asp ligand. Science (New York, NY) *296*, 151-155.

Xu, Q., Bateman, A., Finn, R.D., Abdubek, P., Astakhova, T., Axelrod, H.L., Bakolitsa, C., Carlton, D., Chen, C., Chiu, H.J., *et al.* (2010). Bacterial pleckstrin homology domains: a prokaryotic origin for the PH domain. Journal of molecular biology *396*, 31-46.

Yan, J., Wen, W., Xu, W., Long, J.F., Adams, M.E., Froehner, S.C., and Zhang, M. (2005). Structure of the split PH domain and distinct lipid-binding properties of the PH-PDZ supramodule of alpha-syntrophin. The EMBO journal *24*, 3985-3995.

Yang, Z.R., Thompson, R., McNeil, P., and Esnouf, R.M. (2005). RONN: the bio-basis function neural network tecnique applied to the detection of natively disordered regions in proteins. Bioinformatics (Oxford, England) *21*, 3369-3376.

Ye, F., Hu, G., Taylor, D., Ratnikov, B., Bobkov, A.A., McLean, M.A., Sligar, S.G., Taylor, K.A., and Ginsberg, M.H. (2010). Recreation of the terminal events in physiological integrin activation. The Journal of cell biology *188*, 157-173.

Yoon, H.S., Hajduk, P.J., Petros, A.M., Olejniczak, E.T., Meadows, R.P., and Fesik, S.W. (1994). Solution structure of a pleckstrin-homology domain. Nature *369*, 672-675.

Yu, J.W., Mendrola, J.M., Audhya, A., Singh, S., Keleti, D., DeWald, D.B., Murray, D., Emr, S.D., and Lemmon, M.A. (2004). Genome-wide analysis of membrane targeting by S. cerevisiae pleckstrin homology domains. Molecular cell *13*, 677-688.

Yurchenco, P.D. (2011). Basement membranes: cell scaffoldings and signaling platforms. Cold Spring Harb Perspect Biol *3*, 1-28

Zabel, P., Dorssers, L., Wernars, K., and Van Kammen, A. (1981). Terminal uridylyl transferase of Vigna unguiculata: purification and characterization of an enzyme catalyzing the addition of a single UMP residue to the 3'-end of an RNA primer. Nucleic acids research *9*, 2433-2453.

Zaidel-Bar, R., Itzkovitz, S., Ma'ayan, A., Iyengar, R., and Geiger, B. (2007). Functional atlas of the integrin adhesome. Nature cell biology *9*, 858-867.

Zeng, Y., Forbes, K.C., Wu, Z., Moreno, S., Piwnica-Worms, H., and Enoch, T. (1998). Replication checkpoint requires phosphorylation of the phosphatase Cdc25 by Cds1 or Chk1. Nature *395*, 507-510.

Zhao, Y., Chapman, D.A., and Jones, I.M. (2003). Improving baculovirus recombination. Nucleic acids research *31*, E6-6.

Zhou, M.M., Ravichandran, K.S., Olejniczak, E.F., Petros, A.M., Meadows, R.P., Sattler, M., Harlan, J.E., Wade, W.S., Burakoff, S.J., and Fesik, S.W. (1995). Structure and ligand recognition of the phosphotyrosine binding domain of Shc. Nature *378*, 584-592.

Zhu, J., Luo, B.H., Xiao, T., Zhang, C., Nishida, N., and Springer, T.A. (2008). Structure of a complete integrin ectodomain in a physiologic resting state and activation and deactivation by applied forces. Molecular cell *32*, 849-861.

Ziegler, W.H., Gingras, A.R., Critchley, D.R., and Emsley, J. (2008). Integrin connections to the cytoskeleton through talin and vinculin. Biochemical Society transactions *36*, 235-239.

Zou, L., and Elledge, S.J. (2003). Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. Science *300*, 1542-1548.

CHAPTER 10

10.1. Appendix

10.1.1. Publications Arising from Thesis

Yates L.A.*, Fleurdépine S.*, Rissland O.S., De Colibus L., Harlos K., Norbury C.J., Gilbert R.J. (2012) Structural Basis for the activity of a cytoplasmic RNA terminal uridylyl transferase. Nature Structural and Molecular Biology. 19, 782-787.

(*equal contribution)

Yates L.A., Fuzery A.K., Bonet R., Campbell I.D., Gilbert R.J. (2012) Biophysical analysis of kindlin-3 reveals an elongated conformation and maps integrin binding to the membrane-distal β-subunit NP*X*Y motif. Journal of Biological Chemistry. 287, 37715-37731

Yates L.A., Lumb C.N., Brahme N.N., Zalyte R., Bird L.E., De Colibus L., Owens R.J., Calderwood D.A., Sansom M.S., Gilbert R.J. (2012) Structural and functional characterisation of the kindlin-1 pleckstrin homology domain. Journal of Biological Chemistry. PMID: 23132860 [Epub ahead of print]

10.1.2. Deposited Data PDB Accession Codes

PDB: 4E7X (Apo Cid1 [form I])

PDB: 4E80 (UTP bound Cid1)

PDB: 4E8F (Apo Cid1 [form II])

PDB: 4BBK (Kindlin-1 PH domain)