MRC National Institute for Medical Research

Division of Molecular Structure

Mill Hill, London, UK

Functional analyses of pSer and pThr binding domains

A thesis submitted by

Lasse Stach

In partial fulfilment of the requirements of

University College London

For the degree of Doctor of Philosophy

September 2012

Declaration

I, Lasse Stach, declare that the work presented in this thesis was performed in the laboratory of Dr Steve Smerdon in the Division of Molecular Structure at the MRC National Institute for Medical Resea rch. I confirm that the work presented in this thesis is my own. It has been indicated in the text where information has been derived from other sources. Some of the work presented in Chapter 7 was performed in collaboration with Dr Zuzana Horejsi (Clare Hall, CRUK).

Acknowledgements

First of all I would like to thank Dr Steve Smerdon for giving me the opportunity to work in his lab. His relatively hands-off style of management has allowed me to develop my project according to my own interests with advice and guidance always at hand when needed. Thanks Steve for a memorable Ph.D. experience.

Thanks also to the guys, past and present, from lab 228 for making the last four years not only a productive, but also a thoroughly enjoyable experience, especially Tim for teaching me the ropes at the start of my Ph.D. project.

Simon, I am deeply grateful for your constructive criticisms that have helped me turn this piece of work from an essentially German manuscript into something resembling English. I don't want to know how this thesis would look without you.

I was only able to apply such a large variety of biochemical and biophysical techniques in my Ph.D. project due to the incredibly helpful nature of my colleagues in Molecular Structure and those running the core facilities at NIMR. I would like to express my deepest gratitude to everyone who has helped me carry out experiments.

The encouragement and support from friends and especially my family has been vital in keeping me going throughout. The personal and financial sacrifices made by my family to provide me with the best possible education, without which I would not even have started this Ph.D., are much appreciated.

Abstract

Since the discovery of the phosphotyrosine binding SH2 domain, many classes of phospho-recognition domains have been described which mediate many of the diverse cellular functions of protein kinases.

Among those, FHA domains are unique in their ability to exclusively recognise pThr epitopes. The genome of the human pathogen *Mycobacterium tuberculosis* encodes 5 FHA domains, along with 11 Ser/Thr protein kinases. In the first part of this thesis, it is shown that Ser/Thr protein kinase PknB phosphorylates a threonine residue in an intrinsically unstructured region of protein FhaA. FhaA contains an FHA domain through which it interacts with and presumably inhibits MviN, a muropeptide flippase essential for cell-wall synthesis. Upon phosphorylation, the FHA domain binds the pThr epitope in an intra-molecular interaction occluding the MviN binding surface and alleviating its inhibition. Although the pThr-FHA interaction is relatively weak and nonspecific, the phosphorylated molecule nonetheless assumes a 'closed' conformation 99% of the time and is therefore able to outcompete the 2 orders of magnitude stronger bimolecular FHA-MviN interaction.

In the second part, the phospho-binding capabilities of the human PIH1D1 protein were characterised. PIH1D1 has been shown to interact with a central chaperone assembly comprising the R2TP complex and Hsp90. It has also been shown to interact with co-factor Tel2 in a phospho-dependent manner essential for the stability of the 'giant' PI3-kinase-like kinases mTOR and SMG1. PIH1D1 is shown to function as a novel phospho-reader domain with a consensus binding sequence of D-pS-D-D, agreeing well with the substrate specificity of casein kinase 2. A mutant that abolishes phospho-binding was identified and used in binding experiments which showed that PIH1D1 interacts with the chaperone complex phospho-independently and that its phospho-binding capacity is utilised to recruit a subset of CK2 substrates to the chaperone complex.

Table of contents

List of figures

List of tables

List of abbreviations

Å	angstrom
ATM	Ataxia telangiectasia mutated
ATR	ATM and Rad3 related
µl/ µM	microliter/micromolar
ATP	adenosine triphosphate
CK2	casein kinase 2
CRUK	Cancer Research UK
CD	circular dichroism
Da	Dalton
DNA	deoxyribonucleic acid
DTT	dithiothreitol
ECD	*Homo sapiens* ecdysoneless homolog (*Drosophila*)
EDTA	ethylenediaminetetraacetic acid
FHA	forkhead-associated
FhaA	FHA domain containing protein A
GarA	glycogen accumulation regulator
GST	glutathione-S-transferase
HPLC	high performance liquid chromatography
Hsp90	heat shock protein 90
HSQC	heteronuclear single quantum coherence
IPTG	isopropyl-β-D-thiogalactopyrasonide
K	equilibrium constant
K_B	binding constant
K_D	dissociation constant
LB	Luria broth
MALDI	matrix assisted laser desorption ionisation
MALLS	multi angle laser light scattering
min	minute
ms	mass spectrometry
ms/ms	tandem mass spectrometry
mTOR	mammalian target of rapamycin
nl/nM	nanolitre/nanomolar
NMR	nuclear magnetic resonance
PAGE	polyacrylamide gel electrophoresis

PDB	protein data bank
PCR	polymerase chain reaction
PG	peptidoglycan
PIH1D1	pih1 domain containing protein 1
PIKK	Phosphatidylinositol 3-kinase-related kinase
PKA	protein kinase A
pSer	phospho-serine
pThr	phospho-threonine
pTyr	phospho-tyrosine
RNA	ribonucleic acid
SDS	sodium dodecyl sulphate
SEC	size exclusion chromatography
SH2	Src homology 2
snRNP 116	small nuclear ribonucleoprotein 116
STPK	Serine threonine protein kinase
TCEP	tris(carboxyethyl)phosphine
TPR	tetratricopeptide
TRIS	tris(hydroxymethyl)aminomethane
UV	ultraviolet

In addition, standard one-letter and three-letter abbreviations for natural amino acids were used.

1. Introduction

1.1 Signal Transduction

1.1.1 Overview

In order to survive in a dynamic environment, cells need to sense their surroundings and relay this information into the cellular interior to trigger an appropriate response. The mechanism by which information is transmitted in this context is known as signal transduction. The range of external stimuli is as vast as the number of cellular processes they may govern. The external stimulus may take the form of a change in concentration of a protein, lipid, small molecule such as a hormone or simple ion such as Na^+ or H^+ .To ensure that the stimulus effects the right cellular response, intricate and robust signalling networks have evolved to guarantee efficiency and fidelity. Naturally, signal transduction is more prominent in higher order multi-cellular organisms where cells need not only sense information about their surroundings, but also communicate with each other. Nevertheless, components of signal transduction networks have been reported in all forms of life.

The flow of information within such networks is mainly governed by protein-mediated interactions, and more specifically through the recognition of post-translational protein modifications. These protein modifications are usually the addition of a moiety to a specific reactive group on a protein. Well characterised modifications include methylation, acetylation, glycosylation, phosphorylation and ubiquitination. The addition of said functional groups on the surface of a protein facilitates the flow of information in intra-cellular signalling by altering the properties of a targeted protein. The modified isoform of a protein may have completely different binding properties or altered activity, allowing information to be propagated through a network by consecutive modification events. While the modifications added range from a methyl group to a whole protein in the case of ubiquitination, a common feature of protein modifications is their rapid reversibility, catalysed enzymatically. The addition or removal of functional groups from proteins is an energetically cheaper way to alter the cellular phenotype than synthesising or degrading whole proteins and has thus become a ubiquitous mechanism by which information is relayed inside a cell.

1.1.2 Protein phosphorylation

The post-translational modification that is both best characterised and most abundant is protein phosphorylation. The first mentions of this post-translational modification

were reports of the phosphorylation dependent regulation of metabolic enzymes in the 1950s (reviewed by Krebs and Beavo 1979.[1]). Protein phosphorylation involves the transfer of the γ-phosphoryl of adenosine triphosphate or in some cases guanosine triphosphate onto a hydroxyl group on a target amino acid, efficiently catalysed by protein kinases (figure 1.1). In humans, more than 2% of all genes encode a putative kinase or pseudo-kinase [2]. With more than 500 members in humans alone, kinases are one of the biggest protein families known. Several characteristics of protein phosphorylation can explain its evolutionary success. Apart from the protein to be modified, it only requires ATP as a substrate. ATP serves as the main reservoir of intracellular energy and is highly abundant with concentrations reported for mammalian cells of between 3 and 8 mM [3]. The free energy of phosphorylation of -12 kcal/mol ensures that the post-translational modification is highly stable. The activation energy of the cleavage of a phospho-ester is sufficiently large to stop non-catalysed phosphorylation or de-phosphorylation of proteins at biologically relevant temperatures and time scales. Only the catalytic actions of kinases and phosphatases can facilitate the phosphorylation or dephosphorylation of proteins at significant rates. Using Protein Kinase A as an example, the activation barrier of phosphorylation can be reduced to 20.7 kcal/mol [4].

At neutral or near-neutral pH, the addition of a phosphate introduces a net charge of between 1 and 2 negative charges to the phosphorylated protein. The negative charge as well as the new shape on the protein surface can drastically alter the internal dynamics of a protein and its ability to bind its interacting partners. Phosphorylation can generate or occlude binding surfaces or force a protein into an alternate fold. The consequences of these actions include the assembly or disassembly of protein complexes or an alteration of enzymatic activity. The presence of multi-domain proteins containing catalytic as well as regulatory domains allows for numerous and intricate systems by which phosphorylation can regulate protein function [5]. Virtually all cellular processes are in some way controlled by phosphorylation.

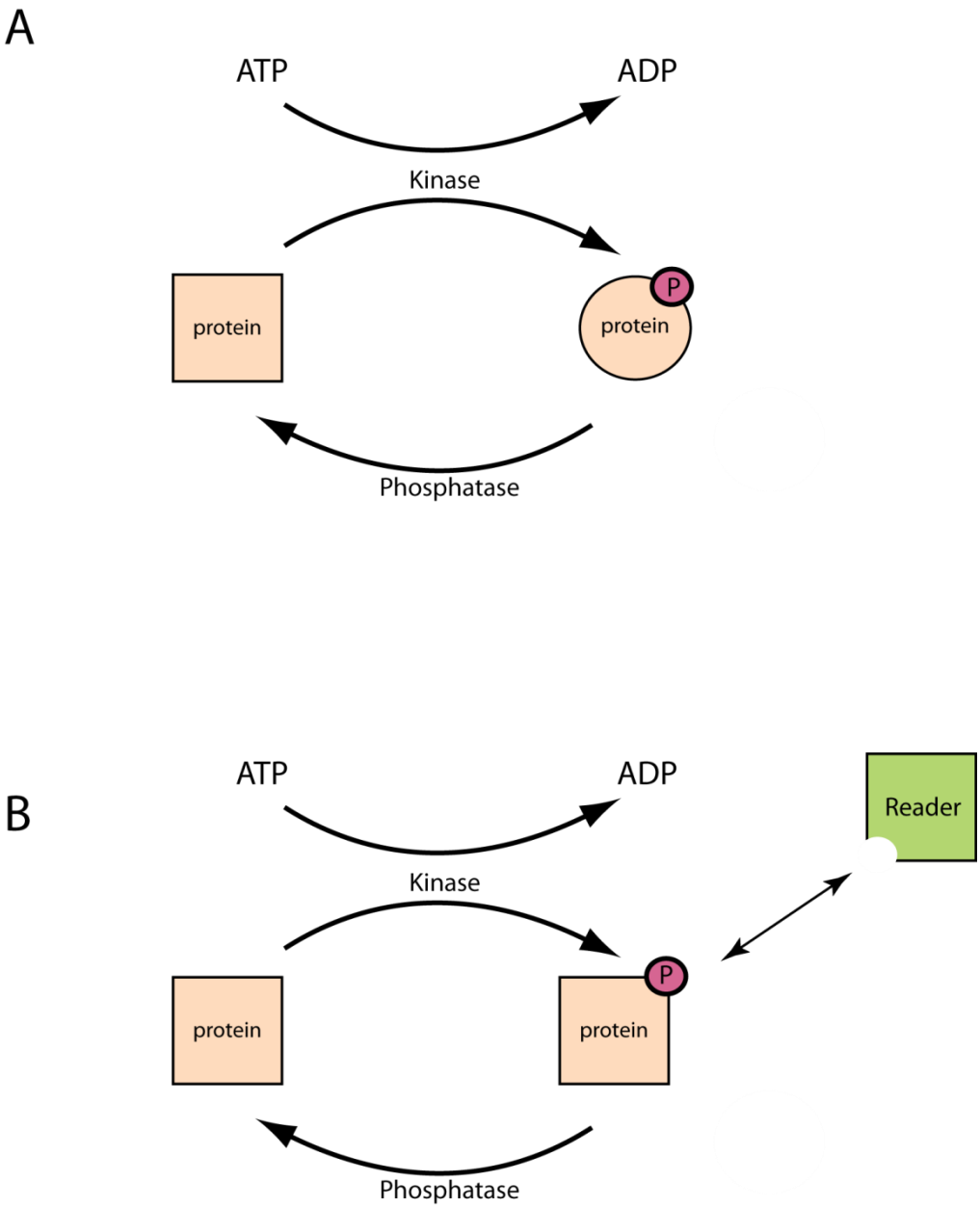


Figure 1.1. Schematic of the two most common forms of regulation by phosphorylation. (A) Allosteric regulation forces a protein (beige square) into an alternative conformation (beige circle). (B) Regulation by modular phospho-protein binding domains; Phosphorylation does not alter the protein conformation, but generates a binding surface for a "reader"-domain, to form a multi-protein complex

In eukaryotes, phosphorylation mediated signal transduction is divided into phospho-tyrosine (pTyr) and phospho-serine/phospho-threonine (pSer/pThr) signalling. pSer/pThr signalling is the more ancient and is conserved in all eukaryotes with many signalling components preserved from *Saccharomyces cerevisiae* to humans. Furthermore, comparatively simpler pSer/pThr signalling networks have been reported in a number of prokaryotic organisms. pTyr-mediated signalling evolved later. It is found exclusively in multi-cellular metazoans and some social uni-cellular protists, but is much less common in plants or lower organisms [6]. Thus, the cellular processes controlled by each system are different. As a general approximation, intra-cellular processes vital to all eukaryotes, such as cellular metabolism, the DNA-damage response and progression through the cell-cycle are controlled by pSer/pThr signalling [1, 7, 8] while inter-cellular processes, such as growth-factor signalling, cell-adhesion and immune responses are governed by pTyr signalling [9]. This is however a crude generalisation and there is ample cross-talk between the two systems. One prominent example is insulin signalling, where the presence of insulin in the extra-cellular space causes a receptor tyrosine kinase (insulin receptor) to autophosphorylate. The active kinase recruits and phosphorylates a number of substrates, amongst them lipid kinase PI3-kinase which is able to convert phosphatidylinositol 4,5-bisphosphate (PIP_2) into phosphatidylinositol (3,4,5)-triphosphate (PIP_3). The increase in PIP_3 concentration then activates a network of Ser/Thr kinases to alter the metabolic fate of the cell. As a result, catabolic pathways are activated and metabolic ones inhibited [10].

1.1.3 Protein domains

Protein domains were first described as folding units of a protein, distinct regions of a protein that may fold and consequently function and evolve independently of the rest of the protein. As it was known that intra-chain hydrogen-bonds between main chain atoms were formed only within certain discrete regions of proteins, it was hypothesised that the folding of some proteins involves multiple folding nucleation events [11]. Since they fold independently, they can be thought of as minimal functional regions of proteins. When a domain is inserted into a protein, it adds its function to that protein. As protein function is highly dependent on its tertiary structure, protein domains are also classified by their fold. In terms of size, domains range between around 30 and 500 amino acids in length with most domains being around 100 residues long. It is the presence of a hydrophobic core that stabilises globular folds and dictates the lower size limit for protein domains. As a consequence, very small domains tend to be stabilised by metal ions at their centre and very large domains may have more than one hydrophobic core. There are large numbers of different folds and domains, the CATH

(Class, Architecture, Topology, Homologous superfamily) protein architecture database lists 49 different folds, split further into 163 superfamilies, 1311 sequence families and 24232 domains (http://www.cathdb.info/). Protein domains may also be classified by their functions. Some are catalytic; they are the functional domains of enzymes, catalysing chemical reactions. The ability of some biological extracts to aid in the breakdown of organic compounds, such as the conversion of starch to sugar in saliva has been known for several hundred years, but it was the 19th century physiologist Wilhelm Kühne who first coined the term “enzyme”. He noticed how trypsin from bovine pancreatic extract was able to digest albumin when isolated from the organism and successfully suggested the name “Enzym” for catalytically active biological extracts [12]. Almost 90 years later a group led by David Chilton Phillips solved the 3 dimensional structure of the enzyme lysozyme. The catalytic activity of an enzyme could then be directly linked to a globular protein fold [13]. Another class of protein domains functions as adaptors. Modular protein domains, usually part of multi-domain proteins, recognize and bind specific epitopes on other biological macromolecules to form complexes and ensure correct localisation of catalytic proteins or protein domains attached to them. The first such interaction domain to be discovered was the SH2 domain which binds with high affinity to pTyr epitopes [14]. In phosphorylation signalling networks three types of protein domains are essential, protein kinases, protein phosphatases and modular phospho-recognition domains. From an evolutionary point of view, it is difficult to envisage how such a system might have evolved spontaneously as it requires 3 classes of protein to be effective. Having only one component of the system but lacking the other 2 does not have any obvious benefits. One is faced with a chicken-egg problem with 3 components. Using tyrosine phosphorylation signalling as an example, a solution to this conundrum was proposed in an essay by Lim and Pawson. Some proteins present in an organism may possess phospho-binding capabilities in addition to their original function, which favours the emergence of tyrosine kinases. These kinases then provide evolutionary incentive for these proto-reader domains to evolve into bona-fide SH2 domains. In a final step, protein tyrosine phosphatases emerge to introduce rapid reversibility to complete the signalling mechanism as observed today [15].

1.1.4 Protein Kinases

Two major classes of protein kinases have been described. These are the prokaryote associated His-kinases and the eukaryotic like Ser/Thr/Tyr kinases. Upon activation, eukaryotic-like kinases phosphorylate target proteins containing specific epitopes or kinase interacting domains. Since a common substrate for kinases are other kinases, this system can facilitate the construction of extensive phosphorylation networks, some of which may have been erroneously oversimplified as cascades [16]. Although initially identified in eukaryotes, these phosphorylation networks have recently been reported to exist in prokaryotes as well [17]. The classical mechanism for prokaryotic signal transduction however is a simpler one. In a so-called 2-component system, a histidine kinase autophosphorylates upon sensing an extra-cellular stimulus and subsequently transfers the phosphoryl group onto an aspartate residue of a “response regulator”. Phosphorylation alters the conformation and activity of the response regulator which in turn causes the repression or activation of target genes [18]. One of the first such systems to be discovered concerned osmoregulation in *Escherichia coli* [19]. Upon sensing high extra-cellular osmolarity, the histidine kinase EnvZ autophosphorylates upon activation and trans-phosphorylates an aspartate reside on transcriptional regulator OmpR. Upon phosphorylation, OmpR switches affinity from the promoter region of a large pore protein to that of a small pore protein the net effect being a reduction in diffusion between the bacterial cell and its surrounding medium. This simplistic and linear pathway of external stimulus, auto-phosphorylation, trans-phosphorylation and gene activation (or repression) does not allow for extensive cross-talk between different kinases. It is found predominantly in prokaryotic organisms. As the work presented in this thesis work focuses exclusively on “eukaryotic like” Ser/Thr and Tyr protein phosphorylation it is only this form of post-translational modification that will be introduced further.

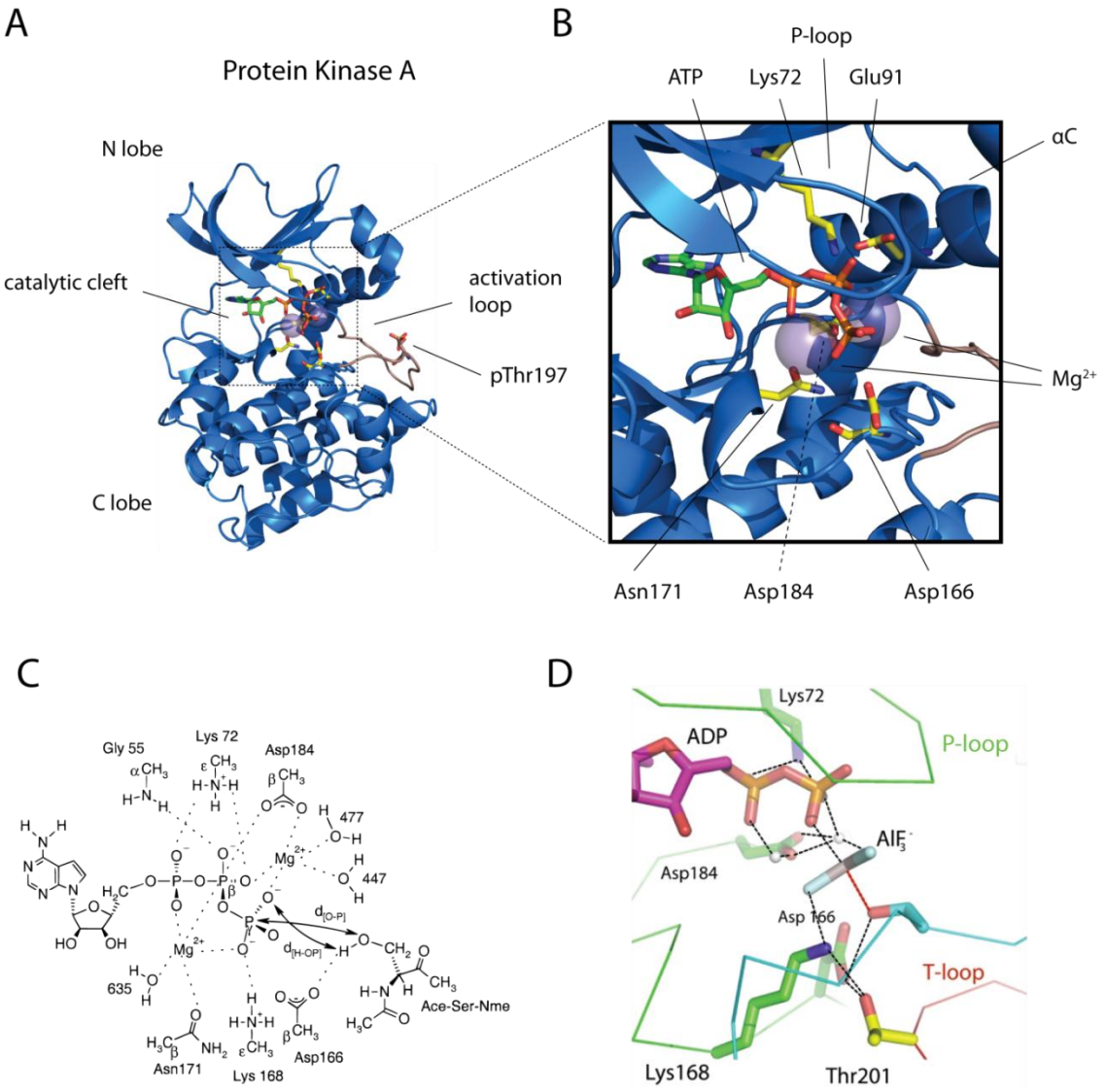


Figure 1.2. Structure of a protein kinase domain (A) Cartoon representation of protein kinase A (PDB: 2CPK), N-lobe and C-lobe are shown in light blue and dark blue respectively and the activation loop is coloured brown. Conserved, phosphorylated Thr197 is shown as sticks. (B) Close up of catalytic cleft of PKA, conserved catalytic residues are shown as yellow, ATP as green sticks and hetero atoms coloured using standard code. (C) Schematic representation of the hydrogen-bond network in the active site of PKA. Adapted from Hutter & Helms, 1999. (D) Integration of the hydrogen bond network into the structural data and the position of a peptide substrate (cyan) in the catalytic cleft. To capture the transition state of the kinase reaction, ATP was substituted with ADP and aluminium fluoride (PDB: 1L3R).

Increased understanding of the regulation and catalytic activity of the kinase domain has been facilitated by the ever expanding array of functional and structural data being made available (figure 1.2). While the first structure of a classical kinase domain was reported in 1991, published kinase structures have increased exponentially and as of March 2010, there were 136 unique human protein kinase domain structures deposited in the Protein Data Bank (www.rcsb.org). Sequences of protein kinase domains can be very divergent, with some examples having only 30% sequence homology with their closest relative. Despite these divergent sequences, the overall fold, and more importantly the catalytic residues are highly conserved. The kinase fold will be introduced using the X-ray structure of protein kinase A as an example (PDB accession 2CPK). Kinase domains fold into a bi-lobal conformation split into the N-terminal lobe and the C-terminal lobe. The N-terminal lobe (N-lobe), consists of a 5 stranded β-sheet and one helix, the αC-helix. The C-lobe is the larger of the lobes and is made up of α-helices. The kinase active site is located in the cleft between the two lobes. In its active state, the kinase domain is found in complex with 1 ATP molecule and 2 divalent cations, usually magnesium or manganese. The metal ions are required as co-factors to recognise and stabilise the bound ATP.

A number of residues essential for catalysis are absolutely conserved among kinase domains. A conserved glycine rich loop (P-loop) is required for nucleotide binding. Its motif is G-X-G-X-Φ-G, where Φ can be either tyrosine or phenylalanine. The glycines allow for sufficient conformational freedom to form a close loop and accommodate the nucleotide. The P-loop and conserved residues Lys72, Glu91, Asp166, Asn171 and Asp184 are essential for kinase function. An ion pair is formed between Lys72 and Glu91 to stabilise and position the α- and β-phosphates of the ATP molecule for catalysis. Asn171 and Asp184 bind the metal cations to ensure correct binding of the nucleotide. Asp166 is directly involved in catalysis through interaction with the substrate hydroxyl. Initially it was proposed to be a catalytic base and to remove a proton from the hydroxyl to generate a more powerful nucleophile [20]. However, the pK_a of the aspartate carboxyl is significantly lower than the pH of the cellular environment and the pH independence of the catalytic activity of PKA suggested a different function for this acidic residue. Its exact function remains contentious, and whether Asp166 acts as a catalytic base or remains neutral throughout has not been determined as yet [21, 22]. The kinase domain has evolved to lower the activation energy barrier of protein phosphorylation. Conserved residues in the catalytic cleft, in concert with the bound magnesium cations, achieve this by holding the two substrates in a favourable orientation and by stabilising reaction intermediates [4].

The appropriately named activation loop is the most common means by which the activity of a kinase is governed. The activation loop, located between the N-lobe and C-lobe, usually contains one or more phosphorylatable residues. In its unphosphorylated form, the activation loop collapses onto the catalytic cleft and prevents substrates from accessing the active site. Upon phosphorylation however it undergoes a conformational change to reveal the cleft [23]. This activation loop phosphorylation may be auto-phosphorylation or carried out by another kinase. It is this property of a kinase to be activated by the activity of another kinase that has enabled the evolution of large networks. The activity of a kinase can be modulated not only by the kinase domain itself but also by accessory non-catalytic domains that are present in many kinases. They recognize a change in concentration of a small molecule such as calcium or the presence of a post-translational modification on itself or another protein. In the example of Src Kinase, the presence of a phosphorylated tyrosine near the C-terminus of the kinase is recognised by an SH2 domain and forces the kinase into a catalytically inactive conformation. Dephosphorylation of this tyrosine alleviates inhibition [24]. In the case of Ca^{2+}/calmodulin-dependent protein kinase II, the kinase consists of a catalytic domain, a regulatory domain, and an association domain responsible for oligomerisation. In its basal state, the regulatory domain inhibits the activity of the kinase by blocking access to the active site using a pseudo-substrate. In the presence of calcium, the interaction between catalytic and regulatory domain is interrupted and the inhibition relieved [25, 26]. As there are more than 500 kinases in the human genome and essentially just one highly conserved kinase fold, additional domains are required to ensure the right kinase is activated at the appropriate time and place.

In order to maintain the appropriate phosphorylation state of all cellular proteins, the activity of a kinase and also its substrate specificity need to be tightly regulated. Regulation of phosphorylation occurs on a number of levels. At a global level, kinase activity is regulated by varying expression profiles of kinases in different cell lines [27]. In a sub-cellular context, it is the properties of the kinase itself that determine which residues they phosphorylate. Different kinases favour different target peptide sequences as they vary in sequence and structure. Individual specificities arise from differences in the catalytic domain itself as well as non-catalytic docking domains that are utilised to recruit substrates. Furthermore, kinases are localised at foci by interactions with scaffold proteins associated with specific intra-cellular macro-molecular assemblies. The substrate-specificity of a kinase had long been thought to be mainly governed by its catalytic domain, but more recently it has been shown that interactions of non-catalytic domains are equally important. When assaying the activity

of native and chimeric mitogen activated kinases (MAPK), it was demonstrated that the correct docking sequence for scaffold recruitment, rather than the identity of the MAPK kinase was sufficient to effect the proper cellular response [28]. Sequence specificity and cellular localisation can work in tandem to achieve considerable specificity. Mitotic kinases, involved in cell-cycle regulation, have evolved to have some overlap in terms of sequence specificity or localisation, but never both. Ambiguity is avoided as there is always at least one characteristic that separates one mitotic kinase from another [29].

Modular protein-protein interaction domains that are essential for protein phosphorylation have evolved not only on kinases, but also on their substrates. One such mechanism is the use of phospho-peptide binding domains such as FHA or SH2 domains that interact with the phosphorylated activation loop of the kinase. For some interactions, docking of such a domain onto the activation loop helps the kinase recruit the substrate and is essential for efficient phosphorylation [30].

Kinase fidelity is achieved through the concerted effects of tight activation regulation, sequence specificity and spatial localisation despite numerous kinases sharing a common fold.

1.2 Modular phospho-protein binding domains

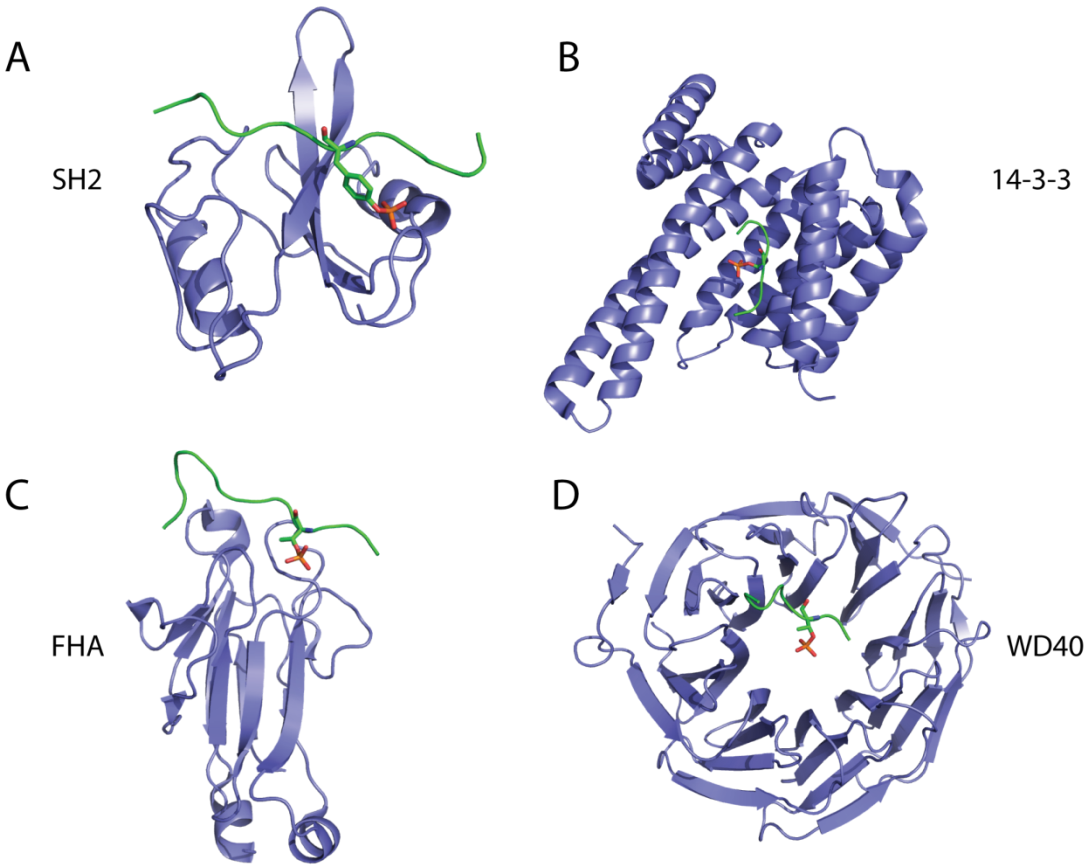
1.2.1 Overview

One means of inducing a change in protein properties upon phosphorylation is by allosteric regulation [31]. In this case the addition of a phosphate stabilises the protein in an alternative conformation which in turn will alter its activity. The regulation of muscle phosphorylase was the first reported instance of such a mechanism. This mechanism however requires a whole protein domain to exist in two separate conformations and it is likely that this mechanism needs to evolve independently for every protein in which it is applied. A second method of effecting cellular changes using protein phosphorylation has evolved; a method that relies on modularity and one where properties are easily transferred from one protein to another [5]. Globular, non-catalytic protein domains that recognise and bind epitopes containing pSer, pThr or pTyr have evolved (figure 1.1). They enable phospho-protein binding properties to be passed from one protein to another by domain shuffling. Protein domains which recognize other post-translational modification, such as arginine methylation or acetylation, have been discovered more recently [32, 33]. As a consequence, our knowledge of these domains lags somewhat behind. All these domains can be thought of as reader-domains as they

“read” protein modifications. The modular nature of protein domains allows the regulation of one protein to be associated with a particular post-translational modification by addition of the appropriate protein-protein interaction domain.

The first modular protein domain recognised to bind phospho-epitopes was the Src homology region 2 (SH2) domain [34], a protein domain roughly 100 amino acids in length. Initially identified as a conserved, non-catalytic region in Src tyrosine kinases it was subsequently shown to bind peptides containing phospho-tyrosines [14]. SH2 mediated binding was shown to be completely phospho-dependent and removal of phosphorylation totally abrogates binding. Initial quantitative studies using surface plasmon resonance determined typical affinities to be between 1 and 10 nM for binding of an SH2 domain to a phospho-peptide derived from an interaction partner [35]. A subsequent study, using in-solution techniques as opposed to immobilised peptides, changed this estimation and places the binding constants for phospho-dependent SH2 domain interactions between 100 nM and 10 μM [36].

Since the discovery of the SH2 domain, around a dozen new folds have been discovered that are non-catalytic and bind specifically to proteins presenting phosphorylated residues on their surface (figure 1.3). Phospho-binding domains can be divided into domains that recognise phospho-tyrosine containing epitopes and those that recognise either phospho-serine or phospho-threonine. Domains that bind pSer and/or pThr motifs are found in all kingdoms of life while pTyr binding domains are predominantly found in, along with pTyr-mediated signal transduction in general, higher multi-cellular or social single celled eukaryotes [6].



E

[redacted]

Figure 1.3. Structural versatility of modular phospho-binding domains. (A-D) Cartoon representation of different phospho-binding domains. Bound phosphopeptide shown as cartoon and phosphorylated residue shown as sticks. (A) SH2 domain of human Src-kinase bound to a pTyr peptide (PDB: 1FMK) (B) 14-3-3 domain monomer of human 14-3-3 zeta protein bound to a pSer peptide (PDB: 1QJB) (C) N-terminal FHA domain of *S. cerevisiae* Rad53 kinase bound to pThr peptide (PDB: 1G6G) (D) WD40 domain of *S. cerevisiae* CDC4 bound to a pThr peptide (PDB: 1NEX) (E) [redacted]

1.2.2 Overview of phospho-dependent interaction domains

A number of different folds that possess the ability to bind phospho-epitopes have been discovered over the last two decades (figure 1.3). While the common characteristic of all such proteins is their ability to bind their interaction partners in a phospho-dependent manner, each protein-binding domain selects for a specific phosphorylated residue. Each category of fold is associated with what type of phosphorylated residue it binds to; SH2 domains for example only bind pTyr epitopes while a number of domains, including POLO-Box domains and BRCT domains have a dual specificity and bind both pSer and pThr containing peptides [37, 38]. In addition, most interaction domains have further sequence preferences N-terminal or C-terminal to the phosphorylated residue. The general consensus sequence for POLO-Box domains is for example S-pS/pT-P. However, not every member of a domain family will have an identical consensus sequence. Each will have a number of absolutely conserved residues involved in binding the phosphorylated amino acid but also a number of variable positions conferring different specificities for the surrounding residues.

Some interaction domains allow for more flexibility than others. 14-3-3 domains for example preferentially bind to pSer epitopes, but also display reduced affinities for pThr epitopes. Two high affinity motifs have been found; R-X-X-X-pS/pT-X-P and R-S-X-pS/pT-X-P [39]. The residues directly C-terminal and N-terminal to the phosphorylated residue do not contribute to the binding event. The POLO-Box however with aforementioned consensus sequence S-pS/pT-P only tolerates a serine in the -1 position. Even a conservative substitution with a threonine will abrogate binding [40]. The availability of *in vitro* peptide library screens has enabled the determination of the optimal binding partners for phospho-interaction domains. Supported by atomic-resolution crystal structures we can now not only determine the binding preferences of phospho-dependent interaction domains, but also rationalise them from a structural point of view. [41]. Despite the vast differences between different interaction domain families in overall fold, some structural features are common, the most prominent being conserved basic and polar amino acids forming hydrogen bonds with the negatively charged phosphate [41].

Phospho-dependent interaction domains have evolved to bind their substrates with roughly similar affinities. Most such domains bind phospho-peptides containing their specific consensus sequences, derived from *in vitro* peptide library screens, with typical binding constants of between 100 and 1000 nM [39, 40, 42].

1.2.3 Functions of phospho-dependent interaction domains

Although the first interaction domain reported, the SH2 domain from Src kinase, functions as an intra-molecular regulator [34], the majority of interactions facilitated by phospho-binding proteins are inter-molecular. They provide the means for cells to form protein complexes induced by kinase activation and are mediated by the phospho-dependent recruitment of proteins to a desired focal point. A wide range of cellular processes are governed by such interactions.

A good example of phospho-binding proteins functioning in such a fashion is the DNA damage response. Upon sensing a single- or double-stranded DNA break or a stalled replication fork, Ser/Thr phosphorylation signalling enables a eukaryotic cell to engage in the appropriate response, namely to activate DNA repair pathways and halt progression through the cell cycle until the DNA lesion has been mended [43]. Effective DNA repair requires extensive multi-protein complexes to be assembled at the site of the break. Naturally, the dynamic assembly of such signalling and repair complexes is highly dependent on protein-protein interaction domains with high affinities for post-translationally modified amino acids. Protein acetylation, methylation, sumoylation and ubiquitination also play an important role in the DNA damage response, but it is phosphorylation dependent signalling that has been studied most extensively [44]. FHA domains and BRCT domains feature most prominently in the formation of complexes maintaining genetic integrity [45, 46]. It is the interplay between activation of the phosphatidylinositol 3-kinase like kinases (PIKKs) ATM, ATR and DNA dependent protein kinase and multi-domain scaffold proteins such as NBS1, MDC1, XRCC1 and XRCC4 that facilitate complex assembly. Histones in the vicinity of a DNA break are phosphorylated by PIKKs on conserved residue Ser139 [47-49] which creates a binding site for the MDC1 BRCT domains [50]. Furthermore, MDC1 also contains an FHA domain which may recruit ATM phosphorylated checkpoint kinase Chk2 or facilitate phosphorylation dependent dimerization of MDC1 [51, 52]. Chk2 also contains an FHA domain, but in this case its function is not to act as a scaffold, but to facilitate Chk2 dimerization [53, 54]. As MDC1 is phosphorylated itself by casein kinase 2, additional scaffold proteins containing FHA or BRCT domains may then be added to the complex being formed at the site of a DNA break [55]. NBS1 for example binds to pS-D-T-D motifs on MDC1 via its BRCT domain and recruits further proteins to the complex [56, 57]. The sole purpose of this brief section is to highlight the different mechanisms through which pSer/pThr signalling is mediated. For a comprehensive review of post-translational modifications governing the DNA damage response see *Polo and Jackson* [44].

Through the generation of phospho-epitopes by protein kinases and multi-domain scaffold proteins associating with such epitopes, large protein complexes can be assembled at desired focal points.

1.3 FHA domains

1.3.1 Overview

FHA (forkhead associated) domains take their name from the proteins in which they were initially discovered. Sequences analysis of a number of proteins involved in nuclear signalling highlighted a conserved motif in some protein kinases and forkhead transcription factors [58]. Subsequently, FHA domains were shown to be protein-protein interaction domains with specific affinity for pThr containing epitopes [59]. On a functional level, FHA domains are most abundant in proteins involved in cellular proliferation. FHA domains are found in proteins regulating the DNA damage response, cellular growth and checkpoint signalling during progression of the cell cycle [60].

The FHA domain is the only phospho-protein binding domain that exclusively binds to pThr epitopes [41]. Other protein domains that can bind pThr epitopes such as 14-3-3, WW and BRCT can also bind to pSer sequences with varying affinities. The BRCT and 14-3-3 domains for example prefer pSer epitopes but also tolerate pThr containing sequences.

1.3.2 Structure of the FHA domain

As of January 2012, there are 24 unique FHA domain containing structures deposited in the Protein Data Bank (www.rcsb.org). They reveal the FHA domain to be an all β protein approximately 100 residues in length. The domain is arranged into a β-sandwich with a total of 11 strands (figure 1.4). In between the 2 sheets there is a hydrophobic core that provides the domain with structural integrity. Differences between FHA domain family members derive from the loop sequences that connect the β-strands. Additionally, some FHA domains contain short helical insertions between the strands. The first FHA domain of RAD53 for example has a helical insertion between β2 and β3 (PDB accession 1G6G). As is usual for modular protein-protein interaction domains, the N-terminus and C-terminus of the domain lie in close proximity, allowing the FHA domain to be inserted into a protein by domain shuffling without interrupting the conformation of other parts of the protein. Strands β1, β2, β7, β8, β10 and β11 form one sheet while β3 – β6 and β9 form the other. The loops between the strands are

directly involved in binding the pThr ligands. A number of conserved amino acids between the β3/β4, β4/β5 and the β6/β7 loop form hydrogen bonds with the ligand and are essential for high affinity binding.

Using numbering from Rad53^{FHA1} from *Saccharomyces cerevisiae* as an example, the role of each conserved amino acid will be introduced. The β4/β5 loop contains a conserved S-θ-X-H motif (where θ is either Arg or Asn) which is directly involved in binding the phosphate moiety and provides structural stability to the domain. The histidine (His88) is highly conserved and provides stability to both the binding loops and the β-sandwich. Mutation of this amino acid results in the FHA domain becoming unstable. Ser85 forms a hydrogen bond with its γ-hydroxyl to one of the OP oxygens of the pThr. Substitution of this serine residue to an alanine abrogates binding to pThr, but does not interfere with the overall fold of the FHA domain making it an attractive target for mutagenesis studies when trying to remove the phospho-dependent binding capabilities of an FHA domain without interrupting the overall fold[61, 62]. Ser85 is most commonly, but not essentially directly succeeded by an arginine or asparagine. For asparagine, one of the OP oxygens of the pThr forms a polar contact with the N_{δ}. In the case of an arginine at this position, the contacts are formed between two of the OP oxygens and both the ε-amine and one of the η-amines. The β3/β4 loop contains two conserved amino acids essential for FHA function. In Rad53^{FHA1} they are Gly69 and Arg70. The glycine has a similar function to His88 as it is not involved in binding directly but is essential for the overall fold. It makes van der Waals contacts with His88 and initiates the turn of the β3/β4 loop. In the β6/β7 loop the conserved Asn107 makes important interactions with the main-chain of the ligand. While it does not select for specific amino acids, it further stabilises the FHA-phospho-peptide interaction. The nitrogen on the side-chain amide of Asn107 makes contact with the main-chain carbonyl of the amino acid directly C-terminal to the pThr in the +1 position relative to the pThr. The side chain carbonyl of Asn107 forms a hydrogen bond with the backbone amide of the amino acid in the +3 position. In summary there are 4 conserved amino acids involved in binding a general pThr containing peptide or protein. There are no large conformational changes in the FHA domain upon peptide binding, but molecular dynamics modelling has suggested a reduction in internal motion of the FHA upon such an event [61].

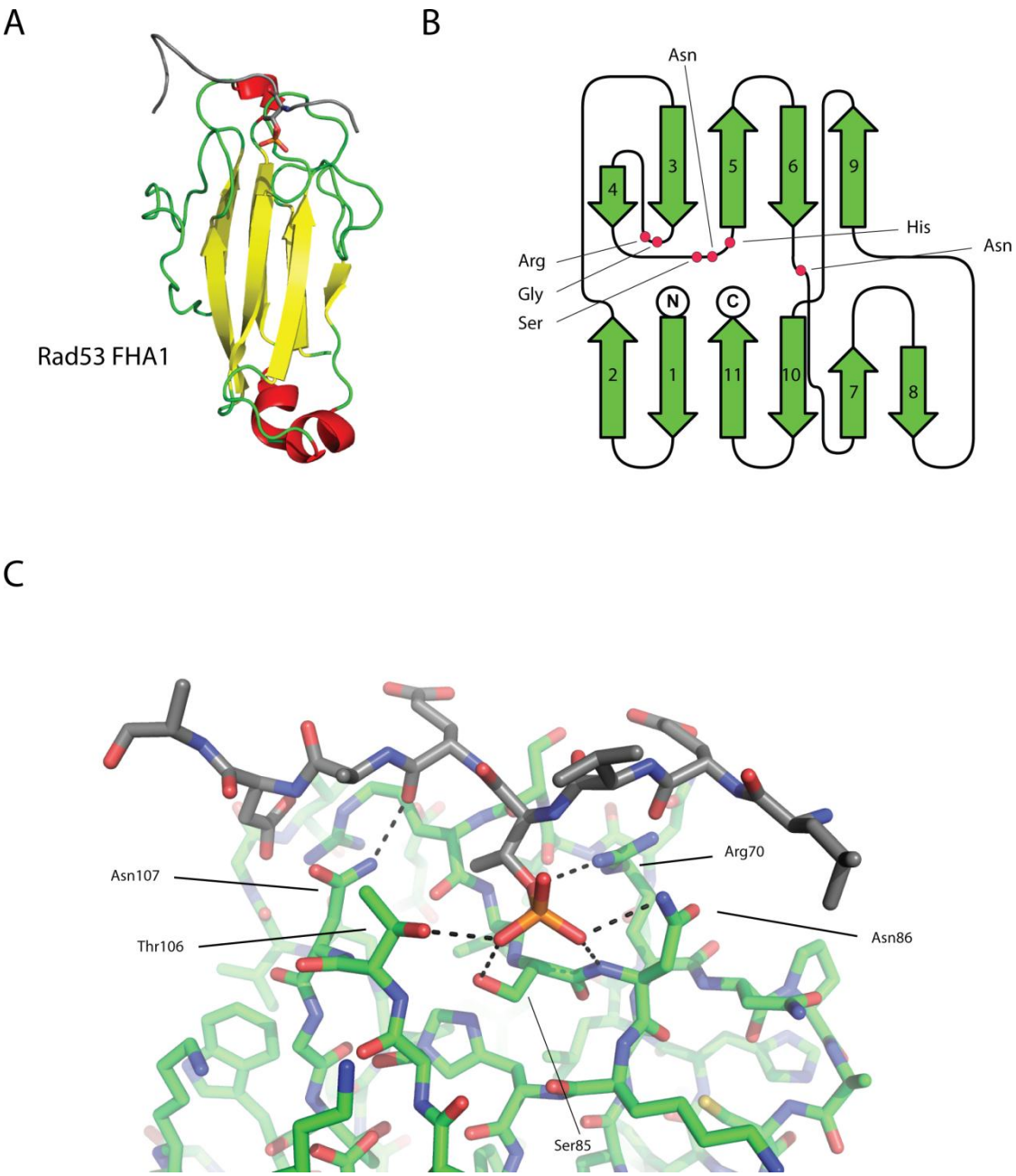


Figure 1.4. Structure and function of the FHA domain (A) cartoon representation of Rad53 FHA1, the first crystal structure of an FHA domain to be solved. Loops are coloured in green, β-strands in yellow and α-helices in red. The bound phospho-peptide is coloured in grey and pThr shown as sticks (B) Schematic representation of the general FHA model. Conserved residues shown as red circles (C) Rad53 FHA1 residues (green) making contact with phosphopeptide (grey) shown in stick representation; hydrogen bonds are shown as dashed lines.

1.3.3 Binding specificity of the FHA domain

A number of factors contribute to the fact that FHA domains exclusively bind pThr containing epitopes. In a study by *Pennell et al*., using the Rv0020c FHA domain of *Mycobacterium tuberculosis* as a model system, molecular dynamics simulation were performed to investigate the strong preference for pThr [61]. It is intriguing how the absence of a single methyl group can reduce the binding affinity of an FHA-phospho-peptide interaction from 100 nM to an affinity not measureable by Isothermal Titration Calorimetry. One explanation is the presence of van der Waals forces between the γ-methyl of the pThr and a conserved methyl binding pocket. The main FHA contributor to those interactions is the conserved Asn of the β6/β7 loop. The absence of relatively weak van der Waals interactions cannot however explain completely the considerable decrease in binding affinity. However, the loss of van der Waals interactions, in addition to the greater conformational freedom of the pSer residue, leads to a collapse in the hydrogen bonding network between the peptide and the conserved FHA domain residues. In some cases, the greater conformational freedom of the unbound phospho-peptide due to the absence of the γ-methyl also increases the entropic penalty upon binding to the FHA domain [61].

The specificity of the FHA domain for a general pThr-containing peptide is mainly governed by the conserved residues introduced above. FHA domains however possess further specificity determinants. Additional residues are involved in the interactions with ligand side chains other than the pThr. These interactions vary between FHA domains and allow for functional versatility amongst the numerous FHA domains found in eukaryotic genomes. They will be introduced in the next sub-section.

Unlike 14-3-3, WD40 domains or most other modular phospho-binding proteins, the residues interacting with the ligand and providing specificity are exclusively located in loops. As a consequence, the part of the protein providing it with a stable hydrophobic core is relatively independent from the functional surface. The length and composition of the interacting loops, most notably β4/β5 and β10/β11, can be altered to provide divergent FHA specificities without disturbing the overall FHA fold [60].

The strong conservation of the FHA amino acids involved in pThr binding is not displayed throughout the whole peptide binding surface, which gives rise to divergent specificities amongst different FHA domains. From the structural data available, we can deduce that the overall interaction between FHA domain and ligand phospho-peptide stretches along the top surface of the domain, approximately 18 Å in length. As a consequence, 6-7 ligand amino acids are in contact with the FHA domain in a

canonical FHA-pThr interaction. Apart from the phosphorylated residue, there are 6 unmodified amino acids (-3 to +3) that may interact with the FHA domain. Of those, it is the +3 position relative to the pThr that contributes most to the binding. When a phospho-peptide is bound, its conformation is such that the side chain of the +3 amino acid points directly towards the FHA domain, providing for strong selectivity at this position. Using *in vitro* peptide library screens to determine optimal binding peptides for some FHA domains, 2 major classes have been reported so far. There are those that select for an aspartate in the +3 position (pT-X-X-D) e.g. Rad53^{FHA1} [63] and those that prefer a medium sized hydrophobic at this position (pT-X-X-I/L/V) e.g. Chk2^{FHA} [42]. When comparing the structures for each FHA-peptide complex it is apparent that the interaction between the FHA-domain and the pThr-moiety are highly similar, but the interacting surfaces for the +3 amino acid show large differences. In the case of Rad53^{FHA1} (PDB: 1G6G), the +3 Asp interacts with Arg83 which is part of the β4/β5 loop. The carboxyl of the aspartate and the guanidinium of the arginine form a salt bridge. In Chk2^{FHA} (PDB: 1GXC) this loop contains a helical insertion and is no longer in proximity of the +3 side chain. Conversely, the +3 Ile makes contact with the β10/β11 loop. Van der Waals interactions with Ser192 and Leu193 contribute energetically to the FHA-ligand interaction. Although the +3 position is the most important for FHA domain selectivity, other amino acids from positions -3 to +3 may contribute somewhat to the overall binding providing the FHA with the ability to fine-tune specificity. In the case of Rv0020c, a proline is favoured in the -1 position [61]. Not all contributions to binding affinities may be due to direct FHA-peptide interactions but come from providing the peptide with necessary amount of conformational freedom. A balance needs to be struck between sufficient flexibility to be accommodated into the binding pocket and entropic loss upon binding.

1.3.4 Non-canonical FHA domain mediated interactions

Several FHA domain mediated interactions that do not conform to the mechanisms described above have been reported.

The 2nd FHA domain of Rad53 has been shown to bind pTyr containing peptides as well as pThr [64]. The pTyr-peptide does not bind in the usual pThr pocket, but to a hydrophobic patch on the same face of the regular FHA binding interface with the orientation of the peptide turned by 90° (figure 1.5). Rad53^{FHA2} binds two pTyr peptides selected from peptide library screens bind with affinities of 1.1 and 5.0 μM [65]. The biological significance of these findings however is questionable. This FHA domain binds a pTyr-peptide derived from its biological interacting partner with an affinity as

weak as 100 μM. Furthermore, the absence of tyrosine kinases in *S. cerevisiae* suggests that these findings are likely to be artefacts.

Antigen protein Ki67 contains an FHA domain that binds a pThr containing peptide from hNIFK (human nucleolar protein interacting with the FHA domain of Ki67) [66]. This FHA domain does not bind to short phospho-peptides, but to a multiply phosphorylated 43 amino acid long peptide derived from hNIFK [67]. The structure of the FHA-peptide complex revealed that in addition to a pThr occupying the canonical binding site, part of the peptide forms a β-strand that is incorporated into the FHA β-sandwich in an anti-parallel conformation next to β4 of the FHA domain. The strand insertion greatly stabilises the interaction. Mutagenesis studies have revealed that substitution of the pThr involved in the regular interaction with the FHA domain with a pSer has less effect than removal of the peptide β-strand[60]. The $Ki67^{FHA}$ – peptide interaction is therefore the result of both phospho-dependent as well as phospho-independent interactions.

The FHA domain containing protein GarA (glycogen accumulation regulator), encoded by gene Rv1827 from *Mycobacterium tuberculosis* engages in a number of interactions of which at least one is entirely phospho-independent [68]. In addition, it is also able to bind phospho-peptides in the canonical fashion [69]. This prokaryotic FHA domain was reported to bind to and alter the activities of 3 metabolic enzymes of the mycobacterial citric acid circle, none of which are thought to be phosphorylated [68]. Using homologous proteins from the close relative *Mycobacterium smegmatis*, a crystal structure was determined consisting of a complex between GarA and one of the enzymes, α-ketoglutarate decarboxylase (PDB: 2YID). Enzymatic assays showed that the GarA homologue inhibits the carboxylase with an IC_{50} greater than 50 nM so it is assumed that the dissociation constant between the two proteins would be equally tight [70]. The X-ray structure of this complex shows that the loops usually involved in binding the phospho-peptide are also responsible for generating the binding surface in this interaction. The same conserved surface can therefore bind in both a phospho-independent and phospho-dependent manner.

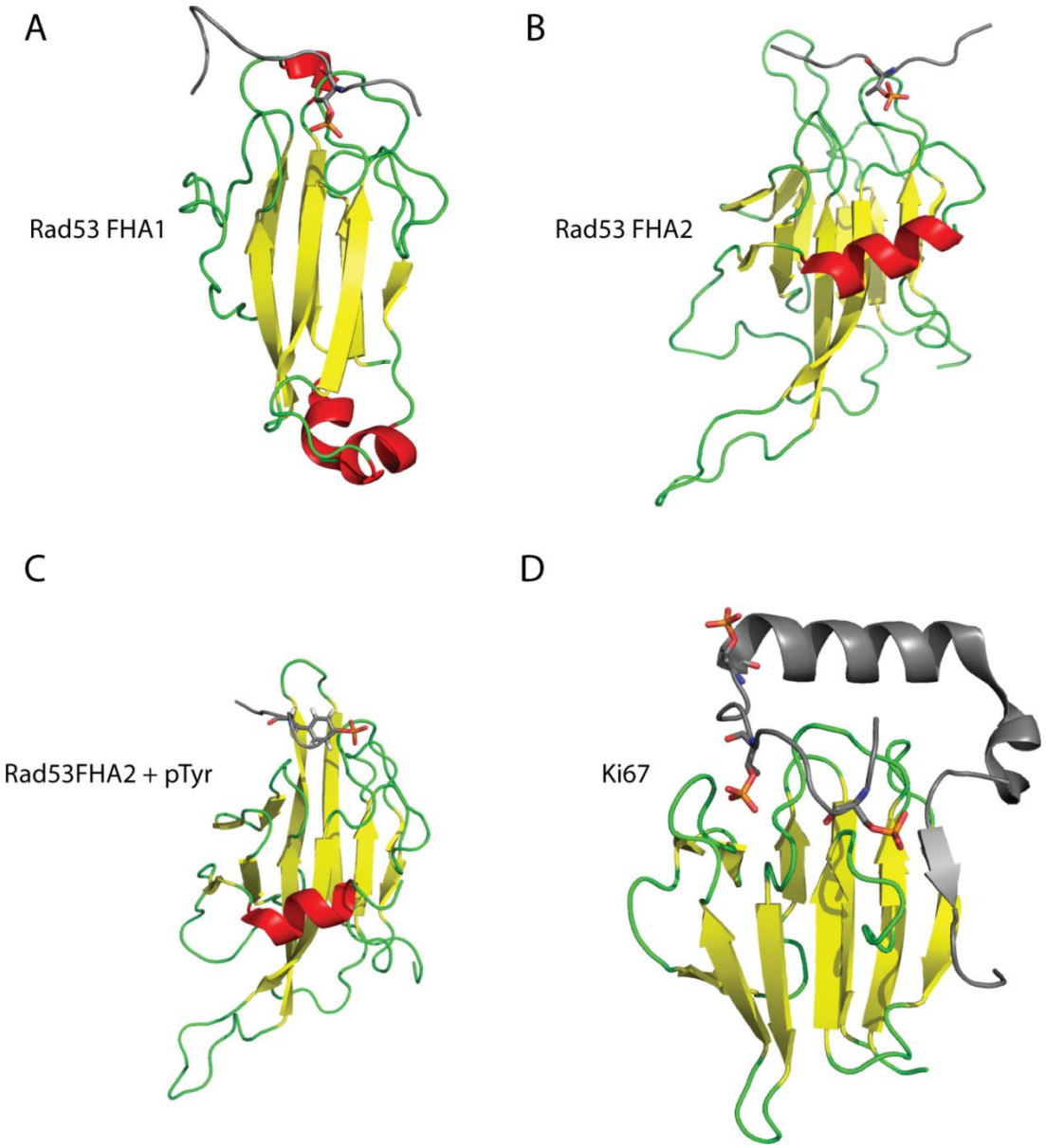


Figure 1.5. Versatility of FHA domain mediated interactions. Cartoon representation of different FHA domains bound to phosphopeptides. FHA loops, α-helices and β-strands are shown in green, red and yellow respectively. The bound phosphopeptide is shown in grey with the phosphorylated amino acids represented as sticks. (A) *S. cerevisiae* Rad53 FHA1 bound to a pThr peptide (PDB: 1G6G) (B) *S. cerevisiae* Rad53 FHA2 bound to a pThr peptide (PDB: 1J4L) (C) *S. cerevisiae* Rad53 FHA1 bound to a pTyr peptide (PDB: 1FHR) (D) FHA domain of human Ki67FHA domain bound to an extended multiply-phosphorylated peptide (PDB: 2AFF)

1.3.5 Modes of FHA domain mediated interactions

The ability of the FHA domain to bind pThr containing epitopes has been utilised by cells in several ways. FHA domains may engage in intra-molecular, inter-molecular or oligomeric associations. The *S. cerevisiae* Rad53 kinase is a useful example for highlighting the functional versatility of the FHA domain [71]. This kinase consists of an N-terminal FHA domain (FHA1), a kinase domain and a C-terminal FHA domain (FHA2). Activation of Rad53 is dependent on its dimerization, which is accomplished via 2 FHA-dependent interactions [72]. Initially, Rad53 is recruited to its parent kinase Mec1 (mitotic entry checkpoint protein 1) via the hyper-phosphorylated adaptor protein Rad9. The inter-molecular association between the phosphorylated Rad9 and Rad53 is facilitated by $Rad53^{FHA2}$ [73]. Mec1 phosphorylation of Rad53 is necessary for kinase activation, FHA domain mediated phospho-dependent interactions to these phospho-sites are needed for Rad53 oligomerisation which then induces kinase activation [72, 74]. The current model is such that Mec1 phosphorylation of Rad53 induces its activation via an FHA1 dependent oligomeric assembly. pThr epitopes on one Rad53 monomer are recognised by an FHA domain on another monomer stabilising an oligomeric association. Once an oligomer is formed, the kinase domains auto-phosphorylate the activation-loops of their partners *in trans* and Rad53 is activated. Protein dimerisation induced by a head to tail FHA-pThr interactions have also been reported for Chk2, the human orthologue of Rad53 and the human protein MDC1 [52, 53].

Intra-molecular associations mediated by FHA domains are somewhat rarer. The mycobacterial FHA domain containing protein GarA as well as its close homologues, whose ability to bind tightly to non-phosphorylated proteins has already been highlighted, display additional unusual behaviour in that their FHA domains bind a pThr epitope *in cis* [68, 75]. See section 1.4.5.3 for a more detailed description of this mechanism.

Overall, FHA domains have been shown to be classical modular phospho-protein interaction domains with considerable affinity and specificity for pThr containing epitopes. Organisms from all kingdoms of life have adopted FHA domains and utilise their pThr-binding potential [58, 76, 77]. Whilst the vast majority of FHA domains bind their ligands via one canonical mechanism, there are several examples of alternative interactions, suggesting the FHA domain may be even more versatile than currently presumed. In the next sections, the role of the FHA domain in a prokaryotic system as

well as the role of an unknown eukaryotic pSer binding domain, will be introduced in detail.

1.4 Phosphorylation signalling in *Mycobacterium tuberculosis*

1.4.1 Overview

Phosphorylation signalling in prokaryotes is not, as initially presumed, limited to 2-component systems. Eukaryotic-like Ser/Thr protein kinases (STPKs) have been described in a growing number of bacteria. Eukaryotic-like systems hold several advantages over 2-component systems. The construction of kinase networks has been mentioned before. In addition, the prokaryotic system generates less stable amino acid modifications than the eukaryotic system. Acyl-phosphates such as phosphorylated aspartate are highly reactive and are prone to undergo nucleophilic substitutions [78]. In 2 component systems, the decay of a signal is dependent on both the non-catalysed hydrolysis of the phosphoesters and the phosphatase activity of a histidine kinase or phosphatase. Only eukaryotic like protein phosphorylation generates a stable chemical modification that is only reversible through the catalytic action of specialised cognate phosphatases. It comes as no surprise therefore that several prokaryotes, including a number of pathogens, have developed or adopted the more powerful Ser/Thr/Tyr kinase signalling networks. One such pathogen possessing extensive phosphorylation signalling machinery is *Mycobacterium tuberculosis*, the causative agent of human tuberculosis.

1.4.2 Tuberculosis, global burden and current strategies

According to the World Health Organisation (WHO), about 1/3 of the world population, more than 2 billion people, are infected with *Mycobacterium tuberculosis* or one of its close relatives and are thus at risk of developing active tuberculosis (TB) [79]. This primarily respiratory, infectious disease, which can also develop into a systemic illness, is devastating in its impact on human life. The number of TB deaths for the year 2010 has been estimated to be around 1.45 million people out of 8.8 million instances of active TB. The considerable danger caused by this disease is also reflected by the US$ 4.4 billion that were spent on the control of TB in 2010 [79]. The majority of the TB burden is carried by India, China and Africa with 64% of all cases. For most types of TB there are effective drug treatments available that can reduce mortality considerably. A combined regimen of Isoniazid, Rifampin, Ethambutol and Pyrazinamide for 6 to 9 months is the most common. As TB becomes asymptomatic after several weeks of

treatment, adherence over the whole course of antibiotics is rare. This non-adherence is detrimental both to the individual patient as the disease resurfaces quickly and to society as exposing the pathogen to a drug without destroying it can lead to the emergence of drug-resistant strains of *M. tuberculosis*. One solution is having the drugs administered directly by a health care professional to ensure adherence and avoid re-lapses and development of drug-resistant strains. Sadly, this treatment regimen is unaffordable for many 3rd world patients and health care systems. TB mortality rates are very much linked to healthcare budgets. The silver-lining of this situation though is that increasing health care standards can greatly reduce the number of TB deaths. China's TB mortality rate has fallen by 80% over the last 2 decades [79]. Its adoption of the WHO's recommended TB eradication program DOTS (directly observed therapy shortcourse) and the subsequent fall in TB fatalities is a model example for other countries suffering from a high TB burden. It should however be noted that the nominal GDP of China has increased more than 10-fold in the last 20 years [80] which may have helped authorities finance the successful implementation of this eradication program.

The large discrepancy of more than 2 billion infected patients and "only" 8.8 million annual instances of TB can be explained by the fact that in a TB infected individual there is an immunological stalemate between host and pathogen. Only when this equilibrium is disturbed in favour of the pathogen does the infection become active. In an individual infected with the human immunodeficiency virus (HIV) the adaptive immune system is greatly compromised and co-infection with TB will more likely result in active TB [81]. TB is one of the most common causes of death for persons infected with HIV [82]. The emergence of drug-resistant strains of *M. tuberculosis* is another cause for alarm. In some countries, more than half of all TB infections are resistant to at least one drug [83].

Although the majority of the deaths caused by TB could be avoided by more universal access to effective health care, the emergence of drug-resistant TB and the long, arduous and costly regimens of current treatments more than justify research into the biology of *M. tuberculosis* and its interaction with the host in order to develop more effective and affordable treatments.

1.4.3 History of medical research on *Mycobacterium tuberculosis*

The causative agent of TB, *M. tuberculosis* has been infecting humans for at least 8000 years. The pathogen has been detected in the remains of Neolithic settlers which display typical tuberculosis lesions and since there is no other natural reservoir for this bacterium humans and *M. tuberculosis* must have co-evolved since at least this time [84]. The earliest written mention is by Hippocrates, who described Phthisis (Greek for TB) as a commonly lethal disease whose symptoms included the coughing up of blood [85]. TB continued to be a devastating disease from Roman times through to the 20th century. The UK Medical Research Council evolved in 1919 out of the "Royal Commission Appointed to Inquire into the Relations of Human and Animal Tuberculosis", a commission whose primary purpose was to increase our understanding of TB [86]. While TB was clearly recognised by its symptoms and had many names, Phthisis, consumption, Pott's Disease and the White Plague among others, it was not until the 19th century that real progress into the research of this disease was made. In an era when the disease was at its peak, Jean Antoine Villemin showed in 1869 that TB was an infectious disease and that healthy laboratory animals could be infected with the infectious remains of previous TB victims [87]. In 1882, Prussian scientist Robert Koch found a way to stain *M. tuberculosis* and was able to show histologically the presence of a bacterium inside tissue infected with TB. TB was thus proven to be a bacterial infection [88]. French microbiologists Albert Calmette and Camille Guerin attempted to develop a vaccine using attenuated bacteria in the 1920s and 1930s but their endeavour never produced an effective vaccine. Although the BCG (Bacille Calmette-Guerin) vaccine they produced protected up to 80% of patients for a maximum of 15 years it is much less effective against TB strains circulating outside Europe [89]. It took until the 1940s before an effective treatment against *M. tuberculosis* was developed. Albert Schatz and Selman Waksman discovered streptomycin, the first antibiotic effective against M*. tuberculosis* [90]. Other antibiotics such as isoniazid and rifampicin followed and for the first time in history an effective treatment for TB became available [79]. Due to this medical revolution, TB became a curable disease and was almost eradicated in the developed world by 1980 [91]. Unfortunately, instances of TB in the developed world have been on the rise over the last 2 decades, partially fuelled by the AIDS pandemic [79, 81]. In addition, the emergence of drug resistant TB, multi drug resistant TB (MDRTB) and extremely drug resistant TB (XDRTB) is a cause for great concern [83]. Currently, TB is in most instances a curable infectious disease whose treatment however is long and costly. As

resistance to drugs has become more and more widespread there has been a renaissance of TB research with the aim to find new anti-mycobacterial agents.

1.4.4 *Mycobacterium tuberculosis* – biology and pathogenicity

M. tuberculosis is the bacterium responsible for the vast majority of human TB cases and has therefore been the subject of extensive medical and microbiological research [92]. It is of the genus *Mycobacterium* which is part of the family Mycobacteriaceae. Phenotypically, it is Gram-positive but due to an unusual cell-wall high in mycolic acid it does not display the characteristic purple colour in a Gram stain. *M. tuberculosis*, as well as the other members of the genus *Mycobacterium*, is called acid-fast as it must be de-stained using dilute hydrochloric acid rather than ethanol. *M. tuberculosis* is a rod shaped obligate aerobe that does not sporulate. It does however possess a complicated life-cycle and has the ability to survive in diverse environments. The bacterium undergoes several distinct stages during the course of an infection and it displays great metabolic flexibility. Even when growing in favourable conditions, it only divides every 15-20 hours which is extremely slow for a bacterium. During some stages of infection, the bacterium slows down even further and brings its growth to a virtual standstill [92].

In the initial stage of infection, a prospective human host acquires a TB infection by inhaling an aerosol containing *M. tuberculosis*. As the bacteria are inhaled, the first anatomical structures they come in contact with are the alveoli, where they are readily engulfed by macrophages as part of the normal immunological response [93, 94]. Inside the macrophages, the bacteria are held within a phagosome. In infections with other pathogens, this phagosome is then fused with a lysosome to produce a phagolysosome. The lysosome is low in pH, contains toxic peroxides and various lyases that aid the breakdown of the ingested pathogen [95]. *M. tuberculosis* however has evolved to survive inside macrophages and is thus able to persist [96]. The most important means of persistence is the inhibition of phagosome lysosome fusion. The unusual mycobacterial cell wall plays a major role in this inhibition [97]. The bacterium is able to survive within macrophages for years as an intra-cellular parasite. The multiplying parasites cause the macrophages to rupture and release the bacteria into the extra-cellular fluid where they are taken up by further macrophages. Eventually the immune-system responds to the non-clearance of the pathogen and it forms a sealed lesion, known as a granuloma, around the infected macrophages to prevent the spread of bacteria [98]. The granuloma is surrounded by activated macrophages which are able to ingest and kill any bacteria escaping. It is via the adaptive immune system,

particularly $CD4^+$ T cells that macrophages become activated and able to destroy the pathogen [98]. Once trapped within the granuloma, the bacterial replication is minimal and the pathogen lies dormant until reactivation or it perishes with the death of the host [99]. There have been reports of reactivation after decades of dormancy but it estimated that only in a minority of cases does *M. tuberculosis* become reactivated during the life time of a host. Reactivation is associated with a compromised immune system caused by for example stress, steroid treatment or co-infection with HIV [98]. Upon reactivation, the replicating bacteria escape from the granuloma. Subsequently, they replicate freely in the lungs. Respiratory failure is the most common cause of death in patients with active TB. The spreading bacteria nucleate liquid droplets in the lungs that are released during bouts of coughing which may then be taken up by another host and the cycle repeats again [98].

1.4.5 Eukaryotic like Ser/Thr phosphorylation signalling in Mycobacteria

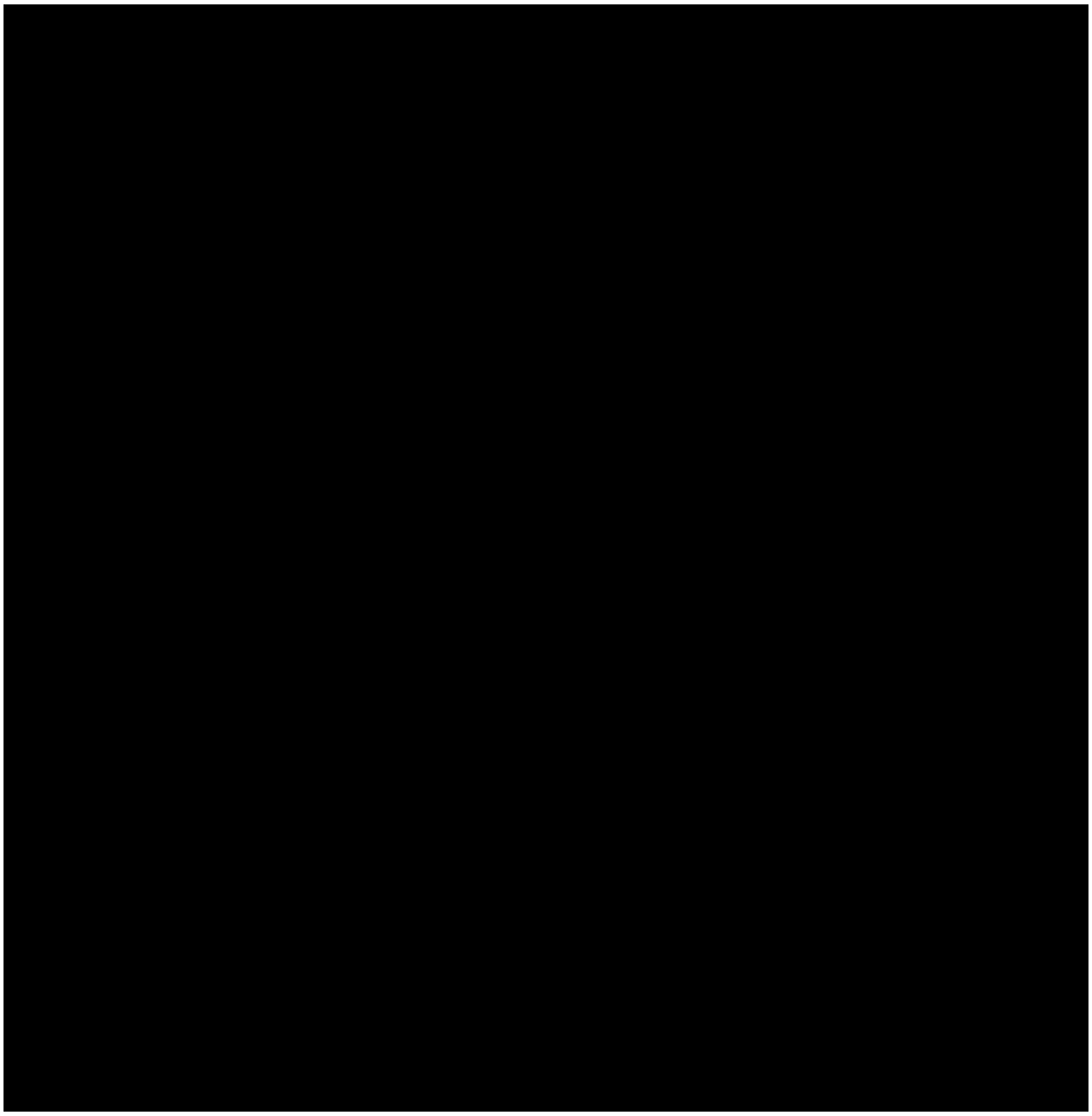
Compared to other prokaryotic genera with similar genome sizes, eukaryotic-like Ser/Thr protein kinases (STPK) are over-represented in the genus *Mycobacterium* (figure 1.6). As these unexpectedly high numbers of STPK are seen predominantly in those members that are intra-cellular parasites, these signalling components are thought to be involved in virulence and regulation of growth [100]. Their implication in virulence and a low sequence identity with human kinases (30%) have marked mycobacterial STPKs as attractive therapeutic targets [77]. The STPKs in mycobacteria appear to have replaced some of the 2-component systems which are less common than would be expected. In the medically most relevant species *M. tuberculosis* there are 11 STPKs (PknA-PknL, except PknC) encoded in the genome as well as 12 histidine kinases. The evolutionary advantage for the replacement of histidine kinases with STPKs in mycobacteria has not been determined, but several general advantages of eukaryotic-like phosphorylation signalling were introduced in section 1.4.1. One characteristic of His-kinases has been retained in a majority of STPKs from *M. tuberculosis*. With the exception of PknG and PknK, all STPKs are membrane-bound with intra-cellular kinase domains which may enable them to function as extra-cellular sensors. There is also one cognate Ser/Thr protein phosphatase, PstP, produced by *M. tuberculosis*. In terms of reader domains, there are 5 FHA domain containing proteins in the *M. tuberculosis* genome containing a total of 6 FHA domains. Interestingly, the substrate specificities of the mycobacterial STPKs studied in detail (PknA, PknB, PknD, PknE, PknF and PknH) reveal motifs highly favoured by FHA domains [101]. These STPKs have so far only been shown to phosphorylate Thr residues and also favour medium sized hydrophobic residues in the +3 position [101].

As the phospho-epitopes generated by the STPK are seemingly tailor-made for FHA-binding and FHA domains are the only recognised modular phospho-biding motifs in *M. tuberculosis* (apart from a single BRCT domain found in a DNA ligase), these domains are likely to play pivotal roles in mycobacterial Ser/Thr phosphorylation signalling [102, 103]. More significantly, 4 out of the 5 mycobacterial FHA domain-containing proteins are transcribed in operons also containing Ser/Thr kinases and have been shown to be substrates of their co-transcribed kinases [104-107]. Although the FHA domain containing proteins are substrates themselves, in the majority of cases the phosphorylation sites reported have not been on the FHA domains themselves, but on flexible linkers N-terminal or C-terminal to the FHA domain. The exact structural and functional consequences of phosphorylation remain to be determined. Only the phosphorylation-induced intra-molecular association of GarA and its functional consequences has been characterised in detail [68, 69, 75]. As only 1 out of 5 FHA domain containing proteins contains a catalytic domain, allosteric regulation is an unlikely mechanism. There are also no reported instances of different mycobacterial FHA domain containing proteins interacting with each other. Oligomeric assembly or an intra-molecular association as reported for GarA could be a common method of FHA domain regulation in mycobacteria. Studies of such FHA domain interactions may not only help to decipher the biology of *M. tuberculosis* but also expand our knowledge of FHA domain versatility in general. The following sections will introduce the currently available information on the FHA domain containing proteins of *M. tuberculosis*.

1.4.5.1 Rv1267 (EmbR)

The first *M. tuberculosis* FHA domain containing protein characterised was EmbR, a transcriptional activator of the EmbCAB operon. The proteins encoded by this operon are involved in the synthesis of arabinogalactan, which is part of the bacterial cell wall. [108]. EmbR itself is not part of this operon but is transcribed directly upstream of STPK PknH. The EmbR crystal structure reveals it to be made up of 3 domains, an N-terminal winged-helix DNA-binding domain, a bacterial transcriptional activation domain and a C-terminal FHA domain [109]. EmbR has been shown to be a substrate of its co-transcribed STPK PknH. For this phosphorylation reaction an intact EmbR FHA domain is essential [105]. Upon phosphorylation by PknH, the affinity of EmbR for the promoter region of the EmbCAB operon increases and leads to increased levels of transcription [100]. In essence, PknH and EmbR function in the same manner as a His-kinase and response regulator. More recently, it has been reported how EmbR2, a homologue of EmbR which cannot be phosphorylated by PknH, can reduce PknH activity by inhibiting auto- and trans- phosphorylation [110]. In addition, there is *in vitro* evidence that PknJ

may also phosphorylate EmbR, providing another layer of EmbR regulation [111]. The proteins of the EmbCAB operon haven been shown to be the target of the antibiotic ethambutol which has made them, as well as their regulators, attractive therapeutic targets.



1.4.5.2 Rv1747

The ABC (ATP-binding cassette) transporter encoded by gene Rv1747 is another example of an FHA-domain containing protein co-transcribed with its cognate kinase. Rv1747 consists of 2 FHA domains, followed by a C-terminal ATP-binding cassette containing a transmembrane region. ABC transporters are ancient ATPases which are found in both prokaryotes and eukaryotes. Their function is the active transport of solutes across cell-membranes utilising energy generated from ATP hydrolysis [112]. Directly upstream in its operon is the STPK PknF which has been shown to phosphorylate Rv1747 [104]. As is the case for EmbR phosphorylation, efficient phosphorylation of Rv1747 requires phospho-binding competent FHA domains [107]. A knock-out of Rv1747 has no effect on *in vitro* growth of *M. tuberculosis*, but it impairs growth *in vivo*. Using mice as a model organism for TB infection, mycobacterial load in the lungs is reduced 10-fold [113]. Tandem mass-spectrometry has revealed that Thr150 and to a lesser extent Thr208 are the main targets of PknF mediated phosphorylation. The same study also found that disruption of the pThr-binding capacity by mutation of a conserved FHA residue (S47A) of $Rv1747^{FHA1}$ impaired mycobacterial growth in cultured macrophages and the lungs and spleen of mice but no effect was seen when disrupting $Rv1747^{FHA2}$ or in a PknF knock out [62]. Furthermore the double mutant T150A/T208A which abrogates Rv1747 phosphorylation also attenuates growth. These results suggest Rv1747 needs to be both phosphorylated and possess a functional FHA1 for function and that Rv1747 may also be phosphorylated by other STPKs [62]. It is unknown whether the role of the FHA domains is purely to recruit PknF to facilitate Rv1747 phosphorylation, or if they have additional functions. It has been hypothesised that the FHA domains may bind the phosphorylated threonines 150 and 208 in order to form oligomers which could be essential for its function [62].

1.4.5.3 GarA

The protein GarA has already been highlighted for its highly versatile FHA domain in previous sections. Here, the biological role of this protein will be introduced in more detail. GarA was initially identified and named as a regulator of glycogen accumulation in *M. smegmatis*. Overexpression of GarA suppressed accumulation of glycogen in overproducing mutant strains [114]. It has also been found in the culture filtrate of mycobacterial cell cultures and was hence named culture filtrate protein 17 (cfp17) [115]. Its close homologue in *Corynebacterium glutamicum* was found to be an inhibitor of 2-oxoglutarate dehydrogenase (called α-ketoglutarate decarboxylase in *M. tuberculosis*) and subsequently named OdhI. The name of this enzyme in *M.*

tuberculosis has been changed several times as its function was originally mis-annotated from the genomic sequence [116-118]. For simplicity, the mycobacterial FHA domain containing protein will be referred to as “GarA” and its corynebacterium homologue as “OdhI”.

Deletion of protein kinase PknG leads to a growth defect in *Corynebacterium glutamicum* which can be alleviated by the simultaneous deletion of OdhI [119]. OdhI was also found to be phosphorylated by the soluble STPK PknG, a mechanism that was shown to be conserved in mycobacteria [70]. The essential mycobacterial STPK PknB also phosphorylates GarA [69]. Phosphorylation by PknB and PknG occurs in a flexible region approximately 30 residues N-terminal to the FHA domain on residues Thr22 and Thr21 respectively. As expected for 2 residues in such close proximity, phosphorylation of either will result in the same response, but simultaneous phosphorylation is disfavoured, presumably due to steric constraints [68, 70]. GarA is the only mycobacterial FHA domain containing protein that does not lie in an operon or putative operon that also encodes a STPK making the identification of cognate kinases less straightforward.

OdhI is a potent inhibitor of 2-oxoglutarate dehydrogenase and its inhibitory effect is abrogated by OdhI phosphorylation. The PknG knock-out strains display increased levels of glutamine and glutamate which can be explained by the presence of unphosphorylated OdhI which persistently inhibits 2-oxoglutarate dehydrogenase. These increased glutamate/glutamine levels impair growth in *M. tuberculosis*, especially in nutrient poor media [120]. In *M. tuberculosis*, two further metabolic enzymes affected by GarA were detected. Glutamate dehydrogenase was found to be also inhibited by GarA while NAD^+-dependent glutamate synthase displayed increased activity when bound to GarA [68, 70]. The combined inhibitory and activating effects of GarA on those 3 metabolic enzymes at the nexus between tricarboxylic acid cycle and nitrogen assimilation push the mycobacterial TCA cycle towards glutamate accumulation. Due to the effects of GarA, α-ketoglutarate, which is an intermediate in the TCA cycle, is mainly converted into glutamate. Phosphorylation by either PknB or PknG alleviates this and conversion of α-ketoglutarate into succinic semi-aldehyde and progression through the TCA cycle is favoured.

At the protein level, the alleviation of the enzymatic regulation upon phosphorylation can be explained by an intra-molecular association. When the flexible linker in the N-terminal region of this protein is phosphorylated, the pThr epitope thus generated associates with the FHA domain on the same poly-peptide chain. Upon GarA

phosphorylation by its parent kinases, the pThr epitope generated outcompetes the FHA-enzyme interaction. A phospho-dependent intra-molecular association replaces a phospho-independent inter-molecular association.

The pThr binding capacities of the GarA and OdhI FHA domains is utilised in 2 ways in the regulation of glutamate levels. Firstly, the FHA domain engages the phosphorylated kinase activation loop for substrate recruitment. Disrupting the conserved FHA Ser residue prevents GarA phosphorylation [121]. The second mechanism involves an intra-molecular association (figure 1.7). Upon phosphorylation, the FHA binds the phosphorylated residue in the flexible linker. The consequence of this binding event is the flexible linker being threaded around the FHA domain occluding the surface required for interactions with the metabolic enzymes [68, 75]. In addition to those 2 phospho-dependent events, the conserved phospho-binding surface of the GarA FHA domain binds 3 metabolic enzymes in a so far unique phospho-independent interaction. The only other reported instances of phospho-independent FHA domain interactions are the FHA dependent oligomerisation of the *Shigella flexneri* protein MxiG and the interaction between human proteins α-centaurin and KIF13B [122, 123]. These FHA domains however do not possess phospho-binding capabilities.

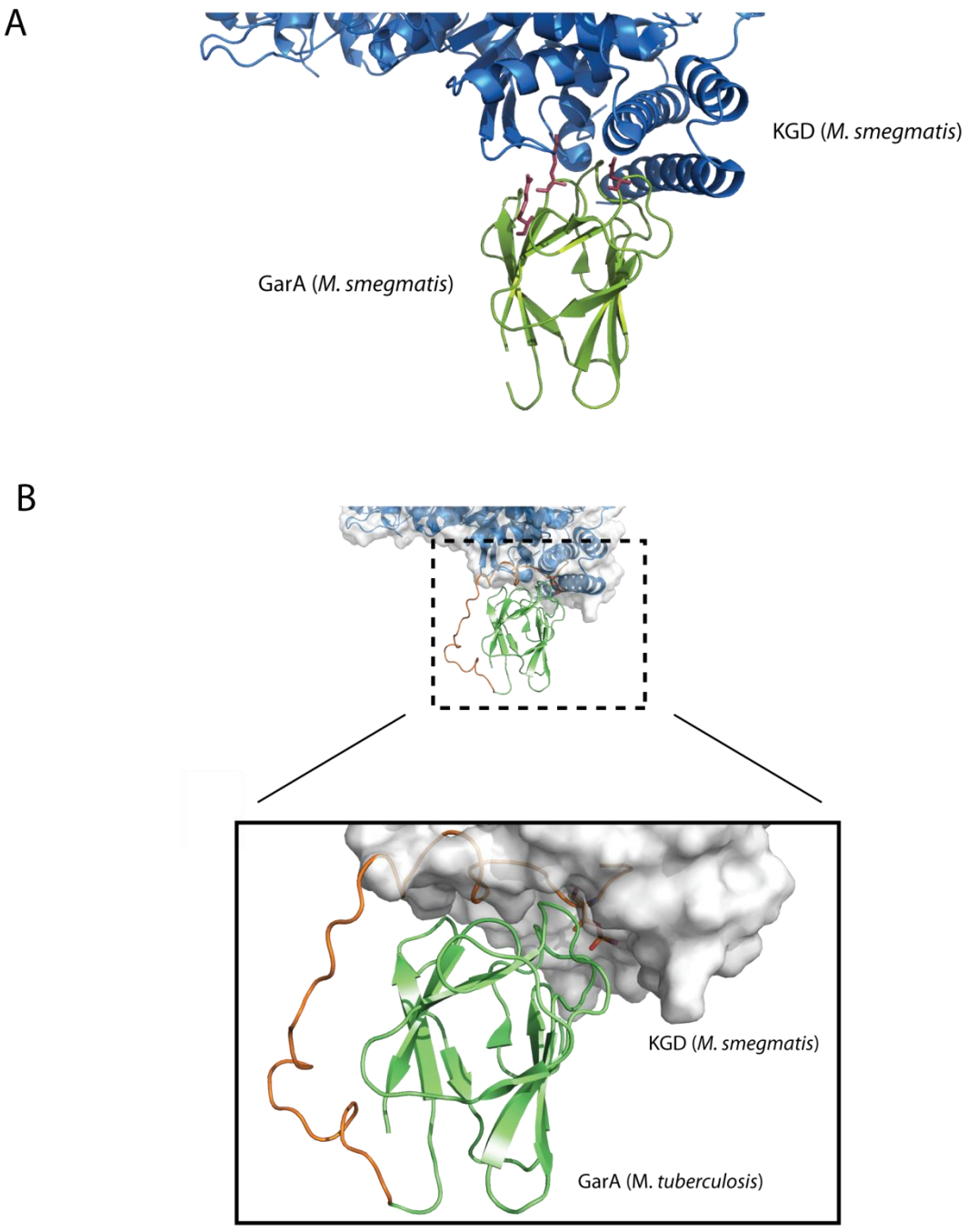


Figure 1.7. The complex between mycobacterial GarA and α-ketoglutarate decarboxylase (KGD) (PDB: 2XT9) (A) cartoon representations of the complex between *M. smegmatis* GarA (green) and KGD (blue) The residues essential for the GarA-KGD interaction identified by Nott et al. are shown as red sticks [68] (B) Cartoon representation of the solution structure of *M. tuberculosis* GarA (lime) (PDB: 2KFU) modelled onto the position of the *M. smegmatis* GarA in the complex from A. FHA domain residues only were used to align the two GarA homologues which have 85% sequence identity. The *M. tuberculosis* GarA flexible linker threaded around the FHA domain is shown in orange. The pThr is shown as sticks.

1.4.5.4 FhaB

The protein encoded by gene Rv0019c is called FhaB, a mycobacterial FHA protein encoded in a putative operon containing a number of proteins involved in Ser/Thr phosphorylation signalling and cell wall synthesis. Seven proteins are organised in this cluster of genes. Upstream there are two FHA domain containing proteins Rv0020c and Rv0019c (FhaA and FhaB respectively), followed by the Ser/Thr phosphatase PstP, cell division protein RodA, a penicillin binding protein pbpA and STPKs PknA and PknB [103]. With the exception of Rv0020c, all proteins encoded by this operon contain at least one transmembrane segment. Both kinases have been shown to phosphorylate FhaA and FhaB with FhaA being the better substrate for both kinases and PknB being the more active kinase [104, 121].

The PknB phosphorylation site on FhaB was identified as Thr36, located in a flexible linker between an N-terminal transmembrane segment and a C-terminal FHA domain [106]. The same study also found that FhaB interacts with PapA5, an acyltransferase implicated in the biosynthesis of virulence enhancing lipid phthiocerol dimycocerosate (PDIM). It is proposed that FhaB binds to PapA5 via its N-terminal linker using a phospho-independent interaction and that PapA5 is phosphorylated by PknB. As the binding assays performed thus far (enzyme linked immune-absorbent linked assay, ELISA) did not generate absolute numbers for binding affinities and neither phosphorylated PapA5 nor FhaB were used, the implications for FhaB phosphorylation or the function of its FHA domain, as well as the significance of the FhaB - PapA5 interaction remain to be determined [106].

FhaB has also been found to interact *in vivo* with FtsZ, a vital protein in Z-ring assembly at the initiation of cell division [124] [125]. The interaction between FhaB and FtsZ requires a functional FHA domain. Furthermore, the FHA domain itself is phosphorylated by PknA on Thr77, which is essential for FtsZ binding. Again, a functional FHA domain is required for phosphorylation by its cognate kinase [124]. A knock-out of the FhaB homologue in *M. smegmatis* slows growth *in vitro*. The cellular roles of the trans-membrane protein FhaB include the localisation of FtsZ to the septum and a function as an adaptor in the interaction between FtsZ and FtsQ [124]. FtsQ is thought to play a role in the initiation of cell division.

1.4.5.5 FhaA

Directly upstream of the Rv0019c gene is FhaA (Rv0020c). FhaA is the only protein in the aforementioned operon which does not contain a transmembrane region and is hence soluble in the cytosol. This FHA domain-containing protein is a substrate of both PknA and PknB [104, 121]. The protein consists of an N-terminal domain of unknown function (DUF3662), a ~300 amino acid low complexity region rich in Gly, Tyr and Pro and a C-terminal FHA domain. An initial study found it to be a PknB substrate and also showed, as in all the cases described above, that an intact FHA domain is required for efficient phosphorylation [104]. The FHA domain of FhaA provides a good example of a minimal FHA domain as it does not contain any helical insertions in the loop regions and has thus been used to investigate FHA binding specificity, especially its absolute discrimination against pSer [61]. An optimal binding peptide derived from *in vitro* peptide library screens binds the FHA domain with an affinity of 100 nM while substitution of the pThr in the optimal peptide with a pSer totally abrogates binding. A poly alanine peptide with a pThr in the centre still binds with an apparent K_D of 7 μM [61]. The residues directly upstream and downstream to the pThr provide binding specificity, but a pThr is essential in every case. Further details of the experiments performed in this study can be found in section 1.3.3.

The biological role of FhaA has remained elusive until very recently. The last six months have seen the publication of 2 independent studies that have both complemented and contradicted some of the experimental results presented in chapters four and five. The solution structure of both DUF3662 and the FHA domain were solved by nuclear magnetic resonance (NMR) and the site for PknB phosphorylation was identified to be Thr116 (figure 1.8) [126]. The role of DUF3662 remains to be determined as its fold did not provide any additional clues as to its function. An affinity of 110 nM was measured between the FHA domain and the phosphorylated juxtamembrane region of PknB. These results suggest that it may be juxtamembrane-mediated, as opposed to activation loop-mediated interactions that play a role in kinase recruitment. The study also showed that DUF3662, which contains Thr116, can be phosphorylated by PknB in isolation. This contradicted earlier findings which suggested an FHA domain dependence for the phosphorylation reaction and also reduces the significance of the tight juxtamembrane-FHA interaction [104].

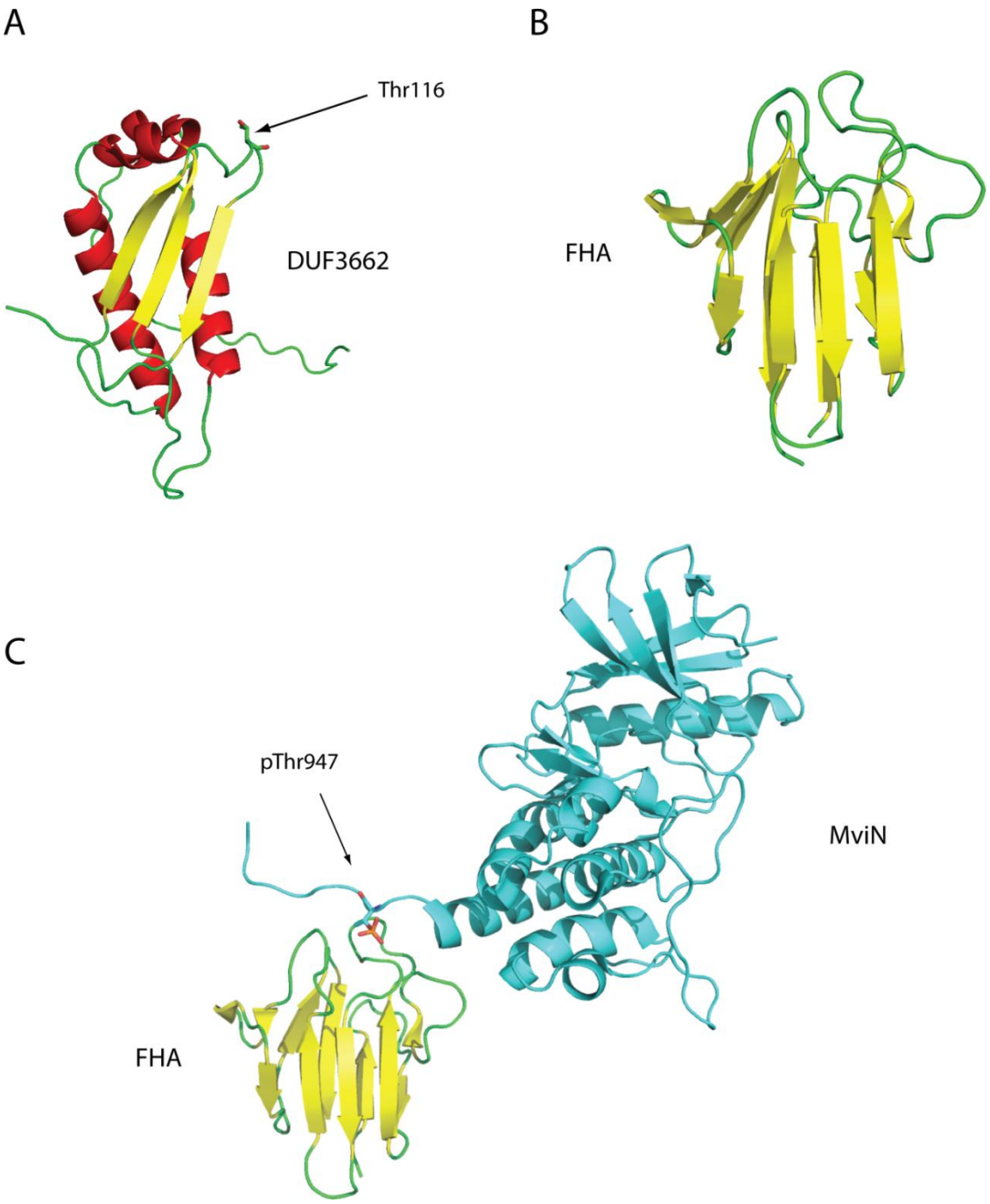


Figure 1.8. Structure of the FhaA domains and their established interacting partner (A) Structure of the N-terminal domain of FhaA in cartoon representation. The Thr phosphorylated by PknB is shown as sticks. (PDB; 2LC0) (B) Cartoon representation of the FhaA FHA domain (PDB: 3PO8)(C) Cartoon representation of the complex between FhaA (standard colour code) and phosphorylated MviN (cyan). The phosphorylated MviN residue pThr947 is shown as sticks. (PDB: 3OUN)

Another study identified MviN (RV3910), a lipid II flippase to be an interacting partner of FhaA [127]. MviN contains a putative pore-forming domain with 14 predicted transmembrane helices followed by an intra-cellular pseudo-kinase domain and a C-terminal extra-cytosolic sugar-binding domain. Flippases are enzymes catalysing the export of peptidoglycan precursors out of the cytosol. Muropeptide precursors are inverted (flipped) and moved into the extra-cellular space where they are incorporated into the bacterial cell wall. Every peptidoglycan producing bacterium produces at least one flippase. Deletion of such enzymes leads to a lethal intra-cellular build-up of muro-peptide precursors [128]. Both FhaA and MviN are substrates of PknB, which is activated by the presence of muropeptides in the extra-cytosolic space (figure 1.9). PknB possesses extracytoplasmic PASTA (penicillin-binding protein and Ser/Thr kinase associated) domains which associate with muropeptides [129]. Muropeptides bound to the PASTA domains cause the kinase to dimerise which in turn induces kinase activation via autophosphorylation [130].

MviN is phosphorylated by PknB on pThr947, directly C-terminal to the pseudokinase domain. In its phosphorylated state, MviN binds the FhaA FHA domain with 80 nM affinity [127]. The crystal structure of the MviN-FhaA complex reveals a canonical FHA-pThr interaction with an additional contact between the C-terminal helix of the MviN pseudokinase domain and the conserved Arg of the β3/β4 loop of the FHA domain. This tight interaction may explain why in *M. smegmatis* FhaA is located at the cell poles and the septum, major sites for cell wall synthesis where MviN is also likely to be located. Depletion of FhaA results in an irregular bulged cell shape and high concentrations of muropeptide precursors at the septum and cell poles but has no effect on growth. Depletion of MviN on the other hand greatly reduced cellular growth [127]. Binding of FhaA to MviN is therefore thought to inhibit the activity of the flippase completing a negative feedback loop for mycobacterial cell wall synthesis. Upon sensing high concentrations of soluble muropeptides, activated PknB phosphorylates MviN. The phosphorylated form of MviN then binds FhaA which inhibits its acitivity, reducing the amount of muropeptide precursors exported [127]. As no interaction was observed between DUF3662 and MviN the role of PknB mediated phosphorylation of FhaA remains to be determined.

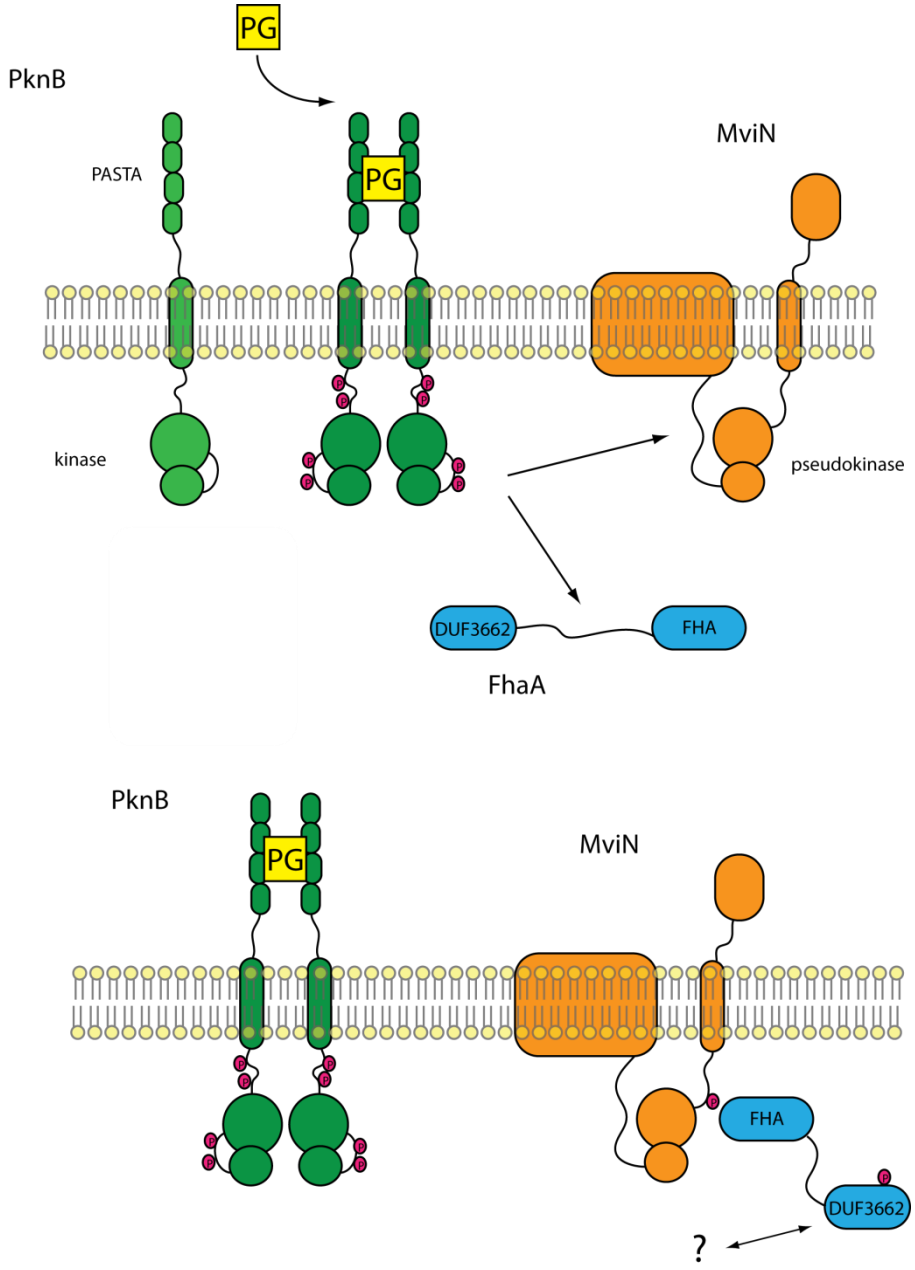


Figure 1.9. Schematic representation of the PknB activation mechanism and the interactions of two of its substrates. PknB (green) phosphorylates MviN (orange) and FhaA (blue). Polar yellow heads and black hydrophobic tails represent the lipid bilayer of the mycobacterial cell separating the cytosol (bottom) from the extracytosolic space (top). Phosphorylated amino acids are shown as the letter "P" in a purple circle. Binding of the PknB pasta domain to peptidoglycan precursors causes the dimerization and subsequent activation of PknB. The kinase then phosphorylates both FhaA and MviN. MviN phosphorylation creates a binding surface for the FhaA FHA domain.

1.5. Phosphatidylinositol-3-kinase like kinases

1.5.1 Overview

The members of the protein superfamily of phosphatidylinositol-3-kinase like kinases (PIKKs) have been implicated as apex regulators of various cellular systems in eukaryotes, particularly those concerned with maintenance of genomic stability. Humans produce 6 PIKKs: ataxia-telangiectasia mutated (ATM), ATM and Rad3-related (ATR), DNA-dependent protein kinase (DNA-PK), mammalian target of rapamycin (mTOR), suppressor with morphological effect on genitalia (SMG1) and transformation/transcription domain-associated protein (TRRAP). The PIKK superfamily displays higher sequence similarity with the lipid kinase phosphatidylinositol-3 kinase than with classical Ser/Thr protein kinases but its members function as bona fide STPKs [2]. The exception to this rule is TRRAP, as it lacks some of the conserved catalytic residues essential for kinase activity (section 1.1.4) [131]. Although TRRAP has close sequence homology with other PIKKs, it lacks kinase activity and can be thought of as a pseudo-kinase. At the protein level, one striking feature of this atypical family of kinases is their large size; they range from 2547 to 4128 amino acids [132] (figure 1.10). The N-terminal region of a typical PIKK is made up of 40-54 α-helical HEAT [Huntingtin, elongation factor 3 (EF3), protein phosphatase 2A (PP2A), and the yeast PI3-kinase TOR1] repeats which do not form globular domains, but may serve as protein-protein interaction surfaces or play a role in intracellular transport. These repeats do not display sequence similarity between the different PIKKs but the C-terminal regions are more conserved. At the extreme C-terminus there is an FRAP-ATM-TRRAP-C-terminal (FATC) domain directly preceded by a PIKK-regulatory domain (PRD). Both FATC and PRD are thought to regulate PIKK activity. Directly N-terminal to the PRD domain is a kinase domain which is catalytically active in all human PIKKs except TRRAP. The kinase domain lies between the PRD and a FRAP, ATM and TRAPP (FAT) domain, another regulatory domain.

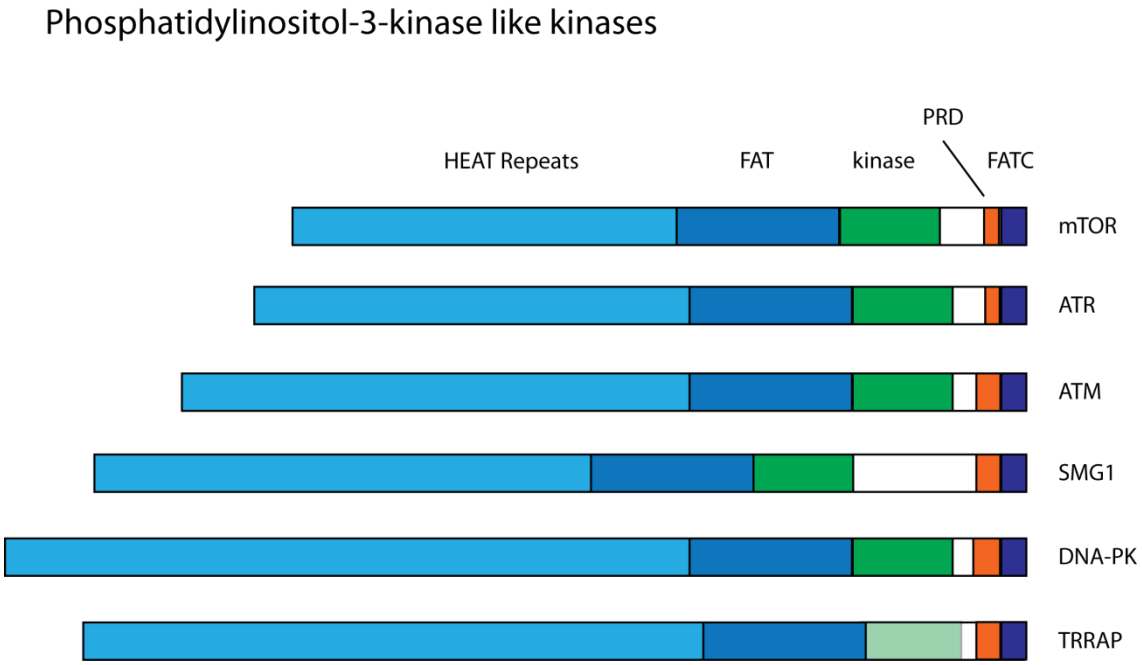


Figure 1.10. Schematic of the domain organisation of human PIKKs. Heat repeats are shown in light blue, FAT domains in dark blue, Ser/Thr protein kinase domains in green, PRD domains in orange and FATC domains in purple. The pseudo-kinase domain of TRRAP is drawn in a lighter shade of green. Sizes of the respective domains are approximate and not drawn to scale.

On a functional level, PIKKs are activated as a response to various types of cellular stress. DNA-PK, and ATM form part of the DNA damage response and aid in the repair of double stranded DNA breaks (DSB). These breaks may be caused by toxic chemicals or radiation, but they also occur during V(D)J recombination in B-cells [133-135]. ATR signalling is triggered by the detection of single stranded DNA breaks and stalled replication forks [136]. mTOR regulates various cellular processes including cell growth and metabolism and has been shown to respond to growth factors and changes in cellular energy levels [137]. SMG1 plays a role in the degradation of prematurely terminated mRNA molecules and has also been implicated in maintaining telomere stability [138, 139]. The pseudo-kinase TRRAP regulates histone acetyltransferases which control transcription and it also associates with the MRN [MRE11, RAD50, and NBS1 (Nijmegen breakage syndrome gene)] complex, part of the cellular response to DSBs [140, 141]. It may act as a link between the DNA damage response and transcriptional regulation. Aberrant PIKK activity has been found in many human diseases, particularly cancers. Inappropriately activated mTOR for example has been found not only in many cancers, but also metabolic disorders and neurological diseases [137].

1.5.2 Expression of PIKKs

Stable expression and folding of PIKKs requires the assembly large of multi-protein complexes. Human telomere length regulating protein 2 (Tel2) was the first protein to be identified as being essential to PIKK stability. It binds to the HEAT repeats of PIKKs and its conditional knock out leads to a dramatic decrease of total PIKK numbers in cells, but no reduction in PIKK mRNA levels. [142]. Deletion of the Tel2 gene leads to embryonic lethality in mice due to genetic defects, reflecting the essential nature of the PIKK proteins, notably mTOR, ATR and TRRAP [143-145]. As Tel2 only binds to newly synthesized PIKKs and the number of Tel2 molecules per cell is at least 20 times lower than the total number of PIKKs it has been proposed that Tel2 acts as a chaperone co-factor to help stabilise immature PIKKs [146]. Tel2 has also been shown to be vital in the assembly of complexes between PIKKs and their essential interacting partners. ATR for example needs to be in complex with ATRIP (ATR interacting protein) for correct ATR signalling. In the absence of Tel2, the levels of newly synthesized ATR associating with ATRIP are diminished. Similarly, Tel2 is required for the assembly of the essential mTOR complexes TORC1 and TORC2.

Tel2 does not possess any catalytic activity, but is thought to act as an adaptor between PIKKs and the ATPase containing R2TP (Rvb1-Rvb2-Tah1-Pih1) complex

[147, 148]. In concert with heat shock protein 90 (Hsp90) and, in the case of mammals, the prefoldin-like complex, this multi-protein complex facilitates the assembly and maturation of a number of large macro-molecular assemblies, namely RNA polymerase II, PIKK complexes and small nucleolar ribonucleoproteins (snoRNP) [148].

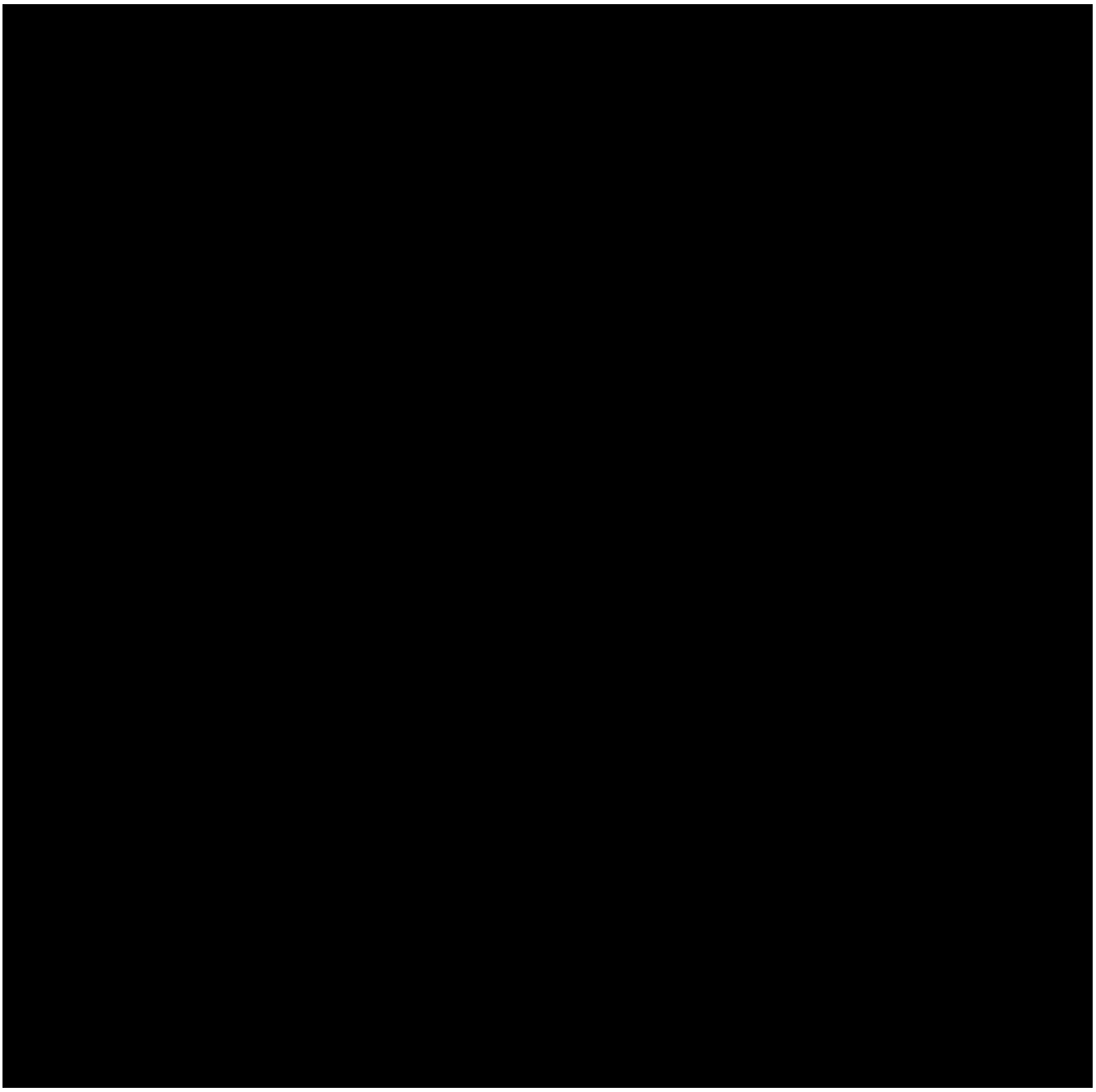
Both Hsp90 and the prefoldin-like complex have chaperone activity [149]. Hsp90 is a highly abundant protein in eukaryotes, making up more than 1% of total cellular protein. Upon encountering cellular stress, particularly heat shock, its levels are elevated [150]. The roles of Hsp90 revolve around the regulation of protein levels in the cell. It is has been shown to bind to immature forms of proteins and has hence been implicated in the folding and maturation of proteins, but it is also prominent in proteasomal degradation [149, 151].

The prefoldin-like complex is also thought to promote protein folding. This complex contains the prefoldin subunits PFDN2, PFDN6 and the prefoldin-like proteins URI, PDRG1, UXT, RPB5 and Monad [152, 153]. The prefoldin family is conserved in archaea and eukaryotes, but not in eubacteria. In eukaryotes, the main function of prefoldin and its relatives is to chaperone the folding of tubular proteins such as actin [154]. Although little is known about this complex, in mammals, prefoldin and prefoldin-like proteins have been seen to interact with members of the R2TP complex.

1.5.3 The R2TP complex

The R2TP complex was first identified in *S. cerevisiae* and comprises four proteins, Rvb1, Rvb2, Tah1 and Pih1. In humans, the proteins making up the core R2TP complex are called RuvBL1, RuvBL2, RPAP3 and PIH1D1 [148]. RuvBL1 and RuvBL2 share close sequence homology and form an AAA ATPase, the only catalytically active proteins of the R2TP complex [155]. Electron microscopy studies have shown that the ATPase is made up of hexameric rings of RuvBL1 and RuvBL2, but there is no molecular information available on how this ATPase utilises the energy from ATP hydrolysis [156].

RPAP3 was initially identified as an Hsp90 co-factor of unknown function [157]. It contains 6 tetratricopeptide repeats (TPR) each of approximately 34 amino acids in length which have been found to act as scaffolds to facilitate binding of Hsp90 to its co-factors [158]. Tah1 binds an MEEVD motif at the extreme C-terminus of Hsp90 via its TPR repeat [159]. As both the TPR repeats and the MEEVD motif are conserved among eukaryotes, the interaction is assumed to occur in humans.



PIH1D1 was also identified as an Hsp90 interacting partner with no other known function [157]. This protein does not, unlike RPAP3, contain any known structural motifs. At a functional level, PIH1D1 has been shown to be essential for R2TP functions such as snoRNP assembly [160]. It is thought to be an adaptor protein connecting the different proteins of the R2TP, Hsp90 and prefoldin-like complex. Tah1 and Pih1 have been shown to form a tight interaction via their C-termini, which is required for PIH1D1 stability [160]. PIH1D1 also binds the RuvBL1/RuvBL2 ATPase and Monad, part of the prefoldin-like complex [160, 161]. While the main interaction between the R2TP complex and Hsp90 is mediated by the TPR domains of RPAP3, there is also a weak direct interaction between PIH1D1 and Hsp90 [159].

PIH1D1 also interacts with Tel2, the protein essential for PIKK stability. This Tel2-PIH1D1 interaction may explain why Tel2 is needed for PIKK maturation. Tel2 and PIH1D1 form the link that connects PIKKs to the catalytically active chaperone complexes essential for correct folding [147]. While the nature of the PIH1D1 mediated interactions have not been determined, there is evidence that the PIH1D1-Tel2 interaction may be phospho-dependent. Two conserved serines on Tel2, Ser487 and Ser491, which are constitutively phosphorylated by casein kinase 2 (CK2), have been shown to be essential [147]. CK2 has been implicated in the regulation of cell proliferation, apoptosis and senescence [162]. Upon mutation of these two serines, the direct interaction between Tel2 and PIH1D1 is no longer observed and at the cellular level, PIKK levels are diminished [147]. Levels of mTOR and SMG1 are decreased dramatically while the effect on the levels of TRRAP, DNA-PK, ATM and ATR is less severe. Deletion of Tel2 decreases the levels of all PIKK while only mTOR and SMG1 are affected severely by the disruption of the Tel2-R2TP link [146, 147]. It can be assumed that Tel2 may bind Hsp90 directly, which is sufficient for TRRAP, DNA-PK, ATM and ATR assembly, but SMG1 and mTOR require the entire R2TP, Hsp90 and prefoldin-like complex for correct assembly. This hypothesis is supported by the fact that inhibition of Hsp90 mainly affects DNA-PK, ATM and ATR levels, but not SMG1 and mTOR levels [142].

1.6. Objectives of this study

This study is focused on the structural and biochemical analysis of 2 phospho-dependent reader domains; FhaA from *Mycobacterium tuberculosis* and human protein PIH1D1. The phospho-binding capabilities of the FhaA FHA domain have already been studied comprehensively and a cellular function has also been assigned to this protein. FhaA binds MviN, a protein that is responsible for export of peptidoglycan precursors across the mycobacterial cell membrane. Its FHA domain binds a pThr epitope on MviN created by mycobacterial kinase PknB. It has been shown that FhaA is also a substrate of PknB, but the effect of this phosphorylation on FhaA function has not been determined. The aim of the first part of this study is to determine the structural and functional consequences of PknB mediated phosphorylation of FhaA.

The second part focuses on PIH1D1. This protein acts as a scaffold at the heart of a large multi-protein complex containing the R2TP complex, Hsp90 and the prefoldin-like complex. One of the functions of this complex is to chaperone the stabilisation and maturation of the Ser/Thr protein kinase superfamily of phosphatidylinositol 3-kinase related kinases (PIKKs). PIH1D1 binds the PIKK co-factor Tel2 in an interaction essential for PIKK stability. This interaction is dependent on two casein kinase 2 phosphorylation sites on Tel2, suggesting that PIH1D1 may contain a phospho-binding domain that recognises the pSer epitopes on Tel2. The aim of the second part of this study is to characterise the phospho-binding capabilities of PIH1D1 and determine the fold of this protein whose sequence is dis-similar to all known phospho-reader domains.

2 Materials and Methods

2.1 Molecular biology

DNA Plasmids containing full length ORFs for FhaA (Rv0020c gene) and PknB (Rv0014c gene) were provided by Ms Debbie Hunt (NIMR). Dr Tom Alber (University of California, Berkeley) supplied an expression clone for amino acids 679-963 of MviN. A plasmid of the human PIH1D1 gene was provided by Dr Simon Boulton (CRUK, Clare Hall). A further plasmid, encoding the first 166 amino acids of FhaA, was purchased (GeneArt). This was a plasmid codon optimised for expression in *E. coli* and used by Mr Vangelis Christodoulou (NIMR) to produce an expression construct of the first 130 amino acids of FhaA fused to a C-terminal 6x His-tag. The plasmids mentioned above were used, where applicable, as expression clones or served as templates to generate the expression clones required. DNA concentrations were measured using a NanoDrop spectrometer (Thermo Scientific) assuming that A_{260} = 1 equated to 50 ng/µl of DNA.

2.1.1 Bacterial strains used

A total of four different bacterial strains were used, shown in table 2.1.

Bacterial strain	**Genotype**
Escherichia coli BL21(DE3) (Invitrogen)	F^- *ompT hsdS*$_B$(r_B–, m_B–) *gal dcm* (DE3)
Escherichia coli TOP10 (Invitrogen)	F^- *mcr*A Δ(*mrr-hsd*$_{RMS}$-*mcr*$_{BC}$) Φ80*lacZ*ΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara leu*) 7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G
Escherichia coli B834 (DE3) (Novagen)	F^- ompT hsdS$_B$(r_B^- m_B^-) gal dcm met (DE3)
Epicurian coli XL10-Gold (Agilent)	TetrΔ (*mcrA*)*183* Δ(*mcrC*$_B$-*hsd*$_{SMR}$-*mrr*)*173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac* Hte [F' proAB *lacI*q*ZΔM15* Tn*10* (Tetr) Amy Camr]

Table 2.1. Genotypes of bacterial strains used

The BL21 (DE3) strain was used for standard protein expression, TOP10 cells were used for plasmid propagation following cloning procedures. XL10-Gold cells are part of the QuikChange kit (Agilent) used for site-directed mutagenesis. The B834 (DE3) strain is a methionine auxotroph and was used for production of selenomethionine-labelled protein.

2.1.2 Plasmid vectors and primers

Following standard cloning techniques (see below), recombinant proteins were expressed using either the pET47b(+) or from the pET49b(+) vector (both Novagen). The multiple cloning sites of these vectors were modified in-house by Vangelis Christodoulou to allow for ligation independent cloning (see below). Custom DNA primers were synthesized and purified by de-salting by Sigma Aldrich and were designed so as to be compatible with the in-house ligation independent cloning (LIC) system (see below). All plasmid DNA was purified from *E. coli* TOP10 cultures using the HiSpeed Mini-prep kit (Qiagen) following the low copy number protocol.

Plasmid vector	Antibiotic marker	Promoter	Purification tag
pET47b(+) LIC	Kanamycin	T7 *lac*	N-terminal 6x-His fusion
pET49b(+) LIC	Kanamycin	T7 *lac*	N-terminal glutathione S-transferase (GST) fusion

Table 2.2. List of plasmids used

2.1.3 Polymerase chain reaction (PCR)

All PCR reactions were carried out on a DNA Engine (DYAD) thermocycler in 200 µl thin-walled PCR tubes (Abgene) using KOD Hot Start DNA Polymerase (Novagen). Reaction volumes were 50 µL (table 2.3) with the lid temperature at a constant temperature of 100 °C. The thermal cycling was programmed to the polymerase manufacturer's instructions with minor modifications (table 2.4).

Reagent	Amount
2x KOD Hot Start Master Mix	25 µl
forward primer (10 µM)	1.5 µl
reverse primer (10 µM)	1.5 µl
template	50 ng
H_2O	to 50 µl

Table 2.3. Reagent mix for a typical PCR reaction

The master mix contains KOD Hot Start DNA Polymerase, two monoclonal antibodies, deoxynucleotides, and reaction buffer with $MgSO_4$. The function of the monoclonal antibodies is to inhibit the DNA polymerase and 3' to 5' exonuclease activities of KOD polymerase at ambient temperature. The concentration of KOD polymerase in the master mix is 0.04 U/µl. As a result, there is 1 U of polymerase present in each 50 µl PCR reaction.

Step	Temperature	time
1	95 °C	2 min
2	95 °C	20 sec
3	60 °C	10 sec
4	70 °C	90 sec/kb
5	cycle to step 2 29 times	

Table 2.4. Typical thermal cycling protocol for PCR

The annealing temperatures varied between 55 °C and 65 °C and were chosen according to primer melting temperatures. Following thermal cycling, primers and dNTPs were removed using a PCR clean-up kit (Qiagen).

2.1.4 Ligation independent cloning (LIC)

Ligation independent cloning was chosen to insert the PCR products generated into the plasmid vectors. This system does not rely on restriction enzymes and DNA ligases, but on the 3' to 5' exonuclease activity of T4 DNA polymerase to generate single stranded overhangs. As is the case with many proofreading DNA polymerases, T4 DNA polymerase not only possesses 5'-3' polymerase activity, but also 3'-5' exonuclease activity which enables the polymerase to excise incorrectly added bases. In the absence of deoxynucleotides, the enzyme functions mainly as an exonuclease as no new bases can be added. If only one type of deoxynucleotide is present, the polymerase removes bases until it encounters a base that can be re-inserted, creating single-stranded overhangs. If the overhangs are sufficiently long, i.e. 10 bases or more, two pieces of DNA with complementary overhangs can be annealed and transformed into competent cells whereupon the nicks are sealed by the bacterial repair machinery.

Reagent	Amount
Purified PCR product	0.2 pmol
T4 DNA polymerase (NEB)	2 µl
25 mM dATP	2 µl
T4 DNA polymerase (1U/µl)	0.5 µl
H_2O	to 20 µl

Table 2.5. Reagent mix for T4 DNA polymerase treatment of PCR products

In the LIC system developed in-house, PCR primers are designed to have 5'-extensions that have their first thymidine base after 10 bases (see figure 2.1). Cleaned up PCR products are incubated with T4 DNA polymerase in the presence of dATP as the only deoxynucleotide, which creates 10 base pair 5' overhangs at their ends. The PCR product was incubated with the polymerase for 30 minutes at room temperature followed by heat inactivation at 75 °C for 20 min. A modified pET47b(+) or pET49b(+) plasmid vector, which had been linearised and treated in a similar way but using dTTP instead of dATP, possessing complementary overhangs, was provided by Vangelis Christodoulou. For the annealing, 0.02 pmol of the treated insert were mixed for 5 minutes with 0.01 pmol of treated vector in a volume of 3 µl. After 5 min 1 µl of 25 mM EDTA was added and the reaction was incubated for a further 5 min. The annealed insert and vector were then used to transform chemically competent *E. coli* TOP10 cells.

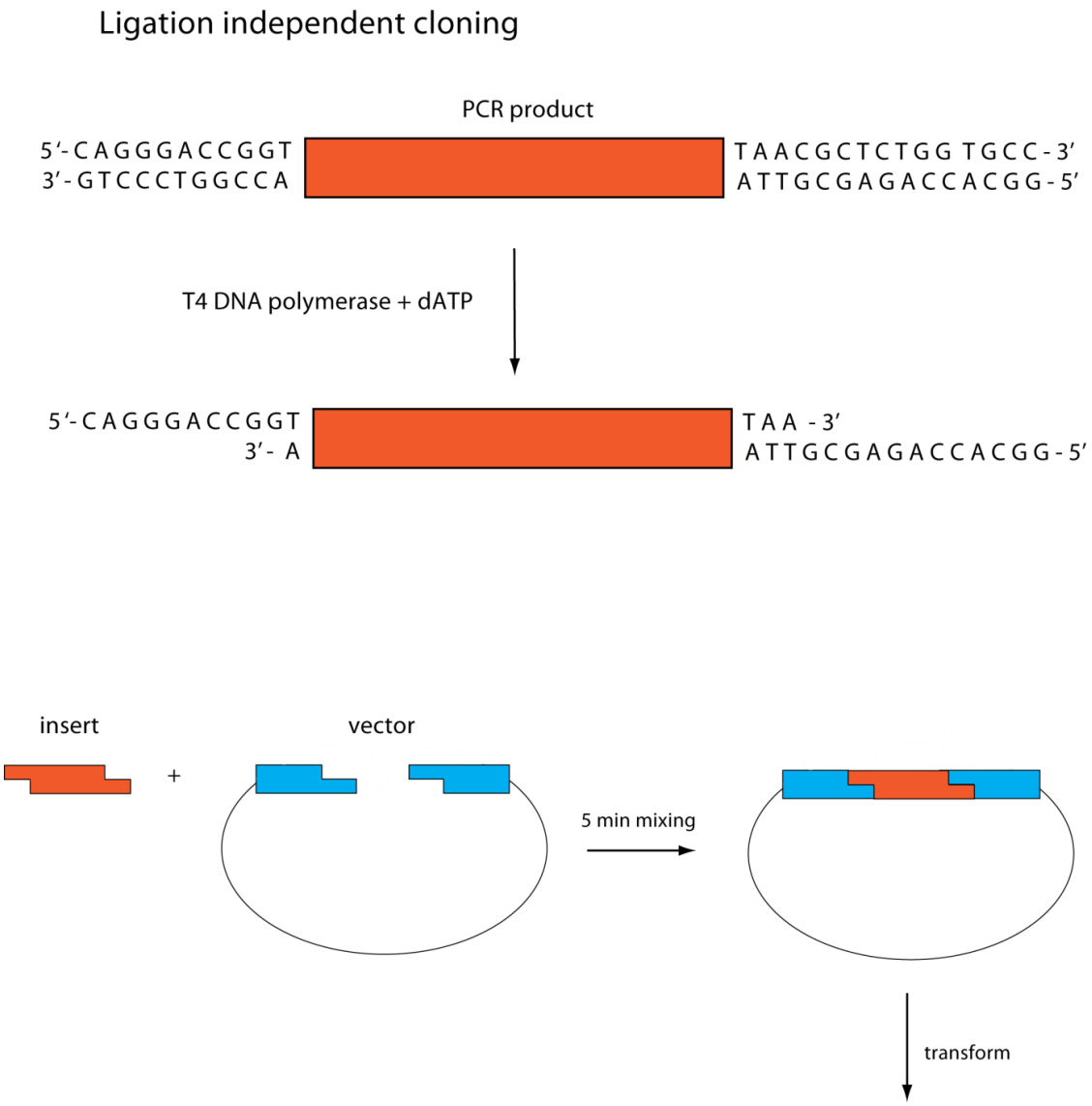


Figure 2.1. The LIC cloning system: A PCR product with the appropriate overhangs is digested by T4 DNA polymerase. The treated product is then mixed with a vector of complementary overhangs. The nicks in the annealed vector are then repaired by bacterial DNA repair mechanisms after transformation

2.1.5 Site-directed mutagenesis

All site-directed mutagenesis was carried out using the QuikChange kit (Agilent). This system relies on the nuclease *Dpn I* which digests methylated DNA. A methylated plasmid from an *E. coli* culture is amplified by PCR introducing a mutation by using slightly mis-matched primers. Upon incubation with *Dpn I*, the methylated template is digested while the newly synthesized, non-methylated plasmid DNA remains intact. See table 2.6 for a typical PCR reaction mix.

Primers were designed using the manufacturer's protocol and were synthesized and HPLC purified by Sigma Aldrich.

Reagent	Amount
template plasmid DNA	50 ng
forward primer (10 µM)	1.5 µl
reverse primer (10 µM)	1.5 µl
reaction buffer	5 µl
QuikSolution	3 µl
Pfu DNA polymerase (2.5 U/µl)	1 µl
H_2O	to 50 µl

Table 2.6. Reagent mix of a PCR reaction for site-directed mutagenesis

The PCR reaction mixtures and the thermal cycling were set up according to the protocol suggested by Agilent.

Step	Temperature	time
1	95 °C	2 min
2	95 °C	50 sec
3	60 °C	50 sec
4	68 °C	1 min/kb
5	cycle to step 2 for 18 times	
6	68 °C	7 min

Table 2.7. Thermal cycling protocol of a PCR reaction for site-directed mutagenesis

The PCR mixtures were incubated for 1h with 10 U of *Dpn I* at 37 °C after which 2 µl of the reaction were transformed into XL10-gold cells.

2.1.6 Transformations and sequencing

Transformations were carried out according to the relevant manufacturer's instructions depending on the strain used. Transformation reactions were plated out on LB agar plates supplemented with 35 µg/ml kanamycin and incubated at 37 °C overnight. Resulting colonies were picked and grown up in 10 ml of LB + kanamycin. Cultures were harvested and plasmid DNA was extracted using the HiSpeed MiniPrep kit (Qiagen) and sent for sequencing (Beckman Coulter Genomics).

2.1.7 Protein expression

Plasmids containing the desired expression constructs were transformed (as above) into BL21 DE3 cells. The next day a single colony was picked and incubated in 250 ml of LB + kanamycin for 4 hours at 37 °C under constant shaking. This starter culture was used to inoculate expression cultures, usually 12 x 500 ml cultures in 2 l conical flasks, at a dilution factor of 1/50. The cultures were grown under constant shaking at 37 °C until an OD_{600} of 0.6 - 0.8 was reached. The temperature was then lowered to 18 °C, the expression of protein induced by the addition of IPTG (Isopropyl-β-D-thio-galactoside) to a final concentration of 0.5 mM, and the culture grown overnight. Cells were harvested by centrifugation at 6000g for 20 min and the pellets frozen at -20 °C. In the case of ^{15}N- and ^{13}C-labelled samples for NMR analysis, the final expression culture was grown up in a minimal medium supplemented with kanamycin. Each litre of minimal media contained Na_2HPO_4 (6 g), KH_2PO_4 (3 g), NaCl (0.5 g), $MgSO_4$ (0.25 g), $CaCl_2$ (0.015 g), $FeSO_4$ (0.015 g), thiamine·HCl (1 g), D-biotin (0.1 g), $^{15}NH_4Cl$ (1 g), U-$^{13}C6$ D-glucose (2 g).

To produce seleno-methionine labelled proteins, the final expression culture was grown in SelenoMethionine Medium Complete (Molecular Dimensions) supplemented with Selenomethionine (Molecular Dimensions) to a final concentration of 40 µg/l. In this case B834 cells were used as they are unable to synthesize methionine and hence only incorporate the selenomethionine from the growth medium into their proteins.

Protein expression was induced via components of the *lac* operon and T7 RNA polymerase. T7 RNA polymerase is a viral polymerase which has specificity for the T7 promoter and only transcribes genes 3' of this sequence. DE3 strains of *E. coli* used in this study are lysogenic for a λ prophage that possess a gene containing the *lac* promoter P_{lac}, the lac operator *lac* O and the gene for T7 RNA polymerase. The gene coding for the lac repressor *lac I*, a protein that binds to the *lac* operator in the absence of lactose and blocks transcription, is located on the bacterial chromosome. It is under

the control of bacterial RNA polymerases and is constitutively expressed. In culture, where glucose is the main energy source, the *lac I* protein binds to the *lac O* DNA sequence and prevents transcription of T7 RNA polymerase. Upon addition of IPTG however, a non-hydrolysable analogue of lactose, the *lac* repressor is displaced at the *lac* operator by IPTG, allowing for transcription of T7 RNA polymerase.

All pET vectors contain a multiple cloning site where the gene of interest can be inserted. This gene is under the transcriptional control of the T7 promoter and a *lac* operator. Following the addition of IPTG into the culture, the repression at the lac operator is alleviated in the plasmid and the rising levels T7 RNA polymerase transcribe the gene of interest. As the viral RNA polymerase and the bacterial RNA polymerases have mutually exclusive specificities, it ensures that the gene of interest is only expressed upon the addition of IPTG and that T7 RNA polymerase does not interfere with endogenous bacterial transcription. Furthermore, the presence of a lac operator upstream of both the genes for T7 RNA polymerase and the recombinant protein puts the expression of the recombinant protein under tight transcriptional control.

2.2 Protein preparation

All protein preparations followed the same general purification protocol of an initial cell lysis and clarification by centrifugation, followed by affinity purification and gel filtration (size exclusion chromatography). In some cases, affinity purification did not produce sufficiently pure samples and ion-exchange chromatography was performed after this step to improve sample quality. Sample quantity and quality were monitored at regular intervals using UV-spectroscopy and SDS-PAGE respectively.

2.2.1 Protein concentration determination

Protein concentration was determined by UV spectroscopic analysis using a NanoDrop spectrometer (Thermo Scientific). A continuous scan from 220 – 350 nm was taken and the A_{280} used to determine protein concentration. Protein extinction coefficients were calculated using the ProtParam tool (web.expasy.org/protparam) assuming individual extinction coefficients of 1280 $M^{-1}cm^{-1}$ per tyrosine and 5500 $M^{-1}cm^{-1}$ per tryptophan.

Construct	Purification tag	Purification strategy	origin
FhaA$^{1-527}$ (FL)	N-GST, 3C-cleavable	GST, Heparin, S200	*de novo*
FhaA$^{1-110}$	N-His, 3C-cleavable	Insoluble protein	
FhaA$^{1-130}$	N-His, 3C-cleavable	Ni^{+}, Heparin, S75	*de novo*
FhaA$^{1-130}$	C-His	Ni^{+}, S75	V. Christodoulou, NIMR
FhaA$^{1-167}$	N-GST, 3C-cleavable	GST, Heparin, S75	*de novo*
FhaA$^{1-166}$	N-GST, 3C-cleavable	GST, Heparin, S75	*de novo*
FhaA$^{155-527}$	N-GST, 3C-cleavable	GST	Dr Dony Patel, NIMR
FhaA$^{354-527}$	N-GST, 3C-cleavable	GST, S75	*de novo*
FhaA$^{367-527}$	N-GST, 3C-cleavable	GST, S75	*de novo*
FhaA$^{430-527}$	N-GST, 3C-cleavable	GST, S75	Dr Dony Patel, NIMR
MviN$^{679-963}$	N-His, N-MBP, TEV-cleavable	Ni^{+}, Resource Q,S200	Dr Tom Alber, UC Berkeley
PknB$^{1-279}$	N-GST, 3C-cleavable	GST, Q, S75	Dr Dony Patel, NIMR
PknB$^{1-331}$	N-GST, 3C-cleavable	GST, S75	*de novo*
PIH1D1$^{1-290}$ (FL)	N-His	Ni^{+}, S200	Dr Simon Boulton, CRUK
PIH1D1$^{1-180}$	N-His, 3C-cleavable	Ni^{+}, S75	*de novo*
PIH1D1$^{40-180}$	N-His, 3C-cleavable	Ni^{+}, S75	*de novo*

Table 2.8. List of expression constructs used, as well as their respective purification tags and strategies (abbreviations used, GST: GST affinity purification, Ni^{+}: Ni^{+}-affinity purification via 6x-His tag, S75/S200: size exclusion chromatography using a Superdex S75 or S200 respectively, Heparin/Q/Resource Q: ion-exchange chromatography using the respective Hitrap column)

2.2.2 SDS-PAGE

All SDS-PAGE experiments were performed in 1x MES Buffer (Invitrogen) on Novex 10% 1.0 mm Bis-Tris gels (Invitrogen) run in an XCell SureLock mini-cell (Invitrogen) for 38 min at a constant 200 V. The SeeBlue Plus 2 (Invitrogen) pre-stained protein standard was used as a size-marker. For each well, 5 µl of protein solution was typically loaded with 5 µl of NuPAGE LDS sample buffer and 1 µl of 1 M TCEP ((tris(2-carboxyethyl)phosphine) pH 8.0. Proteins were boiled in the presence of the sample buffer and reducing agent at 95 °C for 5 min prior to loading. Gels were stained using the InstantBlue (Expedeon) coomassie stain.

2.2.3 Protein buffer exchange, concentration and storage

Proteins were concentrated using 20 ml Vivaspin ultrafiltration concentrators (Sartorius Stedim Biotech) of the appropriate molecular weight cut-offs at 4000g. Sample de-salting was carried out using PD-10 or NAP-5 desalting columns (GE Healthcare) following the manufacturer's instructions. When protein was to be dialysed, mini

dialysis kits (GE Healthcare) of the appropriate molecular weight cut-offs were used. Typically, one dialysis cap with 250 µl of protein solution was placed in 1l of buffer. For long term storage, 20-50 µl aliquots of protein were flash frozen in liquid nitrogen and stored at -80 °C.

2.2.4 Bacterial lysis

Frozen bacterial cell pellets were thawed and resuspended in a buffer of 500 mM NaCl, 50 mM Tris pH 8.0 and 0.5 mM TCEP at a ratio of 10 ml of buffer per 1g of cell pellet. For proteins that were prone to co-purify with nucleic acids, the buffer was supplemented with 20 U/ml of benzonase (Novagen). If further purification involved Ni^+-affinity chromatography, imidazole was added to a final concentration of 20 mM. The suspensions were kept on ice and lysed by sonication (Branson Sonifier 450) until a visible reduction in lysate viscosity indicated effective cell lysis. The lysates were clarified by centrifugation at 45000g for 50 min and a sample of the supernatant removed for analysis by SDS-PAGE.

2.2.5 Ni^+-affinity purification

Some of the recombinant proteins produced in this study possess a 6x-His tag and were purified by Ni^+-affinity chromatography. 1-3 ml of Ni Sepharose 6 Fast Flow (GE Healthcare) were equilibrated in 500 mM NaCl, 50 mM Tris pH 8.0, 20 mM imidazole pH 8.0 and 0.5 mM TCEP pH 8.0 lysis buffer. The cleared lysate was added and the slurry mixed for one hour at 4 °C. After the binding step, the slurry was poured into a gravity flow column. A sample of the flow-through was removed for analysis by SDS-PAGE. The column was washed with 50 column volumes of lysis buffer to remove unbound protein. When a 6x-His fusion protein was desired, the protein was eluted with 2 column volumes 500 mM NaCl, 50 mM Tris pH 8.0, 200 mM imidazole pH 8.0 and 0.5 mM TCEP pH 8.0 elution buffer and the eluate collected and analysed by SDS-PAGE.

In some cases the 6x-His tag was cleaved off using human rhinovirus 3C protease (3C). The protein was bound and the column washed as before and non-cleavable GST-fused 3C protease was added and mixed with the washed beads overnight at 4°C. For each ml of beads, 35 µg of 3C were used. Glutathione Sepharose 4B (GE Healthcare) beads were added to bind the 3C the following day and the cleaved protein eluted from the column using 2 column volumes of lysis buffer before being analysed by SDS-PAGE.

The purification of MviN required a slightly altered protocol as the 6x-His-MBP tag was only cleavable using TEV protease (Sigma Aldrich) and was of a similar size to the cleaved protein. The fusion protein was eluted from the Ni^{+}-column as described before, TEV protease was added at a substrate:protease ratio of 500:1 (w/w) and the reaction incubated at 4 °C overnight. Imidazole was removed by dialysis and the cleaved 6x-His-MBP tag removed by passing it through Ni^{+}-sepharose. The cleaved protein was then further purified using ion-exchange chromatography.

2.2.6 GST-purification

All proteins containing a glutathione S-transferase tag were purified using a matrix of glutathione Sepharose 4B (GE Healthcare). Clarified cell lysate was applied to 10 ml of glutathione sepharose resin equilibrated in lysis buffer (500 mM NaCl, 50 mM Tris pH 8.0 and 0.5 mM TCEP). The slurry was mixed overnight at 4 °C to bind the GST-fused protein to the matrix. It was then poured into a gravity flow column and washed with 400 ml of lysis buffer. When a GST-fusion was required, the protein was eluted from the column using 3 column volumes of 500 mM NaCl, 50 mM Tris pH 8.0, 0.5 mM TCEP and 20 mM glutathione pH 8.0. A sample was taken to be analysed by SDS-PAGE. To produce cleaved protein, the washed column was incubated with 100 µg of non-cleavable GST-fused 3C protease overnight at 4°C. The protein was then eluted from the column using 2 column volumes of 500 mM NaCl, 50 mM Tris pH 8.0 and 0.5 mM TCEP and analysed by SDS-PAGE.

2.2.7 Ion exchange purification

Some protein purifications required an ion-exchange chromatography step. Following affinity purification, cation exchange chromatography using a Q Sepharose Hitrap column (GE Healthcare), or anion exchange using a Heparin Hitrap column (GE Healthcare) was performed. The high resolution Resource Q resin (GE Healthcare) was used to separate phosphorylated from unphosphorylated isoforms of proteins after *in vitro* phosphorylation. Prior to ion exchange, the protein sample buffer was adjusted to 25 mM NaCl, 20 mM Tris pH 8.0 and 0.5 mM TCEP binding buffer. Sample was then applied to the column, pre-equilibrated in the same buffer, using a P1 peristaltic pump (GE Healthcare) with a continuous flow cycle. The column was then attached to an AKTA Prime or AKTA purifier (GE Healthcare) pump and washed with 10 column volumes of binding buffer. Protein was fractionated using a gradient of 25-500 (Q) or 25-1000 (heparin) mM NaCl and fractions analysed by SDS-PAGE. Phosphorylation states were determined by electrospray mass spectrometry.

2.2.8 Size-exclusion chromatography

Size-exclusion chromatography was employed as a final step of the protein purification process to remove contaminants and aggregates. Appropriate fractions from the earlier purification were pooled and concentrated to a final volume <5 ml. A Superdex S75 or S200 16/60 size exclusion column (GE Healthcare) was equilibrated in a buffer of 150 mM NaCl, 20 mM Tris pH 8.0 and 0.5 mM TCEP. The protein sample was then loaded onto the column and run at 0.5 ml/min using an AKTA prime pump. Eluate was fractionated and analysed by SDS-PAGE. Appropriate fractions were pooled, flash frozen and stored.

2.2.9 *in vitro* phosphorylation

To produce stoichiometrically phosphorylated FhaA or MviN, the substrates were diluted to 10-20 µM and the buffer changed to 50 mM NaCl, 20 mM Tris pH 8.0, 5 mM $MgCl_2$, 1 mM $MnCl_2$ and 0.5 mM TCEP by dialysis. PknB kinase, also in the same buffer, was added to a final concentration of 1-2 µM. Typical reaction volumes were 1-2 ml. Upon addition of ATP to a final concentration of 100 µM, the reaction mixture was incubated for 48-96h at room temperature. In the case of FhaA, the kinase was subsequently removed by size-exclusion chromatography using a Superdex S75 column. In the case of MviN, unphosphorylated MviN as well as the kinase was removed by ion-exchange chromatography using a Resource Q column (previous section). Following ion-exchange, the sample buffer was re-adjusted to 150 mM NaCl, 20 mM Tris pH 8.0 and 0.5 mM TCEP using a NAP-5 column.

2.3 Biochemical and biophysical techniques

2.3.1 γ-^{32}P ATP kinase assay

Approximately 20 µg of substrate and 10 µg of PknB kinase were incubated in 100 µL of buffer consisting of 50 mM NaCl, 20 mM Tris pH 8.0, 5 mM $MgCl_2$, 1 mM $MnCl_2$, 0.5 mM TCEP and ~2 µCi of γ-^{32}P-ATP for one hour at room temperature. Unincorporated γ-^{32}P-ATP was removed using a G25 spin column (GE Healthcare) and the proteins separated by SDS-PAGE as described in section 2.2.2. Typically, 10 µl of each reaction were loaded onto each lane. The gel was fixed, stained with instant Blue (Expedeon) and dried before exposure to BioMax light film (KODAK).

2.3.2 Limited proteolysis

Sequencing grade porcine trypsin and chymotrypsin (Promega) were used for all limited proteolysis experiments. 1 µg of protease was incubated at room temperature with 500 µg of substrate in a total volume of 500 µl. For SDS-PAGE analysis, 10 µl aliquots of the reaction mixture were taken at various time points and mixed with 5 µl of LDS loading buffer (Invitrogen) and 1 µl TCEP before being boiled for 5 min to stop the reaction. For analysis by mass spectrometry, 40 µl aliquots were added to 10 µl of 2% trifluoroacetic acid (TFA). Typical time points were 0, 1, 5, 10, 20, 30 and 60 minutes.

2.3.3 Reverse phase-High performance liquid chromatography (RP-HPLC)

For phospho-site mapping, 80 µg of both phosphorylated and unphosphorylated $FhaA^{367-527}$ were subjected to proteolysis using modified trypsin (Promega) at a trypsin/substrate ratio of 1:10 w/w for 2 hours at 37 °C in gel filtration buffer. Urea was added to a final concentration of 8 M and the reaction acidified by the addition of acetic acid to a final concentration of 100 mM. Insoluble material was removed by centrifugation and samples loaded onto a C-18 narrow bore HPLC column (ZORBAX) equilibrated in 5% acetonitrile and 0.05% trifluoroacetic acid. Peptides were eluted using a gradient of 5-95 % acetonitrile, 0.05 % trifluoroacetic acid over 60 minutes at 60°C. Fractions containing peptides were lyophilised and frozen at -20 °C.

2.3.4 Electrospray ionisation mass spectrometry (ESI-MS)

ESI-MS was used to verify the identity of recombinant proteins and, where applicable, their phosphorylation status. In addition, peptides purified from RP-HPLC were also analysed by ESI-MS. Proteins were de-salted using a C_4 Zip-Tip (Millipore). Samples were acidified by adding TFA to a final concentration 100 mM and applied to the Zip-Tip. The Zip-Tip was then washed with a buffer of 0.5% acetic acid and 5% acetonitrile (ACN). Protein was eluted from the matrix in 0.5% acetic acid and 60% ACN. For peptide analyses, lyophilised peptide was redissolved in 50% ACN and 20 mM acetic acid and flow injected directly. Typically, 100-300 pmol of sample at a concentration of 1 µM were used. Molecular weights were determined on a MicroTOFQ (Bruker) mass-spectrometer calibrated using myoglobin. Spectra were acquired and deconvoluted using the MaxEnt program (Bruker) by Dr Steve Howell (Mass spectrometry facility, NIMR).

2.3.5 Multi Angle Laser Light Scattering (MALLS)

Samples of purified protein were applied at a flow rate of 0.5 ml/min to a Superdex 75 10/300 GL column (GE Healthcare) pre-equilibrated in 150 mM NaCl, 20 mM Tris pH 8.0 and 0.5 mM TCEP mounted on a Jasco high-performance liquid chromatography (HPLC) system. A DAWN-HELEOS multiangle laser light–scattering detector (Wyatt Technology Corp., Santa Barbara, CA) recorded the scattered light intensity of the column eluent at sixteen angles. The protein concentration of the eluent was determined from the change in the refractive index (dn/dc = 0.186) detected by an OPTILAB-rEX differential refractometer equipped with a Peltier temperature-regulated flow cell, maintained at 25°C (Wyatt Technology Corp., Santa Barbara, CA). The weight-averaged molecular weight of material contained in the chromatographic peaks was determined with the ASTRA v5.1 software package (Wyatt Technology Corp., Santa Barbara, CA)

In a SEC-MALLS experiment, a protein solution to be analysed is injected into a size exclusion column, where its constituents are separated by size. The eluted fractions are then analysed by three types of detectors. First there is a detector that monitors the UV_{280} absorbance to provide an elution profile of the protein solution. Secondly, a differential refractometer measures the difference in refractive index between the column eluent and a buffer sample. As the differential index of refraction is highly constant among proteins, a refractometer is a useful device of measuring the total concentration of protein in a sample. As a consequence, the total concentration of protein can be calculated for each point of the elution profile. The third detector is a multi-angle laser light scattering detector, in which a laser is directed at the solution and the intensity of the scattered light detected at 16 angles. As long as the wavelength of the incident radiation is considerably larger than the particle size (Rayleigh scattering), the light scattered at each angle is independent of the particle size and shape. While detection at one angle would be sufficient to calculate the total scattering intensity, the scatter is recorded at a number of angles to provide increased signal. The total scattering intensity however is related to the molecular weight of the particles in solution. The individual scattering events of many small molecules are out of phase and thus a large proportion of scattered light will be lost due to negative interference. If those molecules oligomerise, their combined scattering becomes coherent and the scattering intensity increases. By measuring the total scattering of a protein solution and dividing through the total protein concentration (via the refractive index), one can determine the average molecular weight of the protein particles at every point in the elution profile.

2.3.6 Nuclear magnetic resonance

NMR spectra were acquired at 25 °C on Varian Inova 600 and 800 and Bruker Avance 600 and 700 MHz spectrometers, all equipped for $^{1}H/^{15}N/^{13}C$ triple resonance experiments. Purified proteins were dialysed into 150 mM NaCl, 20 mM Tris pH 7.0, 3 mM NaN_3 and 5% D_2O. Typical protein concentrations were 1-2 mM. $^{1}H/^{15}N$ Heteronuclear Single Quantum Coherence (HSQC) spectra were acquired to assess sample quality and for titration purposes. The following three dimensional experiments were performed with $^{15}N/^{13}C$ double labelled FhaA FHA domain to enable back-bone assignment: HNCACB, CBCA(CO)NH, HN(CA)CO, HNCA, and HNCO. All spectra were acquired and processed by Dr Geoff Kelly (MRC Biomedical NMR Centre). Processed spectra were analysed using CARA software (Computer-Aided Resonance Assignment).

2.3.7 Isothermal titration calorimetry (ITC) (summary)

ITC experiments were performed using an ITC-200 microcalorimeter (MicroCal Inc.) and analysed using ORIGIN software as per the manufacturer's instructions assuming a single-site binding model. Proteins and phospho-peptides were dialysed into 150 mM NaCl, 20 mM Tris pH 8.0 and 0.5 mM TCEP and their concentrations determined using UV/visible spectrophotometry. Typical concentrations were 20-80 µM for the cell and 150-800 µM for the syringe. Titrations consisted of 20 injections of 2 µl each into an active cell volume of 200 µl. Measurements were performed at 20 °C unless stated otherwise.

2.3.8 GST pull-down assays

GST pull-down assays were performed to identify biological binding partners of FhaA. Cell-free extracts of *M. tuberculosis* cultures were provided by Ms Debbie Hunt (NIMR). These extracts had been thoroughly homogenized and filtered to remove all intact bacteria and allow for handling in a standard Containment 1 laboratory. Proteins were dialysed into 100 mM NaCl, 20 mM Tris pH 8.0 and 0.5 mM TCEP. 50 µl of each bait protein at 100 µM concentration was incubated with 30 µl of glutathione Sepharose 4B (GE Healthcare) pre-equilibrated in the same buffer. Unbound protein was removed using 3 washes with 400 µl of buffer each. Each time, the buffer was added with force to disturb the bed of resin, the mixture centrifuged for 30 sec at 500g and the supernatant removed. 100 µl of cell free extract were then added and incubated for 2h at 4 °C. Unbound material was removed again by 3 washes with 400 µl of buffer as

before. 20 µl of beads were mixed with 10 µl of LDS sample loading buffer (Invitrogen) and 2 µl of TCEP. 20 µl of these mixtures were loaded onto each SDS-PAGE gel lane.

2.3.9 Nuclease assay

From UV spectrometric analysis, it was thought that the contaminants separated from the FhaA constructs during heparin ion exchange chromatography were mainly nucleic acids. A nuclease assay was performed to determine their properties. 500 µL of the flow through from the heparin ion exchange chromatography step was mixed with 500 µl of proprietary Phenol:Chloroform pH 6.6 (Ambion) to remove protein contaminants and isolate nucleic acids. After 3 cycles of phenol:chloroform extraction, the aqueous phase was washed 3 times with 500 µl of chloroform to remove residual phenol. Phenol contamination could be monitored due to its UV absorbance at 270 nm and its characteristic odour. Ethanol precipitation was performed to concentrate the unknown nucleic acids (XNA). 1/10 volume of 3M NaAc pH 4.8 was added to the solution followed by 5/2 volumes of neat ethanol. The mixture was incubated at -20 °C overnight and centrifuged at 25000g for 10 min at 4 °C. The supernatant was removed and the pellet dried and redissolved in 100 µl of H_2O.

As the quantities of recovered nucleic acids were insufficient for visualisation by ethidium bromide staining, they were 5'-labelled with ^{32}P using polynucleotide kinase (PNK) (Ambion). The 5' phosphates were removed using alkaline phosphatase (Ambion) before labelling with γ-^{32}P ATP. Two reactions were set up (tables 2.9 and 2.10).

Reagent	**Amount**
XNA	40 µl
Phosphatase buffer (Ambion)	5 µl
Alkaline phosphatase (5 U/µl)	1 µl
H_2O	4 µl

Table 2.9. Reaction mix for 5'-dephosphorlation reaction

Following dephosphorylation for 1 h at 37°C, the phosphatase was heat-inactivated at 95 °C for 5 min. The heat-inactivated reaction was taken as the substrate for the second enzymatic step. The end labelling reaction was set up as indicated in table 2.10 and incubated at 37°C for 1 h. Unincorporated nucleotides were removed using G25 spin columns (GE healthcare)

Reagent	Amount
5'-dephosphorylated XNA	50 µl
PNK buffer (Ambion)	6 µl
PNK (10U/µl)	3 µl
γ-^{32}P ATP (2 µCi/µl)	3 µl

Table 2.10. Reaction mix for 5'- labelling

For characterisation, the 5'-labelled XNA was incubated with different nucleases. 10 µl aliquots of XNA, along with RNA and DNA controls, were incubated with 2.5-3 U each of benzonase (Novagen), RNase A and RNase T1 mix (Ambion) and DNase Baseline-ZERO (epicentre) for 1 hour at room temperature in the reaction buffers provided by the respective manufacturers.

Reagent	amount
19:1 acrylamide/Bis mix (Biorad)	16 ml
Urea	21 g
10x TBE	5 ml
Ammonium persulphate (10% w/v)	500 µl
TEMED	50 µl
H_2O	to 50 ml

Table 2.11. Recipe for a 50 ml, 12.5% (w/v), denaturing polyacrylamide gel

The nuclease treated nucleic acids, along with untreated nucleic acids, were then run in 1X TBE (Sigma Aldrich) on a 12.5% acrylamide gel (table 2.11) at 300V for 90 minutes. The gel was fixed and dried before exposure to BioMax light film (KODAK).

2.3.10 Electrophoretic mobility shift assay (EMSA)

Oligonucleotides, 5'-labelled with γ-^{32}P ATP using PNK as described above, were incubated with FhaA$^{1-130}$ for 1 h in 50 mM NaCl and 10 mM Tris pH 8.0. The oligonucleotide concentration was kept constant at 100 nM and the protein concentration was either 0, 1 or 10 µM. The mixtures were then run on a 12.5% non-denaturing polyacrylamide gel and the bands visualised by autoradiography. A shift in the electrophoretic mobility of the oligonucleotide in the presence of the protein would indicate an interaction between the 2 partners.

2.3.11 Protein crystallisation

Protein crystallisation experiments were set up using commercial crystallisation screens (Hampton, Qiagen) as well as screens customised for specific expression constructs. Initial screening was carried out by Dr Lesley Haire and Miss Roksana Ogrodowicz (both NIMR) using an ORYX liquid handling robot (Douglas Instruments) and commercial crystallisation screens. Upon encountering initial crystal hits, optimisation screens were set up around the initial hits to improve crystal size and quality. For the crystal trials set up with the aid of the ORYX robot, the sitting drop vapour diffusion method with a drop size of 100 nl protein solution and 100 nl well solution was used against 60 µl of well solution. The most commonly used plate for these experiments was the MRC 2 Well crystallisation Plate (Swissci). Optimisation trials were set up by hand and used the hanging drop method with a drop size of 1 µl + 1 µl against 1ml of well solution and the EasyXtal (Qiagen) crystallisation plate. In addition to the optimisation of crystallisation conditions, micro-seeding of initial crystal hits was used to improve crystal quality. Details of each crystallisation process can be found in the respective result chapters.

2.3.12 X-ray crystallography (summary)

Crystals were transferred to an appropriate cryo-protectant for 1 minute before being flash frozen in liquid nitrogen. Data were collected at beamlines IO3 and IO4 at the Diamond Light Source using a Pilatus detector (Swiss Light Source Detector Group). Images were integrated, indexed and scaled using the HKL suite of software. Phases were determined, where applicable, by molecular replacement using the CCP4 program Phaser. Rounds of model building with Coot and refinement with Phenix refine (PHENIX) were used to improve models.

2.3.13 Circular Dichroism (CD)

Purified proteins were diluted to ~0.15 mg/ml in 150 mM NaCl, 20 mM Tris pH 8.0 and 0.5 mM TCEP. Spectra were recorded on a Jasco-J715 spectropolarimeter scanning from 200 to 250 nm at a rate of 100 nm/min with 2 nm bandwidth. Thermal denaturation was measured over a temperature range 20-80 °C at a rate of 2 °C/min and a wavelength of 202 nm. Quartz cuvettes with 2 mm path lengths were used throughout. Final spectra, which were the averages of 10 discrete scans, were analysed using Specpro (Dr Steve Martin, NIMR). Spectra were acquired and processed by Dr Steve Martin (NIMR)

2.3.14 Oriented peptide array libraries (OPAL arrays)

Peptides to be studied were synthesised and spotted onto a membrane by Nicola O'Reilly (CRUK, LRI). The membrane was then washed in 150 mM NaCl, 50 mM Tris pH 7.6 and 0.1% (v/v) tween-20 (TBST). Subsequently, the membrane was blocked with TBST and 5% milk at 4°C overnight. Following another TBST wash, the membrane was then incubated with 300 µg PIH1D1$^{1-180}$ in TBST and 5% milk at 4°C overnight. To remove unbound protein, the membrane was again washed with TBST and 5% milk. A mouse monoclonal antibody against PIH1D1 (abcam) at a dilution of 1:1000 was incubated with the membrane at 4°C overnight and the membrane again washed with TBST. A secondary antibody, a mouse anti- IgG antibody conjugated to horseradish peroxidase, was incubated with the membrane and unbound antibody removed with a TBST wash. Bound secondary antibody, and hence bound PIH1D1 was visualised using the Pierce enhanced chemiluminescence (ECL) femto system (Thermo Scientific) and recorded on photographic film (KODAK). As a final step, the location of each peptide spot was identified using Ponceau stain.

3 Biophysical methodologies

3.1 Isothermal titration calorimetry (ITC)

3.1.1 Overview

Isothermal titration calorimetry experiments were carried out to determine binding affinities and stoichiometries of several protein-protein and protein-peptide interactions. ITC is a physical technique that measures the heat changes of binding events in a titration. Typically, several quantities of titrant are injected into the analyte and the heat of binding is recorded for each injection. From the first derivative of these heat changes, the binding constant can be calculated.

3.1.2 Theory

The theory behind ITC was developed and commercialised by Wiseman et al. in 1989 [163]. The so-called Wiseman Isotherm describes how the heat of binding of an injection is dependent on the total amount of ligand present. Indyk et al. ,1998 provides a more comprehensive description of the derivation of the isotherm [164]. An equation needs to be derived that describes the amount of complex formed for each consecutive injection as a function of known quantities such a total ligand and total macromolecule concentration.

From the second law of thermodynamics it can be derived that for any equilibrium at constant temperature:

$$\Delta G = -RTlnK = \Delta H - T\Delta S \quad (1)$$

where K is the equilibrium constant, R the gas constant and T the temperature

In the special case of an equilibrium being a binding event with 1:1 stoichiometry between a macromolecule M and ligand L

$$M + L \rightleftharpoons ML \quad (2)$$

The equilibrium constant K becomes binding constant K_B which can be expressed as

$$K_B = \frac{[ML]}{[M][L]} \tag{3}$$

If $[M_T]$ is the total concentration of macromolecule and $[L_T]$ the total concentration of ligand it applies that

$$[M_T] = [M] + [ML] \tag{4}$$

and

$$[L_T] = [L] + [ML] \tag{5}$$

equation **(3)** can be re-arranged to

$$[ML] = K_B[M][L] \tag{6}$$

using **(6)** to substitute for [ML] in (4) and (5) produces

$$[M_T] = [M] + K_B[M][L] \tag{7}$$

and

$$[L_T] = [L] + K_B[M][L] \tag{8}$$

or

$$[L_T] - [L] = K_B[M][L] \tag{9}$$

solving for [M] one produces

$$[M] = \frac{[M_T]}{1+K_B[L]} = \frac{[L_T]-[L]}{K_B[L]} \tag{10}$$

$$\frac{[M_T]}{1+K_B[L]} = \frac{[L_T]-[L]}{K_B[L]} \tag{11}$$

can be re-arranged into the quadratic equation

$$[L]^2 + \left([M_T] - [L_T] + \frac{1}{K_B}\right)[L] - L_T = 0 \tag{12}$$

which can be solved for [L] using the standard quadratic equation

$$ax^2 + bx + c = 0 \tag{13}$$

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \tag{14}$$

as [L] may not be negative, one solution can be discarded so

$$[L] = \frac{[L_T] - [M_T] - \frac{1}{K_B} + \sqrt{\left([L_T] - [M_T] - \frac{1}{K_B}\right)^2 + 4[L_T]}}{2} \tag{15}$$

which provides an equation to express [L] as a function of total concentrations, which are known, and the binding constant K_B. For clarity, the substitutions $X_R = [L_T]/[M_T]$ and $r = 1/(K_B[M_T])$ will be introduced, and both sides of the equation will be divided by $[M_T]$. In many ITC annotations, X_R is called the molar ratio.

$$\frac{[L]}{[M_T]} = \frac{X_R - 1 - r + \sqrt{(X_R + 1 + r)^2 - 4X_R}}{2} \tag{16}$$

its derivative can be expressed as

$$\frac{d[L]}{dX_R} = [M_T]\left(\frac{1}{2} + \frac{-1 + X_R + r}{2\sqrt{(X_R + 1 + r)^2 - 4X_R}}\right) \tag{17}$$

As the calorimeter measures the 1st derivative of heat produced with respect to the total amount of ligand present $[L_T]$, the derivative $d[L]/d[L_T]$ needs to be determined.

from the substitution for X_R it applies that

$$\frac{dX_R}{d[L_T]} = \frac{d\frac{[L_T]}{[M_T]}}{d[L_T]} = \frac{1}{[M_T]} \tag{18}$$

therefore

$$\frac{d[L]}{d[L_T]} = \frac{d[L]*dX_R}{dX_R*d[L_T]} = \frac{d[L]}{dX_R*[M_T]} \tag{19}$$

substituting with equation **(17)** produces

$$\frac{d[L]}{d[L_T]} = \frac{1}{2} + \frac{-1+X_R+r}{2\sqrt{(X_R+1+r)^2-4X_R}} \tag{20}$$

For the whole titration, the cumulative heat change can be expressed as

$$q = \Delta H * V * [ML] \tag{21}$$

where q is the heat produced, and V is the reaction volume. The derivative with respect to $[L_T]$ is

$$\frac{dq}{d[L_T]} = \Delta H * V * \frac{d[ML]}{d[L_T]} \tag{22}$$

since $[ML] = [L_T] - [L]$

$$\frac{d[ML]}{d[L_T]} = \frac{d[L_T]}{d[L_T]} - \frac{d[L]}{d[L_T]} = 1 - \frac{d[L]}{d[L_T]} \tag{23}$$

$$\frac{d[ML]}{d[L_T]} = 1 - \left(\frac{1}{2} + \frac{-1+X_R+r}{2\sqrt{(X_R+1+r)^2-4X_R}}\right) = \frac{1}{2} + \frac{1-X_R-r}{2\sqrt{(X_R+1+r)^2-4X_R}} \tag{24}$$

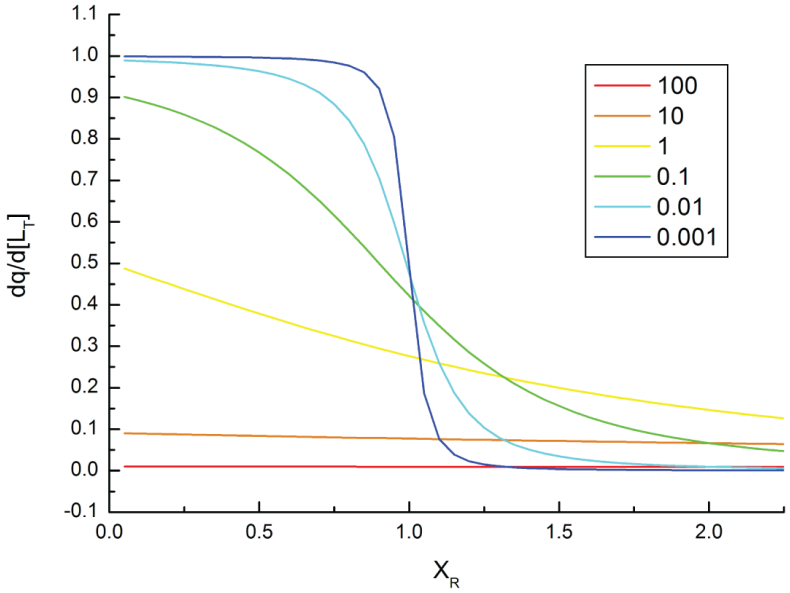
substitution into equation **(22)** produces

$$\frac{dq}{d[L_T]} = \Delta H * V * \left(\frac{1}{2} + \frac{1-X_R-r}{2\sqrt{(X_R+1+r)^2-4X_R}}\right) \tag{25}$$

This equation describes the derivative of heat product q, with respect to $[L_T]$, expressed using ΔH, V, $[L_T]$, $[M_T]$ and K_B. See figure a 3.1 for a graphical representation of $dq/d[L_T]$ against X_R at different values of r. While ΔH and K_B are unknown to the experimenter, V, $[L_T]$ and $[M_T]$ are known. After selection of an appropriate baseline and substractions of heats of dilution, the heat change for each injection is calculated. Furthermore, values of $[L_T]$ and $[M_T]$ are corrected for the dilution effect caused by cumulative injections. Using floating values of ΔH and K_B, the isotherm is described. Numerous iterations are performed until the least squares fit for the isotherm cannot be improved upon. The values which generate the best fit are taken as the best estimates for ΔH and K_B. Further binding parameters can be calculated using equation **(2)**. As this method relies on extrapolation to determine the plateau at which the curve would have originated, which effectively is ΔH, it is clear from the graph that values of r larger than 0.1 will not lead to accurate estimates of binding parameters.

It should be noted however that operation of a modern microcalorimeter does not require knowledge of the theory behind the measurements. Most calculations are performed within the ORIGIN software.

A



B

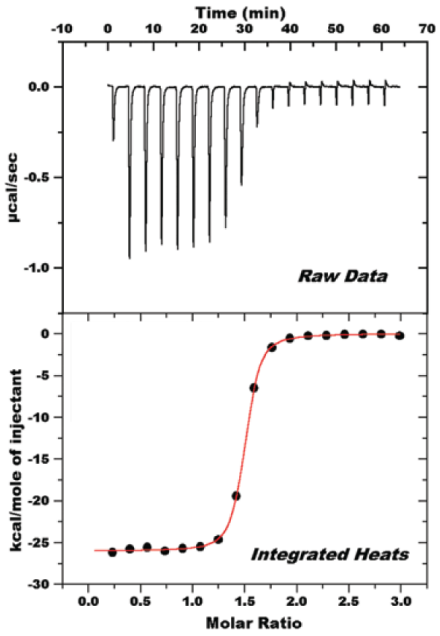


Figure 3.1. Theoretical isotherm curves. (A) Plots of theoretical isotherms dq/d[L_t] against X_R for various values of r. The legend indicates which colour refers to which value of r. (B) Example of practical ITC data for an exothermic binding event. Top: differential heat between reference and sample cell. Every trough indicates an injection. Bottom: Integrated heats for each titration with the heat of dilution subtracted (black dots) and its least squares isotherm curve (red line).

3.1.3 ITC in practice

A schematic of the overall build of a calorimeter can be seen in figure 3.2. There are two cells of equal volume and properties, a reference cell and a sample cell, each within an isothermal jacket and thermostated at the desired temperature for the experiment. The space surrounding the cells is cooled slightly below the cell temperatures so that a constant amount of energy is needed to maintain the cell temperatures. The reference cell is filled with a liquid comparable to the one used in the sample cell. For most aqueous buffers, water is used in the reference cell. A syringe attached to a stirrer carries out the injections into the sample cell. If there is an exothermic binding event upon an injection, less energy from the calorimeter is needed to maintain the temperature of the sample compared to the reference cell. For an endothermic reaction, the opposite is the case. From the differences in energy needed, the total heat change inside the sample cell is calculated.

The strengths of ITC are that it is label-free and both analyte and titrant are free in solution. Furthermore, it allows us to determine the enthalpy (ΔH), entropy (ΔS), Gibbs free energy (ΔG) and stoichiometry of a binding event using one experiment.

One disadvantage of this technique is that it relies on a net change of heat upon binding, which for some interactions may not be sufficiently large. A tight interaction may not necessarily be accompanied by a large heat of binding as several events may contribute to this net heat, such as interactions formed between titrant and analyte and rearrangements of hydration shells, all of which may be exothermic or endothermic. As the heat capacity of a binding event is temperature dependent, one way to overcome this problem is to perform ITC experiments over a range of temperatures. However, the temperature ranges over which biological macromolecules such as proteins remain stable are rather small. Despite advances in instrument sensitivity, another inconvenience is that ITC requires considerable quantities of the substances to be analysed which may be a problem when working with recombinantly expressed proteins. Nonetheless, the fact that ITC works without the aid of labels and does not rely on the immobilisation of analytes has made it a popular and powerful technique to study interactions between biological macromolecules.

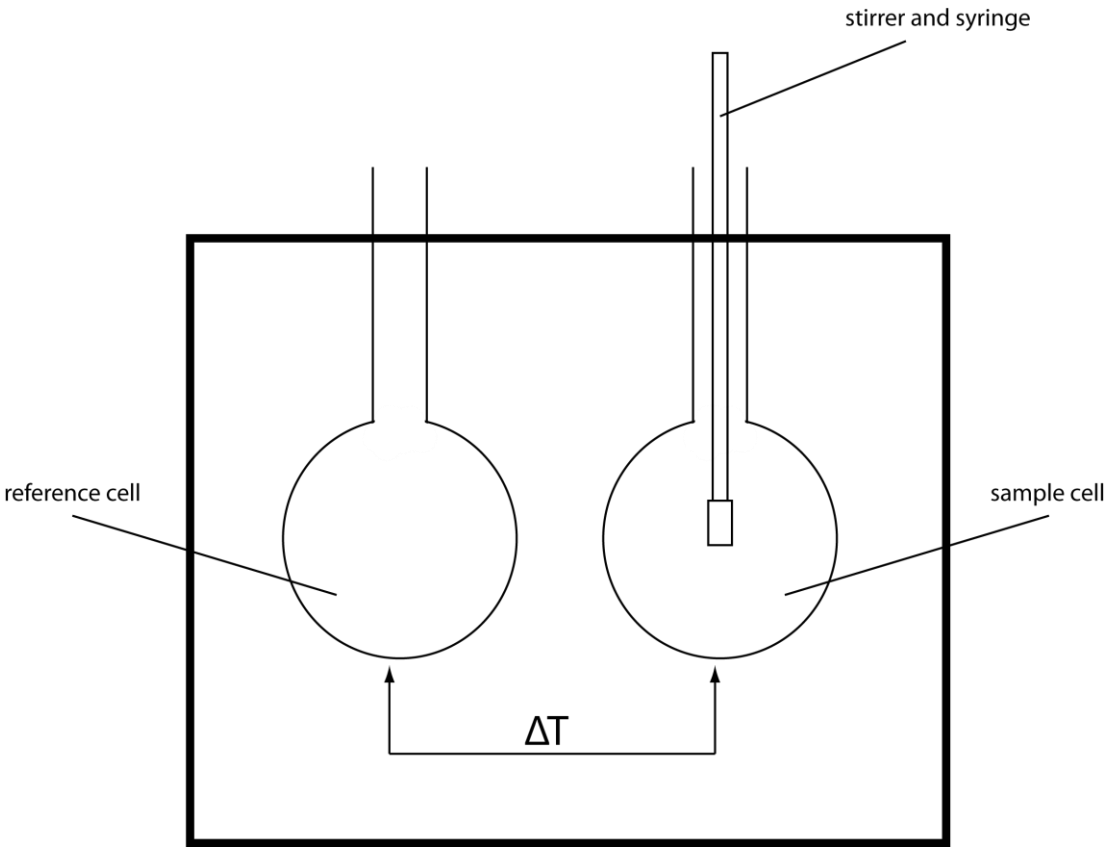


Figure 3.2. Schematic of an ITC calorimeter. Thermal jackets keep both sample and reference cell at a constant temperature. The environment around the cells is kept slightly colder than the cells so a constant energy input is required. Upon injection of a titrant, the energy required to thermostat the sample cell may increase or decrease depending on whether an endothermic or exothermic binding event takes place. This energy differential is then recorded to determine the energy released after every injection

The ITC experiments in this study were carried out using an ITC 200 microcalorimeter (MicroCal). In this instrument, the active volumes of reference and sample cells are 200 µl each and titrations consisted of 20 injections of 2 µl each. The temperature difference between the cell and its surroundings was adjusted to require a constant energy output of 5 µcal/sec to maintain equilibrium. Despite being discouraged by IUPAC (International Union of Pure and Applied Chemistry), non-SI units such as calories remain the standard in the field of calorimetry [165]. Concentrations of analyte and titrant were chosen so that ideally, $0.01 < r < 0.05$ and final $X_R \approx 2.0$. An analyte concentration that is between 20 and 100 times higher than the expected equilibrium constant and a final 2-fold excess of titrant are desirable as they produce a sigmoidal curve that allows for an accurate estimation of the binding parameters. Limitations in available protein amounts and protein solubility meant that some titrations were performed at lower concentrations.

3.2 X-ray crystallography

3.2.1 Crystals

One advantage X-ray crystallography has over techniques such as electron microscopy, where the signal is recorded for one particle at a time, is that it makes observations not of a single molecule, but of a crystal which contains a vast number of molecules which greatly amplifies the signal. Even a comparatively small crystal of a biological macromolecule, e.g. a 50 µm cube of a 40 kDa protein, contains ~10^{13} copies (assuming a protein density of 1.4 g/cm^3 and 50% solvent content). A crystal can be described as a motif, which is repeated in all three dimensions. Crystals commonly contain a vast number of atoms, but knowledge of the motif, and the geometrical relationship between the individual motifs, is sufficient to determine the position of every atom in the entire crystal. When discussing the geometrical relationship between motifs, i.e. the crystal lattice, as opposed to the contents of the motif itself, it is common designate an arbitrary, but fixed point called the lattice point. The spatial relationship between different lattice points is the same as between different motifs. The different arrangements which the lattice points can take are classified by the symmetries they contain. Following the determination of the lattice geometry, a unit cell is chosen by the crystallographer to represent the minimal repeating unit of the crystal. As these unit cells, which can be imagined as small “boxes”, need to tessellate in three dimensions, they can take the shape of the following polyhedrons: cubes, hexagonal prisms, rectangular prisms with and without

square bases, rhombohedrons and parallelepipeds with and without a rectangular face (from highest symmetry to lowest). It should be noted that any arrangement of lattice points can be described by a unit cell that is a parallelepiped, but that some arrangements contain higher symmetries. The lattice in figure 3.3, simplified to 2 dimensions, can be described as a lattice with a parallelogram as the unit. Such a lattice where each unit cell corresponds to one lattice point is called a primitive lattice. In the same arrangement of lattice points the unit cell can also be described as a rectangle with one lattice point at the corners and one additional lattice point at the centre of the unit cell. This type of lattice, whether in two or three dimensions, is classified as an I (inner) or body centred lattice. Similarly, some types of unit cells in three dimensions may have additional lattice points at their faces. In C-centred lattices there are additional lattice points on two opposing faces of the unit cell and F-centred lattices have lattice points at the centres of all six faces. In a three dimensional primitive lattice, there is one lattice point at each of the 8 corners of the parallelepiped. Each lattice point though is shared between 8 unit cells that the number of lattice points per unit cell is one (8 * 1/8 = 1). In a body-centred lattice, the presence of an additional lattice point increases the number to two. From the 7 different polyhedrons available, 14 different lattices can be constructed. While the symmetry operations relating adjacent unit cells are limited to translations, several identical motifs, called asymmetric units can be related by translations, rotations and reflections to make up the unit cell. These symmetry operations within the unit cell are called point groups. For crystals containing only one type of enantiomer, such as protein crystals made up entirely of the naturally occurring L-amino acids, no reflections are allowed. As a consequence, only rotations are allowed of which there are proper rotations around a fixed point and screw rotations, where a rotation is accompanied by a translation along the axis of rotation. The number of point groups with which each of the 14 lattices can be combined is dependent on the symmetry of the respective unit cells. A cubic lattice for example will have more rotational symmetries than a lattice whose unit cells are parallelepipeds. In total, there are 65 combinations, or space groups, for protein crystals. For crystals where reflections are permitted, this number increases to 230.

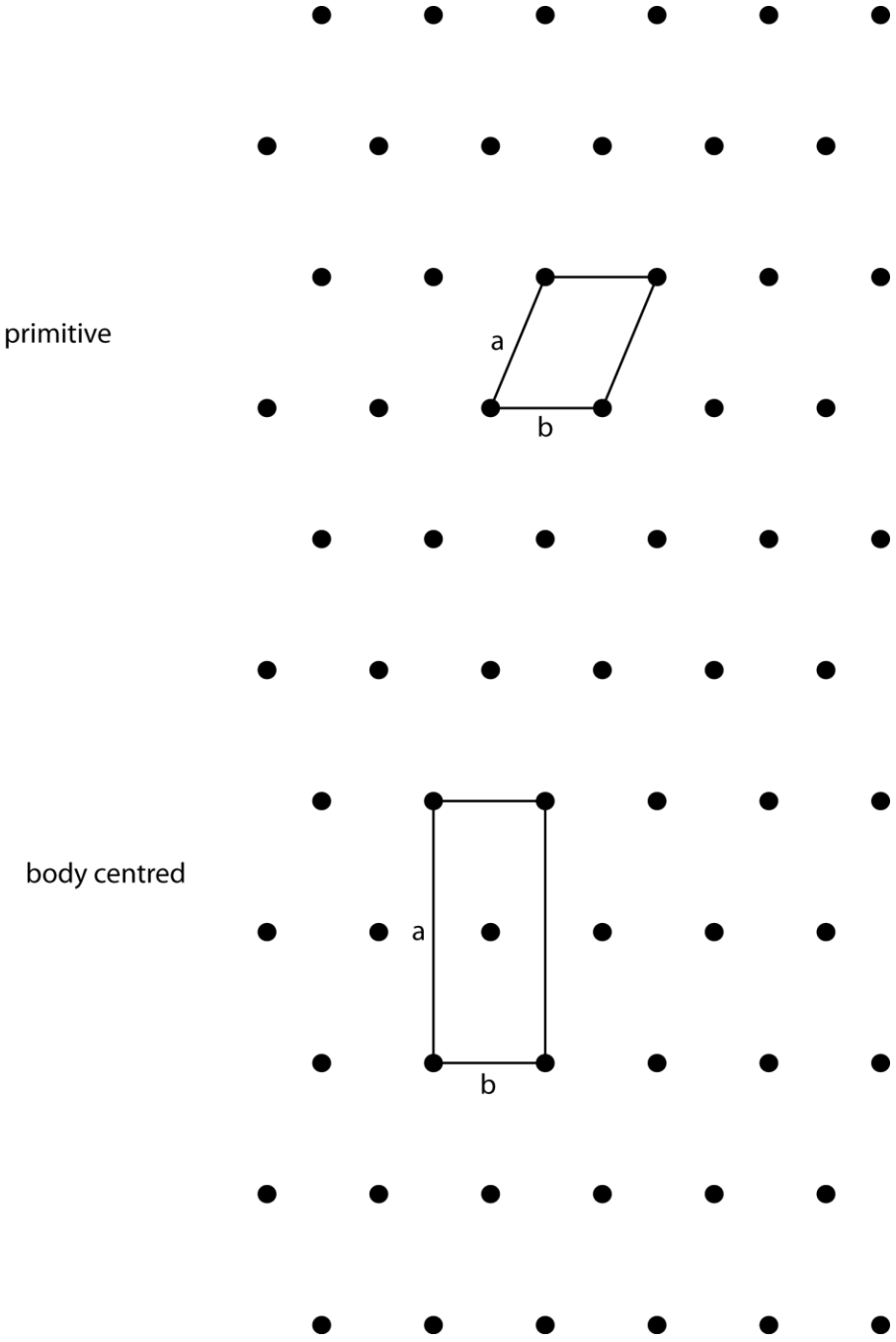


Figure 3.3. Schematic of a crystal lattice: In the presented crystal lattice, the unit cell is defined by the vectors a and b. A primitive unit cell does not display the symmetry inherent to the lattice, instead a body-centred unit cell is chosen which represents the rectangular nature of the lattice.

3.2.2 Diffraction theory

When a monochromatic X-ray beam is directed at a crystal, a diffraction pattern is produced. Since the wavelengths of X-rays have roughly the same size as the interatomic distances of molecules (~1 Å), and therefore the spacing of the crystal planes, they generate strong diffraction patterns at atomic resolution. The diffraction pattern is produced exclusively by the electrons in the crystal as they scatter electro-magnetic radiation to the strongest extent. Neutrons carry no charge and therefore do not interact with the radiation. Electrons and protons carry charge of equal magnitude, but since the scattering power is also inversely related to the square of the mass of the scattering particle, electrons scatter more intensely by a factor of ~3 million. Most scattered X-rays are annihilated by negative interference, but in some cases as defined by Bragg's law (see figure 3.4), constructive interference peaks are created. In the case of a diffraction pattern being generated by a crystal, every spot of the diffraction pattern refers to one interference peak. In theory, one molecule would be sufficient to generate an X-ray interference pattern from which the structure of the molecule could be reconstructed. For a measurable pattern however, the signal needs to be amplified. Furthermore, a crystal does not produce a continuous interference pattern, but a diffraction pattern with discrete peaks corresponding to the positions for which Bragg's Law is satisfied. The symmetry of the crystal means that each of the vast numbers of unit cells contributes to the same diffraction pattern. One could describe a crystal as a three-dimensional, periodic function of electron density through real space with the diffraction pattern being its Fourier transform. It should be noted that for every spot, or reflection, Bragg's law needs to be satisfied for all atoms in a unit cell. Every atom contributes to every diffraction spot and the phase of every contributing scattered X-ray is identical. The planes of a crystal, which define the intensity and phase of diffraction spots can be described by their Miller indices. The Miller indices describe any given crystal plane using the three orthogonal planes h, k and l. To collect reflections for planes in every direction, the crystal needs to be rotated through the beam and a diffraction image is collected for every increment, i.e. 0.5°. The positions of the reflections are determined by the unit cell dimensions and the space group of the crystal. The intensity provides information about the contents of the unit cell. A reflection resulting from crystal plane hkl can be described mathematically using the structure factor equation

$$F_{hkl} = \sum_{j}^{n} f_j e^{2\pi i(hx+ky+lz)}$$

For a unit cell with n atoms where j is the jth atom, xyz are the fractional coordinates of each atom in the summation and f_j is the scattering factor of atom j. The derivation of this equation from first principles however would exceed the scope of this brief chapter. For a comprehensive description of crystallography and diffraction theory, see Sherwood & Cooper, 2010 [166].

3.2.3 The Phase problem

The structure factor is a complex number with contributions from both the intensity and phase of the reflection. The intensity can be measured easily, but it is impossible to measure the phase of a reflection directly. The problem is not mathematical, but technical. Coherent X-ray beams are difficult to produce but more importantly, it is hard to imagine an X-ray detector with a response time better than the frequency of X-rays and positioning of such device with an error smaller than the wavelength of this radiation. Without the phase information, the diffraction pattern cannot be used to relate to the real-space electron density. For a small molecule, phases can be calculated theoretically from the expected structures determined using the valence shell electron pair repulsion (VSEPR) rules. For biological macromolecules with several thousand atoms, this is not feasible. There are three commonly used methods to overcome this “phase problem”. They are multiple isomorphous replacement, anomalous dispersion and molecular replacement (MR). As MR is the only technique relevant to the results presented here, only this method will be introduced in detail.

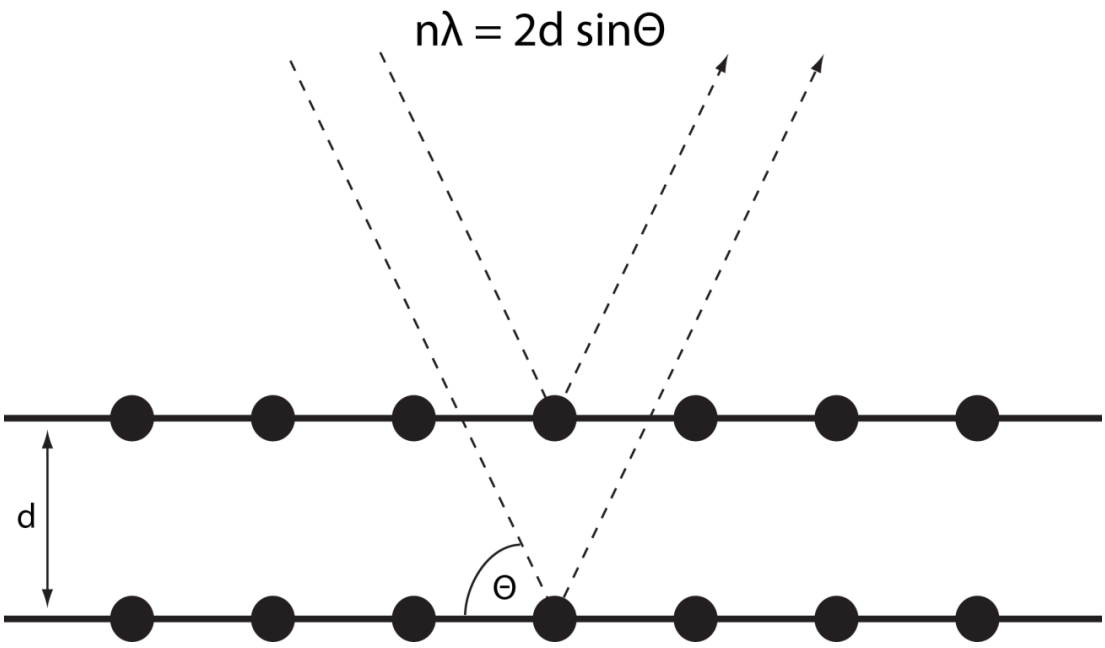


Figure 3.4. Schematic of Bragg's law. For constructive interference, the differential distances travelled by scattered X-rays of different crystal planes need to be a multiple of the radiation wavelength. The scattered waves need to be in phase at the reflection. The relationship nλ = 2d sinΦ, where n is any positive integer and λ is the wavelength of the radiation, needs to be obeyed to ensure the differential path is appropriate.

3.2.4 Molecular Replacement

Molecular replacement (MR) relies on the coordinates of a related protein to determine the phases of a molecule with unknown structure. Even when the structures of two homologous proteins are highly similar, these two proteins rarely crystallise in the same unit cell. Therefore one cannot simply copy the phase information from a solved structure and insert these into an experimental data-set. To avoid using phases when comparing the solved and unsolved datasets, a Fourier transform of the reflection intensities only, a Patterson function, is used. The resulting map is called a Patterson map, which can be thought of as a map of all interatomic vectors in the molecule.

The intensity I of a wave is related to its amplitude by the equation,

$$I = |A|^2$$

Therefore the Patterson function in the three dimensions u, v and w is defined as

$$P_{uvw} = \sum_{hkl} |F_{hkl}|^2 e^{-2\pi i(hu+kv+lw)}$$

For molecular replacement, a model of the known structure is rotated in the unit cell of the unsolved dataset using a rotational function to find the correct orientation of the molecule in the unit cell. The Patterson maps of the search model positions are matched against a Patterson map calculated from the experimental data to find the position of the molecule in the unit cell. As intra-molecular vectors are independent of the position of the molecule, only these vectors are matched in the rotation function. To obtain a Patterson map for the search model with negligible inter-molecular vectors, the search model is placed into a large, artificial P1 unit cell and the vectors calculated from this lattice. In a second step, the search model, in the orientation determined by the rotation function, is moved around the real unit cell using a translation function which matches both intra-molecular and inter-molecular vectors to find the correct position of the molecule in the unit cell. Once a likely position is found, phases can be calculated from the search model and used to determine the real-space electron density map of the experimental data. An atomic model can then be built into the electron density. The program Phaser from the CCP4 suit of programs can be used to perform an automatic rotation search function followed by a translational search. One advantage of this technique is that it does not rely on derivatising the crystal using anomalous scatterers or heavy atoms as is the case for isomorphous replacement and anomalous dispersion. A disadvantage is that using a homologous model may bias the solution towards the search model. MR also places an upper limit to the number of

molecules in the unit cell. The amount of possible positions each search model may take in the unit cell increases exponentially with the number of molecules in the unit cell. Too many degrees of freedom decrease the chance of a solution being found.

3.2.5 Structure validation and refinement

The initial atomic models obtained using any phasing methods can usually be improved upon considerably. They represent a starting point from which the model can be refined. In the refinement process, the model is matched against both a database of known atomic bond angles and lengths such as the CCP4 monomer library, as well as the experimental data. These sets of restraints, experimental data and ideal bond angles and lengths, may lead to different solutions. The influence of each set of restraints on the refinement process must therefore be varied with the resolution of the diffraction data. Only if suggested by high resolution data may one depart from ideal bond angles and lengths in their model.

One of the most widely used tools to determine to determine the goodness of fit between the experimental data and the atomic model is the R-factor or R_{work}. It is defined as

$$R = \frac{\sum ||F_{obs}| - |F_{calc}||}{\sum |F_{obs}|}$$

Theoretical structure factors are calculated from the atomic model and compared to experimental factors. A perfect model would have an R_{work} of zero. Acceptable R_{work} values for well refined biological macromolecules are usually less than 0.3 for resolutions between 2 and 4 Å. One weakness of the R_{work} parameter is that it can be lowered not only by producing a correct model, but also by over-fitting data to match the noise in the experimental data. The same structure factors that are used in the R_{work} work calculation are also used to produce the model for the next round of model building. To avoid this problem, a randomly selected subset of reflections, usually 5%, is taken out of the refinement process. An R-factor is calculated as before using only these 5% of reflections. This parameter is called R_{free}. It indicates a goodness of fit with the experimental data that has not been biased by previous rounds of refinement.

For proteins there are several libraries of coordinate standards that the model can be assessed against. There is for example the CCP4 monomer library which contains a

list of all commonly found bond lengths and angles in proteins, as well as allowed side chain rotamers. Furthermore, the Ramachandran plot indicates a range of favoured, allowed and disallowed Φ and Ψ angles for a polypeptide.

Programs such as Phenix-refine can be used to match an atomic model against these data-bases. In addition, it can also calculate R_{work} and R_{free} values against experimental data and will check for other errors in the atomic model such as steric clashes.

3.2.6 Experimental set up

The X-rays used in crystallographic experiments are Bremsstrahlung; they are produced when high velocity electrons collide with a metal anode and decelerate or by the angular acceleration of electrons round a synchrotron ring. Synchrotron radiation has the advantage of being brighter by several orders of magnitude as well as having an easily tuneable wavelength. The beam of X-rays is passed through vertical and horizontal slits and a monochromator to produce a narrow, collimated X-ray beam of a very narrow band of wavelengths. In the IO3 beamline at the Diamond Light Source, which was used to acquire the data presented here, the beam is 70 µm by 20 µm in size with a wavelength of 0.98 Å, which refers to an energy of 12.7 keV per photon. Using an aperture, the flux is kept at $1.5 * 10^{12}$ photons per second and therefore an energy output of ~3 mW. Crystals are kept at cryogenic temperatures during exposure to minimise the amount of radiation damage suffered due to the X-rays. As preparation, the crystal is moved into a solution containing a cryo-protectant such as glycerol and then plunged into liquid nitrogen. From this point on the crystal is kept under liquid nitrogen to prevent re-annealing. During exposure, the crystal is moved into an X-ray beam, while also being in a stream of N_2 gas at 100K with the resulting diffraction pattern recorded by a PILATUS detector [167]. The sensitivity and response time of this detector allows the operator to rotate the crystal through the x-ray beam continuously and record images after every desired increment, for example every 1°.

4 Characterisation of the FhaA N-terminal domain

The role of FhaA and its phospho-dependent interaction with MviN have been described in section 1.4.5.5, but the effect of FhaA phosphorylation and the function of its N-terminal domain remain to be determined. From a ClustalX (www.clustal.org) sequence alignment of FhaA homologues of different mycobacterial species (figure 4.1), it is apparent that the C-terminal FHA domain of FhaA is highly conserved, the N-terminal domain slightly less so and that there is a linker rich in hydrophilic residues in between. The individual residues of the linker region are not conserved, but the almost complete absence of hydrophobic residues for a ~300 residue sequence rich in Gly, Pro and Tyr is conserved among *Mycobacteria*. In the more distantly related *Corynebacterium glutamicum*, this region is only ~100 amino acids long. A previous study characterising FhaA had shown that full length FhaA is not a soluble protein when expressed in *E. coli* (D. P. Patel, PhD Thesis 2006) [121]. An N-terminally deleted construct comprising the linker region and the FHA domain produced a soluble protein that is an *in vitro* substrate of the mycobacterial kinase PknB. Furthermore, the FHA domain had been shown to be essential for kinase activity, but was not a substrate itself [121, 127]. No attempts however had been made to characterise the N-terminal domain.

4.1 Designing expression constructs

At the commencement of the experiments described below, the N-terminal domain, named DUF3662 (domain of unknown function) displayed no apparent similarity to any known protein fold based on sequence homology. Experiments were carried out to determine the three-dimensional structure of DUF3662 to provide further insights into the function of the FhaA protein. As minimal globular domains without any flexible extensions are usually most amenable to crystallisation, the domain boundaries of DUF3662 needed to be determined. From an alignment of FhaA orthologues however, the boundaries of DUF3662 are not apparent (figure 4.1). As the amino acids at the N-terminus of FhaA show some conservation, the first amino acid of FhaA was assumed to also be the first residue of the DUF3662 domain. The position of the C-terminal boundary of DUF3662 was less clear from the FhaA sequence alignment. Sequence conservation decreases sharply after amino acid 130 (…PDVE), but there are still some hydrophobic amino acids C-terminal up to amino acid 155 (…VAPM). The C-

terminal boundary was therefore estimated to be somewhere between residues 130 and 160. While an expression construct containing an incomplete globular domain would most likely be insoluble, a construct containing more than a minimal domain may retain solubility and could provide a starting point for proteolytic experiments to determine domain boundaries experimentally. Taking the above information into consideration, two expression constructs containing DUF3662 and probably some part of the linker region, FhaA$^{1\text{-}157}$ and FhaA$^{1\text{-}166}$, were cloned into the pET49b vector to produce N-terminal GST-fusions. Both constructs were soluble and could successfully be purified from *E. coli* BL21 cultures. After cell lysis, GST-affinity purification and removal of the purification tag using 3C protease, the 260/280 UV absorbance ratio of the protein solution was >1, indicating contamination with nucleic acids. A heparin ion-exchange chromatography step was used to separate the nucleic acids from the protein. After this chromatography step, the 260/280 ratio was reduced to 0.5, consistent with the absorbance profile of a protein solution. In a final step, aggregates were removed using size-exclusion chromatography.

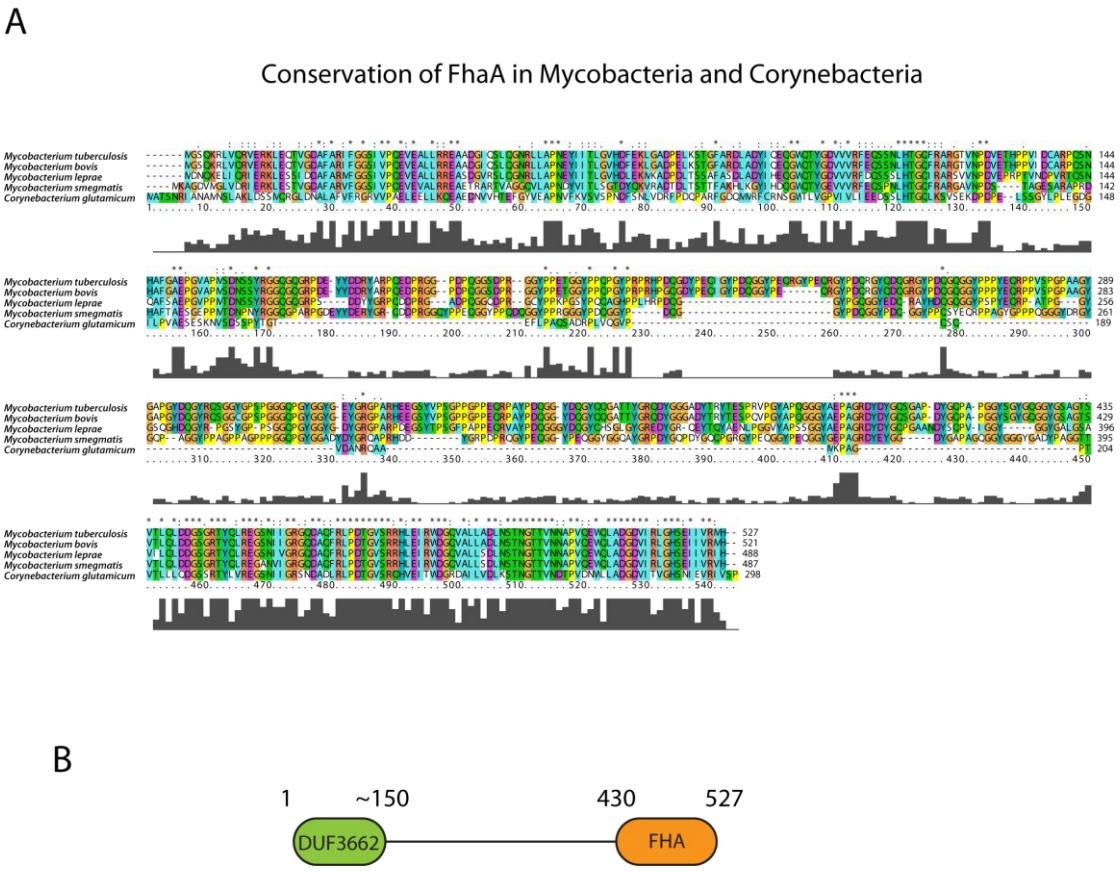


Figure 4.1. Domain organisation and conservation of FhaA. (A) Sequence alignment using ClustalX of FhaA and its homologues from the species indicated. Amino acids are coloured according to their properties. Hydrophobic: cyan, polar: green, basic: red, acidic: purple, tyrosine or histidine: dark cyan, glycine: beige, proline: yellow. The size of the bars under the alignment indicates the level of sequence conservation for that particular amino acid. The dots and stars on top of the alignment also indicated relative levels of conservation. A star indicates absolute conservation while two dots or one dot refer to a less strongly conserved amino acid. (B) Domain organisation of FhaA inferred from the sequence alignment and the crystallographic data available on the FHA domain (3POA).

4.2 Limited proteolysis

To determine the C-terminal domain boundary, limited tryptic proteolysis was carried out on FhaA$^{1-166}$ (figure 4.2). Trypsin cleaves peptide bonds C-terminal to lysine and arginine residues that are not followed by a proline. Cleavage sites within stable domains however are usually protected from proteolysis while sites within unstructured and flexible regions are highly susceptible to proteolysis due to their accessibility. In many cases, proteins can be reduced to their constituent minimal domains by treatment with trypsin or related proteases. In this case, treatment with trypsin reduced the apparent size of the FhaA$^{1-166}$ construct (actual molecular weight 18462 Da) by PAGE from approximately 20 kDa to 15 kDa. While the size of the main band was reduced by approximately 5 kDa, there were no abundant smaller bands visible and the intensity of the main band remained constant suggesting that it may represent the minimal DUF3662 domain. From the apparent molecular weight it was estimated that the proteolytic fragment would be around 120 amino acids in length. As the C-terminal region of FHA$^{1-166}$ is much less conserved than the N-terminal region, it was assumed that this region would also be less structured and hence more prone to proteolysis. It was therefore likely that the reduction in size after proteolysis occurred mainly at the C-terminus of the protein. Two new expression constructs, FhaA$^{1-110}$ and FhaA$^{1-130}$, were cloned into the pET47b expression vector to produce N-terminal 6x-His fusions as it was possible that the GST-protein had been responsible for the large amounts of nucleic acid contaminants seen during the initial stages of protein purification of FhaA$^{1-166}$. Both constructs expressed, but only FhaA$^{1-130}$ was soluble. It could be purified using the same protocol as for the longer constructs. Despite the introduction of the 6x-His tag, nucleic acid contamination was still present.

In an attempt to learn about the function of DUF3662, pull-down experiments with GST-fused FhaA$^{1-166}$ were carried out to identify protein binding partners. The GST-fused protein was immobilised on glutathione sepharose beads and used as bait to pull out FhaA interacting partners from *Mycobacterium tuberculosis* cell lysate. No protein interacting partners could however be identified. As the nucleic acid contamination was independent of the purification tag and no protein interacting partners could be identified, it was thought that the nucleic acid contamination may be indicative of a genuine function of DUF3662 and the nucleic acid contaminants were investigated further.

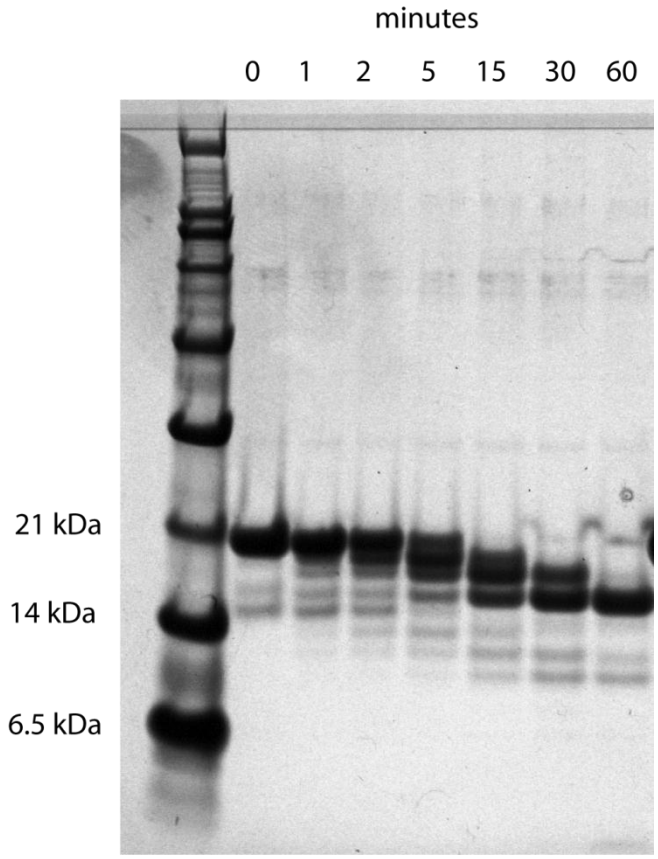


Figure 4.2. Limited proteolysis of DUF3662. SDS-PAGE analysis of a tryptic limited proteolysis experiment of FhaA$^{1-166}$ and porcine trypsin at a ratio of 500:1 (w/w). The numbers on the left indicate the sizes of the respective molecular weight marker bands.

4.3 Nucleic acid characterisation

As a starting point, the identity of the nucleic acids was to be identified. Flow through from the heparin purification step was used as a starting point for nucleic acid purification as this solution was already depleted in protein. Residual protein was removed by phenol:chloroform extraction and the purified nucleic acids 5′ labelled with ^{32}P for ease of visualisation. To determine what type of nucleic acid was present, the sample was subjected to treatment with a range of nucleases with different target specificities. The unknown nucleic acids, along with a 25 base DNA oligonucleotide and a 24 base RNA oligonucleotide were treated with RNase, DNase and benzonase, which digests both DNA, RNA and DNA/RNA hybrids. The treated and untreated nucleic acids were then run on a 12.5% denaturing polyacrylamide gel. From the photographic film visualisation of the gel (figure 4.3), it can be seen that the unknown nucleic acids are degraded by RNase and benzonase, but remain intact in the presence of DNase. They run as a smear indicating a range of sizes with the majority of fragments running just below the 24 and 25 base control oligonucleotides. The nucleic acid contaminants were therefore shown to be RNA oligonucleotides of varying lengths, with the majority between 15 and 30 bases. Both the DNA and RNA controls were degraded by the expected nucleases confirming the activity and fidelity of the enzymes used. Electrophoretic mobility shift assays using single-stranded and double-stranded DNA and RNA oligonucleotides failed to identify any nucleic acid that bound to DUF3662 (figure 4.4). The RNA contamination was subsequently shown to be an artefact since an $FhaA^{1-130}$ expression construct with a 6x-His tag at the C-terminus did not co-purify with any nucleic acids.

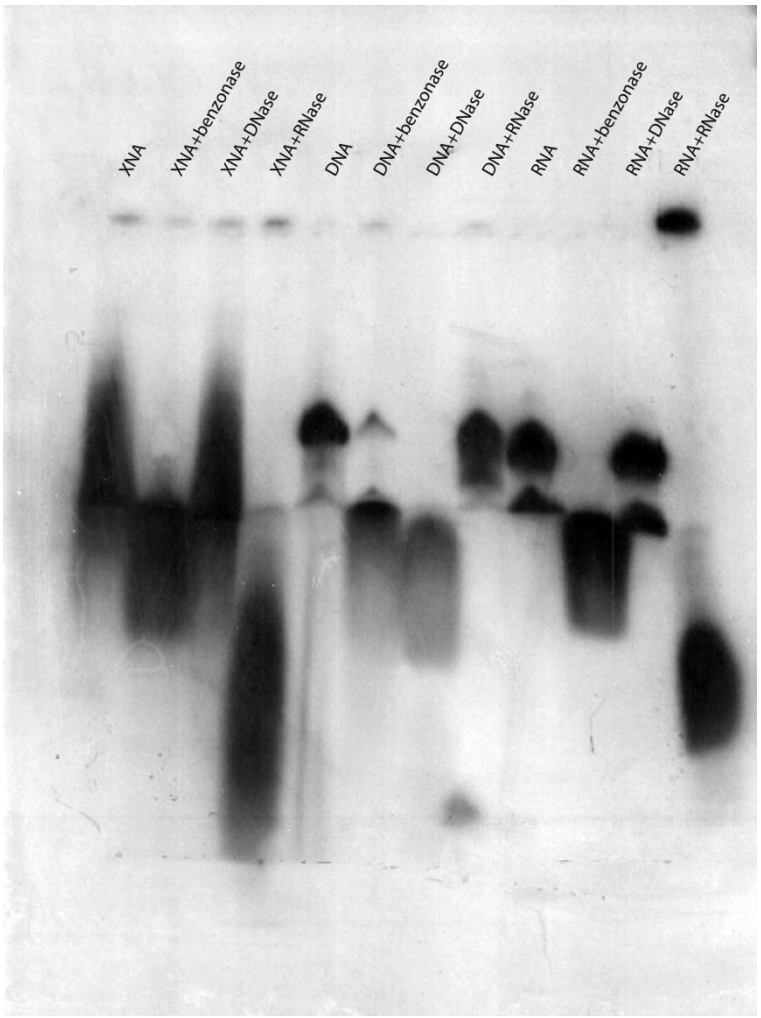


Figure 4.3. Nuclease digest of unknown nucleic acid contaminants of FhaA$^{1-166}$. The unknown nucleic acid (XNA), a 25 base DNA oligonucleotide (DNA) and a 24 base RNA oligonucleotide (RNA) were 5′-labelled with ^{32}P and incubated with 2.5 units of the nucleases indicated. The resulting digests were separated by denaturing PAGE and visualised by autoradiography.

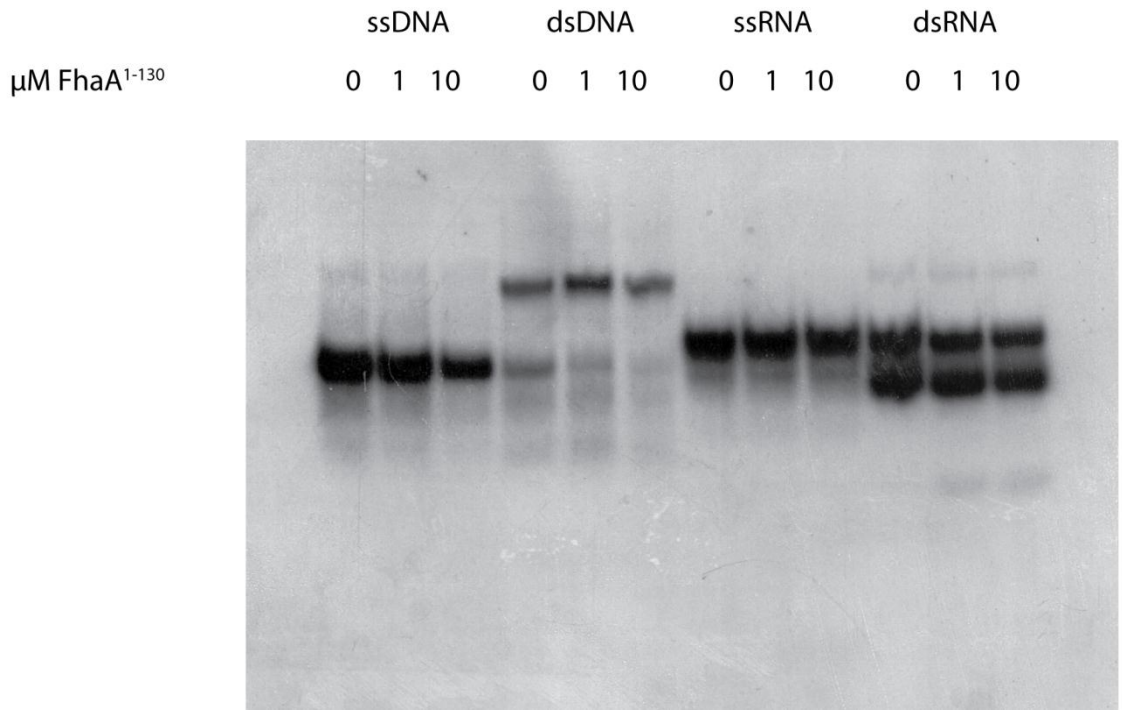
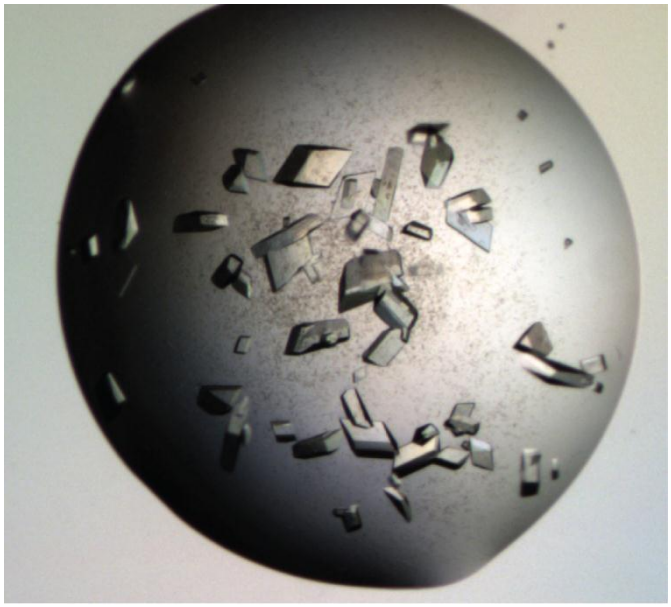


Figure 4.4. EMSA: Autoradiography of an EMSA with FhaA$^{1-130}$. A 24 base DNA oligonucleotide, a 24 bp double-stranded oligonucleotide, a 25 base RNA oligonucleotide and a 16 bp double stranded RNA oligonucleotide at 100 nM were incubated with FhaA$^{1-130}$ at the concentrations indicated and then separated by non-denaturing PAGE.

4.4 Crystallography of DUF3662

4.4.1 Crystallisation

As the function of the DUF3662 domain could not easily be determined biochemically, attention shifted back towards determining its structure. Crystallisation trials of $FhaA^{1-130}$ constructs with either N-terminal or C-terminal 6x-His tags were set up. The protein was concentrated to 30 mg/ml and the buffer changed to 150 mM NaCl, 20 mM Tris pH 8.0 and 0.5 mM TCEP and hanging drops with 1 µl protein solution and 1 µl well solution were set up in EasyXtal 24 well plates (Qiagen). Protein with an N-terminal tag did not crystallise, but $FhaA^{1-130}$ with a C-terminal 6x-His tag still intact produced crystals from a single condition. Single crystals grew in 4% (w/v) polyethylene glycol 4000 (PEG4000) and 0.1 M sodium acetate. To increase the crystal size, a 2D optimisation grid was set up around the original crystallisation condition. Different protein concentrations (18-30 mg/ml) were used and the PEG4000 concentration varied between 2-7%. The largest crystals were obtained at a protein concentration of 28 mg/ml and 3-5% PEG4000 (figure 4.5). 0.1 M sodium acetate was used as a buffer throughout. Crystals were moved to a drop containing mother liquor plus 25% (w/v) glycerol as a cryo-protectant and then flash-frozen in liquid nitrogen.



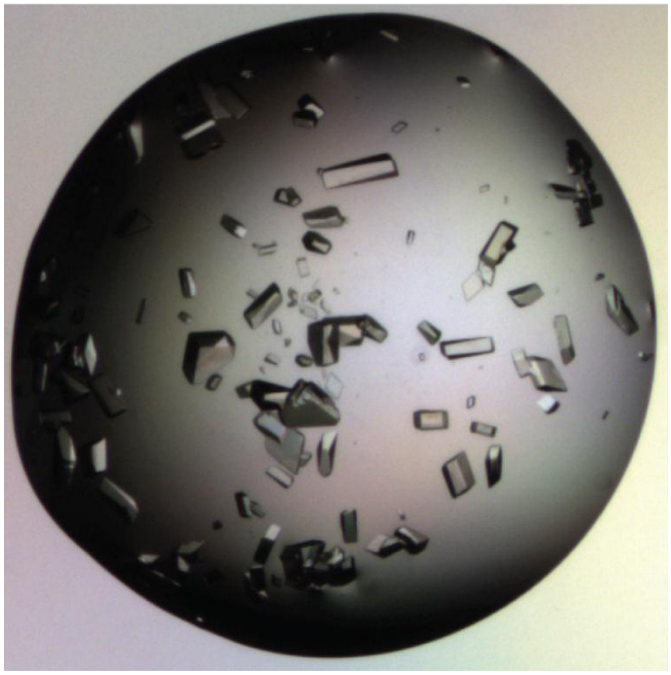


Figure 4.5. DUF3662 crystals. Photographic images of drops containing crystallised FhaA$^{1-130}$. The total drop size is 2 µl.

4.4.2 Data processing

X-ray diffraction data were collected at the IO3 beamline at Diamond Light Source. The crystals were indexed as space group P2 or $P2_1$ and a complete 180° data-set of the best-diffracting crystal was collected. The diffraction images were processed using the HKL suite of programs (HKL research Inc.) The processing statistics can be found in table 4.1. Following integration of the images, systematic absences for reflections with indices 0, k, 0 for which k = 2n +1 showed the space group to be $P2_1$.

wavelength	0.9763 Å
data collected	180° (360 x 0.5° oscillation)
space group	$P2_1$
unit cell	a = 91.62 b = 145.29 c = 101.64 α = 90 β = 116.34 γ = 90
Resolution range	30.0 – 2.50 Å
total reflections observed	312498
unique reflections	81690
average redundancy	3.8 *(3.7)*
Completeness (%)	99.4 *(98.2)*
I/σ	14.4 *(3.0)*
R_{symm}	0.063 *(0.44)*

Table 4.1. Processing statistics of $FhaA^{1-130}$ crystal (highest resolution shell, 2.61 - 2.50 Å, in brackets)

The diffraction data was processed to a resolution limit of 2.50 Å. At this resolution limit, even the highest resolution shell had an acceptable signal/noise ratio of three. One notable feature of the data-set is the relatively large unit cell for a 140 amino acid protein. Using the Matthews coefficient, it was calculated that there would be 16 copies per asymmetric unit at 50% solvent content [168].

4.4.3 Phasing attempts

As no homology was shared between the sequence of the DUF3662 domain and the sequences in known protein structures, molecular replacement could be excluded as a way to overcome the phase problem. Attempts were made to produce seleno-methionine substituted protein for crystallisation in order to acquire phase information.

Phases can be determined from crystals containing ordered selenium atoms since those atoms scatter X-rays both elastically and anomalously as selenium absorbs electromagnetic radiation in the X-ray range. Since the valency electron shell of a

selenium atom is equivalent to the valency shell of a sulphur atom, selenium can be introduced into proteins by substituting the sulphur atom in methionine residues with a selenium atom. For recombinant proteins, this is achieved by supplementing bacterial growth media with synthetically produced seleno-methionine.

Although there are no methionine residues in the FhaA$^{1-130}$ construct, the sequence alignment of FhaA in figure 4.1, shows that in some closely related species there are methionines in the DUF3662 domain. In *Mycobacterium leprae*, where the DUF3662 domain is conserved, there are methionines at positions 26 and 74. Site-directed mutagenesis was performed to introduce methionines at these positions into the *Mycobacterium tuberculosis* DUF3662 domain as these methionines were assumed to not interfere with the fold of this domain. The two mutants I26M and L74M were introduced in a step-wise process. After a first round of mutagenesis reactions, the FhaA$^{1-130}$ I26M construct was used in a crystallisation experiment to test whether crystals could be reproduced for the mutant protein. The mutant protein did not crystallise under the same condition used for the wild-type and re-screening was also unsuccessful.

To investigate whether the overall DUF3662 fold was still intact, ^{15}N-labelled FhaA$^{1-130}$ I26M and FhaA$^{1-130}$ wild type were produced and HSQC spectra for each were acquired (figure 4.6). Both spectra are those of folded globular proteins, but there are numerous chemical shift differences indicating changes in the electromagnetic microenvironments of several amino acids. Overall, these changes were considered more severe than might be expected for a relatively conservative point-mutation. The introduction of only one methionine was enough to alter the protein domain sufficiently to prevent crystallisation. As a rule of thumb, one structured seleno-methionine is needed per 10 kDa of protein [169]. It was thought that multiple seleno-methionines were expected to be needed to obtain phases and phasing by anomalous dispersion was dismissed as a viable option to solve the structure of DUF3662. Although L74M, a more conservative mutation than I26M, might not have prevented crystallisation of DUF3662 one seleno-methionine would most likely not have been sufficient to solve a 17 kDa protein domain.

While alternative methods to obtain the phases needed to solve the structure of DUF3662 were considered, such as heavy-metal soaking of crystals and multiple isomorphous replacement, an NMR solution structure of DUF3662 was reported at this point (PDB: 2LC0, figure 1.8). An attempt was made to use the ensemble of solution structures as a molecular replacement search model in PHASER, but no credible

solution was obtained. The lack of success could be explained by several reasons. The large asymmetric unit contains approximately 16 copies of $FhaA^{1-130}$. A less likely explanation could be that the conformation of DUF3662 in the crystal was fundamentally different to the one reported in the solution structure. Solution structures are not necessarily good models for molecular replacement as the resolutions of such structures are not comparable with crystallographic models.

4.5 Summary

Despite the combined biochemical and structural approach to characterise DUF3662, no information was ultimately obtained that could provide any insights into the function of this domain. No conclusions could be drawn about the function of FhaA, or the N-terminal domain in particular, in its parent organism *Mycobacterium tuberculosis* or in the human host. Using bio-informatics and limited proteolysis, the approximate location of DUF3662 was determined to be between amino acids 1-130 of FhaA. An expression construct of these amino acids produced protein that could be purified and crystallised to produce well diffracting crystals. The lack of methionines in DUF3662 and its sensitivity to mutations hindered phase determination. Nucleic acids that had co-purified with different DUF3662 expression constructs were identified to be RNA oligonucleotides but eventually proved to be an artefact of the affinity purification method.

It should be noted that in spite of the structural data now available, there is still no function assigned to DUF3662 [126]. Its unusual topology of two α-helices lying next to a three stranded β-sheet (figure 1.8) has no clear homologue structures. The structure though reveals that the first 20 amino acid of FhaA are unstructured and that the N-terminal domain boundary assumed in this thesis was incorrect. As the structure of DUF3662 had already been reported elsewhere and attempts to find interacting partners, namely electrophoretic mobility shift assays and GST pull-downs, did not provide any information about potential functions of the FhaA N-terminal domain, work on DUF3662 was abandoned. Instead, focus shifted exclusively towards investigations into the role of PknB mediated FhaA phosphorylation.

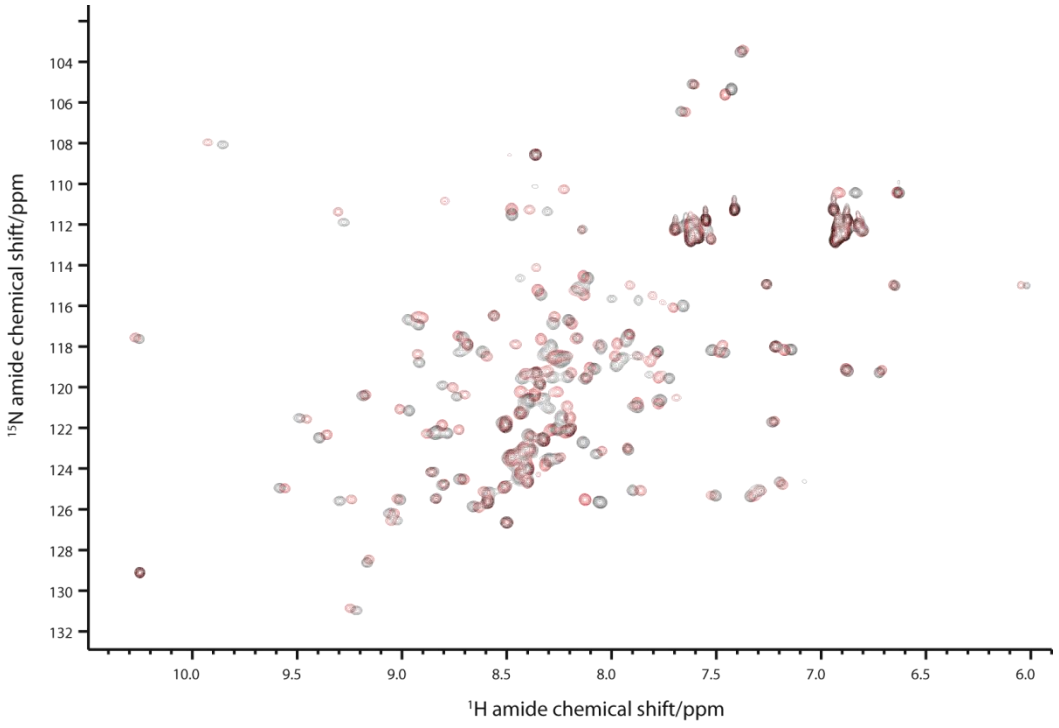


Figure 4.6. Comparison of FhaA$^{1\text{-}130}$ wild type and I26M by NMR. Overlay of ^{1}H-^{15}N-HSQC spectra of FhaA$^{1\text{-}130}$ wild type (black) and I26M (red).

5 Characterisation of PknB mediated phosphorylation of FhaA

While the fact that FhaA is a substrate of STPK PknB had been reported previously, limited information was available on the mechanism of phosphorylation or its consequences [104]. It had been shown that a phospho-binding competent FHA domain was required for efficient phosphorylation and that the phosphorylated activation loop of the kinase domain engages in a low micromolar interaction with the FHA domain. Furthermore, an FhaA construct comprising amino acids 155 to 527 had been shown to be a substrate of PknB, but the minimal FHA domain ($FhaA^{430-527}$) could not be phosphorylated [104, 121]. The FHA domain is essential for phosphorylation, but is not a substrate itself [104]. From this information, it could be deduced that there was likely to be one or more PknB-phosphorylation sites between amino acids 155 and 430. It was thought that this phosphorylation event may lead to a phospho-dependent oligomerisation event as seen for Chk2 or an intra-molecular association similar to GarA [53, 68]. As a first step towards characterising FhaA phosphorylation, the phosphorylated residue, or residues, needed identifying.

5.1 Identification of the PknB phosphorylation site

To identify the phosphorylated amino acids of a protein by mass-spectrometry, it is desirable to have a pure, stoichiometrically phosphorylated protein sample. The $FhaA^{155-527}$ construct that had been shown to be a substrate though was only sparingly soluble and not easily purified, presumably due to the long linker region. In the study that identified FhaA as a substrate of PknB, the phosphorylation could only be shown by ^{32}P kinase assays in combination with SDS-PAGE [104]. Initial attempts to identify the phosphorylation site of $FhaA^{155-527}$ by mass-spectrometry failed for several reasons. The low solubility of the $FhaA^{155-527}$ protein prevented effective purification. Furthermore, the long flexible linker was very susceptible to proteolysis which prevented prolonged incubation of $FhaA^{155-527}$ with PknB. The flexibility and solvent exposure of the linker region in combination with considerable amounts of impurities, which may have contained bacterial proteases led to protein instability with significant sample degradation before it could be phosphorylated to a measurable extent.

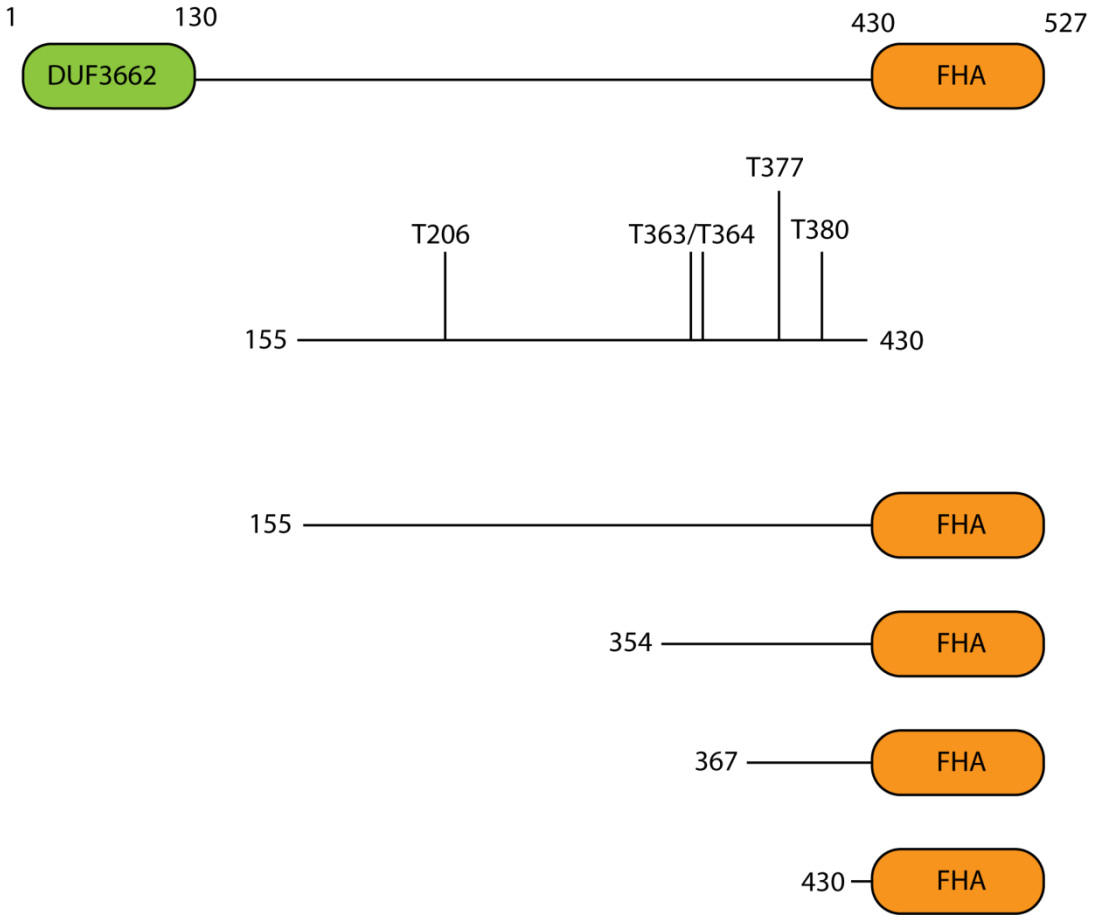


Figure 5.1. Schematic of FhaA domain organisation and the locations of all threonines in the linker region. In the linker, which is known to be phosphorylated, there are 5 threonines which may be phosphorylated by PknB. Bottom: constructs used in *in vitro* kinase assays to identify the phosphorylation site.

5.1.1 Kinase assays

Upon closer inspection of the FhaA amino acid sequence, it can be seen that there are only five threonines between amino acids 155 and 430 (figure 5.1), the highly preferred type of amino acid that PknB is reported to phosphorylate [101]. Protein constructs extending from the FhaA C-terminus to different points in the linker region were produced to determine which of the 5 candidates are phosphorylated. As the potential sites were threonines 206, 363, 364, 377 and 380, intermediate constructs FhaA$^{354-527}$ and FhaA$^{367-527}$ were generated. No constructs were generated that started between 363 and 364 or 377 and 380 as native protein sequence needed to be present N-terminal to the potential phosphorylation site to ensure substrate recognition by the kinase. By monitoring the activity of PknB on the proteins FhaA$^{155-527}$, FhaA$^{354-527}$, FhaA$^{367-527}$ and FhaA$^{430-527}$, it could be determined which threonines are phosphorylated by PknB. The different FhaA constructs were incubated with either the PknB kinase domain (PknB$^{1-279}$) or the whole intra-cellular part of this kinase (PknB$^{1-331}$) in the presence of γ-^{32}P ATP. It has been reported that the juxtamembrane region of PknB, which comprises residues 280 to 332, interacts strongly with FhaA in a phospho-dependent manner [126]. It was therefore investigated whether the presence of this region has any effect on the PknB mediated phosphorylation of FhaA.

In a first experiment, the four different FhaA constructs were incubated with PknB$^{1-279}$, the reaction mixtures separated by SDS-PAGE (figure 5.2A) and phosphorylated protein visualised autoradiographically. PknB$^{1-279}$, which runs as a doublet on the gel, phosphorylates FhaA$^{155-527}$, FhaA$^{354-527}$ and FhaA$^{367-527}$, but does not phosphorylate FhaA$^{430-527}$. The phosphorylated species in the lane 1 is PknB$^{1-279}$. From ESI-MS it was known that PknB purified from E. coli was already phosphorylated, but in the lanes where there is PknB, but no protein substrate, PknB appears to use the radiolabelled ATP from the kinase assay reaction mixture (section 2.3.1) to autophosphorylate. This is particularly prominent in lane 1 of the gel, where several oligomeric states of PknB are visible. Control lanes 5-8 with no kinase added were set up to ensure the phosphorylation of FhaA is due to the catalytic activity of PknB. In these lanes, no phosphorylation could be detected. The exceptions are lanes 4 and 8, where a contaminant is present that is able to autophosphorylate. This contaminant, which migrates faster than the protein of interest, is phosphorylated both in the presence and the absence of PknB$^{1-279}$. The fact that this contaminant is phosphorylated very well even in the absence of PknB suggests that a significant amount of ATP utilising enzymes from the *E. coli* expression culture was present. These enzymes may either autophosphorylate and or phosphorylate a substrate that is also present in this gel

band. In a phosphoproteomics screen, 79 phosphoproteins have been identified in *E. coli*. Furthermore, several STPKs have been identified in this organism with examples being isocitrate dehydrogenase kinase/phosphorylase YihE kinase [170]. Any of these kinases may have co-purified with FhaA and been responsible for the strong signal in the kinase assay.

A second round of kinase assays and SDS-PAGE analysis was carried out to determine whether the presence of the PknB juxtamembrane region has any effect on FhaA phosphorylation (figure 5.2 B-top). It had previously been shown that the FhaA FHA domain has nanomolar affinity for the phosphorylated juxtamembrane region of PknB [126]. Mycobacterial GarA, a known substrate of PknB, was used as a positive control [69]. The molar amounts of protein loaded were similar for each lane, so one could compare the rates of phosphorylation semi-quantitatively. From the autoradiography, it can be seen that both $PknB^{1-279}$ and $PknB^{1-331}$ phosphorylate $FhaA^{367-527}$. The level of phosphorylation is slightly higher when the juxtamembrane region is present. In addition to FhaA, GarA is phosphorylated by both constructs of PknB. In the case of FhaA, the longer PknB construct is a more active kinase. For GarA, which is a better substrate, no comparison can be made due to the intensity of the labelled bands. Following exposure to the photographic film, the SDS-PAGE gel was stained with a coomassie stain to confirm the amounts of protein loaded in each gel are equivalent (figure 5.2 B-bottom). The kinase assays thus confirmed that there is at least one PknB phosphorylation site between amino acids 367 and 430 on FhaA. Due to the specificity of PknB for threonine residues [126], the two potential candidates for phosphorylation were threonines 377 and 380.

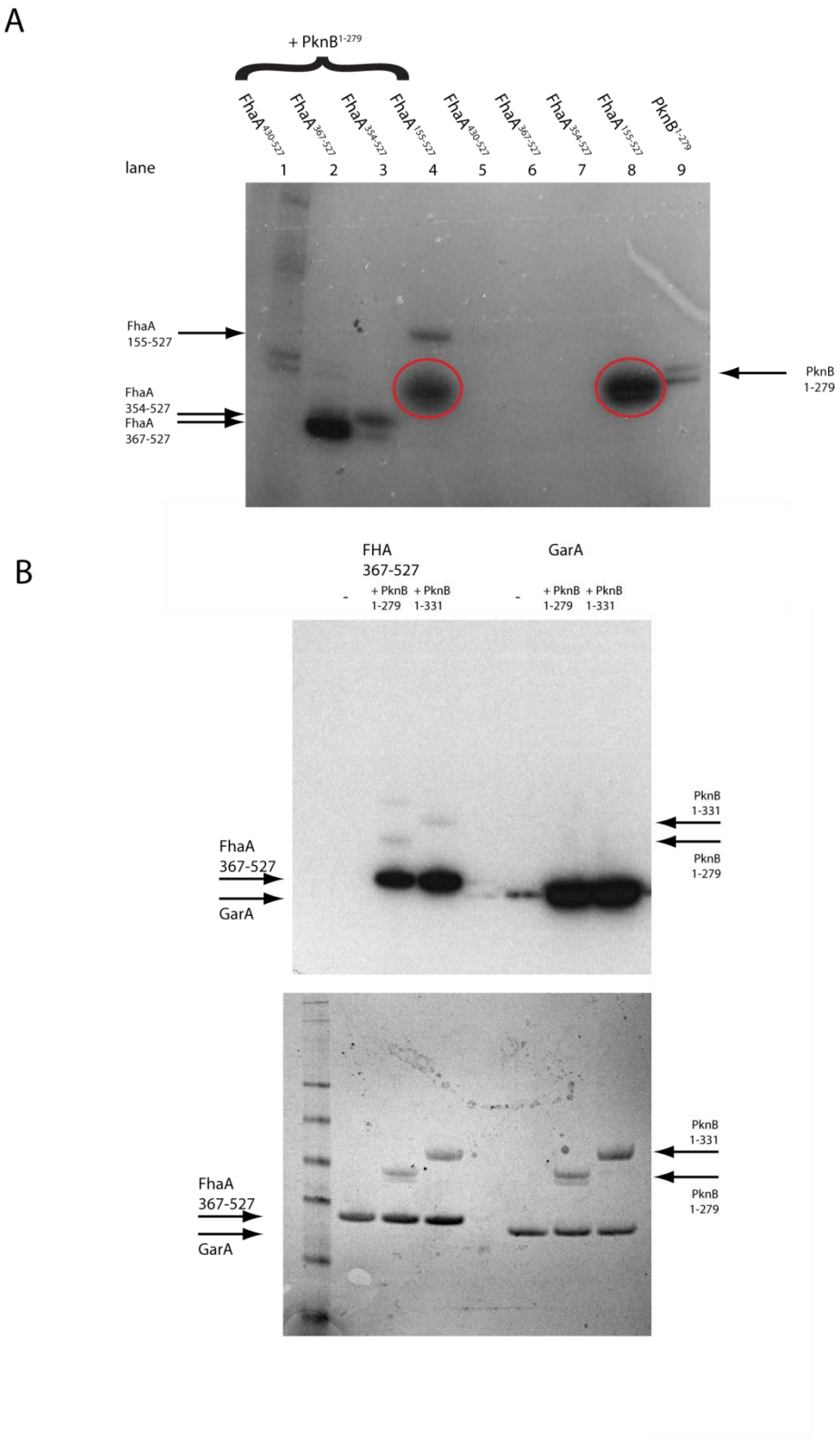


Figure 5.2. PknB mediated phosphorylation of FhaA. *In vitro* kinase assay using γ-^{32}P-ATP and the kinases and substrates indicated. Bands were visualised by autoradiography (A and upper panel of B) and Coomassie staining (lower panel of B)

5.1.2 Generation of stoichiometrically phosphorylated FhaA$^{367-527}$

As the FhaA$^{367-527}$ protein could be purified easily and was more stable than the longer constructs, an attempt was made to identify the PknB phosphorylation sites on this protein by mass-spectrometry. FhaA$^{367-527}$ was incubated with PknB$^{1-279}$ at room temperature in the presence of ATP to produce stoichiometrically phosphorylated protein. After 4 days a sample of the reaction mixture was analysed by ESI-MS. The mass spectrum (figure 5.3) reveals that the majority of the protein had been phosphorylated. There is a peak at 17398.7 Da corresponding to unmodified FhaA$^{367-527}$ (calculated mass 17399.7 Da) and a significantly larger peak at 17478.8 Da. The difference in mass of 79.1 Da matches closely the 80 Da increase expected for a phosphorylation event. In weakly acidic conditions as found in the mass spectrometer the phosphorylation of a hydroxyl is accompanied by the loss of one hydroxyl (-17 Da), offset against the gain of one molecule of phosphoric acid (H_2PO_4 97 Da). The minor peaks that can be found at position n(+22 Da) are thought to be sodium adducts (calculated mass 23 Da). In the mass-spectrometer, the protein is ionised using protons. If there is Na^+ present in the sample these ions can be used to ionise the sample instead, leading to sodium adducts. Despite the de-salting of samples prior to analysis by ESI-MS, some residual sodium ions remained. The exact proportion of FhaA$^{367-527}$ that was phosphorylated could not be determined as the different phosphorylation states of a protein do not necessarily produce equally intense mass spectrometry signals. It was however assumed that the majority of protein had been phosphorylated. The mass spectrum also reveals that there is only one phosphorylation site on this protein, as no peak for doubly-phosphorylated FhaA$^{367-527}$ (calculated mass 17559.7) is visible. From the autoradiographies and the mass spectrometry data it could be assumed that PknB phosphorylates threonine 377 or threonine 380 or that it may be phosphorylate either in mutually exclusive events.

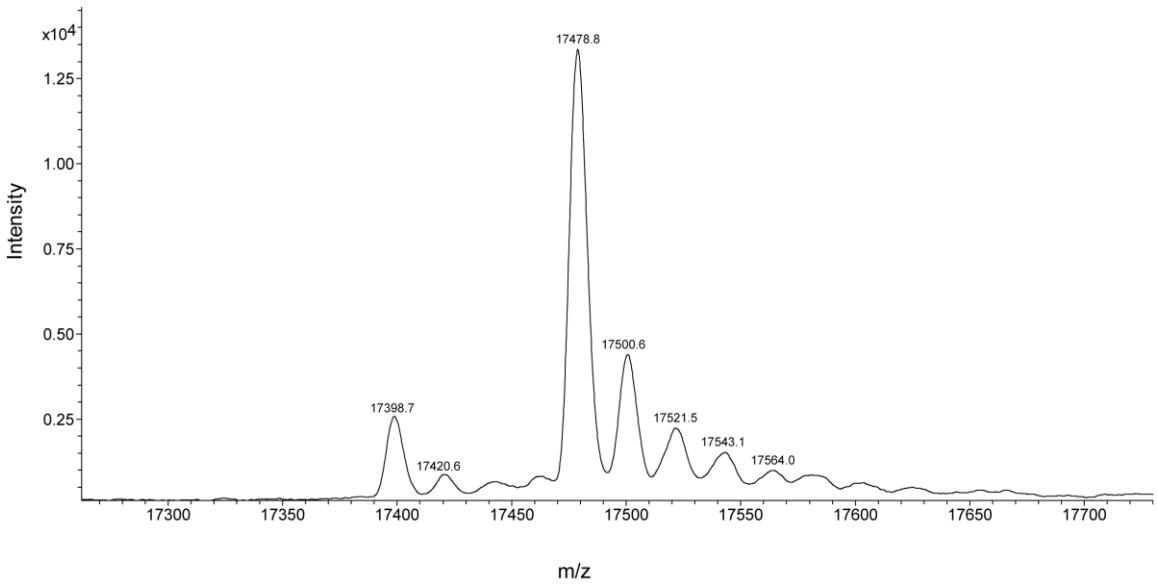


Figure 5.3. Deconvoluted mass spectrum of phosphorylated FhaA$^{367-527}$. A small proportion of native (17398.7 m/z) is present, but the majority of protein is mono-phosphorylated (17478.8 m/z). Several Na^+ adducts can be seen.

5.1.3 Separation and sequencing of peptides

Tandem mass-spectrometry (ms/ms) was performed on the intact protein in an attempt to distinguish which of the two sites was phosphorylated, but due to the close proximity of the two candidate residues the phosphorylation site could not be assigned unambiguously. It was therefore necessary to digest the protein proteolytically, separate the resulting peptides by RP-HPLC and then analyse the peptides by ms/ms to identify the residue phosphorylated by PknB. In an ms/ms experiment, one ion from a primary spectrum is isolated and further fragmented into a secondary spectrum. This causes a reduction of signal as the signal of one parent ion is used to generate a whole new spectrum, especially if the parent ion is of low abundance. It is therefore desirable to isolate the peptide to be analysed prior to mass spectrometry. If there is only one dominant ion in the primary spectrum, the reduction of signal in a secondary spectrum generated from this dominant primary ion is greatly reduced. Another reason for digesting a protein to be studied into smaller peptides is that peptides with lower mass/charge ratios (m/z) generate clearer signals in a mass spectrometer than proteins.

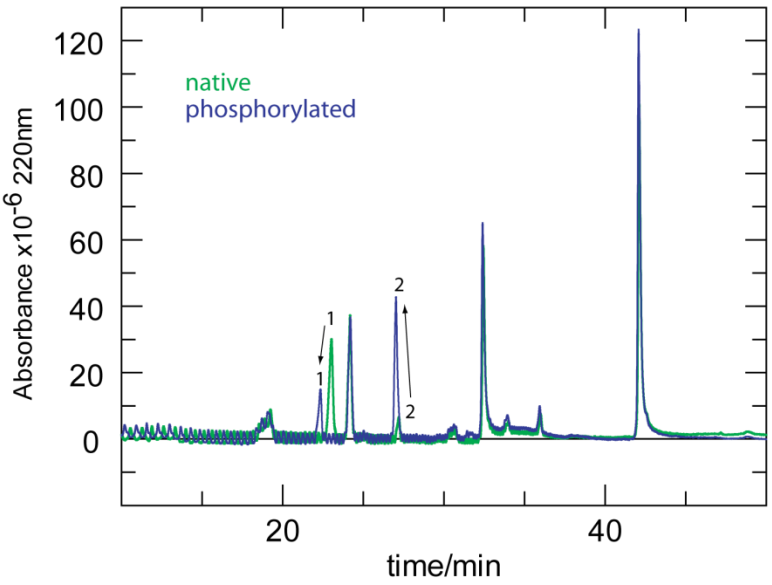
As preparation for the mass spectrometric analysis of fragments of $FhaA^{367-527}$, tryptic digests of both the phosphorylated and the unphosphorylated isoform of $FhaA^{367-527}$ were carried out. Unlike the mild conditions of limited proteolysis described in the previous chapter, where a relatively trypsin resistant fragment was required, the aim here was to digest the protein to completion. 10% (w/w) trypsin was used over an incubation period of 4 hours at 37 °C in order to produce small fragments. As there is a tryptic site in the protein sequence between the two candidate phospho-threonines (A-D-Y-**T**-*R*-Y-**T**-E-S-P), it was thought that through thorough digestion, cleavage could be achieved at this position and the phosphorylation site determined without sequencing the ion by ms/ms. An increase in mass of either the N-terminal or C-terminal peptides measured by standard ESI-MS would be sufficient to assign the phosphorylation site to one threonine.

Peptides generated by the tryptic digests were separated by RP-HPLC. In this experiment, the peptides were first loaded onto a C18 reverse phase column. This column contains long (18 carbon atoms), aliphatic, and hence hydrophobic, chains. In an aqueous buffer, the tryptic peptides bind to this matrix in preference to an aqueous solvent shell. After this binding step, a solvent gradient was applied to the column ranging from 5% ACN and 95% H_2O to 95% ACN and 5% H_2O to elute the peptides. As the solvent gradually changes from aqueous to organic, the hydrophobic force

keeping the peptides bound to the matrix decreases and the peptides elute, with more hydrophobic peptides eluting at higher ACN concentrations.

The digests of phosphorylated and unphosphorylated FhaA$^{367-527}$ produced chromatograms with well resolved peaks (figure 5.4). An overlay reveals the two chromatograms to be highly similar, as would be expected for different isoforms of the same protein. Both display 5 major peaks, significantly fewer peaks than would be expected from a simulated digestion of FhaA$^{367-527}$. Although no investigation was carried out to determine which tryptic sites resisted cleavage or which peptides failed to bind the column matrix, one could speculate that the core FHA domain may have remained intact. The major differences between the 2 chromatograms are observed for 2 peaks, labelled peak 1 and peak 2. It was assumed that peak 1, as well as peak 2, contains identical peptide sequences in both chromatograms. The abundance of the peptide eluting in peak 1 decreases in the phosphorylated form of the protein and the peptide itself appears to become more hydrophilic. It was assumed that the peptides in peak 1 and peak 2 contained a phosphorylation site that made the peptide more hydrophilic and hence elute earlier. The loss of abundance of peak 1 is at least partially offset by the increase in size of peak 2. Since the retention times were shortened and the relative abundances of these two peaks were considerably affected by phosphorylation, fractions containing these peaks were lyophilised and analysed my ESI-MS as they would most likely contain the phosphorylation site.

A



B

GPGMQDYGGGADYTR-YTESPR-VPGYAPQGGGYAEPAGR-
DYDYGQSGAPDYGQPAPGGYSGYGQGGYGSAGTSVTLQLDDGSGR-
TYQLR-EGSNIIGR-GQDAQFRLPDTGVSR-R-HLEIR-
WDGQVALLADLNSTNGTTVNNAPVQEWQLADGDVIR-LGHSEIIVR-MH

C

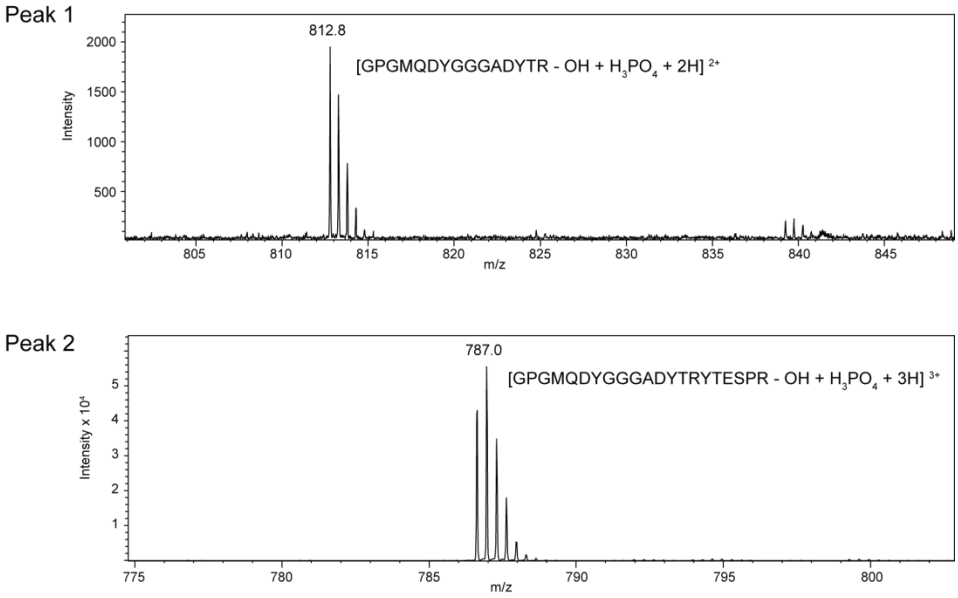


Figure 5.4. Analysis of tryptic digests of FhaA. (A) Reverse-phase HPLC chromatogram of a tryptic digest of native (green) and mono-phosphorylated (blue) FhaA$^{367-527}$ .The numbers indicate the two peaks where differences could be seen between the two isoforms. Peaks 1 and 2 were further analysed by ESI-MS. (B) Simulated tryptic digest of FhaA$^{367-527}$. The dashes indicate tryptic sites. (C) Deconvoluted mass spectra of the lyophilised peptides eluted in peaks 1 and 2 (phosphorylated form). The measured m/z ratio of the most abundant ion is indicated as well as the calculated composition of each ion.

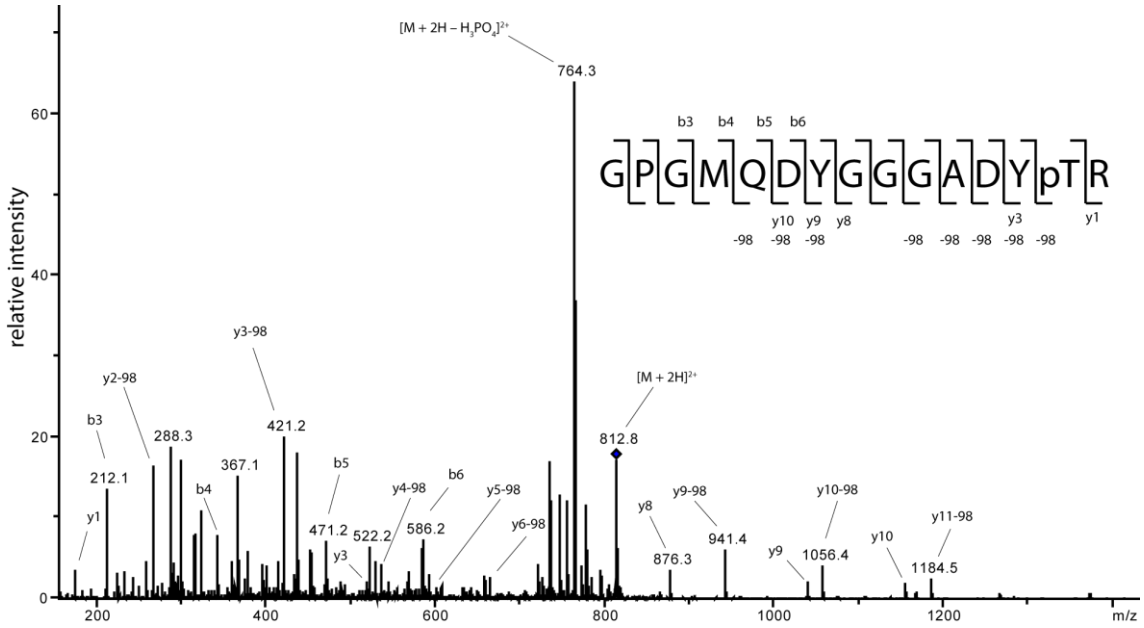


Figure 5.5. Tandem mass-spectrometry analysis of peak 1 (812.8 m/z). The b and y ion series are shown in the upper right hand corner. -98 denotes a loss of phosphoric acid

This analysis revealed peak 1 to be the first 15 amino acids of FhaA$^{367-527}$ containing one phosphorylation site. The m/z ratio of the 2+ ion of such a peptide was calculated using Protein Prospector (UCSF) to be 813.3, which agrees with the observed value of 812.8. The peptide that eluted in peak 2 consists of the first 21 amino acids of this protein and also contains one phosphorylation site. For the longer peptide, the 3+ ion was the most abundant. Again the observed m/z ratio of 786.9 matches closely the calculated value of 787.1. Peak 2 is therefore the result of an incomplete digest. Cleavage after Arg378 (first cleavage site indicated in figure 5.4 B) was not efficient. It can also be seen that phosphorylation interferes with cleavage after this residue. Phosphorylation decreases the abundance of the peptide in peak 1, which is a result of Arg378 cleavage, but increases the size of peak 2, containing the peptide that is a result of cleavage at Arg384 but an intact tryptic site at Arg378. More importantly however, the presence of a phospho-peptide consisting of the first 15 amino acids of FhaA$^{367-527}$ provides strong evidence that PknB phosphorylates this protein construct at Thr377, the only threonine in this peptide. The presence of a phosphorylated amino acid in the -1 position to the tryptic site may also be responsible for the reduced cleavage at Arg378 in the phosphorylated form of FhaA$^{367-527}$. The reduction of hydrophobicity of the peak 1 peptide upon phosphorylation, indicated by its earlier elution, can be explained by the addition of the highly hydrophilic phosphate group. In the longer, peak 2, peptide, the shift in elution time is much less pronounced. The change in hydrophobicity upon phosphorylation has a reduced effect in the context of a longer peptide.

Although there already was significant evidence pointing towards phosphorylation of Thr377, ms/ms was performed with the peptide of peak 1 to confirm these findings (figure 5.5). When sequencing a peptide by ms/ms, several ions can be observed, representing the full length molecular (M) ion and its fragmentation products. When the peptide is cleaved at a peptide bond, the resulting fragments are called b ion and y ions. If after the fragmentation, the charge is kept on the N-terminal product, the resulting ion is called a b ion. If it is retained C-terminal to the cleavage, a y ion is produced. Sequencing of a peptide that is phosphorylated will also yield ions that correspond to a loss of phosphate. These ions were labelled b-98/y-98 as the molecular weight of fully protonated phosphoric acid, which has been lost in such ions, is 98 Da.

In this particular spectrum, the 2+ ion of the phospho-peptide of peak 1 (m/z of 812.8), was isolated and further fragmented with an increased acceleration voltage. The ions observed of the intact peptide were 2+ ions, but the b and y ion series observed were

1+ ions. A small proportion of the phospho-peptide remained intact, as there is still a peak at 812.8. The most common fragmentation was a loss of phosphoric acid, which could be observed for the intact peptide (m/z 764.3) as well as several y ions. The most abundant peak in the spectrum is the intact peptide with loss of phosphoric acid. No b-98 ions were detected as the phosphorylation site was on the second to last residue of the peptide and no b ions of peptides longer than 6 amino acids could be detected. Overall, a total of 4 b ions, 5 y ions and 8 y-98 ions could be seen. The presence of the y2-98 indicates that there is a phosphorylation site on the last two amino acids of the peptide, confirming that Thr377 is phosphorylated by PknB. Arginine phosphorylation could be discounted for such a well-characterised STPK. No other phosphorylation sites could be found on the peptide, agreeing with the mono-phosphorylated state of the molecular ion.

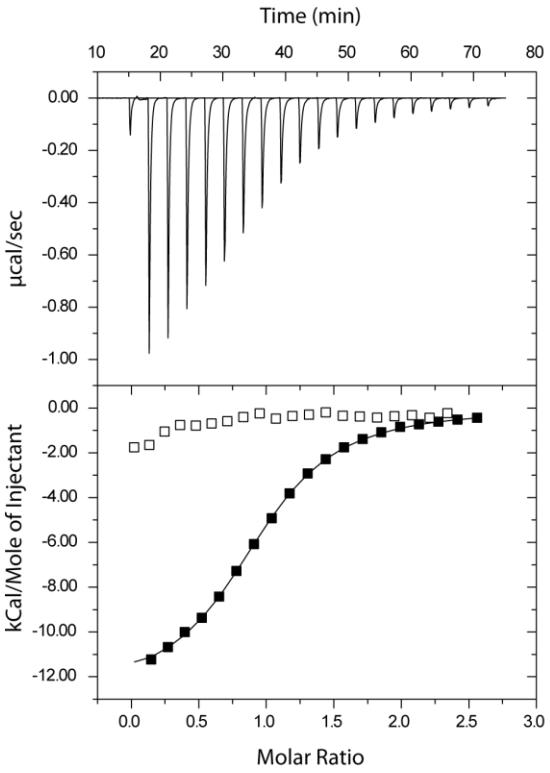
5.2 pThr377 associates with the FHA domain and impairs its phospho-binding capabilities

With the phosphorylation site determined to be Thr377, the consequences of FhaA phosphorylation were investigated. The phospho-binding capabilities of both native and phosphorylated FhaA$^{367-537}$ were characterised by ITC to further investigate the effects of phosphorylation on FhaA. It was thought that the pThr377 epitope may associate with the FHA domain. A chemically synthesized phospho-peptide was designed to mimic the region around Thr377 phosphorylated by PknB. The sequence of this peptide was A-D-Y-pT-R-Y-T-E-S. The unphosphorylated form of FhaA$^{367-527}$ bound the phospho-peptide with an affinity of 7.3 μM (figure 5.6), 2 orders of magnitude weaker than the interaction of the FhaA FHA domain with an optimal peptide derived from peptide-library screens [61]. It is however comparable to the reported 7.0 μM affinity of this FHA domain for a peptide with sequence A-A-A-A-pT-A-A-A-A-Y-K-K Furthermore, the energetic contributions to binding are also similar with the binding event being mainly driven by enthalpy. For both equilibria, the ΔH is approximately -12 kcal/mol and an entropic penalty of 5 kcal/mol is incurred. One can therefore assume that no interactions are made between peptide side-chain atoms and the FHA domain. Such a weak signal suggests that the interaction between the pThr377 and the FHA domain is non-specific and most pThr-containing epitopes will bind with such affinities.

The affinity of the minimal FHA domain for the pThr377 phospho-peptide was also determined (figure 5.6, raw data not shown) to find out whether the presence of the

amino acids 367-430 interferes with peptide binding. From the table, it can be seen that binding capabilities of FhaA$^{367-527}$ and FhaA$^{430-527}$ are comparable. They allow direct comparison of the binding affinities of the FhaA FHA domain with those of FhaA$^{367-527}$ and suggest that the non-phosphorylated linker does not interact with the FHA domain. The phosphorylated form of FhaA$^{367-527}$ however does not bind to the phospho-peptide. The small signal present for the first four injections is most likely due to the fact that not all FhaA$^{367-527}$ may have been phosphorylated. Residual, native FhaA$^{367-527}$ may have been present and responsible for the signal. Since the presence of a pThr377 epitope that is covalently bound to the FHA domain prevents binding of a phospho-peptide with an identical sequence, it is likely that the FHA domain binds pThr377 via an intra-molecular association. The intra-molecularly attached pThr can outcompete the phospho-peptide at the FHA binding pocket due to its much higher rate of sampling. For other possible FHA mechanisms, such as an oligomerisation event as seen for Chk2 the rates of sampling would be identical for the phospho-peptide and the protomers and the loss of phospho-binding capabilities upon phosphorylation would be less severe when identical phospho-epitopes compete for one site [53].

A



B

	K_D (µM)	N	ΔH (kcal/mol)	TΔS (kcal/mol)
$FhaA^{367-527}$ + A-D-Y-pT-R-Y-T-E-S	7.3	0.95	-12.7	-5.8
$FhaA^{367-527}$ pThr377 + A-D-Y-pT-R-Y-T-E-S	N. D. B.			
$FhaA^{430-527}$ + A-D-Y-pT-R-Y-T-E-S	6.3	0.82	-13.4	-6.4
$FhaA^{430-527}$ + A-A-A-A-pT-A-A-A-A	7.0	0.97	-11.5	-4.7

[61]

Figure 5.6. ITC analysis of the phospho-binding capabilities of the FhaA FHA domain. (A) Raw calorimetric data, heat changes and fitted binding curves for the interaction between peptide ADYpTRYTES and $FhaA^{367-527}$ (filled squares) and $FhaA^{367-527}$pThr377 (empty squares. (B) Thermodynamic parameters for the ITC experiments shown above and for the interaction of the minimal FHA domain with the phospho-peptide.

5.3 FhaA engages in an intra-molecular association

Multi angle laser light scattering coupled to size exclusion chromatography (SEC-MALLS) was used to ascertain whether an intra-molecular association regulates FhaA or if phosphorylation changes the oligomeric states of this protein (figure 5.7).

From the SEC-MALLS experiment it can be seen that both FhaA$^{367-527}$ and FhaA$^{367-527}$ pThr377 are monomeric with their apparent molecular weights measured to be ~17 kDa, agreeing with the calculated monomeric mass of 17.4 Da. Both proteins also elute in well resolved peaks indicative of a single species. Phosphorylation neither promotes nor hinders oligomerisation of FhaA$^{367-527}$. The peak profiles of the differential refractive indices (solid lines) and the scattering at 90° (dashed lines) are highly correlated, indicating the offset between the different detectors is correct. Interestingly, the phosphorylated isoform of FhaA$^{367-527}$ eluted approximately 1 minute later than the unphosphorylated protein, indicating a structural compaction of the protein. The fact that the protein experiences a structural compaction upon phosphorylation provides additional evidence for an intra-molecular association between the FHA domain and the pThr-epitope on the linker. A very similar mechanism had already been reported for GarA, also a mycobacterial FHA domain containing protein [68].

5.4 Conclusions

The combined data described in this chapter show that PknB phosphorylates FhaA$^{367-527}$ at amino acid Thr377 and that this phosphorylation event causes an intra-molecular association between the phosphorylated threonine and the FHA domain.

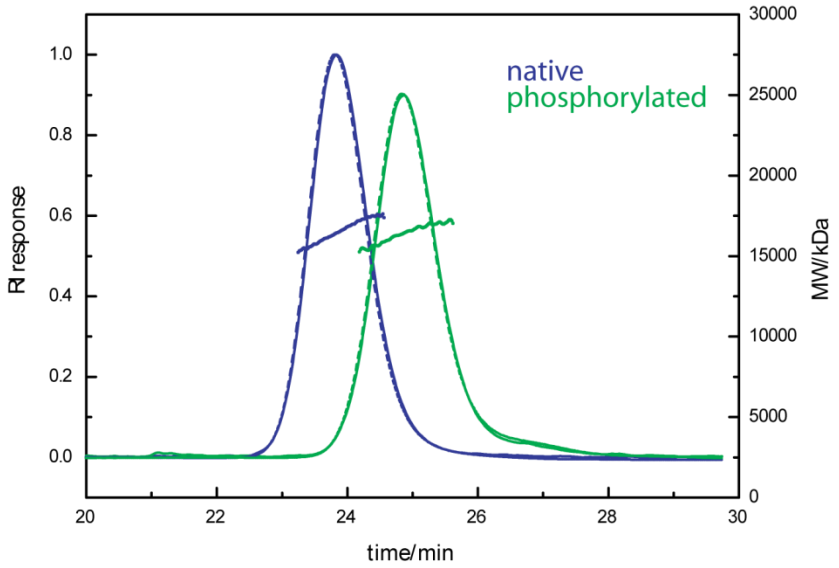
Kinase assays demonstrated that there is a phosphorylation site in the linker region of FhaA. The mass spectrometry revealed that for the fragment FhaA$^{367-527}$, the only site that gets significantly phosphorylated is Thr377. There may however be further phosphorylation sites N-terminal to amino acid 367. Furthermore, the phosphorylation event was not characterised in the context of the full length protein due to the instability or apparent insolubility of longer FhaA constructs. The substrate specificities of PknB may be different for FhaA$^{367-527}$ compared to full length FhaA. If it is assumed that they are not different, further conclusions can be drawn about the role of FhaA phosphorylation. From the SEC-MALLS experiment, a conformational compaction can

be seen upon phosphorylation, but both the phosphorylated and unphosphorylated forms of FhaA$^{367-527}$ are monomeric.

The ITC data reveals that phosphorylation at Thr377 abolishes the pThr-binding capabilities of the FHA domain. One may conclude that in the phosphorylated form of the protein, the pThr binding pocket of the FHA domain is occupied by intra-molecular association with pThr377 and therefore no peptide can be bound. In a mechanism similar to GarA, phosphorylation may function as a switch that inhibits the binding capabilities of the FHA domain. As the pThr attached to the FHA domain by the linker has an increased local concentration, it can out-compete the inter-molecular association with the phospho-peptide. Due to the intra-molecular link between the pThr and the FHA domain, even a weak interaction with a dissociation constant of 7 μM is sufficient for the regulation to be effective. This mechanism allows the phospho-binding capabilities of FhaA to be regulated by phosphorylation. Furthermore, it would also abrogate any potential phospho-independent interactions on the part of the FHA surface that is covered by the linker. The phospho-dependent interaction of FhaA and MviN is one that could be regulated by intra-molecular association of FhaA [127].

One difference between this mechanism and the one reported for GarA is that in the mechanism presented here the presence of the non-phosphorylated linker does not interfere with the phospho-binding of the FHA domain. The affinities of the phospho-peptide for FhaA$^{367-527}$ and FhaA$^{430-527}$ (minimal FHA domain) are identical within acceptable margins suggesting the non-phosphorylated linker does not interact with the FHA domain. In the case of GarA, phosphorylation only stabilises an already existing, phospho-independent interaction between an FHA domain and an N-terminal linker [68].

A



B

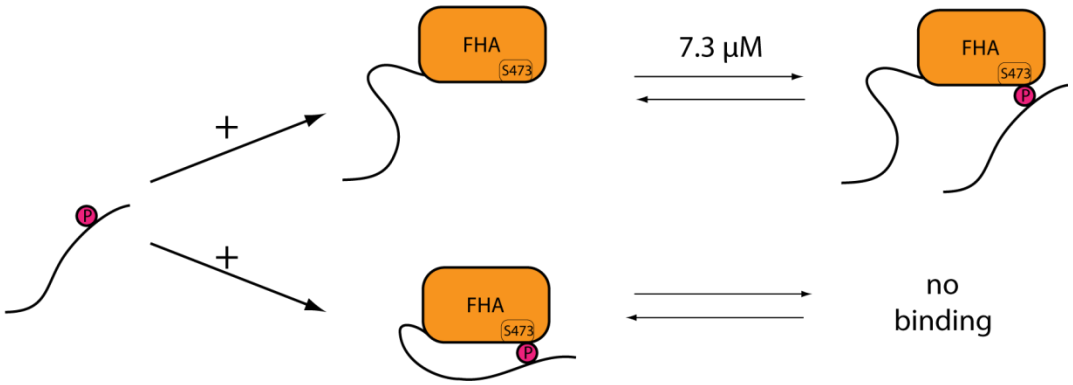


Figure 5.7. SEC-MALLS of FhaA and model of intra-molecular association (A) Size exclusion chromatograms of FhaA$^{367-527}$ (blue) and FhaA$^{367-527}$ pThr377 (green) coupled to multi angle laser light scattering. Differential refractive indices are shown in solid lines while light scattering at 90° is shown in dotted lines. (B) Proposed model for the structural consequences of phosphorylation of Thr377. Unphosphorylated FhaA binds the peptide with an affinity of 7.3 µM. In the phosphorylated form, the peptide is unable to bind as the phospho-binding residues, for example the conserved Ser473, are already occupied by the pThr377.

6. Structural and functional consequences of FhaA regulatory phosphorylation

With the overall mechanism of phospho-dependent intra-molecular association established, the structural details of the intra-molecular interaction remained to be determined. In addition, the effects of phosphorylation on the interaction between FhaA and MviN needed to be characterised. Chapter one introduced the fact that phosphorylated MviN binds the FhaA FHA domain in a canonical FHA-pThr interaction. It was also hypothesised that the putative flippase MviN, responsible for peptidoglycan precursor export, may be inhibited by FhaA [127]. As a first step, a structural description of the interaction between FhaA pThr377 and its FHA domain was sought using a two-pronged approach of solution NMR and X-ray crystallography. Subsequently, the interactions of MviN with phosphorylated and unphosphorylated FhaA was probed. The pThr377 epitope differs significantly in sequence from the optimal FHA substrate sequence reported previously [61]. An atomic description of this interaction would highlight any changes in FHA residues important for binding and would allow direct comparison to other FHA structures. Investigating how phosphorylation of Thr377 affects the binding of the FHA domain to MviN would indicate what effect the intra-molecular association reported in the previous chapter has in the context of the MviN-FHA protein complex.

6.1 Mapping the pThr377 interaction surface of the FhaA FHA domain by NMR

The interaction between the phospho-peptide and the FHA domain, described in the previous chapter, was assumed to be a canonical FHA-pThr interaction. This needed to be confirmed experimentally and the amino acids involved identified. To this end, $^{13}C/^{15}N$-doubly labelled $FhaA^{430-527}$ was produced and HSQC spectra of the protein in the free state as well as bound to 4 molar excess of unlabelled phospho-peptide, were recorded. By mapping the chemical shift differences between the free and the bound state onto the known structure of the FHA domain (3POA), the interaction surface could be determined. In an HSQC experiment, the chemical shifts of every amide proton and nitrogen are recorded. A change in the chemical shift of an amide upon addition of a ligand, caused by an altered magnetic environment, indicates that said

amide is involved in ligand binding or has experienced a conformational change upon binding.

Both the free and pThr-bound FHA domain produced spectra with well dispersed peaks indicative of a properly folded protein domain (figure 6.1). A total of 155 and 157 peaks were observed for the two forms respectively. The expression construct was 111 amino acids in length containing amino acids 430-527 of FhaA as well as the sequence G-P-L-G-S-A at the N-terminus and P-L-E-R-P-H-R-D at the C-terminus derived from the expression vector. HNCACB, HNCOCA, HNCA and CBCA(CO)NH spectra of the free protein were recorded in order to assign the backbone amide peaks of the HSQC spectrum. In these triple-resonance spectra, the C_α, C_β, and CO carbon chemical shifts of two consecutive residues are correlated with the amide chemical shift of the C-terminal amide. From this information, it can be determined which peaks in the HSQC are adjacent in the protein sequence. Using the order of the peaks and characteristic chemical shifts for each type of amino acid, each protein residue can be unambiguously assigned to its back-bone amide HSQC peak. In this case, a total of 80 amino acids were assigned their back-bone amide peak out of a total of 106 non-proline amino acids. Unassigned amino acids were mainly at the termini and between amino acids 474 and 484. Some of the unassigned peaks are also resonances from side-chain amides, specifically the δ amides of asparagines and the ε amides of glutamines. As these are primary amides, they appear as horizontal pairs in the spectrum with one ^{15}N shift correlated with two ^{1}H shifts. The protein also contains 11 arginine residues, whose ε nitrogens form secondary amines. The resonances from these arginine side chains are sufficiently far upfield to be outside the spectral range and appear out of phase in the spectrum. Amino acid Ser473, the conserved serine residue essential for pThr binding, had the most downfield ^{1}H-chemical shift at 11.8 ppm. This amino acid displays high chemical shifts in many FHA domains. In the FHA domains of proteins Ki67, Rv1827 and RNF8 for example, the amide proton chemical shift of the conserved **S**-X-X-H serine is approximately 11 ppm (BMRB accession codes 5959, 16248 and 11247).

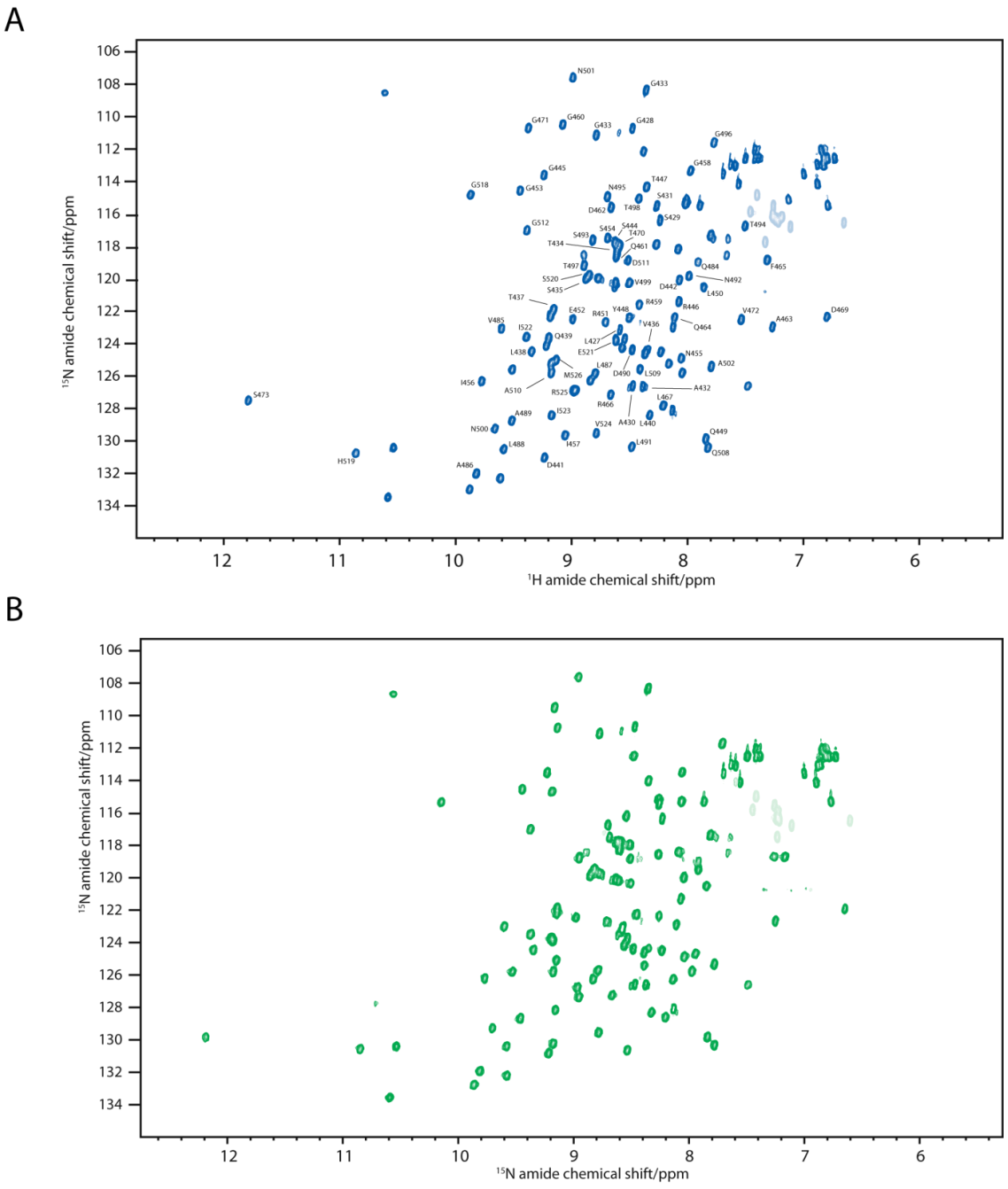


Figure 6.1. ^{1}H-^{15}N-HSQC spectra of FhaA$^{430-527}$ in the free form (A-blue) and bound to a 4 molar excess of peptide ADYpTRYTES (B-green). Both spectra were recorded at pH 7.0. For the free form, the assigned main chain amide peaks are labelled with the appropriate residue number. Peaks that are out of phase and have been folded in the ^{15}N-dimension are drawn semi-transparently.

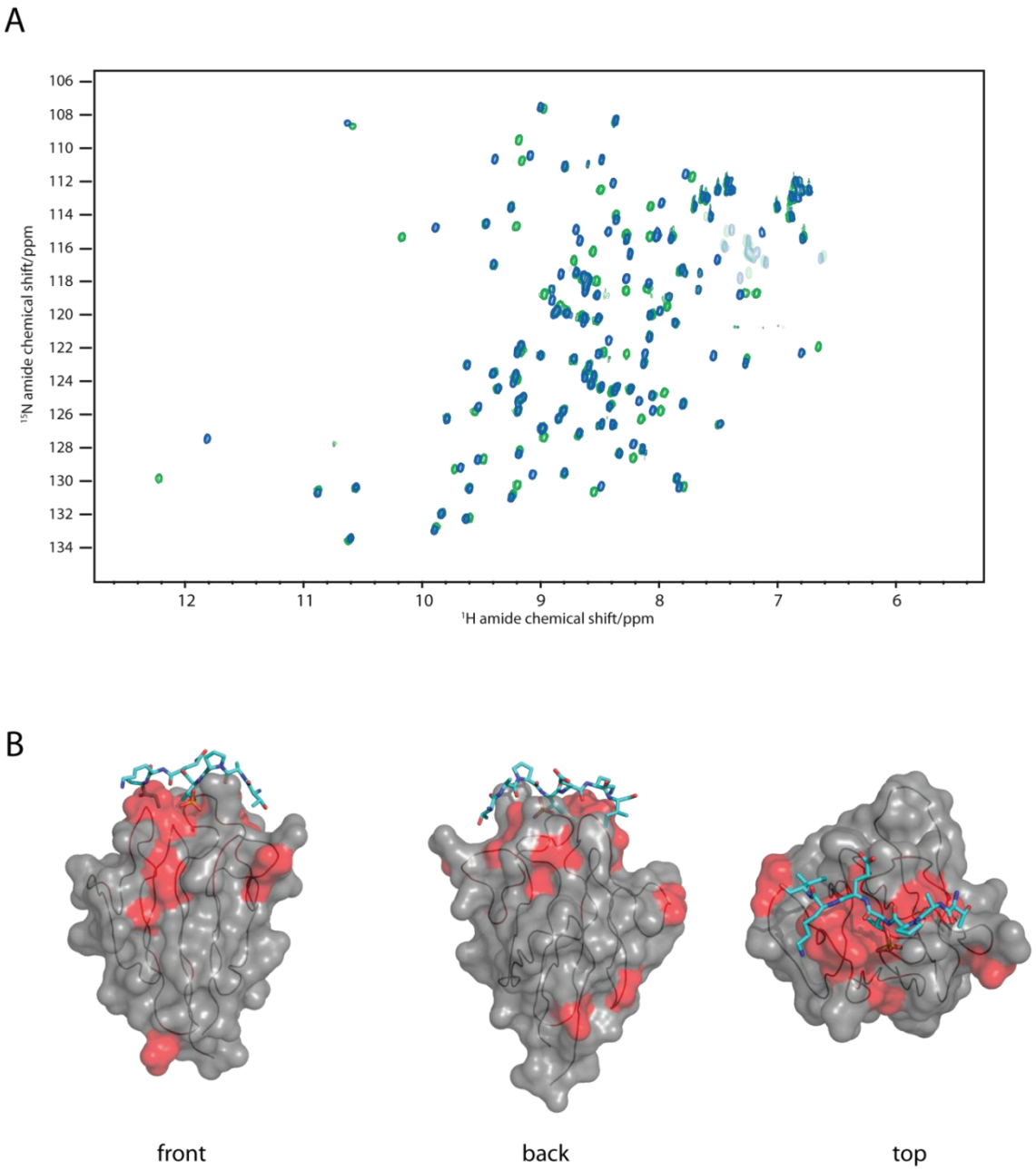


Figure 6.2. Overlay of ^{1}H-^{15}N-HSQC spectra of free FhaA$^{430-527}$ (blue) and bound to the phospho-peptide (green). Mapping of the residues that have altered chemical shifts upon peptide binding (red)onto the structure of the FHA domain in complex with its optimal peptide (PDB: 3POA).

There are significant differences in chemical shift for approximately one third of the peaks between the free and bound spectra of the FHA domain (figure 6.2). With around two thirds of the amides retaining their resonances, one can assume that the overall fold has not been altered, which would be expected for a canonical FHA-pThr interaction. The positions of the residues that have altered chemical shifts have been mapped onto the known crystal structure of the FHA domain. It is apparent that the majority of the residues that have moved are in the vicinity of where a phospho-peptide would be expected to bind. It is therefore likely that the FHA-pThr377 interaction is a canonical FHA binding event. As the spectrum of the bound form was not assigned, some peaks may have moved and swapped positions with each other and their shift perturbations will not have been detected. There are most likely several false negatives in figure 6.2B. While one can be certain that the amide peak of every amino acid coloured in red has moved, not all the residues coloured in grey will have unchanged chemical shifts.

6.2 Crystal structure of the FHA domain in complex with the phospho-peptide

In addition to NMR experiments, crystallisation trials of the FHA domain in complex with phospho-peptide were set up with the aim being a description of the binding event at atomic resolution to determine which FHA and peptide residues are involved in binding.

6.2.1 Crystallisation

The FhaA FHA domain in complex with the phospho-peptide produced well diffracting single crystals following a two-step crystallisation process. FhaA$^{430-427}$ was dialysed in a buffer containing 50 mM Tris pH 8.0, 300 mM NaCl, 1 mM DTT and 1mM EDTA, the same buffer in which this expression construct had been crystallised before [61]. The synthetic phospho-peptide with sequence A-D-Y-pT-R-Y-T-E-S was dissolved in the same buffer and mixed with the protein at a ratio of 1.2 : 1 (peptide: protein) to a final protein concentration of 8 mg/ml. The crystal trials of the protein-peptide complex were set up using the sitting drop vapour diffusion method at 18 °C. Crystalline plates formed after 6 weeks in 30% w/v PEG 4000, 100 mM sodium acetate pH 4.6 and 200 mM ammonium acetate (Qiagen Classics Screen condition H2). This crystal condition was fundamentally different to the crystals reported for the complex between the FHA

domain and its optimal peptide. There, the well solution consisted of 10% v/v 2-propanol, 0.1 M sodium cacodylate pH 6.5 and 0.2 M zinc acetate [61]. The plates were formed of multiple crystals and unsuitable for data collection. Therefore, the drop containing the crystals was diluted in 50 µl of well solution, homogenised by vortexing with glass beads for 30 seconds and used as a seed solution for microseeding. This microseed solution was used to set up further crystallisation trials that focused around the initial crystal hit. A 2D grid was set up where the precipitant concentration was varied along with the pH. Single crystals grew after 10 days in drops containing 1.5 µl protein/peptide, 1.5 µl well solution and 1 µl seed solution with 27% w/v PEG 4000, 100 mM sodium acetate pH 4.5 and 200 mM ammonium acetate. 30% PEG 4000 was needed to induce crystal nucleation, but in a solution where nucleation seeds were already present, 27% were sufficient for crystal growth. The crystals were flash frozen in liquid nitrogen using 20% w/v glycerol as a cryo-protectant.

6.2.2 Data collection and indexing

Diffraction data were collected at the IO3 beam line at Diamond Light Source. Following indexing of the crystal and integration of the data-set, the crystal was shown to belong to space group $P2_1$. The data was processed using the HKL suite of software (table 5.1)

Wavelength	0.9763 Å
Data collected	180° (360 x 0.5° oscillation)
Space group	$P2_1$
Unit cell	a = 41.89 b = 47.12 c = 51.72 α = 90° β = 112.34° γ = 90°
Mosaicity	0.8°
Resolution range	30.0 – 1.10 Å
Total reflections observed	251640
Unique reflections	72295
Average redundancy	3.5 *(2.6)*
Completeness (%)	96.1 *(76.6)*
I/σ	20.0 *(2.4)*
R_{symm}	0.064 *(0.37)*

Table 6.1. Processing statistics of FHA domain crystal (highest resolution shell, 1.15 – 1.10 Å, in brackets)

The protein crystal diffracted exceptionally well and data could be processed to a resolution of 1.10 Å. At this cut-off, the signal/noise ratio of the highest resolution bin was still acceptable at 2.4 and the R_{symm} lower than 40%. Calculation of the Matthews coefficient suggested there were 2 copies of the protein in the asymmetric unit

corresponding to a solvent content of 38% consistent with such a well-diffracting crystal.

6.2.3 Molecular replacement

The phases needed to solve the native data-set were obtained by molecular replacement. The crystal structure of the free FhaA FHA domain (PDB: 3PO8) was used as a search model. It had previously been shown that the differences in atomic positions between free form of the FhaA FHA domain and one bound to an optimal peptide were minimal with an rmsd of 0.3 Å for the Cα positions. It could therefore be assumed that the structures of free FHA domain and the FHA domain bound to the phospho-peptide would be sufficiently similar to allow molecular replacement to succeed. The search was carried out by the program Phaser (CCP4 suit of programs), which automatically calculates Patterson maps from the search model and from the experimental data. Then it performs an automated rotational and translational search in which the search model is moved around the unit cell and the calculated Patterson maps for each position compared against the Patterson map of the experimental data. The solutions are then checked and scored according to whether the resultant crystal packing conforms to the unit cell dimensions and if any clashes are observed. The search was performed looking for 2 copies per asymmetric unit and the program found one unique solution with 2 copies of the search model per asymmetric unit and high z-scores for both the rotation and the translation function. The z-score is an indicator of the probability phaser has found a correct solution with scores higher than 6 meaning a solution may have been found and scores higher than 8 meaning a solution has most likely been found. The z-scores of the rotation and translation functions were 16.3 and 17.7 respectively, indicating that the correct solution had almost certainly been found.

6.2.4 Model refinement

The unrefined model generated from the molecular replacement, containing just 2 copies of the FHA domain, but no peptide or waters, had an R_{work} of 40.4%. The $|F_O| - |F_C|$ difference maps displayed a large amount of unaccounted density near conserved residue Ser473 for both copies of the FHA domain where a bound phospho-peptide would be expected. This indicated that peptide was likely to be bound in the crystal. As a first step in the refinement process, the phospho-peptides were built into the density manually using Coot. Coot was also used for real-space refinement of the atomic model into the density, as the complex structure of the FHA domain was very similar, but not identical, to the one used as a molecular replacement search model. Several alternating rounds of manual model improvement with Coot and automated refinement

with Phenix_Refine were applied. Phenix_Refine was used to refine the atom coordinates and optimise the individual B-factors and atomic occupancies. Additionally in the first round of automated refinement, simulated annealing from 5000K to 300K was also used to lower the overall energy of the model. Due to the high resolution of the data, alternative conformations for several amino acid side-chains were observed and subsequently added to the model. Furthermore, individual, anisotropic B-factors were refined for each atom..

Protein atoms	1711
Waters	338
R_{work} %	16.1
R_{free} %	18.7
Rmsd bond-lengths (Å)	0.007
Rmsd bond angles (°)	1.4
Rotamer outliers %	2.2
Average B-factor (Å^2)	17.8
Ramachandran plot	
Most Favoured %	96.9
Additional allowed %	3.1
Disallowed %	0

Table 6.2. Refinement statistics of FHA domain crystal

6.2.5 The Model

At the end of the refinement process, the overall R_{free} was reduced to 18.7%, indicating that disagreement between the observed reflections and the reflections calculated from the model had been reduced to an acceptable level for data at this resolution. There were 2 copies of the peptide-bound FHA domain arranged side by side in the asymmetric unit (figure 6.3). For the FHA domains, amino acids 433-527 are visible. In addition, two of the residues C-terminal to the native protein sequence that were added during cloning could also be built into the model. Of the peptide sequence A-D-Y-pT-R-Y-T-E-S, amino acids D-Y-pT-R-Y-T could be built into the density for one copy and D-Y-pT-R-Y-T-E for the other. The average B-factors for the two FHA domains were 14.6 Å^2 and 14.2 Å^2 respectively. For the two peptides, these values were 27.4 Å^2 and 23.7 Å^2, indicating that the positions of the peptide atoms were less well defined (figure 6.3). For both copies of Gln484, the side-chain was truncated to C_α and C_β only as no density could be observed for the side-chain amide. Similarly, the side chain of Arg378 of chain C was truncated to a C_α as no electron density could be detected for these atoms. These three amino acids aside, all other atoms expected from the amino acid

sequences were built into the model. Furthermore, 338 water molecules with an average B-factor of 31.0 $Å^2$ could also be detected.

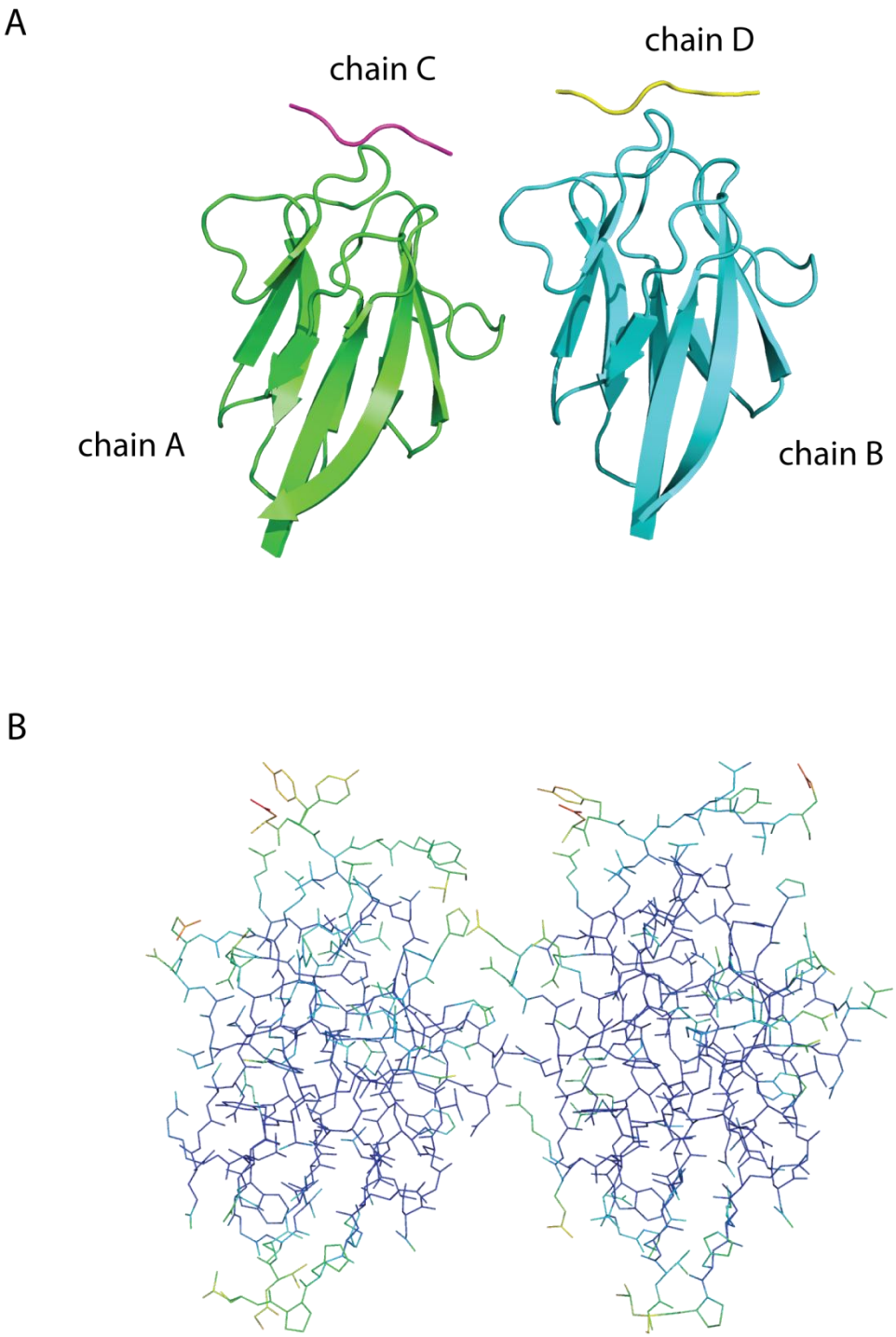


Figure 6.3. Model of the refined asymmetric unit of the FHA domain crystal. A: cartoon representation of the asymmetric unit. There are 2 copies of the FhaA FHA domain (green and cyan) and 2 copies of the phospho-peptide (magenta and yellow) bound to the FHA domains. B: Stick representation of the asymmetric unit with every atom coloured according to its B-factor. The B-factors range from 10 Å^2 (blue) to 50 Å^2 (red).

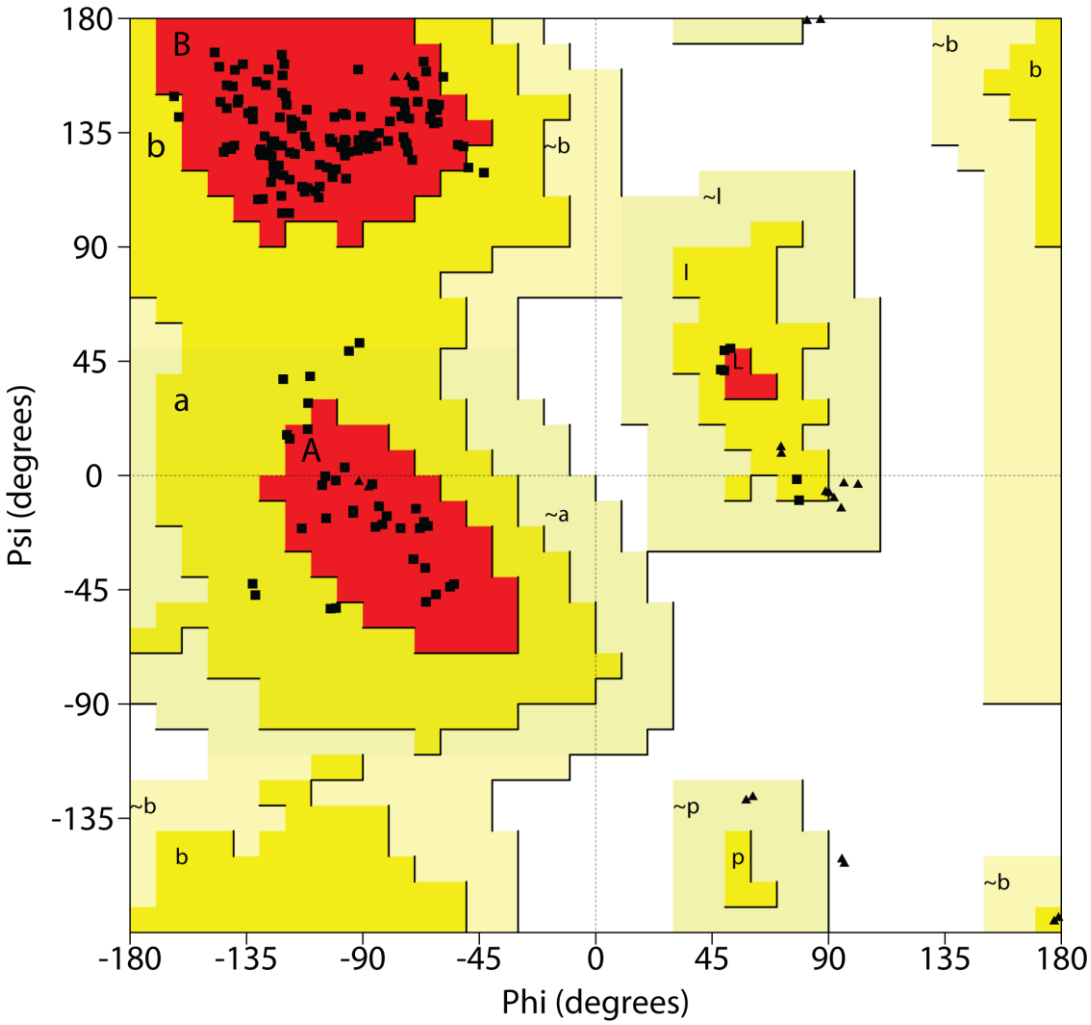


Figure 6.4. Ramachandran plot of the protein model with the most favoured regions (A,B,L) coloured red, additionally allowed regions yellow (a,b,l) and generously allowed regions beige (~a,~b,~l). The A, B and L regions correspond to α-helices, β-sheets and loops respectively. Non-glycine amino acids are represented by squares and glycine residues are shown as triangles.

Residue	Ψ	Φ	
Gln484	-47.1 -42.6	-131.6 -132.6	chain A chain B
Leu491	49.9 52.2	-95.4 -91.3	chain A chain B
Asn501	-9.8 -2.2	78.6 78.8	chain A chain B

Table 6.3. Dihedral angles of Ramachandran plot outliers

From the Ramachandran plot it is apparent that the dihedral angles of the majority of residues are in the “favoured” regions with only a small minority in the “additionally allowed” regions (figure 6.4). No residues are in the “disallowed” regions. There are three residues per FHA domain which lie in regions of the Ramachandran plot which are only ‘additionally allowed’ (table 6.3). Considering the high resolution of the data and the fact that the dihedral angles of these residues are consistent between chains A and B and the residues fitted well into the density it was assumed that these residues do, in fact, take these conformations in the protein and were not built into the model erroneously.

6.2.6 The interaction of the phospho-peptide with the FHA domain

From a functional point of view, the structure shows that the interaction between the FHA domain and the phospho-peptide is indeed a canonical FHA-pThr interaction. The FHA domains are folded into the expected β-strand sandwich topology (figure 6.5). The differences between this structure of the FhaA FHA domain and the one reported in 3POA are minimal. For C_α of the FHA domain, the average (XYZ) rmsd between 3POA and this structure is 0.49 Å. When taking into account all atoms of the FHA domain, this value increases to 1.07 Å. The structural similarities are apparent from a C_α ribbon overlay and demonstrate that even the loops follow virtually identical paths (figure 6.4). At this resolution and with such small B-factors, the position of every atom could be assigned and the model could be built into the electron density map without any ambiguity (figure 6.6). From this image, it is apparent how pThr377 sits in between the side-chains of Ser473 and Arg474. The interatomic distances, particularly between the FHA domain and the peptide-ligand could also be determined with a high degree of accuracy. As predicted from the ITC experiments, where the interaction of the FHA domain with the peptide was highly similar to that reported for a pThr flanked with poly-alanines, no interactions are made between side-chain atoms of the peptide and the FHA domain with the exception of pThr377. The main contributors to binding are hydrogen bonds, of which the majority are formed by the phosphorylated threonine (figure 6.7). The phosphate moiety hydrogen bonds with FHA residues Arg459, Ser473, Arg474 and T494. Of these residues, Arg459 and Ser473 are absolutely

conserved in FHA domains. An arginine at the +1 position relative to the conserved serine is not absolutely conserved in FHA domains, but can be found in many examples and particularly all such domains in the species *Mycobacterium tuberculosis*. Additional hydrogen bonds are made between the back-bone of the peptide with residues Arg459, Thr470 and Asn495.

The residue in the +3 position relative to the pThr, which is usually a main contributor for peptide-binding specificity, does not make any substantial contacts [61]. The side chain of Thr380 does not interact with the FHA domain.

The crystal structure also explains the relatively weak dissociation constant between the FHA domain and the pThr377 peptide when compared to the other FhaA FHA domain structures (figure 6.8). In the two previously reported complex structures of this domain, the residues in the +3 position relative to the bound pThr were leucine and an isoleucine respectively (PDB: 3OUN, 3POA). The affinities of these peptides for the FHA domain were 0.1 μM each, consistent with data from peptide library screens that identified medium size hydrophobic amino acids (Iso, Leu, Val, Phe) in the +3 position as strongly preferred [61] (section 1.3.3). When comparing the three structures, it is apparent that a threonine in the +3 position, with its comparatively smaller and polar side-chain cannot make the hydrophobic interactions that the isoleucine and leucine make that greatly stabilise the interaction. The overall weaker binding of the pThr377 peptide of 7.0 μM is most likely mainly caused by the lack of a hydrophobic residue at the +3 position. An overlay of the structures reveals that the conformations of the FHA domain residues are relatively independent of the peptide bound, the main exception being His519 which can be found in a different position for each of the three peptides.

Overall, the structure reveals the FHA-peptide binding to be a canonical FHA domain interaction, but with no extensive contributions from any peptide residues except for the phosphorylated threonine.

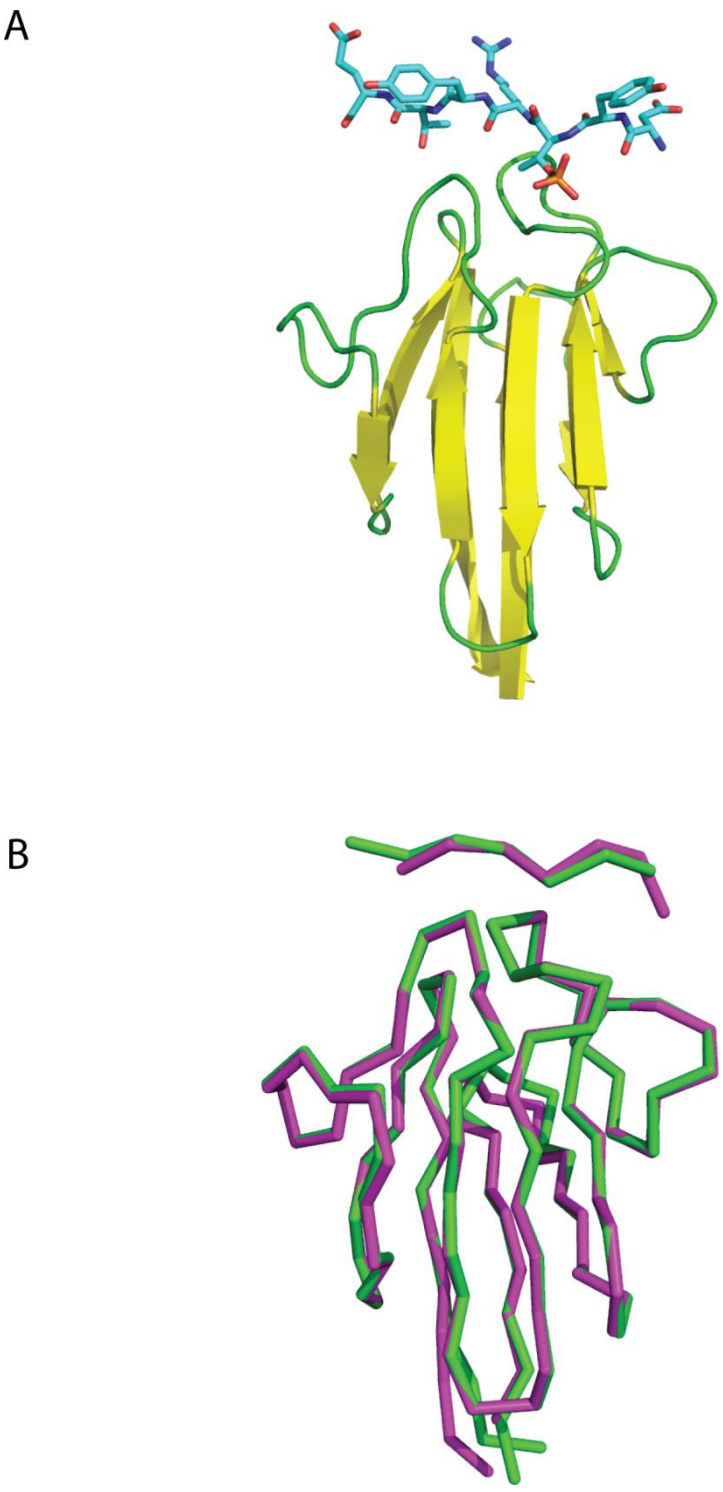


Figure 6.5. Cartoon representation of FHA domain complex and comparison with optimal peptide structure A: Cartoon representation of chain B using standard colouring and chain D shown as sticks. B: Ribbon overlay of B and D (green) against the structure of the FHA domain bound to its optimal peptide (magenta). Optimal peptide-structure taken from PDB, accession number 3POA.

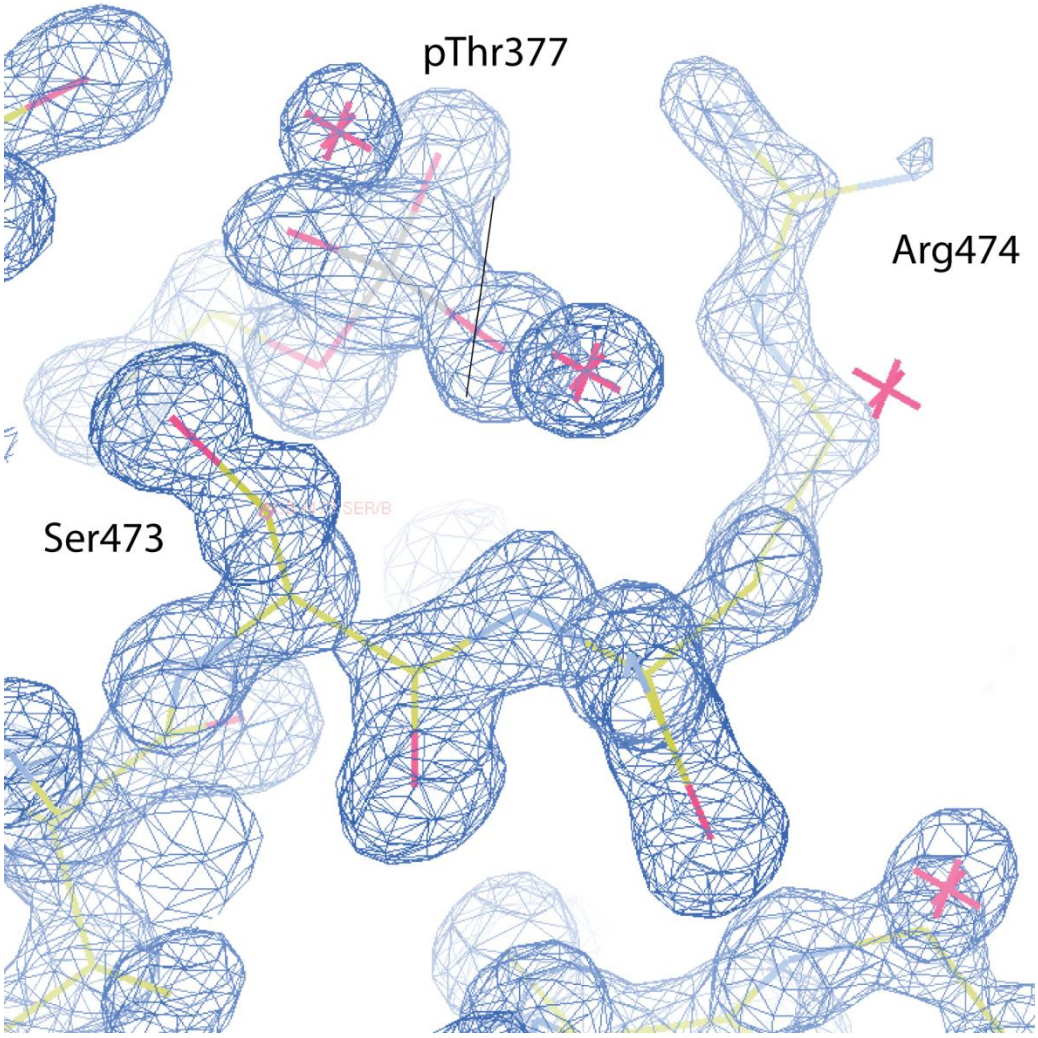


Figure 6.6. Image of the FHA domain and peptide models built into the experimental electron density showing pThr377 of chain D bound in the pThr binding pocket of chain B. The electron density was contoured at 1.5 σ. Red crosses represent water molecules

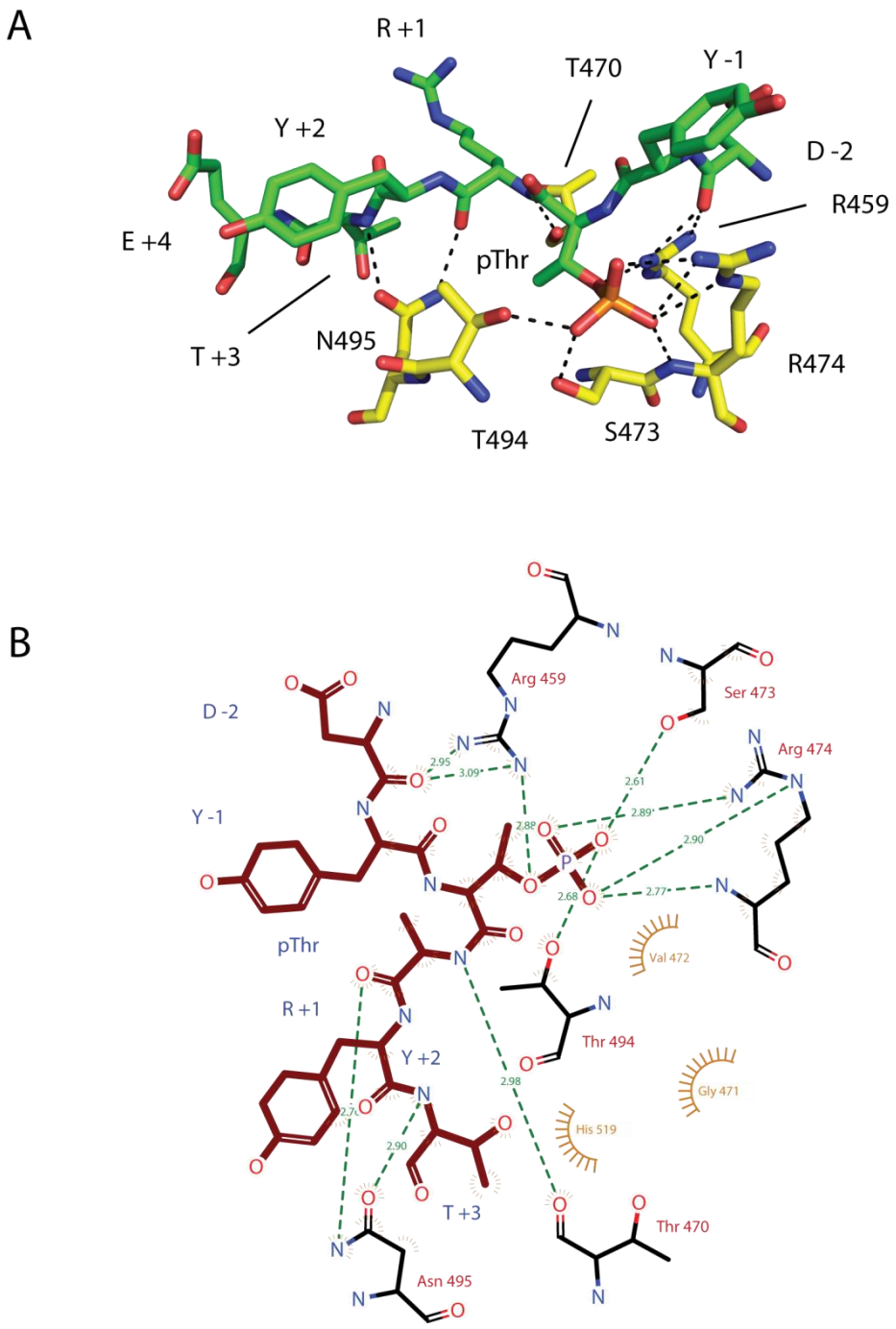
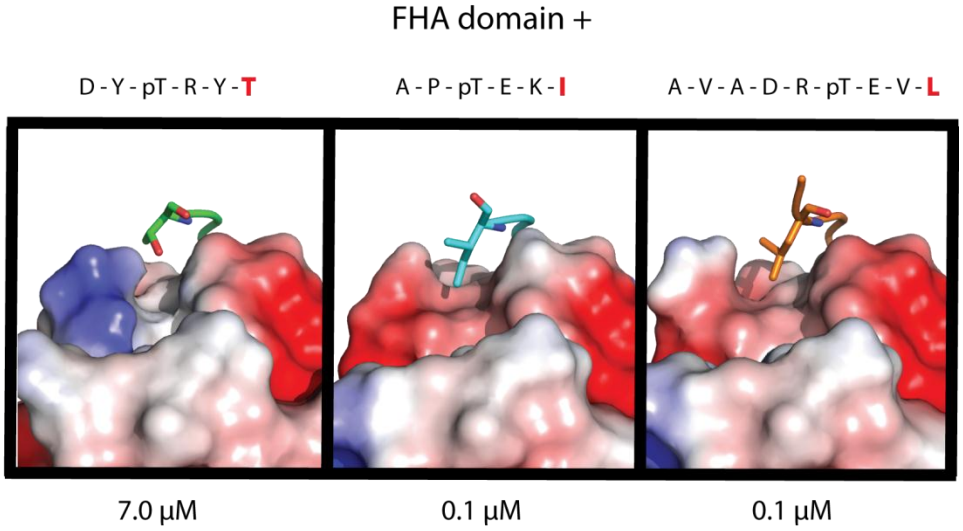


Figure 6.7. The interaction of the peptide with the FHA domain. A: Structure of the phospho-peptide (chain D - green) and the conserved residues of the FHA domain (chain B-yellow). The residues of the peptide are numbered relative to the pThr. B: Schematic of the FHA-pThr interaction generated using the program LIGPLOT following the same residue numbering as above. Green lines represent hydrogen bonds with their respective lengths given in Å. FHA domain residues forming hydrophobic interactions with the peptide are shown as orange semi-circles with spikes.

A



B

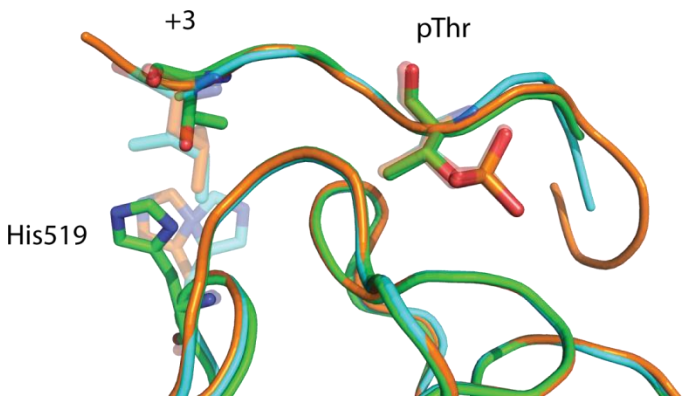


Figure 6.8. The interaction of the FhaA FHA domain with different phospho-peptides. (A) Crystal structure of the FHA domain in complex with the ADYpTRYTES peptide (left panel, green), its optimal peptide (central panel, cyan) (PDB: 3POA) and the MviN pThr epitope (right panel, orange) (PDB: 3OUN) The phospho-peptides are drawn as ribbon with the extreme C- and N- termini removed for clarity and the FHA domains are drawn using a surface representation with the colours referring to the calculated charges on the surface of the protein. The +3 residue relative to the pThr, which is known to be important for ligand specificity, are drawn as sticks. Below are the dissociation constants of the respective phospho-peptides for this FHA domain. (B) Overlay of ribbon representations of the FHA domain in complex with the different pThr epitopes. Colours as above. The pThr and the +3 peptide position are shown as sticks. His519 of FhaA, in a different conformation in each of the three structures, is also shown as sticks.

6.3 The effect of FhaA phosphorylation on the interaction between FhaA and MviN

Since it had been reported that FhaA binds MviN in a phospho-dependent interaction, it seemed likely that phosphorylation of FhaA residue pThr377 may have an effect on the interaction between the two proteins [127]. If the intra-molecularly bound pThr377 occupies the pThr binding pocket of the FHA domain the ability of the FhaA FHA domain to bind MviN via phosphorylated Thr947 could be impeded. To investigate any such effects, the binding parameters for the interactions between the phosphorylated and unphosphorylated isoforms of both proteins were measured by ITC.

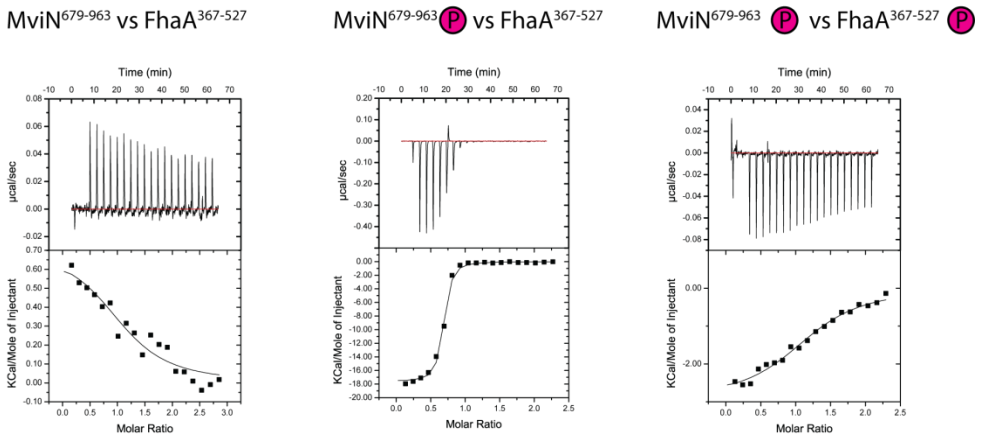
6.3.1 Measuring the binding affinities of the FHA-MviN interaction by ITC

The proteins used in the ITC experiments were expressed and purified as described in Chapter 2. In preparation for the titrations, all proteins were dialysed into 150 mM NaCl, 20 mM Tris pH 8.0 and 0.5 mM TCEP. Since the FhaA construct was known to be soluble to high concentrations, it was decided to use MviN as the analyte and FhaA as the titrant. For FhaA, the $FhaA^{367-527}$ construct was used. The intra-cellular pseudo-kinase domain including a small C-terminal extension was used to represent MviN. This construct spanned amino acids 679-963 and included the PknB phosphorylation site Thr947

$FhaA^{367-527}$ bound $MviN^{679-963}$ with an affinity of 4.5 μM. Phosphorylation of $MviN^{679-963}$ on Thr947 greatly stabilised the interaction (figure 6.9). Confirming the findings of Gee et al. [127], phosphorylated $MviN^{679-963}$ binds $FhaA^{367-527}$ with an association constant of 0.08 μM. Phosphorylation of $FhaA^{367-527}$ on the other hand greatly destabilised the interaction with the affinity reduced to 1.5 μM, similar to that observed for the unphosphorylated forms. The apparent 3 fold increase in affinity is within the expected error for such noisy titration curves. Binding between $FhaA^{367-527}$ and $MviN^{679-963}$ is therefore mediated by 2 interactions, a weak phospho-independent interaction and a tight phospho-dependent one which can be interrupted by phosphorylation of FhaA itself. While the weak phospho-independent interaction is largely driven entropically, the tight phospho-dependent interaction has a strong (-17.6 kcal/mol) enthalpic element. Sharp decrease in affinity upon phosphorylation of pThr377 indicates that the intra-molecular association is able to outcompete an inter-molecular interaction despite it being 2 orders of magnitude tighter. Although the stoichiometries of the interactions vary between 0.64 and 1.23, it was assumed that all interactions occur with at a ratio of 1:1. The varying stoichiometries of the titrations were thought to have been due to

protein aggregation which can reduce effective protein concentrations. In addition, the proteins may not have been phosphorylated to completion giving rise to inaccurate stoichiometries. The sequence surrounding the MviN phosphorylation site is A-V-A-D-R-**T_{947}**-E-V-L-G-P. As it contains a medium size hydrophobic residue in the +3 one would expect this motif to bind tightly to the FhaA FHA domain. The optimal binding peptide for this domain derived from peptide-library screens was D-T-A-P-pT-E-K-I and bound with an affinity of 0.10 μM [61]. Although the MviN sequence is similar to the optimal peptide and also contained a favoured medium sized hydrophobic residue in the +3 position, it was thought that further contacts may be made between the pseudo-kinase domain and the FHA domain that contribute to binding. The enthalpic contribution of -17.6 kcal for MviN binding was significantly greater than the -14.3 kcal measured for the optimal peptide. The crystal structure reported for the MviN-FhaA complex shows that the NH1 of Arg459 of FhaA is within 3 Å of the main chain carbonyls of Ala492 and Asp495 indicating that hydrogen bonds may be formed between these residues. (figure 6.9, PDB: 3OUN).

A



	K_D (µM)	N	ΔH (kcal/mol)	TΔS (kcal/mol)
MviN $^{679-963}$ + FhaA $^{367-527}$	4.5	1.14	0.7	7.8
MviN $^{679-963}$ pThr 947 + FhaA $^{367-527}$	0.08	0.64	-17.6	-8.1
MviN $^{679-963}$ pThr 947 + FhaA $^{367-527}$ pThr377	1.4	1.23	-2.9	4.9

B

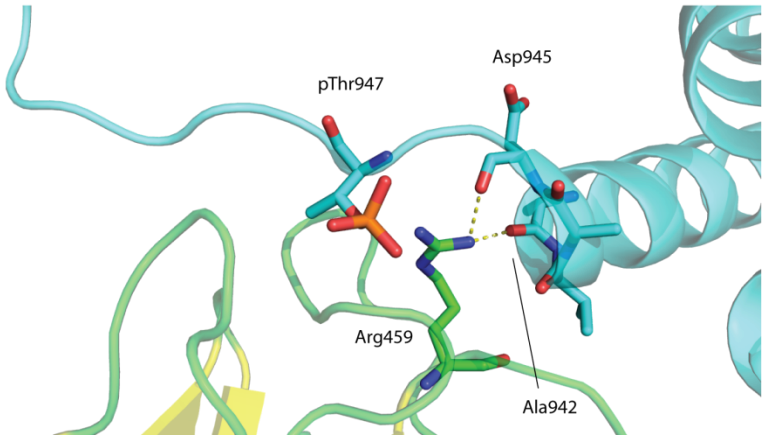
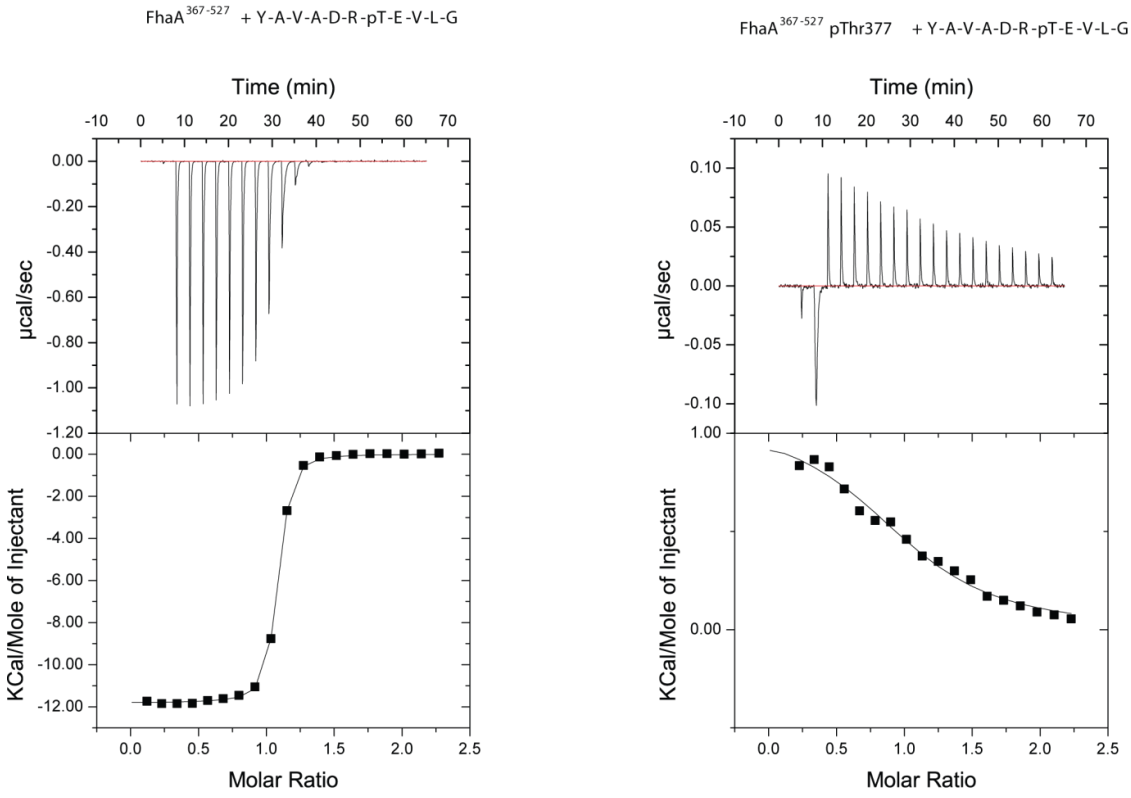


Figure 6.9. ITC analysis of the interaction between FhaA and MviN. (A) Raw calorimetric data, heat changes and fitted binding curves for the interaction between the respective constructs and isoforms of FhaA and MviN and their thermodynamic parameters. (B) Cartoon representation of MviN-FhaA complex (PDB: 3OUN) with the FHA domain coloured in green and yellow and MviN in cyan. pThr947, Asp945 and Ala942 of MviN and Arg459 of FhaA shown as sticks. Putative hydrogen bonds between Arg459 and MviN shown as yellow lines.

6.3.2 The affinity of the FHA domain for a MviN derived phospho-peptide

To ascertain whether the high affinity of the MviN-FhaA interaction was partly caused by contacts due to the secondary structure of MviN, a synthetic phospho-peptide mimicking pThr947 was used instead of MviN in repeat titrations. This peptide had the sequence Y-A-V-A-D-R-pT-E-V-L-G and was assumed to be free of any secondary structure. Generally, the peptide-protein ITCs followed the same pattern as the protein-protein ITC experiments (figure 6.10). The affinity of native FhaA for the MviN phosphorylation site is approximately 0.1 μM and phosphorylation at Thr377 weakens the apparent affinity of FHA domain for the peptide considerably. Upon closer inspection however there are some differences. The protein-protein interaction is accompanied by a large enthalpic component (-17.6 kcal/mol), but incurs an entropic penalty of 8.1 kcal/mol at 293 K. For the peptide-protein interaction, both the favourable enthalpic component and the unfavourable entropic component are decreased. Since they are both decreased by similar amounts, the overall equilibrium constant remains unchanged. This does not however provide evidence whether MviN secondary structure contributes to the binding event. One could deduce that the decrease in enthalpic contributions may be caused by a loss of hydrogen bonds formed between FHA domain and ligand but that is highly speculative. The highly unfavourable entropic component of the protein-protein interaction may be due to a loss of conformational freedom of the MviN pThr epitope upon binding. The phospho-independent interaction between FhaA and MviN which had a ΔH of only 0.7 kcal/mol is unlikely to contribute significantly to the large enthalpic component of the strong protein-protein interaction. One can only conclude that the two binding events, despite their similar binding constants, are different and that the pThr947 motif has an affinity for the FHA domain as great as the MviN pseudo-kinase domain and the library-derived optimal peptide. Another apparent difference between the two sets of ITC experiments is that while the protein-protein interaction was weakened 20-fold, the peptide-peptide interaction decreased by a factor of 100. As the binding curves for the interactions between the phosphorylated isoforms (or its mimics) are not sigmoidal due to the weak binding, the binding constants calculated from these curves are likely to be inaccurate and a 5-fold difference may be within experimental error.



	K_D (µM)	N	ΔH (kcal/mol)	TΔS (kcal/mol)
FhaA$^{367-527}$ + Y-A-V-A-D-R-pT-E-V-L-G	0.09	1.03	-11.8	-2.4
FhaA$^{367-527}$ pThr377 + Y-A-V-A-D-R-pT-E-V-L-G	8.4	1.03	1.1	7.8

Figure 6.10. ITC analysis of the interaction between FhaA and a MviN derived phospho-peptide with sequence YAVADRpTEVLG. Raw calorimetric data, heat changes and fitted binding curves for the interaction between the respective constructs and isoforms of FhaA and the peptide and the thermodynamic parameters of these interactions.

6.4 Modelling the intra-molecular association

From the two sets of ITC experiments, it is possible to calculate the dissociation constant of the intra-molecular association between the pThr377 motif and the FHA domain. As this interaction is intra-molecular and hence independent of protein concentration, this dissociation constant is dimensionless. The constant measures the probability of the intra-molecular association to be either in an open or closed conformation and determines by how much the effective concentration of FhaA is decreased upon phosphorylation at Thr377.

The interaction between unphosphorylated FhaA and phosphorylated MviN was modelled as a simple equilibrium whose parameters are known from the ITC titration (figure 6.11). It was shown in Chapter 5 that the affinity of the FHA domain for a phospho-peptide is independent of the presence of a non-phosphorylated N-terminal linker. The probability of this non-phosphorylated linker being in a closed conformation is thus negligible. For the interaction between phosphorylated FhaA and phosphorylated MviN, the system becomes more complex. Here, there is an intra-molecular association as well as inter-molecular binding. For the two proteins to interact, the FhaA needs to be in an open conformation. The energy of the intra-molecular association needs to be overcome before FhaA can interact with MviN. The Gibbs free energy of the intra-molecular association can therefore be calculated as the difference between the energies of the interactions of MviN with native and phosphorylated FhaA. From the peptide-protein ITCs, the ΔG of the intra-molecular association was calculated to be 2.7 kcal/mol in favour of the closed conformation corresponding to an equilibrium constant of 0.01 ($\Delta G = -R*T*\ln K$). This means that there is only a 1% chance of FhaA being in an open conformation at any one time. The energetic contributions to this equilibrium show that the open conformation is entropically more favourable, while the closed conformation is enthalpically more favourable. The enthalpic component can be explained by the extensive hydrogen bonding found between the pThr377 epitope and the FHA domain (figure 6.7). The entropic penalty upon binding may be explained by the loss of conformational freedom experienced by the peptide upon binding. In support of this, it has been shown that the FHA domain suffers a reduction in internal mobility upon optimal peptide binding [61].

Using the same model, but the data from the protein-protein ITCs, the intra-molecular association can be calculated to be in open conformation 5% of the time. Using these data, the ΔG of the equilibrium was measured to be 1.7 kcal/mol with similar enthalpic

and entropic contributions. It was thought that this estimation may be less accurate than the 1% calculated previously as the protein-protein ITCs were generally noisier than the protein-peptide ITCs. In addition, there is also a phospho-independent interaction between FhaA and MviN that was not incorporated into the model. A titration of phosphorylated FhaA into native MviN would have given some indication of whether the phospho-independent interaction is dependent on the intra-molecular association. However, it was thought that the small signal generated by the phospho-independent interaction would not have been sufficient to obtain good quality data that would enable such conclusions to be made. The peptide-protein ITCs may not fully describe the interaction between FhaA and MviN, but for the purpose of modelling the intra-molecular association they represent a more rigorous experimental system.

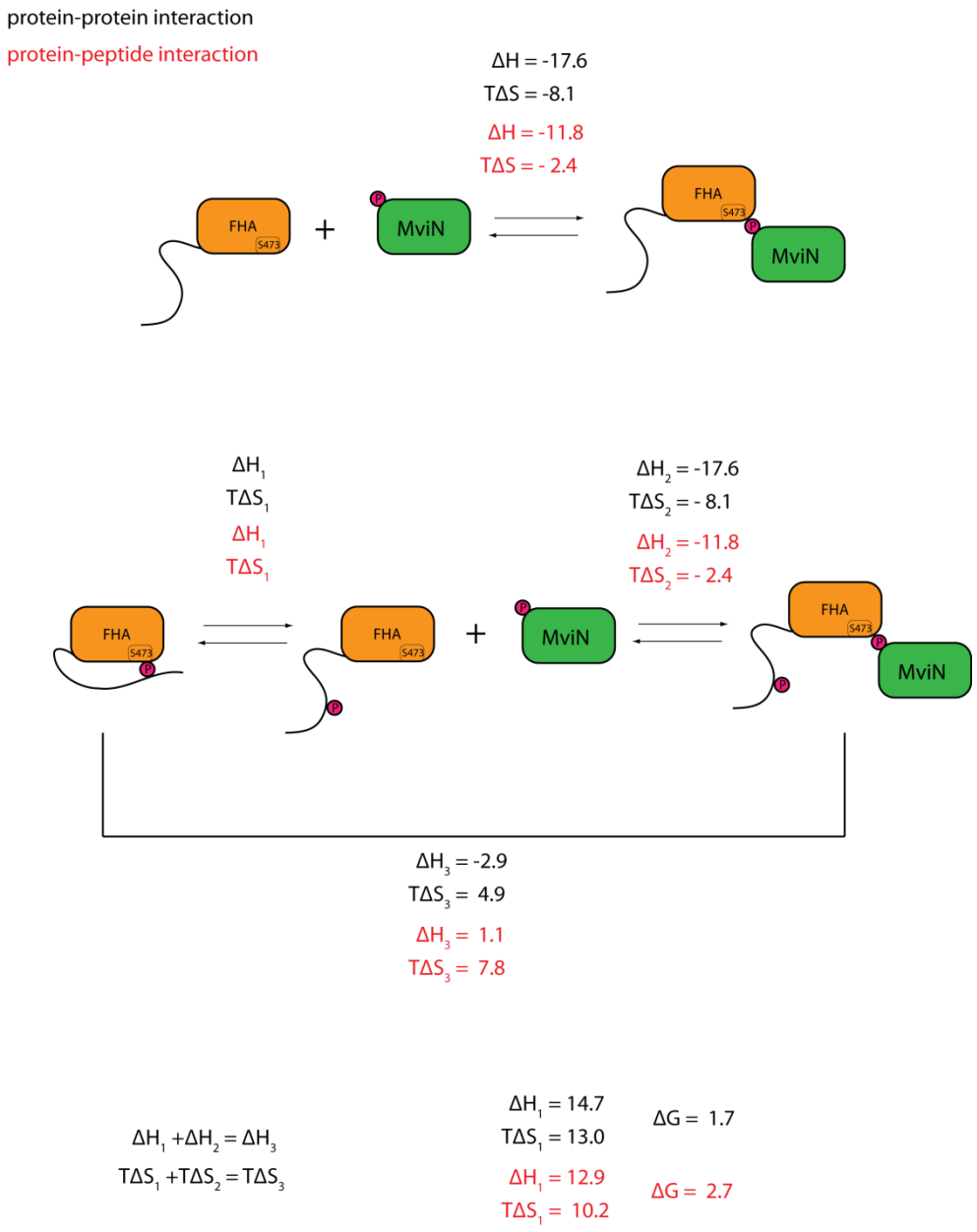


Figure 6.11. Modelling of the FhaA intra-molecular association and the interaction between FhaA and MviN. The binding parameters taken from the protein-protein ITCs are in black and those from the peptide-protein ITCs are in red. In the first system, there is an equilibrium between FHA and MviN, and the complex they form. In the second sytem, the interaction of phosphorylated FhaA with MviN is modelled as two consecutive steps. There is one equilibrium between the open and closed form of FhaA and one between the open form of FhaA and its complex with MviN. Since we know the thermodynamic constants for the second equilibrium and the overall system, the constants for the intra-molecular equilibrium can be calculated.

6.5 Conclusions

From the experiments described concerning phosphorylation of FhaA by PknB, one can deduce that an intra-molecular association, triggered by the phosphorylation event, regulates FhaA function. Upon phosphorylation at pThr377, the FHA domain binds the pThr motif and subsequently loses its ability to bind its interaction partner. Unphosphorylated FhaA binds unphosphorylated MviN with a dissociation constant of 4.5 μM via a so far unknown mechanism. Once MviN is phosphorylated by PknB at Thr947, the two proteins interact in a canonical pThr-FHA interaction with a binding affinity 0.08 μM as reported previously [127]. Phosphorylation of FhaA decreases the affinity of FhaA for MviN to a level comparable to the unphosphorylated isoforms (1.5 μM). Taking into consideration the mechanism reported in chapter 5, one concludes that through this intra-molecular association, FhaA loses its ability to interact with MviN. An intra-molecular interaction with an apparent K_D of 7 μM is able to outcompete an inter-molecular interaction that is 2 orders of magnitude tighter. The crystal structure shows this FHA-pThr377 interaction to be a canonical FHA interaction. It also confirms that none of the side-chain atoms of the pThr377 motif, other than pThr377 itself, contribute to the binding. The intra-molecular association is therefore independent of the motif surrounding the phosphorylated residue. An auto-inhibitory mechanism that is mediated through an intra-molecular association does not need to be very specific as the rate of sampling for the intra-molecular association is greatly increased. The differences in amide chemical shifts of the FHA domain upon peptide binding provide additional evidence that the intra-molecular association is a canonical FHA binding event.

It has been hypothesised elsewhere that FhaA may act as an inhibitor of MviN, a flippase that inverts and exports peptidoglycan precursors across the mycobacterial cell-membrane [127]. If this is true, then phosphorylation at Thr377 provides an additional level of control at this step of mycobacterial cell-wall synthesis, with the phosphorylation event relieving inhibition of MviN.

FhaA may function similarly to GarA, where a phospho-dependent intra-molecular association is utilised to relieve the regulatory effect this protein has on a number of enzymatic targets. One major difference is that the GarA domain is able to bind and inhibit unphosphorylated enzymes [68]. The pThr binding pocket of GarA can bind unphosphorylated binding partners as well as the intra-molecularly bound pThr.

Conversely, FhaA requires its enzymatic target to be phosphorylated for efficient binding and engages in a canonical pThr-FHA interaction.

There are several remaining questions concerning this model of FhaA regulation. Firstly, there is some conservation of the Thr377 site (figure 4.1), none of the residues surrounding the threonine are involved in the interaction with the FHA domain. Essentially, any phosphorylated threonine on a linker which is covalently linked to an FHA domain could function in such a fashion. Furthermore, the A-D-Y-pT-R-Y-T motif is not one that is favoured as a substrate by PknB. The exact reasons for any conservation of this motif remain to be determined.

Secondly, it remains unclear why the action of one kinase can both stabilise one interaction, by phosphorylating MviN, but also destabilise the same interaction by phosphorylating FhaA. One possible explanation is that phosphorylation acts a timer, where the membrane-bound STPK PknB phosphorylates the membrane-bound flippase MviN, which then recruits the soluble protein FhaA via pThr947. Once FhaA is bound, it is also attached to the cell-membrane and becomes a more attractive substrate for PknB. While FhaA will inhibit MviN for a while, prolonged activity of PknB will cause the majority of FhaA to be phosphorylated and hence relieve the inhibition. From the binding constants calculated for the intra-molecular association, we can assume the effective concentration of FhaA to be reduced by a factor of 100 upon phosphorylation.

7 The human protein PIH1D1 contains a pSer/pThr-reader domain

The second part of this thesis focuses on the characterisation of the human protein PIH1D1. In section 1.5 it was introduced how this protein, being part of the R2TP complex, is essential in the assembly and maturation of PIKKs. As part of the R2TP complex, it forms a link connecting PIKKs mTOR and SMG1 with the chaperones essential for their stable expression, namely Hsp90 and the prefoldin-like complex. While the PIKKs TRRAP, DNA-PK, ATM and ATR can most likely interact with Hsp90 directly, mTOR and SMG1 require PIH1D1 for their maturation. It had been shown how the PIKK co-factor Tel2 binds to PIH1D1 directly in an interaction that is essential for mTOR and SMG1 expression. This interaction is dependent on two constitutively phosphorylated serine residues on Tel2 suggesting the PIH1D1 protein contains a phospho-binding domain [147]. The phospho-dependent interactions of PIH1D1 were investigated in a collaborative effort where the *in vivo* search for interacting partners was carried out by Zuzana Horejsi and Simon Boulton (CRUK, Clare Hall, authors of the study mentioned above) whilst the characterisation of these interactions *in vitro* and the structure determination of PIH1D1 were the subject of the second part of this thesis. If more proteins were shown to bind PIH1D1 in a phospho-independent fashion, it would suggest that this protein is involved in the stable expression of more proteins than just PIKKs. In addition to the conserved CK2 sites on Tel2 reported previously, PIH1D1 was shown to also bind to snRNP 116 (small nucleolar ribonucleoprotein 116 kDa) and ECD [ecdysoneless homolog (Drosophila)], a putative transcriptional regulator, [171]. The protein snRNP 116 is part of the spliceosome responsible for pre-mRNA splicing [172].

The interacting proteins were initially identified by mass-spectrometry following immunoprecipitation. Conserved CK2 sites similar in sequence to the Tel2 site were then identified from phospho-proteomics databases and phospho-peptides generated containing these sequences. Using the phospho-peptides as bait, binding to PIH1D1 in HeLa cytoplasmic and nuclear extracts was then investigated using pull-down assays. Armed with this information, these interactions could be investigated quantitatively using ITC as part of this thesis.

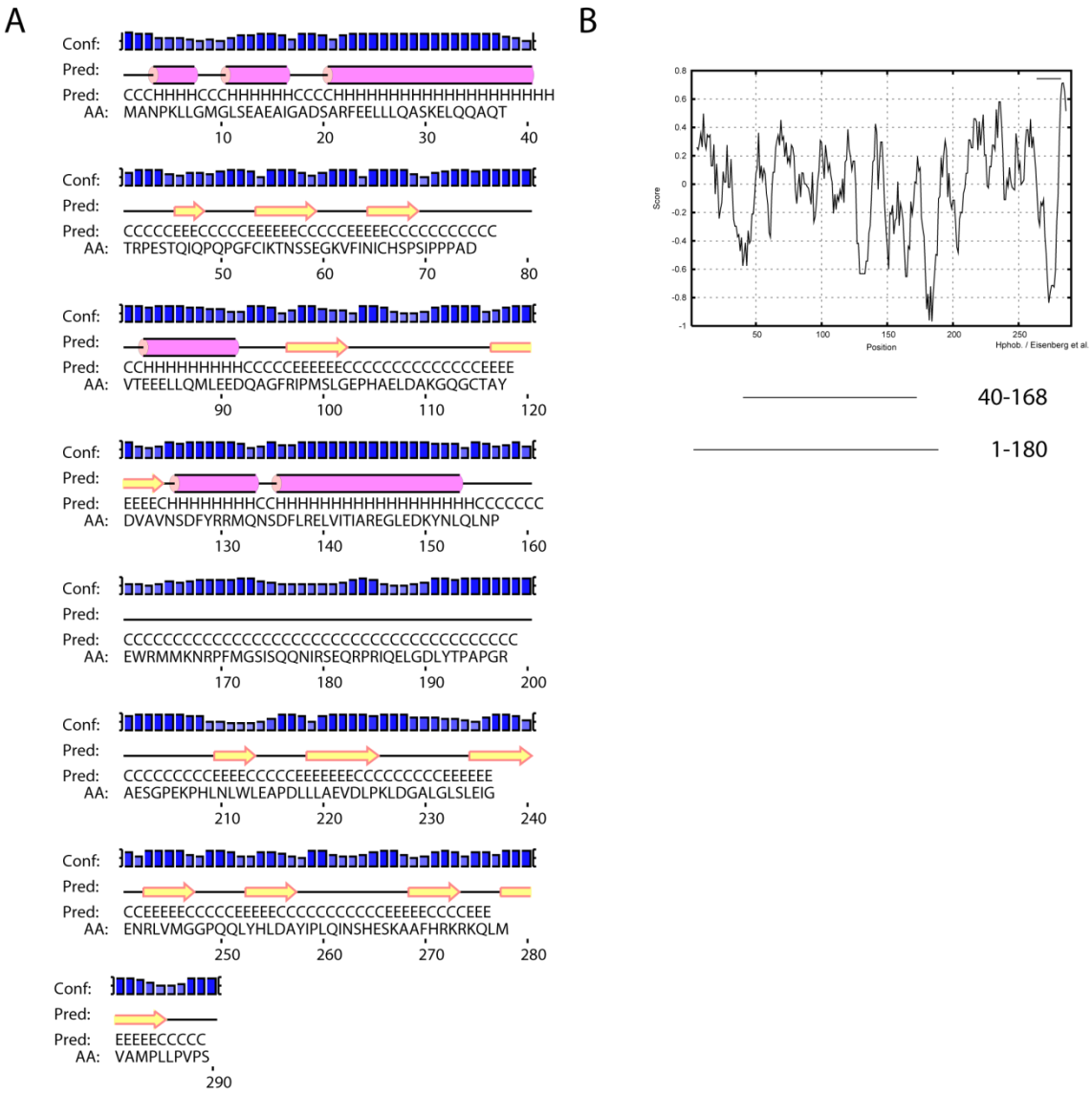


Figure 7.1 Secondary structure prediction of human PIH1D1 as predicted by the program PSI-PRED (UCL) with purple cylinders, yellow arrows and a black line corresponding to predicted α-helices, β-strands and random coil respectively. The height of the blue bars on top of the prediction indicates the confidence with which these predictions could be made. (B) Hydrophobicity plot (Eisenberg et al.) of PIH1D1 with the positions of the stable proteolytic fragments indicated below.

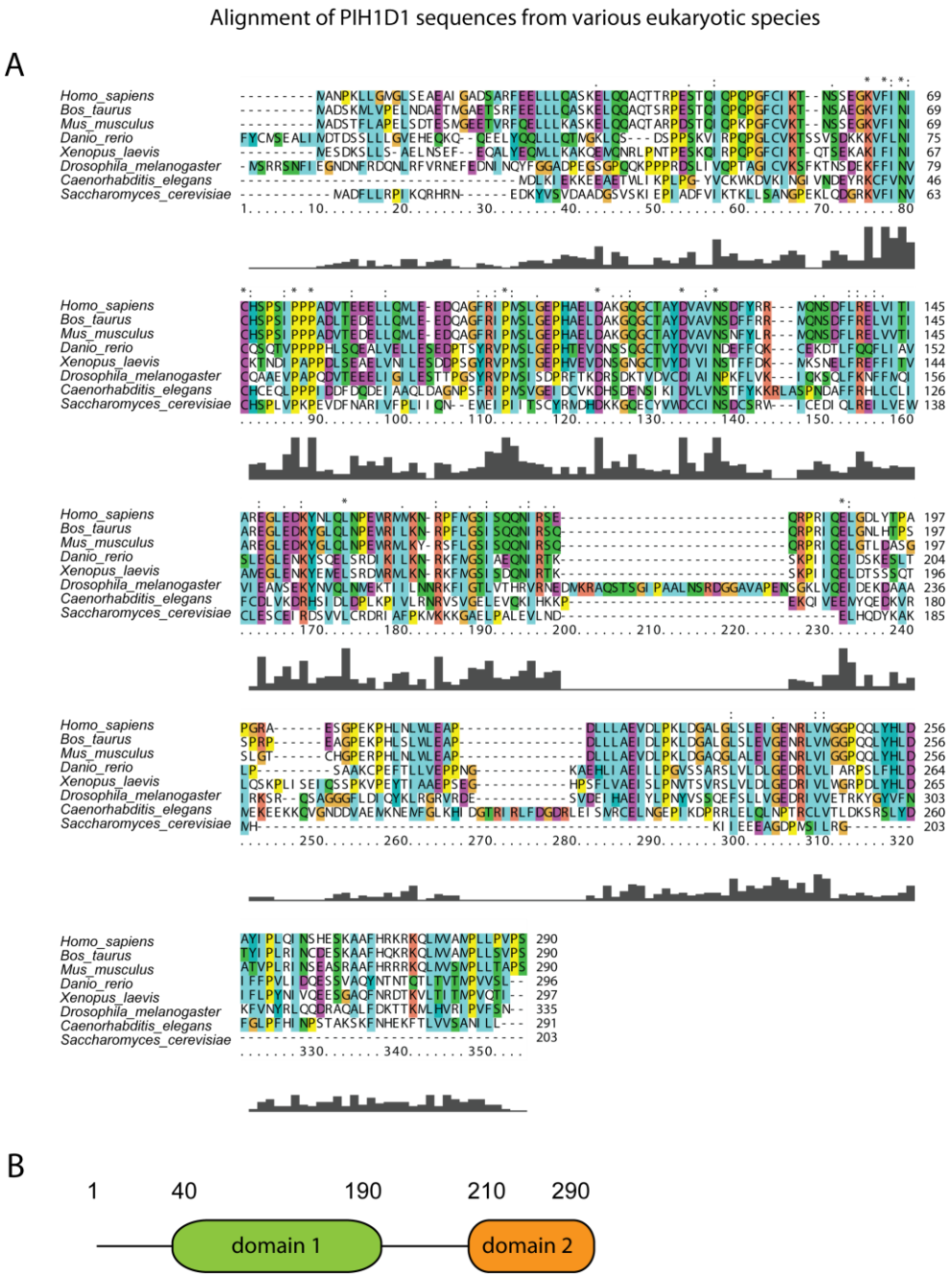


Figure 7.2. Conservation of PIH1D1 and its probable domain organisation. (A) Sequence alignment using ClustalX of PIH1D1 and its homologues from the species indicated. Amino acids are coloured according to their properties. Hydrophobic: cyan, polar: green, basic: red, acidic: purple, tyrosine or histidine: dark cyan, glycine: beige, proline: yellow. The size of the bars under the alignment indicates the level of sequence conservation for that particular amino acid. The dots and stars on top of the alignment also indicated relative levels of conservation. A star indicates absolute conservation while two dots or one dot refer to a less strongly conserved amino acid. (B) Domain organisation of PIH1D1 inferred from the sequence alignment, ITC and the stable fragments identified by limited proteolysis.

7.1 Limited proteolysis

In preparation for binding experiments, recombinant, full-length PIH1D1 protein was expressed as an N-terminal 6x-His fusion in *E. coli* and purified using Ni^{+}-affinity and size exclusion chromatography. The full length protein was soluble, but prone to proteolysis and hence not a good subject for biochemical and biophysical experiments. As a consequence, a more, well characterised stable expression construct was sought. From tryptic limited proteolysis, a stable fragment was identified which upon analysis by mass-spectrometry was found to encompass amino acids 1-180 of the protein. In a second round of limited proteolysis, the full length protein and $PIH1D1^{1-180}$ were subjected to proteolysis by chymotrypsin and the resulting fragments separated by SDS-PAGE (figure 7.3). From the gel it can be seen that full length PIH1D1 was already degraded during the protein purification and that it migrates, along with its degradation products, as a smear on the gel. $PIH1D1^{1-180}$ is fully intact at the start of the experiment. Following digestion with chymotrypsin both expression constructs are reduced to a fragment with an apparent molecular weight of 10 kDa. Following analysis by electrospray mass-spectrometry, this fragment was identified to span amino acids 40-168 of PIH1D1.

The PIH1D1 sequences of different eukaryotic species were aligned using ClustalX to determine which parts of the protein are conserved and to put the results of the limited proteolysis experiments into perspective (figure 7.2). The protein itself is conserved across the eukaryotic domain with homologues found in species ranging from human to *Saccharomyces cerevisiae*. The protein is however not conserved in the fission yeast *Schizosaccharomyces pombe*. A core homology region can be found between amino acids 40-190, but the extreme N-terminus does not seem to be conserved. The amino acids directly C-terminal to the core homology region are not conserved but a region of sequence similarity at the extreme C-terminus suggests an additional domain at this position. This putative second domain is not present in budding yeast so one can assume the core homology region is likely to be able to fold independently. From the secondary structure prediction it also seems likely that PIH1D1 consists of two separate domains (figure 7.1). The first 160 residues of this protein make up a mainly α-helical domain which is then followed by a flexible region approximately 50 amino acids long and an all β-sheet C-terminal domain comprising of amino acids 210-280. Generally, the sequence alignment agrees well with the secondary structure prediction, but there are some exceptions. Although the first 40 amino acids of the protein are not well conserved, they are confidently predicted to form two helices. Conversely, the

linker between the two putative domains appears longer when taking into account the secondary structure predictions than the sequence alignment. Some of the residues in this linker are not thought to have any secondary structure.

The fragments generated by limited proteolysis somewhat support the assumptions made from the sequence alignment and the secondary structure alignment. The tryptic fragment 1-180 (…NIR) consists of the core homology domain plus the N-terminal extension. There is no tryptic cleavage after Lys5 and Lys33 which are both predicted to be part of α-helices but are not in regions of strong sequence similarity. The chymotryptic fragment 40-168 (TTR… - …KNR) suggests a smaller domain and agrees more with the sequence alignment than the secondary structure prediction. The different protease specificities result in slightly different sizes for the stable proteolytic fragments. While the C-terminus of the first domain is relatively well defined, there is some ambiguity about its N-terminal boundary. Nevertheless, both proteolytic fragments broadly agree with both bioinformatics approaches. There is a stable domain at least 120 residues long contained in the first 180 amino acids of PIH1D1. The presence of a region with low sequence conservation and no predicted secondary structure between the two putative domains indicates this region may be flexible and presumably explains the proteolytic sensitivity of the full length protein when purified from *E. coli*. Since a stable expression construct had been found in PIH1D1$^{1\text{-}180}$, it was used to confirm and characterise the binding of PIH1D1 to Tel2, ECD and snRNP116 *in vitro*.

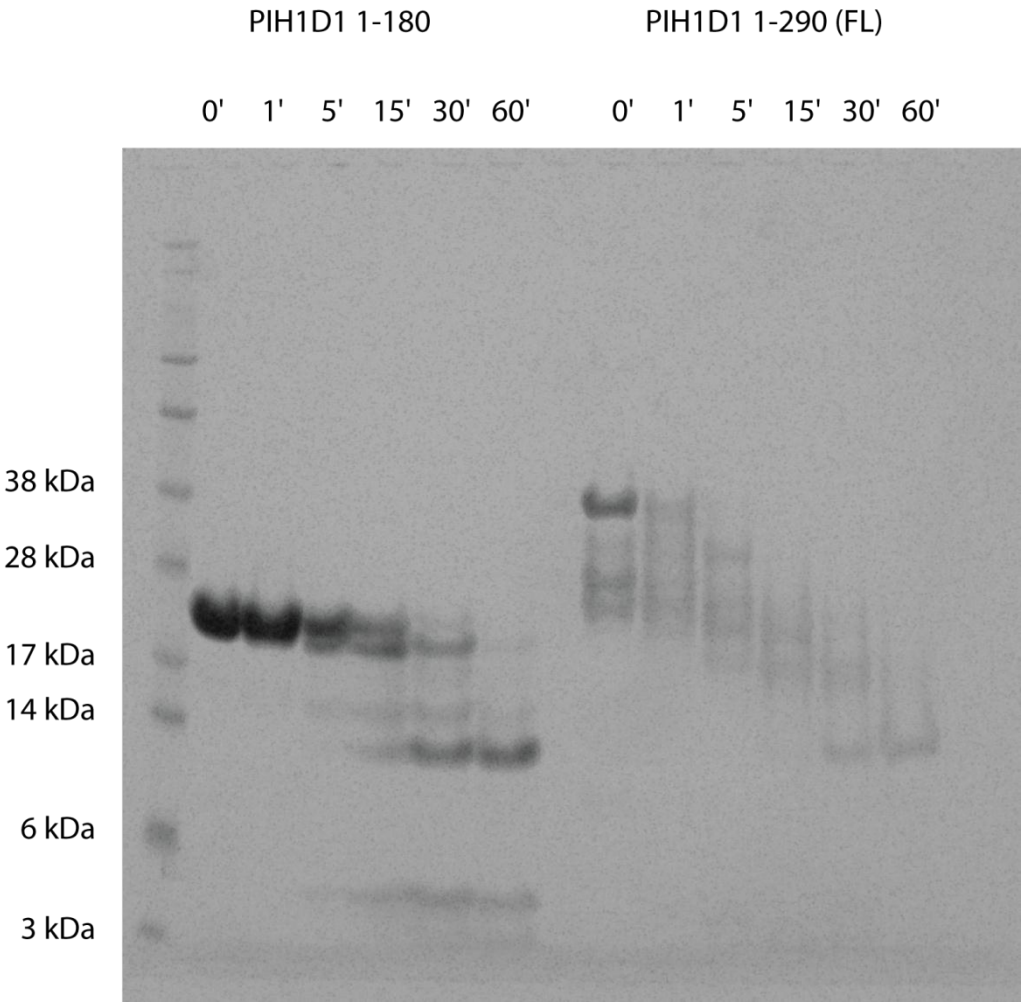
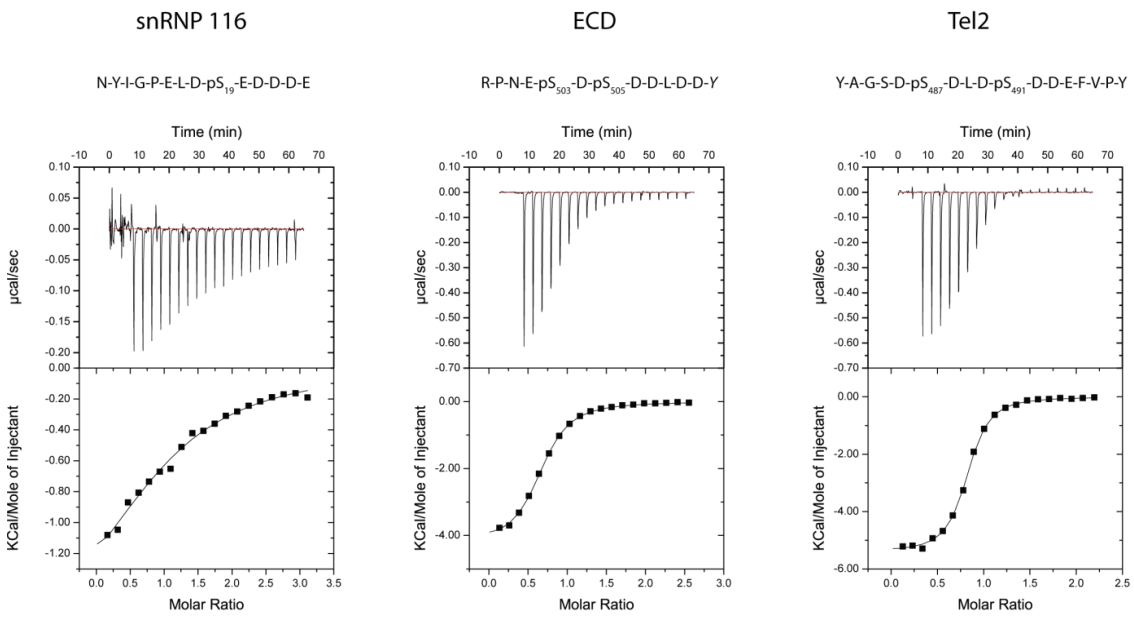


Figure 7.3. Limited proteolysis of PIH1D1. SDS-PAGE analysis of a chymotryptic limited proteolysis experiment of PIH1D1$^{1-180}$ and full length PIH1D1 using a substrate/protease ratio of 500:1 (w/w). The numbers on the left indicate the sizes of the respective molecular weight marker bands.



	N	K_D (µM)	ΔH (kcal/mol)	TΔS (kcal/mol)
PIH1D1 $^{1-180}$ vs snRNP 116	1.07	~60	-1.9	3.7
PIH1D1 $^{1-180}$ vs ECD	0.65	5.6	-4.2	2.8
PIH1D1 $^{1-180}$ vs Tel2	0.79	1.3	-5.4	2.5

Figure 7.4. ITC analysis of the interaction between PIH1D1$^{1-180}$ and phospho-peptides derived from the proteins indicated. The numbering of the phosphorylated residues in the peptide sequences corresponds to the amino acid numbering in the context of the respective full length proteins. Raw calorimetric data, heat changes and fitted binding curves for the titrations. The table below contains the thermodynamic parameters of the respective interactions.

7.2 Binding of PIH1D1 to Tel2, ECD and snRNP116

The binding of PIH1D1 to the interacting partners identified by pull-down assays was investigated using ITC to enable quantitative measurements in an in-solution system. In these experiments, the PIH1D1$^{1-180}$ construct was placed in the cell of the ITC calorimeter with synthetic phospho-peptides as titrant. To investigate binding of the three proteins to PIH1D1, synthetic phospho-peptides mimicking the respective CK2 phosphorylation sites on each protein were generated (figure 7.4). The sequences of the phospho-peptides for the ITC experiments were same as the ones that had been shown to bind in the pull-down assays but with some excessively long peptides shortened. Furthermore, a C-terminal Tyr was added to the ECD peptide to allow for concentration determination by UV spectroscopy.

The three peptides mimicking CK2 phosphorylation sites were shown to bind PIH1D1$^{1-180}$ with varying affinities. The binding parameters for each interaction can be seen in the table of figure 7.4. All interactions are driven both entropically and enthalpically. While a heat change could be observed for all three interactions, only the two double phosphorylated peptides bound with affinities high enough to make reliable estimates about the binding parameters. For the PIH1D1-snRNP 116 interaction, the analyte concentration was not sufficiently large compared to the dissociation constant for a sigmoidal binding curve to form. The expression construct is not soluble to such high concentrations to allow such a weak interaction to be measured by ITC.

For the Tel2 and ECD interactions with PIH1D1 however the binding parameters could be determined accurately. The energies contributing to the binding were roughly similar for Tel2 and ECD. In both cases the enthalpic contribution to binding was around 5 kcal/mol and the interaction was also driven entropically, with a contribution of approximately 2.5 kcal/mol. It is likely that the ECD and Tel2 derived peptides bind with the same mechanism as the energies of the two events are comparable. Compared to the FHA dependent phospho-binding measured in the previous chapters, both the favourable enthalpic contribution and the entropic penalty were considerably smaller. The respective dissociation constants also suggest that the first 180 amino acids of PIH1D1 are sufficient for a tight phospho-dependent interaction. The instability of full length protein did not allow binding experiments to be carried out. As a consequence, any contributions of the PIH1D1 C-terminus towards binding of phospho-peptides could not be investigated. Overall, these data show that PIH1D1 is a new, bona-fide phospho-reader domain that recognises pSer epitopes.

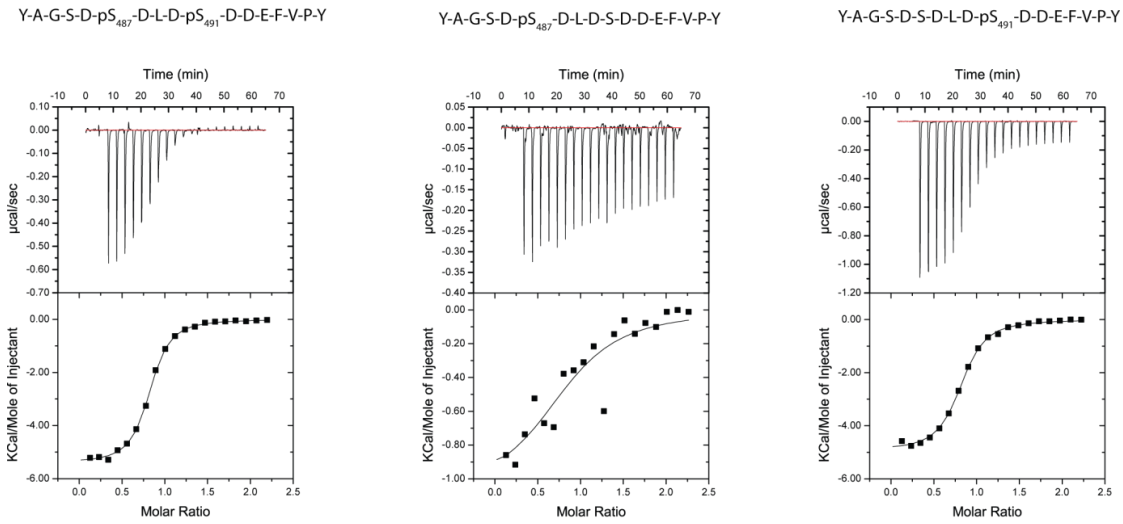
7.3 Binding specificity of PIH1D1

With PIH1D1 having been established as a phospho-reader domain, the next step was to further characterise its binding capacities, such as its preference for different phospho-epitopes and the number of binding surfaces. In the previous round of experiments, it was shown that the Tel2 and ECD peptides, both doubly-phosphorylated, interact much stronger with PIH1D1 than the mono-phosphorylated snRNP 116. It was therefore hypothesised that PIH1D1 possesses two phospho-binding pockets. To ascertain whether the stronger affinities for Tel2 and ECD were due to the presence of two pSer epitopes or to the differences in surrounding sequences, different phosphorylation states and mutants of the Tel2 peptide were synthesised. Binding to the Tel2 peptide had produced the largest signal and the dissociation constant was the lowest. Two mono-phosphorylated variants of the Tel2 peptide were synthesised with either Ser487 or Ser491 phosphorylated respectively (figure 7.5). The stoichiometries for each interaction were consistently around 0.8, suggesting that PIH1D1 has a single phospho-binding pocket. Since pSer491 is already sufficient for efficient binding and an additional phospho-site in the peptide only increases the affinity from 3.5 μM to 1.3 μM, it is likely that PIH1D1 only possesses one primary phospho-binding pocket and that the contribution to binding from a second phospho-epitope is much smaller than the primary interaction. The variable affinities for the different phospho-peptides in section 7.2 most likely arose from the different sequences surrounding the phospho-serines, not from the number of phospho-epitopes per peptide. Phosphorylation at Ser487 does not produce as tight an interaction as phosphorylation at Ser491 suggesting the PIH1D1 phospho-binding pocket has stringent sequence specificity. When pSer491 is present, this residue occupies the primary binding pocket of PIH1D1 and pSer487 only makes secondary contacts. The low affinity pSer487 only occupies the primary binding pocket when no higher affinity phospho-epitopes are available.

The relative affinities of PIH1D1 against a large number of peptides were investigated using OPAL (oriented peptide array library) arrays. While this technique allows only relative comparisons to be made between different peptides, it enables binding to a number of peptides to be tested simultaneously. The peptide arrays were synthesised and spotted onto a nitro-cellulose membrane by Nicola O'Reilly (CRUK, Lincoln's Inn Fields). PIH1D1$^{1-180}$ was incubated with the membrane and any bound protein detected by immunoblotting. The immunoblotting was carried out at the CRUK laboratories in Clare Hall under the supervision of Zuzana Horejsi, CRUK. In a first experiment, a

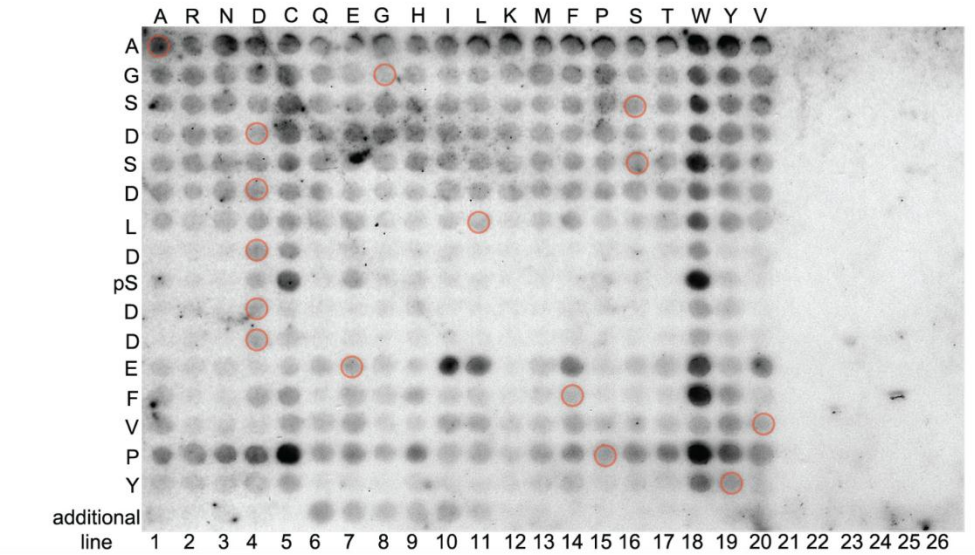
degenerate peptide library with a pSer fixed at the central position (sequence X-X-X-X-pS-X-X-X-X) was synthesised. For each spot, one amino acid was fixed to determine the binding specificity for PIH1D1. Unfortunately, no significant binding could be detected for any spot so it was deduced that PIH1D1 requires more than one conserved amino acid for efficient binding. Subsequently, a second OPAL array was synthesised using the Tel2-pSer491 peptide as the starting point. For each spot, one amino acid was mutated to determine which amino acids are essential for PIH1D1 binding (figure 7.6). As all possible mutations were spotted for each amino acid, one spot in each row corresponds to the wild-type Tel2 peptide. These wild-type spots have similar intensities throughout and hence confirm the consistency between the different rows and columns of peptide spots. The columns containing mutations to cysteine and tryptophan residues were ignored as they display consistently strong binding for virtually every position. Most importantly, this second OPAL array confirms that several amino acids need to be present for efficient binding. The aspartate residues in the -1, +1 and +2 positions, as well as the pSer, are essential for binding. Mutation of these amino acids severely decreases binding and even a conservative mutation to a glutamate severely weakens binding, particularly for the +1 and +2 Asp. In those rows, only very low levels of binding can be observed for all mutations. These data therefore suggest a short consensus sequence of D-pS-D-D. PIH1D1 binding is less sensitive to mutation at other positions in the peptide, suggesting the side-chains of these amino acids are not involved in the interaction. There are several amino acids that may contribute additionally to the binding event. Glutamate is favoured in the +3 position, but medium sized hydrophobic residues seem to bind even better. Positions -2, +4, +5 and +7 also show sensitivity to some mutations but there the data are not sufficiently clear to draw credible conclusions.

The spots in the bottom line of the OPAL array provide more evidence for a D-pS-D-D consensus sequence. Binding is only observed for spots 6-11, in which positions -1 to +2 have not been mutated. In all other cases, PIH1D1 does not bind to the spotted peptides. The fact that the protein does not bind to spots 19-22, where the D-pS-D-D sequence is present suggests that further peptide residues are important to the binding event. The lack of binding to spot 19 in particular suggests a role for either the proline in the +6, or the tyrosine in the +7 position in PIH1D1 binding.



	N	K_D (µM)	ΔH (kcal/mol)	TΔS (kcal/mol)
PIH1D1 $^{1-180}$ vs pSpS	0.79	1.3	-5.4	2.5
PIH1D1 $^{1-180}$ vs pSS	0.86	~30	-1.1	5.0
PIH1D1 $^{1-180}$ vs SpS	0.79	3.8	-4.9	2.3

Figure 7.5. ITC analysis of the interaction between PIH1D1$^{1-180}$ and phospho-peptides imitating different phosphorylation states of Tel2. The numbering of the phosphorylated residues in the peptide sequences corresponds to the amino acid numbering in the context of the respective full length protein. Raw calorimetric data, heat changes and fitted binding curves for the titrations. The table below contains the thermodynamic parameters of the respective interactions.

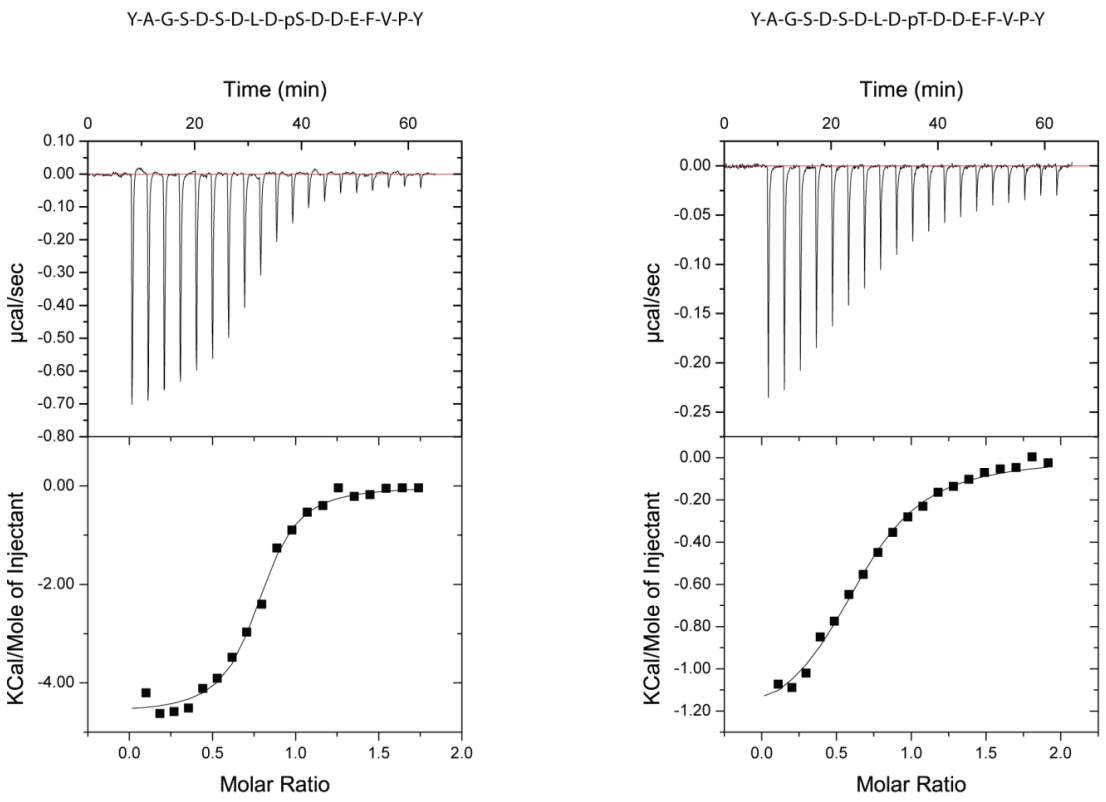


Additional line:

1	A-G-S-D-S-D-A-A-pS-D-D-E-F-V-P-Y
2	A-G-S-D-S-A-A-A-pS-D-D-E-F-V-P-Y
3	A-G-S-D-A-A-A-A-pS-D-D-E-F-V-P-Y
4	A-G-S-A-A-A-A-A-pS-D-D-E-F-V-P-Y
5	A-A-A-A-A-A-A-A-pS-D-D-E-F-V-P-Y
6	A-A-S-D-S-D-L-D-pS-D-D-E-F-V-P-Y
7	A-A-A-D-S-D-L-D-pS-D-D-E-F-V-P-Y
8	A-A-A-A-S-D-L-D-pS-D-D-E-F-V-P-Y
9	A-A-A-A-A-D-L-D-pS-D-D-E-F-V-P-Y
10	A-A-A-A-A-A-L-D-pS-D-D-E-F-V-P-Y
11	A-A-A-A-A-A-A-D-pS-D-D-E-F-V-P-Y
12	A-G-S-D-S-D-L-D-pS-A-A-E-F-V-P-Y
13	A-G-S-D-S-D-L-D-pS-A-A-A-F-V-P-Y
14	A-G-S-D-S-D-L-D-pS-A-A-A-A-V-P-Y
15	A-G-S-D-S-D-L-D-pS-A-A-A-A-A-P-Y
16	A-G-S-D-S-D-L-D-pS-A-A-A-A-A-P-Y
17	A-G-S-D-S-D-L-D-pS-A-A-A-A-A-A-Y
18	A-G-S-D-S-D-L-D-pS-A-A-A-A-A-A-A
19	A-G-S-D-S-D-L-D-pS-D-D-E-F-V-A-A
20	A-G-S-D-S-D-L-D-pS-D-D-E-F-A-A-A
21	A-G-S-D-S-D-L-D-pS-D-D-E-A-A-A-A
22	A-G-S-D-S-D-L-D-pS-D-D-A-A-A-A-A
23	A-G-S-D-S-D-L-D-pS-D-A-A-A-A-A-A
24	A-G-S-D-S-D-L-A-pS-A-D-E-F-V-P-Y
25	A-G-S-D-S-D-L-A-pS-A-A-A-F-V-P-Y
26	A-G-S-D-S-D-A-A-pS-A-A-A-F-V-P-Y

Figure 7.6. Binding of PIH1D1$^{1-180}$ to a spotted peptide array visualised by electrochemiluminescence. At the left of the array the wild-type sequence of the Tel2 pSer491 peptide is shown. For every spot, this sequence has been mutated to the amino acid indicated at the top at the position indicated on the left. The spots that correspond to the wild-type sequence have been highlighted using red circles. Below are the sequences for the peptides in the additional line.

All the peptides tested for PIH1D1 binding thus far had been pSer-containing peptides, so a pThr-containing peptide was synthesised to establish whether PIH1D1 prefers either pSer or pThr or whether it has absolute specificity for one, such as an FHA domain. A version of the Tel2 pSer491 peptide was synthesised which had pSer491 mutated to a pThr. The pSer491 and pThr491 peptides and $PIH1D1^{1-180}$ were all dialysed into the same buffer and binding for each peptide was tested by ITC (figure 7.7). Although binding against pSer491 had already been measured previously, the experiment was repeated that both titrations were performed with the same batch of protein and in identical buffers to ensure the two results are comparable. The ITC titrations reveal PIH1D1 to prefer pSer over pThr epitopes. The pSer491 peptide bound with a K_D of 2.6 μM while the pThr mutation only binds with an affinity of 15 μM. The 6 fold difference in binding is significant, suggesting a preference for pSer. Nonetheless, the experiments show that PIH1D1 can bind both types of epitopes with biologically relevant affinities.



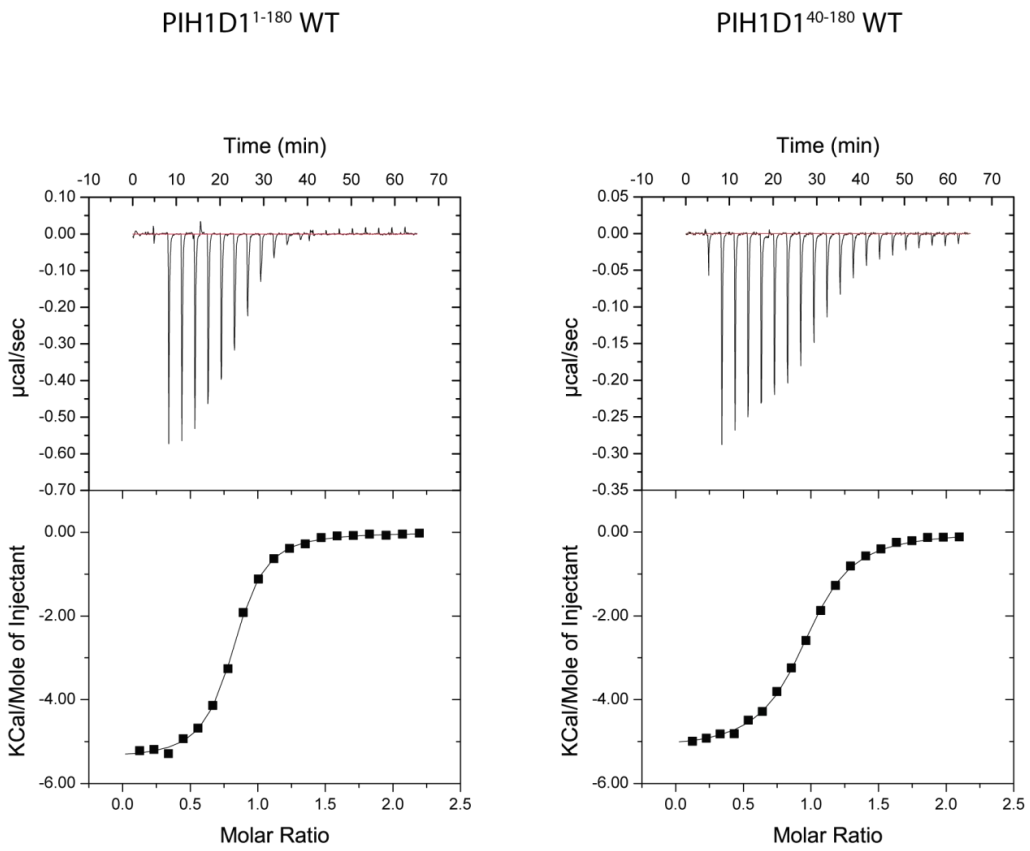
	N	K_D (µM)	ΔH (kcal/mol)	TΔS (kcal/mol)
PIH1D1 $^{1-180}$ vs SpS	0.76	2.6	-4.6	2.9
PIH1D1 $^{1-180}$ vs SpT	0.66	15	-1.3	5.2

Figure 7.7. Comparison of the relative affinities of PIH1D1$^{1-180}$ for pSer and pThr containing epitopes using ITC. The Tel2 pSer491 wild-type peptide and its pS491pT mutant were used in the experiment. Raw calorimetric data, heat changes and fitted binding curves for the titrations. The table below contains the thermodynamic parameters of the respective interactions.

7.4 Essential residues for pSer binding

After establishing which residues of the phospho-peptide are essential to the interaction, experiments were carried out to determine which parts of the PIH1D1 protein are needed for phospho-peptide binding. A truncated PIH1D1 construct was cloned, expressed and purified that contained amino acids 40-180 based on the data in section 7.1. The ability of this protein to bind the Tel2 pSer487/491 peptide was tested by ITC (figure 7.6). Both proteins bind the peptide with an affinity of approximately 1 μM. The titrations with PIH1D1$^{40\text{-}180}$ showed an estimated stoichiometry of 0.95, much closer to the assumed theoretical value of 1 than the earlier PIH1D1$^{1\text{-}180}$ titrations which had calculated stoichiometries of around 0.7. This difference in apparent stoichiometry suggests that PIH1D1$^{40\text{-}180}$ is more stable and that there was less misfolded or aggregated protein present which would reduce the active fraction. The experiment reveals the shorter expression construct to bind the protein as well as PIH1D1$^{1\text{-}180}$ and demonstrates that amino acids 1-40 of PIH1D1 do not contribute significantly to phospho-peptide binding.

Y-A-G-S-D-pS$_{487}$-D-L-D-pS$_{491}$-D-D-E-F-V-P-Y vs :



	N	K_D (µM)	ΔH (kcal/mol)	TΔS (kcal/mol)
PIH1D1 $^{1-180}$ vs pSpS	0.79	1.3	-5.4	2.5
PIH1D1 $^{40-180}$ vs pSpS	0.95	1.2	-5.2	2.7

Figure 7.8. ITC analysis of the binding of both PIH1D1$^{1-180}$ and PIH1D1$^{40-180}$ to the Tel2 pSer487/491 peptide to compare the pSer binding competence of each construct. Raw calorimetric data, heat changes and fitted binding curves for the titrations. The table below contains the thermodynamic parameters of the respective interactions.

A PIH1D1 residue was sought whose mutation could abrogate phospho-binding but leave the PIH1D1 fold intact. Such phospho-specific mutations enable one to determine which interactions of a protein are phospho-dependent and are which not. They are also useful when phospho-reader domains are part of a multi domain protein and one wants to inhibit a phospho-dependent interaction but have an otherwise intact protein. One such example is the conserved arginine residue in the β3/β4 loop of FHA domains. A mutation of this amino acid to an alanine has been used in a number of studies to knock out the phospho-binding capabilities of FHA domains [53, 62, 68]. An equivalent of this arginine residue was sought in PIH1D1. As it is usually basic amino acids that bind the negatively charged phosphate moiety (figure 7.9), an alignment of PIH1D1 sequences was searched for a conserved arginine or lysine residue (figure 7.2). The most promising candidate was Lys64, which is conserved in all PIH1D1 sequences. It is present from humans to *Saccharomyces cerevisiae*. The 5 amino acids directly C-terminal to Lys64 are also highly conserved (K_{64}-V-F-I-N-I). Site-directed mutagenesis was performed on the PIH1D1$^{1-180}$ construct to produce PIH1D1$^{1-180}$ K64A.

The fact that the mutant eluted from a size exclusion chromatography column at the same point as the wild-type protein suggested the mutant protein was properly folded and not aggregated. To confirm the mutant protein was properly folded, far-UV CD was carried out on the K64A and the wild-type protein which showed that the proteins had identical amounts of secondary structure (figure 7.9) Both spectra indicate that the majority of amino acids in the respective proteins form α-helices. The slightly larger signal for the K64A mutant is due to a higher concentration of protein in the cuvette. The similar thermal denaturation profiles show that both variants of PIH1D1 are equally stable. Once it had been established that the K64A mutation did not interfere with the PIH1D1 fold, the phospho-binding capabilities of this mutant were tested by ITC (figure 7.10). The Tel2 pSer487/491 peptide was used to test for PIH1D1 binding as this peptide had shown the highest affinity for the protein. No binding could be observed between the mutant protein and the peptide, indicating that Lys64 is essential for pSer binding. The experiment was carried out at two different temperatures to find out whether the lack of signal was due to a lack of binding or an enthalpically neutral binding event at a particular temperature. As no peptide binding was observed at either temperature, there is good evidence that K64A abrogates phospho-binding.

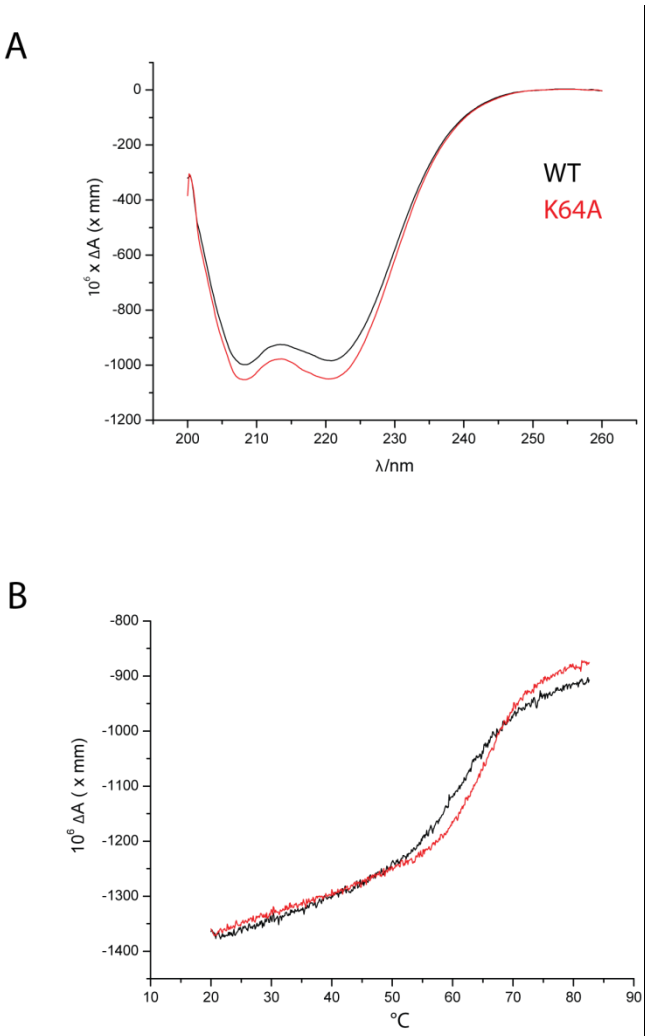


Figure 7.9. CD analysis of PIH1D1$^{1-180}$ WT and PIH1D1$^{1-180}$ K64A. (A) Far UV spectrum of the wild-type and mutant proteins, showing the majority of the constructs is α-helical and the amounts of secondary structure unchanged in the K64A mutant. (B) Thermal denaturation of both proteins measured using the higher differential absorption of α-helices as opposed to random coil at 222 nm.

Y-A-G-S-D-pS$_{487}$-D-L-D-pS$_{491}$-D-D-E-F-V-P-Y vs :

PIH1D1$^{1-180}$ WT

PIH1D1$^{1-180}$ K64A

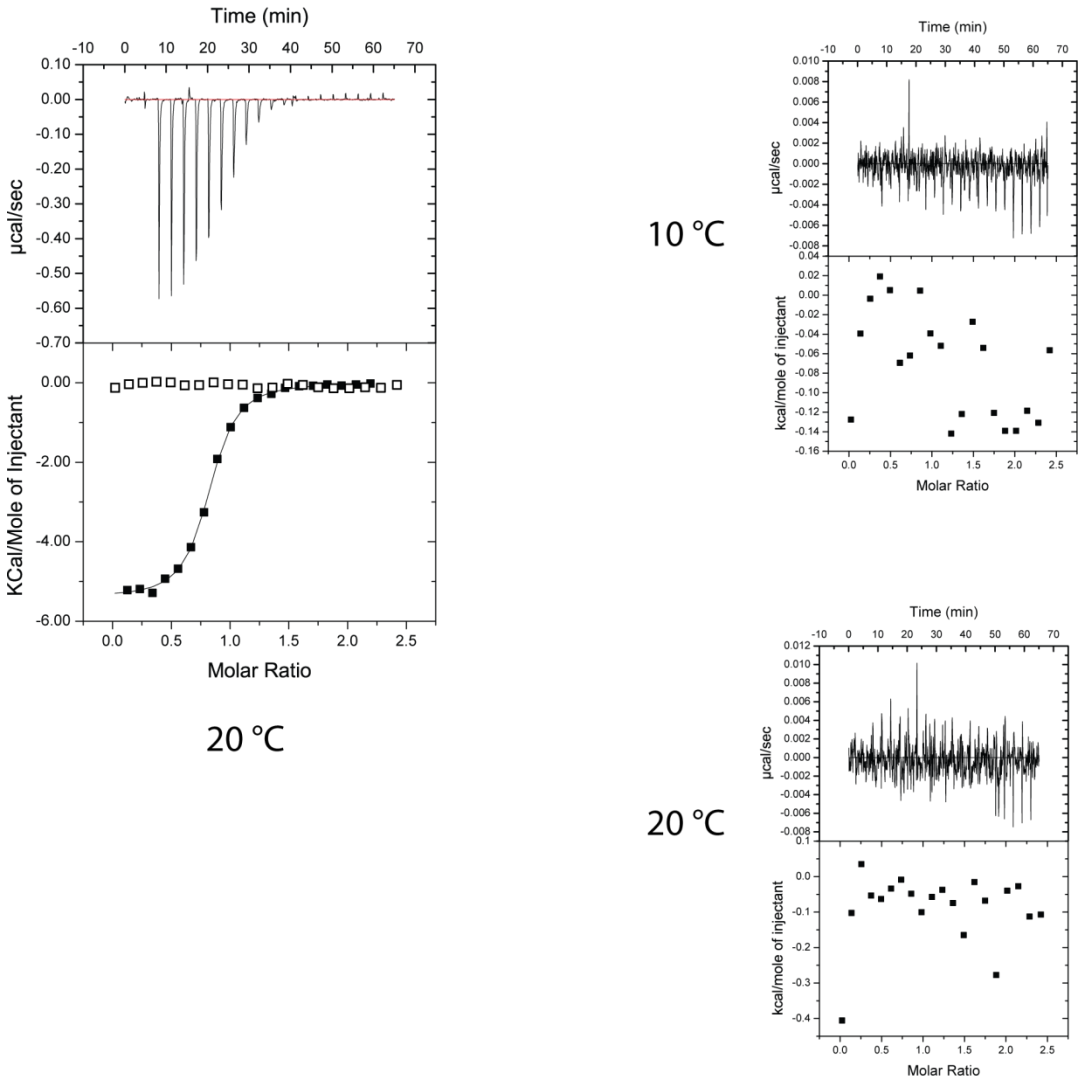


Figure 7.10. Comparison of the phospho-binding capabilities of PIH1D1$^{1-180}$ WT and PIH1D1$^{1-180}$ K64A. To the right are the raw calorimetric data and heat changes of the interaction between PIH1D1$^{1-180}$ K64A and the Tel2 pSer487/491 peptide at 10 °C and 20 °C. To the left is an overlay of the heat changes and fitted binding curves of the wild-type protein binding to the Tel2 pSer487/491 peptide (filled boxes) and the heat changes of the interaction of the K64A mutant with this peptide at 20°C (empty boxes).

This information was passed onto the collaborating lab which confirmed these ITC experiments *in vivo*. They showed that in HEK293 cells, the K64A mutation knocks out the established phospho-dependent interaction between Tel2 and PIH1D1, but the interaction between PIH1D1 and RPAP3, RUVBL1 and RUVBL2, the other constituents of the R2TP complex, was unaffected by this mutation (figure 7.11) (personal communication). The interaction between the *S. cerevisiae* homologues of PIH1D1 and RPAP3 had previously been shown to occur between native, unphosphorylated proteins and to be mediated by the C-terminal region of the PIH1D1 homologue [159]. This suggests the PIH1D1 N-terminal domain has phospho-reader capability and that the C-terminal region of this protein can mediate non-phospho protein-protein interactions. The abrogation of phospho-binding due to the K64A mutation that had initially been shown *in vitro* using synthetic phospho-peptides was confirmed *in vivo* using native protein interacting partners. Furthermore, this mutation enables one to distinguish between the phospho-dependent and phospho-independent contributions to interactions mediated by PIH1D1.

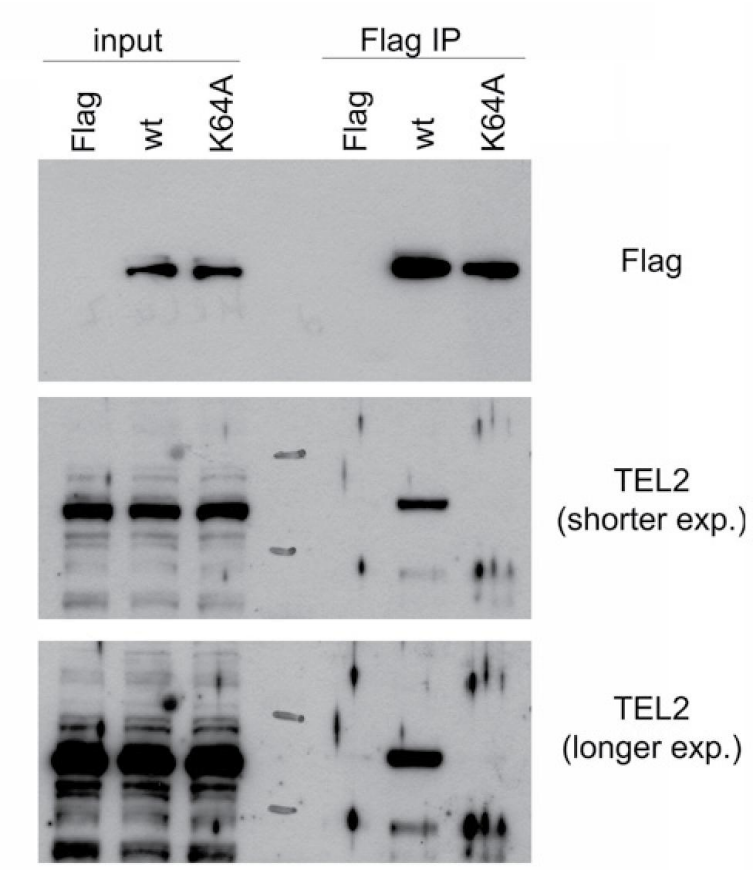


Figure 7.11. Western blot analysis of anti-flag immunoprecipitates of HEK293 cells expressing empty flag vector, Flag-PIH1D1 WT and Flag-PIH1D1 K64A. The blots were stained with antibodies against Flag (top panel) and Tel2 (central and lower panel). No binding can be observed between Tel2 and the PIH1D1 K64A mutant. This experiment was carried out by Zuzana Horejsi in the lab of Simon Boulton (CRUK)

7.5 Crystallisation of PIH1D1

It was attempted to crystallise PIH1D1 in order to solve its structure to determine the molecular basis for its specificity for acidic phospho-peptides. The sequence of this protein is dis-similar to any protein with a solved structure and the highly specific consensus sequence of D-pS-D-D has not been reported previously. Although the PIH1D1 phospho-binding domain had already been shown to belong to a new class of phospho-binding modules in terms of sequence specificity, it remained to be determined if it also belonged to a new class of fold. Not all members of a protein domain family can be identified through a shared sequence homology. Following determination of its crystal structure, the *S. cerevisiae* protein Spt6 for example was shown to contain a bona-fide SH2 domain that had not be identified by its sequence [173]. Similarly, PIH1D1 may contain a “cryptic” BRCT or Polo-Box domain.

Initially, the PIH1D1$^{1-180}$ construct was used in crystallisation trials but no crystals could be produced. Subsequently, the PIH1D1$^{40-180}$ construct was used instead as it was known that this smaller construct also had phospho-binding ability. The protein was dialysed into 150 mM NaCl, 20 mM Tris pH 8.0 and 0.5 mM TCEP and crystallisation trays were set up using commercially available crystallisation screens (Hampton Research, Qiagen) as well as ones that had been developed in-house. The protein was concentrated to 8 mg/ml and mixed with a 1.2:1 molar excess of pSer487/491 peptide. Drops containing 100 nl of protein peptide mix and 100 nl of well solution were set up and incubated at 18 °C in MRC plate II crystallisation plates (Swissci) using the sitting drop method.

Single crystals formed after several days in drops where the well solution was 0.2 M $MgCl_2$, 0.1 M MES pH 6.5 and 35% (w/v) PEG 4000 (Qiagen, PEGs II D3). There were also crystals in 0.1 M MES pH 6.5 and 20% PEG 4000 (Qiagen, PEGs II F2). These crystals were flash frozen in liquid nitrogen using 20% glycerol as a cryo-protectant and tested for diffraction at the IO3 and IO4 beamlines at Diamond Light Source. However, none of these crystals diffracted.

Crystalline needles, which were not of sufficient size to collect data, formed in 0.1 M sodium citrate pH 5.6, 20% PEG 4000 and 20% (v/v) 2-propanol (Qiagen, Classics A7). A seeded optimisation matrix was then set up around the condition varying the concentration of PEG 4000 and 2-propanol using 0.1 M sodium citrate pH 5.6 as the buffer for every well solution. The crystalline needles were used to prepare a microseed solution as described in section 6.2.1. The seeded drops contained 1.5 µl

protein solution, 1.5 µl well solution and 1 µl seed solution. Crystals formed in seeded drops in 0.1 M sodium citrate pH 5.6, 20% PEG 4000 and 15% 2-propanol after one week. Unlike the previous cases where glycerol had been used as a cryo-protectant, 1,2-propanediol was used instead as it is similar to 2-propanol and was thought to alter the environment of the crystal less than the addition of glycerol. The cryo-buffer contained the constituents of the protein buffer as well as 0.1 M sodium citrate pH 5.6, 20% PEG 4000, 10% 2-propanol and 15% (v/v) 1,2-propanediol.

7.6 Data collection and processing

The frozen crystals were tested for diffraction at the IO3 beamline at Diamond Light Source and diffraction was observed. The crystals were indexed and following integration assigned space group $I4_122$ or $I4_322$. The handedness of the screw axis cannot be determined from a data set without any phase information.

wavelength	0.9795 Å
data collected	90° (180 x 0.5° oscillation)
space group	$I4_122$/ $I4_322$
unit cell	a = b = 107.30 c = 54.73
mosaicity	0.6°
resolution range	30.0 – 2.8 Å
total reflections observed	24528
unique reflections	3971
average redundancy	6.2*(3.6)*
completeness (%)	96.0 *(73.2)*
I/σ	23.3 *(2.0)*
R_{symm}	0.069 *(0.33)*

Table 7.1. Processing statistics of PIH1D1$^{40\text{-}180}$ crystal (highest resolution shell, 2.93 – 2.80 Å, in brackets)

The crystal diffracted the X-rays sufficiently well that data could be processed up to a resolution of 2.8 Å. At this cut-off, the I/σ still exceeded two and hence indicated reliable data with an acceptable noise level. The Matthews coefficient suggested 1 copy of the protein in the asymmetric unit, which would correspond to a solvent content of 45%.

7.7 Generation of seleno-methionine substituted protein crystals for phasing by anomalous dispersion

Several options were considered as to how to obtain the phases needed to solve the native data-set. The easiest method, molecular replacement, could be discounted as there was no protein in the PDB sharing significant homology with PIH1D1. Single anomalous dispersion (SAD), using seleno-methionine substituted protein appeared to be the most promising as the PIH1D1$^{40\text{-}180}$ construct contained six methionine residues and this method requires data to be collected from one crystal only. The seleno-methionine substituted protein was expressed in the B834 *E. coli* methionine auxotroph in seleno-methionine supplemented minimal media. The expression levels were comparative to standard PIH1D1$^{40\text{-}180}$ and it could also be purified using the same protocol.

The seleno-methionine substituted protein was prepared for crystallisation trials in the same way as the native protein. Identical concentration, protein/peptide ratios and protein buffers were used. Initially, crystallisation trials were set up around the condition that the native protein had crystallised, but no crystals formed. As a consequence, screening for crystallisation conditions was started again using commercially available sparse crystallisation screens. After extensive screening, one condition was found that produced crystals, 1.8 M $NaKPO_4$. The crystals were frozen using glycerol as a cryo-protectant and tested for diffraction. The crystals did not diffract well enough to collect a complete data-set. The first images collected from the best crystal showed diffraction up to 4.5 Å and were indexed as R3. Unfortunately the crystal suffered radiation damage and the diffraction faded quickly after 30 images.

As no better diffracting crystals could be produced using this construct in complex with the Tel2 pSer487/491 peptide, new crystallisation trials were set up as before but without peptide. Again, a number of sparse crystallisation matrixes were used and a number of conditions produced crystals (figure 7.12). While the buffers used for these successful conditions were variable, the use of ethylene glycol polymers as a precipitant was common to all these conditions. None of these crystals produced any diffraction when tested. Although the protein crystallised more readily without the peptide, it did not improve diffraction.

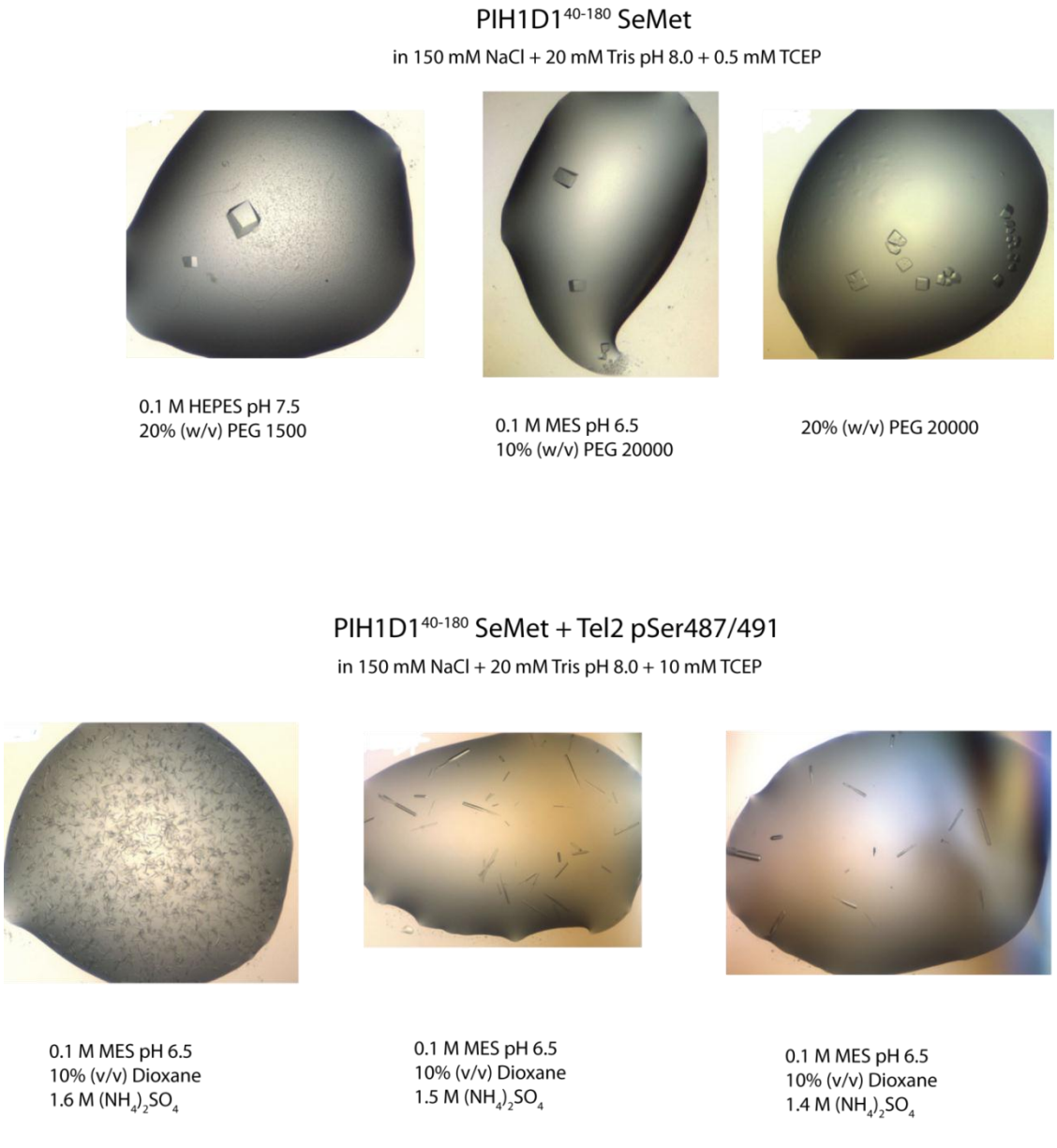


Figure 7.12. Images of PIH1D1$^{40-180}$ SeMet crystals. The compositions of the respective protein buffers are indicated above the drops and the well solutions used to crystallise the protein are shown below the drop images.

The native as well as the seleno-methionine substituted protein tended to crystallise when PEG was used at around neutral pH. The only diffracting crystals were produced when both PEG and propan-2-ol were used as precipitants to crystallise the native protein. These crystals could not be reproduced using seleno-methionine substituted protein. It was hypothesised that oxidation of the selenium might hinder ordered crystal packing and new crystallisation trials were set up using the Tel2 pSer487/491 peptide and an increased concentration of reducing agent, 10 mM TCEP instead of 0.5 mM. Sparse crystallisation matrixes were used again and crystalline needles formed in 0.1 M MES pH 6.5, 10% (v/v) dioxane and 1.6 M $(NH_4)_2SO_4$ (figure 7.12). An optimisation screen was set up for which a range of lower ammonium sulphate concentrations were used with the intention to reduce the number of nucleation events and subsequently increase crystal size. At concentrations between 1.4 and 1.5 M ammonium sulphate, small single crystals could be obtained that produced some limited diffraction, but not of sufficient quality for structure solution. To solve the structure of the PIH1D1 protein, a different expression construct may need to be cloned or an additional crystal form of the existing construct may need to be found. Alternatively, soaking of native crystals with heavy-metal derivatives may enable one to solve the PIH1D1 structure by multiple isomorphous replacement. Despite crystallising in several different forms, with peptide bound and without, no well-diffracting crystals were formed by the seleno-methionine substituted PIH1D1 protein. Nonetheless, the limited diffraction shown from the crystals grown in 1.4 and 1.5 M ammonium sulphate is promising. Given that there was some limited diffraction from such small crystals, bigger crystals, which could be obtained by optimising the crystallisation condition, may lead to well diffracting crystals. Another reason for optimism is the fact that these crystals were indexed to have the same space group as the diffracting native crystals.

7.8 Conclusions

The experimental results in this chapter show that PIH1D1 contains a phospho-reader domain with a high specificity for acidic phospho-epitopes.

Initially, a core homology region of amino acids 40-180 was identified from sequence alignment of PIH1D1 orthologues in eukaryotes, which was then shown to be a stable and soluble domain that could be purified from *E. coli*. The full length protein is not stable *in vitro*. Its ability to bind phospho-epitopes was probed using ITC, a method that had previously been used to characterise phospho-dependent protein-protein interactions of a number of reader domains [39, 40, 59]. The ITC experiments confirmed the interactions reported *in vivo* between PIH1D1 and Tel2 and ECD. but the binding of PIH1D1 to a snRNP 116 derived phospho-peptide was shown to be considerably weaker than the other two interactions. One explanation for the weak interaction between snRNP 116 and PIH1D1 in the ITC experiments could be the fact that residues 1-180 of PIH1D1 only were used and that the interaction may not necessarily be entirely phospho-dependent or that for example the semi-conserved C-terminal region of PIH1D1 may also contribute to the snRNP116 binding that been observed in IPs.

Peptide sequence	K_D (μM)	Peptide derived from
Y-A-G-S-D-**pS**-D-L-D-**pS**-D-D-E-F-V-P-Y	1.3	Tel2 pSer487/491
Y-A-G-S-D-S-D-L-D-**pS**-D-D-E-F-V-P-Y	3.5	Tel2 pSer491
Y-A-G-S-D-**pS**-D-L-D-S-D-D-E-F-V-P-Y	30	Tel2 pSer487
Y-A-G-S-D-S-D-L-D-**pT**-D-D-E-F-V-P-Y	15	Tel2 pSer491
R-P-N-E-S-D-**pS**-D-D-L-D-D-Y	16	ECD pSer505
R-P-N-E-**pS**-D-**pS**-D-D-L-D-D-Y	5.6	ECD pSer503/505
N-Y-I-G-P-E-L-D-**pS**-E-D-D-D-E	60	snRNP 116 pSer19

Table 7.2. Dissociation constants of PIH1D1$^{1-180}$ for various phospho-peptides measured by ITC

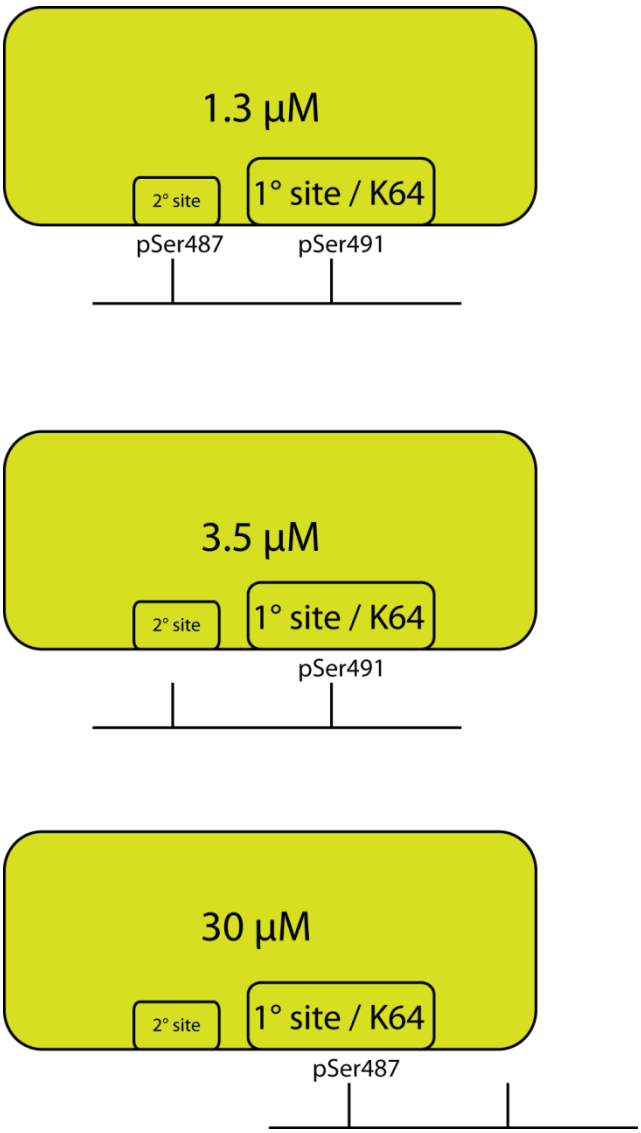


Figure 7.13. Binding of Tel2 peptides to PIH1D1. pSer491 occupying the main binding surface is sufficient for a tight interaction. The additional presence of pSer487 stabilises the interaction through a secondary interaction. When pSer487 is the sole pSer available, this residue presumably engages with the main binding surface, albeit with much lower affinity.

The relative binding affinities of PIH1D1$^{1-180}$ for the different phospho-peptides (table 7.2) as well as the OPAL array (figure 7.5) suggest a sequence of D-pS-D-D to be essential for a tight interaction. Although it can be seen from the different affinities of the Tel2 pSer491 and the ECD pSer505 peptide that D-pS-D-D is not sufficient for a tight interaction by itself, it is a requirement for tight binding. The low micromolar affinities of PIH1D1 for its favoured binding partners are similar to other established phospho-reader domains such as FHA and SH2 domains [36, 59]. One pSer residue is sufficient for efficient binding, but it was shown that the presence of a second pSer in the -2 or -4 position increases binding by approximately 3 fold. It was hypothesised that PIH1D1 possesses two phospho-binding surfaces, a major and a minor site. When two phospho-residues are available, the higher affinity epitope will occupy the primary site and the lower affinity one the secondary site (figure 7.13) PIH1D1 has dual specificity for pSer and pThr containing epitopes like all other pSer and pThr binding proteins apart from the FHA domain [59]. The slightly increased affinity for pSer epitopes over pThr ones is reminiscent of BRCT domains [174]. It was also shown that amino acids 40-180 are sufficient for strong pSer binding. Since the full length protein was unstable, it could not be determined whether the C-terminal part of PIH1D1 is an autonomous domain, with for example a role in phospho-independent protein-protein interactions or whether this region is part of the phospho-reader domain and contributing towards even stronger pSer binding in the context of the full length protein.

Residue Lys64 was shown to be essential for phospho-binding and mutation of this amino acid to alanine disrupted phospho-dependent interactions *in vitro* and *in vivo*. The K64A mutation enables one to distinguish the phospho-dependent from the phospho-independent interactions of PIH1D1.

Ultimately, no crystal structure of PIH1D1 could be obtained, but the results so far have been encouraging. The PIH1D1$^{40-180}$ construct could be crystallised in the native form as well as a seleno-methionine substituted derivative. While the native protein produced crystals that diffracted well enough to probably solve the structure of this protein, the seleno-methionine protein however did not produce any strongly diffracting crystals. Single crystals could be obtained that showed some diffraction, but not enough to obtain a useful electron density map.

PIH1D1 has only formed diffracting crystals in the presence of phospho-peptide or a large concentration of phosphate, Since binding of a phosphate-ligand appears essential to crystal formation, it can be hypothesised that the binding event may reduce the internal mobility of the crystal.

Due to its distinct primary sequence and the failure to obtain a three-dimensional structure it remains to be seen whether this protein belongs to an already established class of phospho-reader domains or constitutes a separate class entirely.

8 Discussion

8.1 Overview

Since the phospho-binding capabilities of the SH2 domain were discovered almost 20 years ago, other folds that can facilitate protein-protein interactions in a phospho-dependent manner, such as BRCT, 14-3-3 and FHA domains have been discovered [14, 39, 59]. Through the ability to associate with specific phospho-epitopes, they can facilitate intra-molecular associations, oligomerisation events and the formation of hetero protein complexes as a response to the reversible phosphorylation of serine, threonine or tyrosine residues by protein kinases. The data presented in the first part of this thesis show how intra-molecular association may be a common mechanism for FHA domain regulation using the mycobacterial protein FhaA as an example. In the second part, the human protein PIH1D1, which had previously been reported to bind the protein Tel2 in a phospho-dependent manner that is essential for the maturation of several PIKKs [147], was shown to function as a bona-fide phospho-reader domain with strong sequence specificity.

8.2 FhaA

Experiments were carried out to characterise the mycobacterial protein FhaA with the aim of determining its biochemical properties and the structural and functional consequences of its phosphorylation by the STPK PknB. PknB has been shown to be essential for *Mycobacterium tuberculosis* viability and its trans-membrane localisation has made PknB an attractive therapeutic target for novel anti-mycobacterial drugs [77, 175]. The effects of PknB mediated phosphorylation were studied with the aim of expanding our knowledge of the cellular changes effected by PknB activity. In turn, elucidation of the functional significance of this phosphorylation event had the potential to throw further light on FHA domain function in general.

8.2.1 The N-terminal domain

Initially, the N-terminal domain of FhaA, a domain of unknown function (DUF3662), was characterised and a stable and soluble fragment spanning amino acids 1-130, was identified using limited proteolysis. This fragment could also be crystallised suggesting the majority of this fragment formed a folded protein domain. As this protein did not contain any methionines and introduction of methionines by site-directed mutagenesis

disrupted crystal formation, the structure could not be solved by anomalous dispersion. While other methods of structure solving were considered, a solution structure was reported elsewhere [126]. This structure, shown in figure 1.8, consists of two parallel α-helices adjacent to a β-sheet comprising three strands. In addition, the solution structure demonstrates that the first 25 amino acids of FhaA are unstructured. The DUF3662 domain therefore covers residues 25-130 of FhaA, shorter than initially assumed. In terms of function of DUF3662, no conclusions could be drawn from either the data presented here, or by the authors of the NMR study. One expression construct containing DUF3662 co-purified with RNA oligonucleotides which suggested an RNA binding function for this domain. As *Mycobacterium tuberculosis* had been shown to contain an abundance of non-coding RNAs, this was a potentially promising lead [176]. RNA binding could however not be reproduced using EMSAs and was ultimately shown to be an artefact caused by the 6x-His purification tag of the expression construct.

8.2.2 FhaA FHA domain phospho-regulation

In order to determine what function FhaA phosphorylation may have in *Mycobacterium tuberculosis*, the phosphorylated residues as well as the biochemical consequences of the phosphorylation event needed to be determined. Using mass spectrometry it was shown that PknB phosphorylates FhaA on Thr377 and causes an intra-molecular association between the FHA domain and pThr377. A number of studies have shown FhaA to be phosphorylated by PknB, but there is conflicting evidence about the exact site and mechanism. The full-length protein had been shown to be phosphorylated by PknB [104]. Separately, it had been shown that an expression construct containing a deletion of the first 155 amino acids could also be phosphorylated [121]. Both studies also showed that while the FHA domain was required for efficient phosphorylation, it was not itself phosphorylated. A third study showed that Thr130, part of the DUF3662, is phosphorylated exclusively and that the FHA domain is not required for this to occur [126]. Conversely, it was also shown that the FHA domain binds to the juxtamembrane region of PknB with high nanomolar affinity. It was decided that the FHA dependent phosphorylation event would be characterised further as there had been several similar mechanisms reported for other mycobacterial FHA domain containing proteins [69, 107]. Following *in vitro* phosphorylation of an FhaA expression construct by PknB, Thr377 was found to be phosphorylated by tandem mass spectrometry. As the protein could be phosphorylated stoichiometrically and specifically, the structural and/or biochemical differences between the phosphorylated and non-phosphorylated isoforms of FhaA could be determined. The ability of native FhaA to bind a phospho-peptide

derived from the pThr377 epitope with high micromolar affinity, and the inability of the phosphorylated isoform to bind this phospho-peptide suggested that the FhaA FHA domain binds to its own pThr377 epitope. A conformational compaction upon phosphorylation, detected by size exclusion chromatography, supported this hypothesis. Furthermore, both isoforms were shown to be monomeric which suggested an intra-molecular association as reported for the mycobacterial FHA domain containing protein GarA and not a phospho-dependent dimerization or oligomerisation event as shown for Chk2 or TIFA (TRAF interacting protein with an FHA domain) respectively [53, 68, 177]. For such events mediated by FHA domains, two different mechanisms have been reported. The kinase Chk2 and its *S. cerevisiae* orthologue Rad53 dimerise upon phosphorylation. The intrinsically dimeric protein TIFA however forms higher oligomers when phosphorylated. Of these mechanisms, the dimerisation event will probably require a rather specific interaction as both binding partners are free in solution although the cooperative binding of two pThr-FHA interactions will most likely benefit from an avidity effect. The oligomerisation of TIFA dimers will benefit from a similar effect and require an interaction with a similar level of specificity. The intra-molecular association however will only require a weak interaction as the two binding partners have an intra-molecular link and the pThr will sample the FHA binding surface at an increased rate compared to two independent particles in solution.

Although the biochemical measurements of the FhaA isoforms are conclusive, the question of which FhaA residue is phosphorylated *in vivo*, and in the context of the full-length protein, remained to be answered. Due to the long unstructured linker between the N-terminal domain DUF3662 and the C-terminal FHA domain it was not possible to produce stoichiometrically phosphorylated full-length protein. Proteolysis between the two domains also prohibited mass spectrometric analysis of partially phosphorylated protein. When stoichiometrically phosphorylated $FhaA^{367-527}$ was produced, only the Thr377 phosphorylation site could be detected which suggests that PknB has some preference for this site and that Thr377 phosphorylation is unlikely to be an artefact caused by the boundaries of the expression construct.

8.2.3 Structural basis of intra-molecular association

The crystal structure of the FHA domain – pThr peptide complex, along with the model of the intra-molecular equilibrium, shows that intra-molecular associations may be a common FHA domain regulatory mechanism. The crystal structure shows that no contacts are made between the peptide side chains other than the pThr and the FHA domain supporting the relatively weak dissociation constant of 7 µM between them

measured by ITC. This suggests that the intra-molecular association is largely independent of the sequence surrounding the pThr epitope. Despite the weak bi-molecular interaction, the intra-molecular association provides such an increased sampling rate of the FHA binding surface that the dimensionless dissociation constant for the intra-molecular equilibrium was calculated to be between 20 and 100. In the context of its interaction with MviN, this means that the FhaA FHA domain phospho-binding surface is accessible only 1-5% of the time. Through phosphorylation at Thr377, the effective cellular concentration of FhaA phospho-binding activity can be reduced by a factor of 20 to 100. The fact that the intra-molecular association is relatively independent of the sequence of the pThr epitope and only requires a moderately strong affinity means that this mechanism may be utilised as a regulatory mechanism for many FHA domain containing proteins which are phosphorylated. One assumption made here is that the sampling rate would be similar for the full length protein. For the ITC measurements, a linker was used that was only slightly longer than required for the pThr to be included. In full length FhaA, the distance between pThr377 and the FHA domain is identical, but extends further beyond the phosphorylated residue and is also attached to the additional N-terminal domain. The sampling rate may be lower due to the slower diffusion of the much larger linker. On the other hand, the intra-molecular association may explain why FhaA has evolved to contain such a long, unstructured regions since this may permit the movement of the pThr377 epitope to be relatively independent of the N-terminal domain.

8.2.4 FhaA regulation of MviN

A phospho-dependent intra-molecular association of FhaA appears to add another layer of control to the regulation of peptidoglycan export by PknB. As described in chapter 1, FhaA associates with PknB phosphorylated MviN, a flippase which exports peptidoglycan precursors across the mycobacterial cell membrane [127]. The fact that both PknB and FhaA are transcribed from the same operon that also encodes other proteins involved in cell division, such as RodA is consistent with the notion that FhaA is involved in cell wall synthesis. The FhaA FHA domain binds MviN via pThr947, a phospho-epitope generated by PknB. No direct data describing how this interaction affects MviN catalysis are available, but it has been shown that depletion of FhaA causes mycobacterial cells to bulge [127]. Although there may be alternative pathways in which FhaA could contribute to these phenotypes, it now seems likely that FhaA regulates mycobacterial cell wall synthesis through the inhibition of MviN. It has also

been proposed that PknB is activated upon binding to peptidoglycan via its extra-cellular PASTA domains [130]. This completes a negative feedback loop in which high levels of extracellular peptidoglycan activate PknB, which phosphorylates, among other substrates, MviN causing FhaA to bind and inhibit the flippase.

If one assumes the pThr377 site identified in this thesis to be a bona fide PknB phosphorylation site in vivo, a picture of how this intra-molecular association regulates the cellular function of FhaA emerges. The effect of Thr377 phosphorylation on the interaction between FhaA and MviN was investigated in chapter six. In addition, the interaction between the pThr377 epitope and the FHA domain was described structurally and modelled using the relative dissociation constants between the different FhaA and MviN isoforms. It was shown that the phospho-dependent intra-molecular association of FhaA considerably weakens binding to MviN. While the interaction between native FhaA and PknB phosphorylated MviN had been reported previously [127], the data presented here describe the relative affinities for the different isoforms of both FhaA and MviN. Weak phospho-independent binding with a dissociation constant of ~5 μM could be measured between the two proteins. Phosphorylation of MviN by PknB greatly stabilises binding and reduces the dissociation constant to 100 nM. This phospho-dependent binding event however could be disrupted by phosphorylation of FhaA. The data therefore suggest a model for FhaA regulation comparable to that described previously for GarA [68]. In both cases, an FHA domain recognises a single enzyme or set of enzymes via protein-protein interactions with affinities tighter than 100 nM which can then be interrupted by outcompetition through an intra-molecular FHA phospho-peptide interaction. The fact that the peptide is covalently linked to its binding partner means that the effective affinity is far greater than that observed in bimolecular ITC titrations due to an increased rate of sampling of the FHA binding site. A significant difference between the systems is the nature of the binding events with the enzymes. GarA binds to its partners in a phospho-independent interaction. On the other hand, FhaA can also bind its partner via a weaker, phospho-independent interaction, but requires MviN to be phosphorylated for a tight, sub micro-molar affinity. While the interaction abrogated by the FhaA intra-molecular association is a canonical FHA domain interaction (figure 1.7), the enzyme-FHA interaction that is regulated by GarA phosphorylation is an unusual, phospho-independent FHA interaction (figure 1.8). In both cases however, the FHA interaction surface is at the “top” of the FHA domain, where a pThr would expect to bind, to enable the interaction to be regulated by the intra-molecular association. From a functional point of view, the

intra-molecular association adds another layer of control to the regulation of cell wall synthesis by FhaA.

When integrated with the information available about mycobacterial cell wall synthesis, and particularly the roles of FhaA and MviN, some conclusions on how the FhaA intra-molecular switch may function in a cellular context can be drawn. It was shown that phosphorylation of MviN by PknB stabilises its interaction with FhaA which presumably inhibits the enzyme. Phosphorylation of FhaA by the same kinase however destabilises this interaction and thus relieves the inhibition. It is therefore unclear why the action of one kinase may stabilise and destabilise the same one binding event since de-phosphorylation of MviN would be a simpler way of relieving MviN inhibition. As FhaA is not as good a PknB substrate as MviN *in vitro* and unlike MviN, is not attached to the cell membrane, it is likely that MviN will be phosphorylated at a much faster rate than FhaA. Therefore, FhaA phosphorylation could act as a molecular 'timer' to slowly relieve MviN inhibition (figure 8.1). Importantly, using fluorescent labelling, both FhaA and PknB have been found to localise to the cell poles and the septum, the main sites of cell wall synthesis [127, 129]. Furthermore, it is known that, upon sensing high concentrations of peptidoglycan precursors, PknB phosphorylates MviN to recruit FhaA to inhibit MviN activity [127, 129]. Once FhaA is bound to MviN and hence located at the inside of the mycobacterial cell membrane, it may be phosphorylated by PknB to disrupt the FHA/MviN complex. Once the majority of FhaA at a particular location has been phosphorylated, inhibition of MviN is relieved. For rod-shaped bacteria, the localisation of cell wall synthesis alternates between the cell septum and the cell poles during their cell cycle [178]. Following cell division, the septum of the parent cell becomes the cell pole of the daughter cells. The temporospatial organisation of the switch between septal growth and elongation may therefore be where FhaA timer acts. Upon activation of PknB, inhibition of muropeptide export would only persist as long as there is sufficient non-phosphorylated FhaA available. The Ser/Thr phosphatase PstP, transcribed from the same operon as FhaA and PknB, may reset this signalling event by dephosphorylating PknB, MviN and FhaA. As the STPK PknA is also encoded in this same operon, it is not unlikely that this kinase also plays a role in FhaA regulation, potentially integrating signals from other pathways.

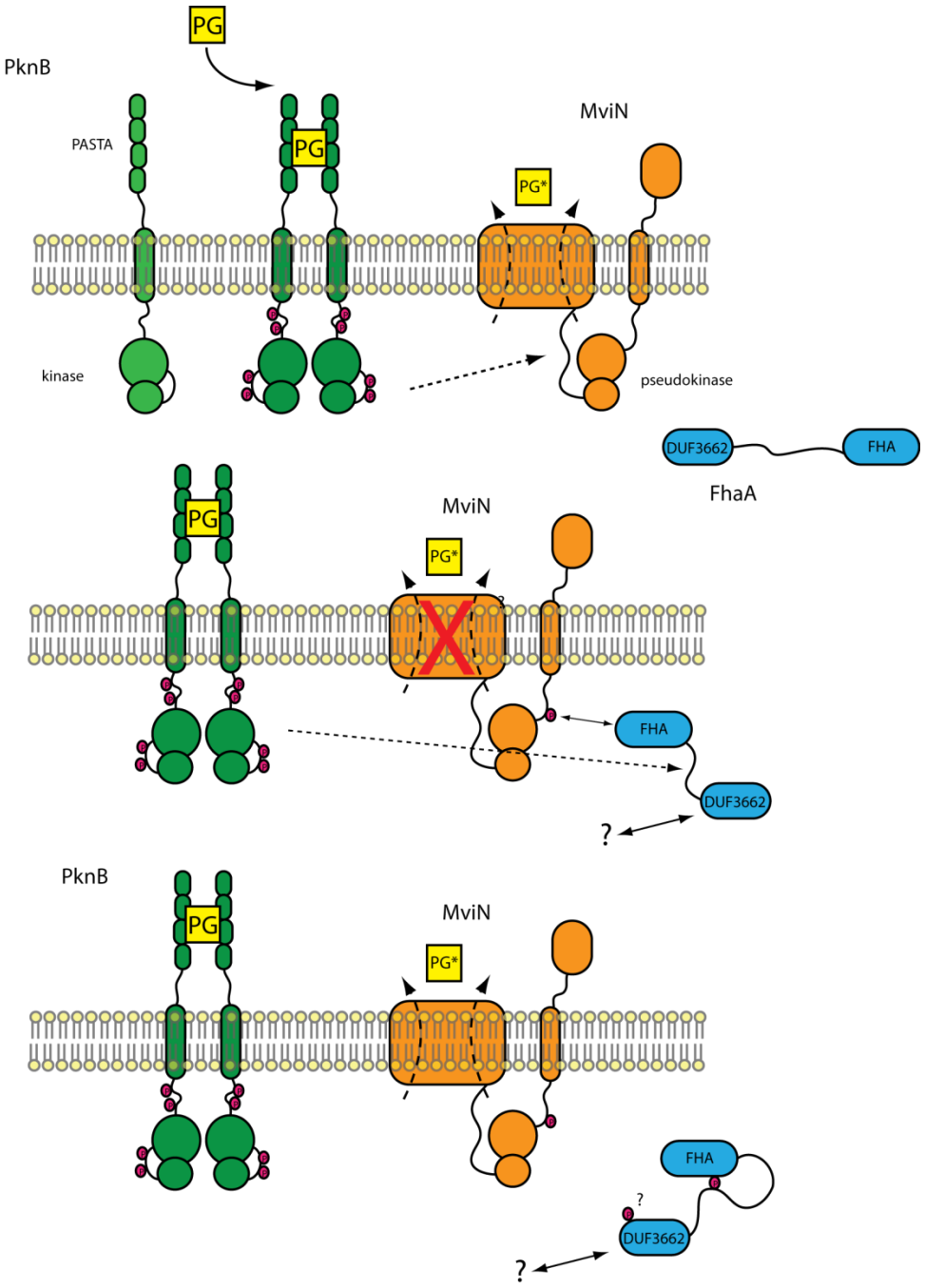
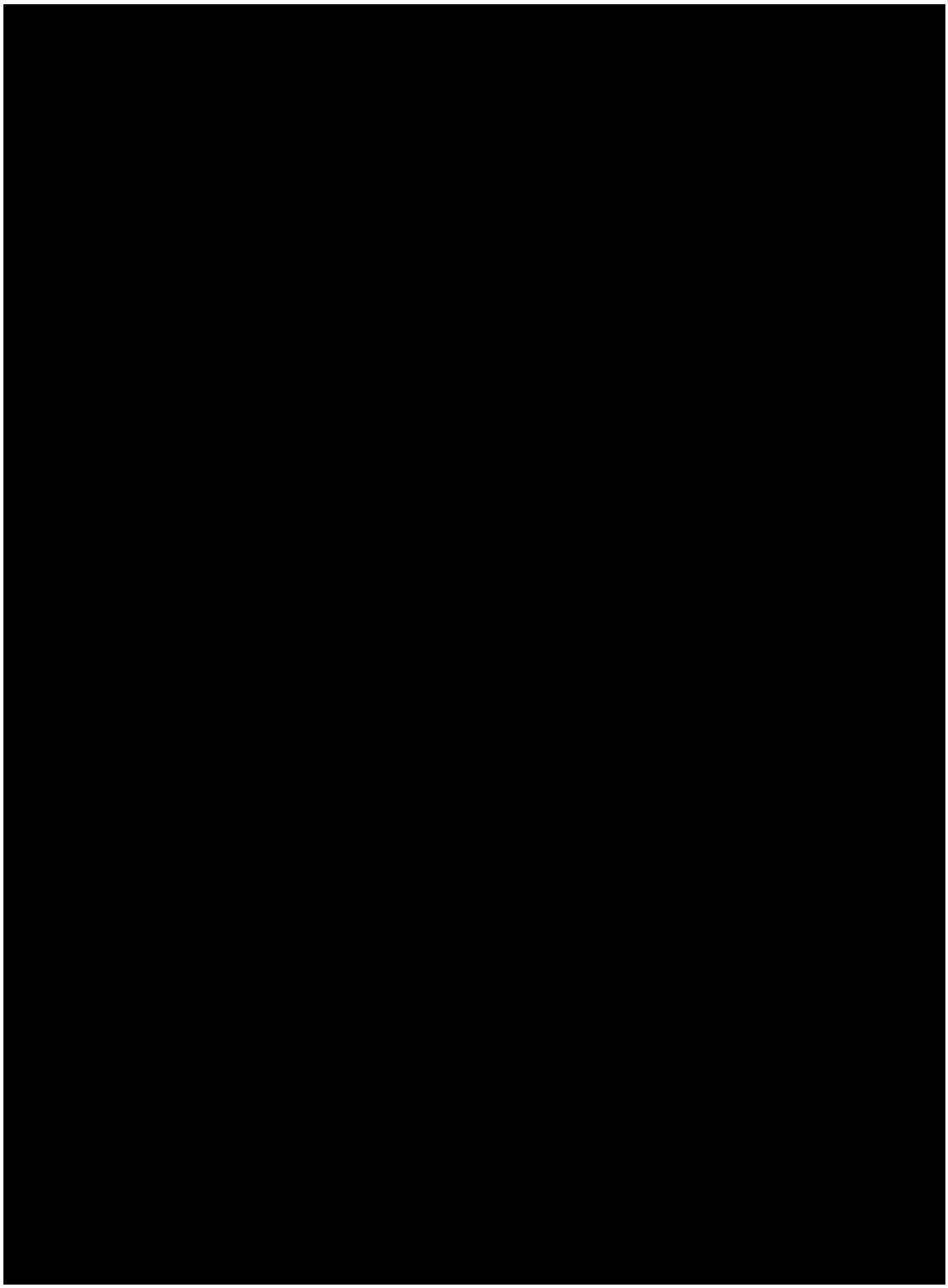


Figure 8.1. Schematic representation of the PknB activation mechanism and the proposed role of FhaA phosphorylation. Polar yellow heads and black hydrophobic tails represent the lipid bilayer of the mycobacterial cell separating the cytosol (bottom) from the extracytosolic space (top). High levels of Peptidoglycan activate PknB (green) which in turn phosphorylates MviN (orange). FhaA (blue) binds the pThr epitope on MviN and inhibits the enzyme. Once FhaA is recruited to the cell membrane, it may be phosphorylated by PknB. The inhibition is then increasingly relieved as more and more FhaA is phosphorylated.

From a phylogenetic and evolutionary viewpoint, it seems that that this extra layer of control of cell wall synthesis might be limited to slow-growing mycobacteria. All of the slow growing mycobacterial species of which there are complete genome sequences available contain the conserved Thr377 motif (figure 8.2). This motif is not found in fast-growing mycobacteria. In the FhaA orthologues of these species, no threonines can be found for 200 residues N-terminal to the FHA domain. Unfortunately, full genome sequences are only available for less than a dozen mycobacterial species outside the medically relevant pathogenic species so this trend might be modified as more genome data become available. Indeed it could be that the lack of a Thr377 epitopes may only be due to the overall larger genetic distance between *Mycobacterium tuberculosis*, itself a slow-grower, and the fast-growing mycobacteria. That this regulation occurs only in slow growing mycobacteria is perhaps counterintuitive as FhaA phosphorylation relieves inhibition of MviN and would therefore most likely increase the rate of cell wall synthesis. It may therefore be that there are additional, as yet unknown factors that contribute to overall MviN regulation *in vivo*.

The data presented here add considerably to the information available about STPK signalling in *Mycobacterium tuberculosis*, particularly how FHA domains interact with the phospho-epitopes generated by STPKs and the role of PknB in the regulation of cell wall synthesis. While it had been long established that misregulation of PknB leads to morphological aberrations in mycobacterial cells and that most of its primary substrates are involved in cell wall regulation, the mechanisms through which these defects are effected have been less clear [179]. Previously, two roles for the FhaA FHA domain had been established, namely the inhibition of MviN and the recruitment of PknB via activation loop and juxtamembrane region mediated interactions which are essential for FhaA phosphorylation [104, 126, 127]. Here, a third role for the FhaA FHA domain has been found which describes how it engages in an intra-molecular association to bind to its own pThr377 epitope and abrogate MviN inhibition. Reminiscent of the GarA FHA domain, the FhaA FHA domain also engages in an intra-molecular association with a pThr epitope to relieve inhibition of target enzymes.



8.3 PIH1D1

The data presented in chapter seven describe the investigation of a novel STPK signalling pathway in human cells with a focus on a putative phospho-binding domain within the human protein PIH1D1 that is essential for maturation of PI3-kinase-like kinases.

Experiments were designed to quantitatively measure the affinities of PIH1D1 for phospho-peptides derived from its phosphorylated binding partners alongside mutagenesis studies based on bioinformatic identification of the residues essential for phospho-binding. As for all modular phospho-reader domains, knowledge of the binding specificity is vital in determining which kinase substrates PIH1D1 may bind and how it performs its cellular function. To this end, the consensus sequence of PIH1D1 was determined using peptide spot arrays. Although the kinase substrates Tel2, snRNP 116 and ECD had already been shown to bind PIH1D1 in a phospho-dependent manner, a clearer description of the PIH1D1 consensus sequence would potentially allow identification of further interacting partners using sequence searches of the protein sequence database. A structural description of this putative phospho-binding domain was pursued to determine the mechanism through which PIH1D1 binds phospho-peptides and how this compares to other phospho-binding proteins.

8.3.1 Identification of a stable PIH1D1 construct

As a first step, a stable expression construct of PIH1D1 needed to be identified since, although it was possible to express soluble full-length PIH1D1 from *E. coli*, the protein degraded quickly during the purification process. From an alignment of PIH1D1 sequences, it is apparent that there are two regions of high sequence homology, which lie between residues 20-190 and 210-290 (figure 8.3). In contrast, the extreme N-terminus and the region between 190 and 210 are not highly conserved. This suggested that PIH1D1 may consist either of two separate domains or a single domain with an extended loop around 190-210. Limited tryptic proteolysis identified a stable fragment consisting of residues 1 to 180. As this fragment was found to be soluble when expressed from E. coli, it was assumed that PIH1D1 may indeed be made up of two separate domains, or that at least the N-terminal part of a putative single large domain can fold independently. Limited chymotryptic proteolysis largely confirmed these results but suggested a fragment consisting of residues 40-168. While the extreme N-terminus of PIH1D1 may not have been essential for the stability of the first

domain, the 1-180 construct was soluble and stable and hence used for binding experiments and crystallisation trials.

8.3.2 PIH1D1 binds phospho-peptides

Simon Boulton's laboratory had previously shown by immunoprecipitation that proteins Tel2, ECD and snRNP 116 all bind to PIH1D1 in a phospho-dependent manner. These three proteins all contain similar casein kinase 2 phosphorylation sites which appeared to be sufficient for binding to PIH1D1 as evidenced by the fact that PIH1D1 can be pulled down from HEK293 cell lysate by phospho-peptides derived from these CK2 sites [147] (personal communication).

It was attempted to reproduce these binding events *in vitro* and to determine equilibrium binding constants for each interaction. To these ends, synthetic peptides derived from the phospho-epitopes were used. The use of synthetic peptides instead of proteins in phospho-binding experiments has been widely adopted since the pTyr binding potential of the SH2 domain, the first known phospho-reader domain was discovered [36]. Although many different methods for in-solution binding measurements are available, ITC has become the established method for accurately determining the dissociation constants in these systems largely because of the substantial heat changes observed upon phosphopeptide binding. Some accessory contributions towards a phospho-dependent binding event from the overall folds of the binding partner may be lost in a protein-peptide binding experiment. The Ki67 FHA domain for example binds hNIFK via a combination of phospho-dependent and phospho-independent interactions (figure 1.5). In addition to the canonical FHA-pThr interaction, part of the peptide forms an additional β-strand which is hydrogen-bonded to the FHA domain. Such interactions are exceptions however and it is considerably easier to synthesise or purchase a peptide with the desired sequence than cloning, expressing, purifying and stoichiometrically phosphorylating a protein-binding partner. The sequence specificities of most kinases are such that when incubated with a substrate *in vitro*, they will phosphorylate several residues with varying efficiency. As a consequence it is very difficult to produce proteins that are stoichiometrically mono-phosphorylated on the desired residue. The fact that it was possible to generate stoichiometrically phosphorylated FhaA and MviN for ITC measurements, as described in Chapters 5 and 6, is more the exception rather than the rule.

Another reason why phospho-peptides may be used instead of phosphorylated proteins in ITC measurements is the fact that in the majority of crystal structures of phospho-reader domain phospho-peptide complexes, the phospho-ligand is in an

extended conformation and does not contain any secondary structure. The Ki67 interaction with a phospho-peptide containing an α-helix and an extended interaction surface is again an exception. Since in many cases the crystal structures showing phospho-reader domains in complex with phospho-peptides in extended conformations can explain the sequence specificities of these domains, it is likely that these conformations are not artefacts of looking at short motifs out of the context of full-length proteins [39, 41, 61]. Furthermore, phosphoproteomics has shown that the majority of phosphorylation sites of proteins are outside structured domains and predominantly found in intrinsically unstructured regions, presumably to allow kinases, phosphatases and phospho-reader domains access to these phospho-sites [180]. In the case of Tel2 specifically, the phospho-site that PIH1D1 binds to (pSer487/pSer491) is in an unstructured region since the crystal structure of the *S. cerevisiae* Tel2 homologue shows that the extended loop in which this conserved phospho-site is located could not be seen in the electron density and is hence most likely disordered [146].

In some cases, phospho-peptides can be ligated to a protein in order to generate a specific phosphorylation state of a protein. Human Chk2 kinase is highly phosphorylated by several kinases including ATM, ATR, DNA-PK and itself on a number of residues not all of which have been fully characterised [181, 182]. In order to investigate the role of a single phospho-site, pThr68, part of Chk2 was expressed recombinantly and dephosphorylated before it was ligated to a phospho-peptide with the pThr68 epitope to generate mono-phosphorylated Chk2 [53]. The limitations of this method however are that it requires a cysteine for the chemical ligation on the N-terminus of the protein and that the recombinantly expressed part of the protein needs to be able to fold independently.

8.3.3 Sequence specificity of the PIH1D1 domain

As it could be assumed that the phospho-dependent binding of PIH1D1 to its interacting partners can be safely approximated using phospho-peptides derived from the respective phospho-epitopes, the interactions were measured by ITC. The doubly phosphorylated Tel2 pSer487/pSer491 phospho-peptide bound to the PIH1D1$^{1-180}$ with an affinity of 1.3 μM, similar to the affinities reported for other established phospho-reader domains. Although many well characterised phospho-binding domains such as SH2 and FHA domains bind optimal peptide substrates selected from oriented peptide libraries or library arrays with dissociation constants stronger than 0.1 μM, many interactions that have been shown to biologically relevant were measured to be considerably weaker with affinities as low as 10 μM [36, 59, 61, 69]. A peptide derived

from two casein kinase 2 sites in ECD (pSer503/pSer505) bound with an affinity of 5.6 μM. This binding event was weaker than the Tel2 interaction, but still well within a physiologically significant range. The weak interaction with snRNP 116 (pSer19) which had a dissociation constant of approximately 60 μM suggests the phospho-dependent interactions of PIH1D1 with this protein are substantially weaker, but additional phospho-independent interactions with the PIH1D1 C-terminal region may contribute to the PIH1D1-snRNP 116 binding event *in vivo*. Regardless, all three peptides had previously been shown to pull down PIH1D1 from HEK293 cell lysate and the ITC data show that these results are explained by phospho-dependent interactions in vitro, albeit with a wide range of apparent affinities.

To test whether PIH1D1 contains two phospho-binding pockets that both need to be occupied for efficient binding, mono-phosphorylated variants of the Tel2 peptide were synthesised. It could be seen that PIH1D1 binds to the pSer491 peptide with nearly the same affinity as to the pSer487/pSer491 peptide and only displayed weak binding to the pSer487 peptide. All stoichiometries were measured to be around 0.8, suggesting PIH1D1 has a single major phospho-binding surface. The protein binds preferentially to pSer491 over pSer487 with pSer487 in an auxiliary role only for the doubly-phosphorylated peptide. All peptides that showed strong binding contained a D-pS-D-D sequence. Although mono-phosphorylated peptides are sufficient for binding to most phospho-reader domains, additional phospho-sites have been found to stabilise binding in many cases. The PIH1D1 preference for doubly-phosphorylated peptides is not unlike the PNK (polynucleotide kinase) FHA domain where one phosphorylated threonine is sufficient for ~10 μM binding and the addition of further phospho-epitopes incrementally increases the affinity up to 100 nM for a peptide with 4 phospho-residues [183]. Similarly, the Ki67 and aprataxin FHA domains prefer poly-phosphorylated peptides over mono-phosphorylated ones [183, 184].

The clustering of phosphorylatable residues does not only improve binding to phospho-reader domains, but may also increase the rate at which a protein is phosphorylated. For acidophilic kinases, the phosphorylation of one residue and thus addition of two negative charges will increase the chances of nearby residues being phosphorylated. The protein Rad51 for example has been shown to require priming phosphorylation on Ser14 by the cell-cycle regulated Polo-like kinase 1 before it can be phosphorylated on Thr13 by the constitutively active CK2 that is required for binding of the FHA domain of Nbs1 [185]. Thus, this system has evolved to link the recruitment of a CK2 site binding protein to the regulation of a different kinase. In the case of Tel2, Ser487, Ser491 and to a lesser extent Ser485 are phosphorylated by the same kinase, but it may be that

the presence of several serines in close proximity, which can all be phosphorylated, increases phosphorylation of Ser491.

PIH1D1 preferentially binds to pSer epitopes over pThr ones with an approximately 6 fold stronger dissociation constant. Apart from the FHA domain which has absolute specificity for pThr, the majority of pSer and pThr binding proteins have dual specificity for both types of epitopes. The slight preference for pSer epitopes is, for example, also displayed by BRCT domains [174].

Attempts were made to determine the consensus sequence of PIH1D1 by oriented peptide array libraries and to find out which additional residues are also contributing to binding. Although the OPAL array does not generate absolute affinities, it is useful in screening large numbers of peptides. Initially, a degenerate peptide library was spotted with a fixed pSer at the centre (X-X-X-X-pS-X-X-X-X) and an additional position fixed for each pool. Strong binding could not be detected for any pool of peptides suggesting PIH1D1 requires more than two conserved amino acids for efficient binding. Thus a second array was printed using the Tel2 pSer491 peptide as a base sequence with a point mutation for each position. Here, a clear binding pattern could be detected and the D-pS-D-D consensus sequence that had been suggested from the ITC experiments could be largely confirmed. It appears the PIH1D1 phospho-binding domain is much less promiscuous than most other phospho-binding domains. Most phospho-binding domains only have specificity for one phospho-peptide residue in addition to the phosphorylated one, most often in the +3 position. Many FHA domains for example have pT-X-X-I/L/V specificity and some BRCT domains prefer pS-X-X-F containing peptides [186, 187]. In terms of selectivity, the most similar domain to the PIH1D1 domain is the POLO-Box domain, which also has strong specificity for the residues directly N- and C- terminal to the phospho-residue (S-pS/pT-P).

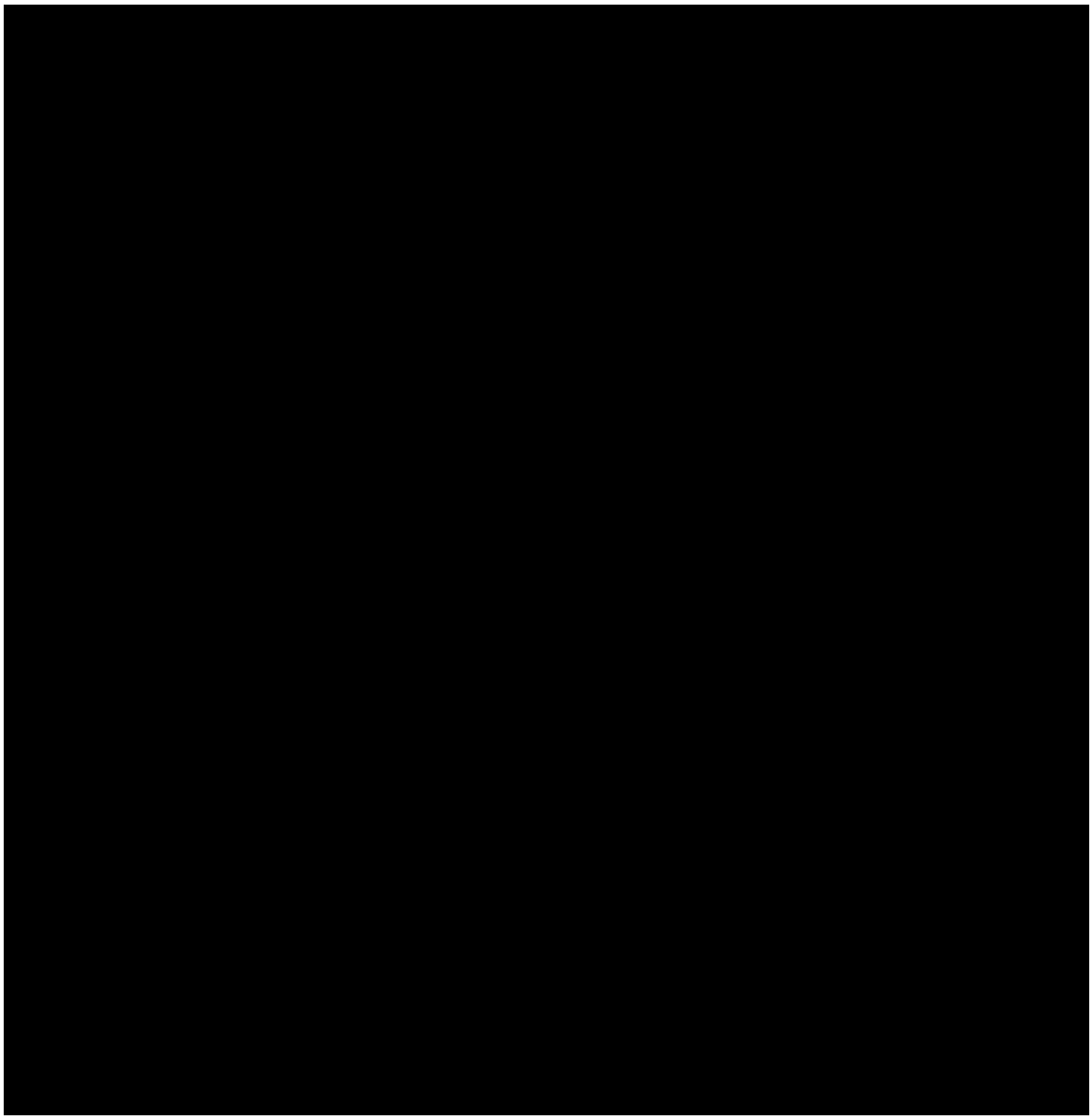
8.3.4 The K64A mutant abrogates phospho-dependent PIH1D1 interactions

Using sequence homology, a single lysine residue (Lys64) was identified whose mutation was found to abrogate phospho-dependent PIH1D1 interactions, but not interfere with its phospho-independent interactions. The fact that the PIH1D1 K64A mutant only disrupts a sub-set of interactions suggests this protein possesses two or more binding surfaces. The phospho-dependent interaction surface is clearly located on the N-terminus of the protein and the phospho-independent one likely closer to the C-terminus. Knowledge that this mutation interrupts only one binding surface enables investigation of the cellular phenotypes specifically induced by the loss of the phospho-binding capabilities of PIH1D1 without disrupting the whole protein. Such mutations

provide us with the tools to determine the functions of each domain in multi-domain proteins. They have for example been utilised to determine the respective roles for the two FHA domains of Rad53 kinase (section 1.3.5) [71]. Furthermore, for phospho-reader domains such as some FHA domains which have both phospho-dependent and phospho-independent binding-surfaces, such a mutation will enable one to distinguish between the two types of interaction as long as these surfaces do not overlap [68]. The phospho-independent GarA-KGD interaction (figure 1.7) for example was initially erroneously annotated as phospho-dependent when the mutation of a conserved serine essential for phospho-binding abrogated binding to KGD [75]. In fact the interaction is phospho-independent but both types of interactions of the GarA FHA domain are mediated by the same FHA domain surface.

8.3.4 Crystallisation of PIH1D1

While the 1-180 expression construct of PIH1D1 could not be crystallised, a 40-180 construct, which displayed comparable phospho-binding capabilities, crystallised in several forms, suggesting the 40-180 core homology region of PIH1D1 forms a highly ordered protein domain agreeing with the stability displayed by the protein during the biophysical experiments. Seleno-methionine containing crystals were obtained but did not diffract sufficiently for the crystal structure of PIH1D1 to be solved. Nonetheless, the progress described indicates that structure solution is likely to be possible in the near future



8.4 Future perspectives

8.4.1 FhaA

The data presented in this thesis have shown that the novel mechanism for FHA domain phospho-binding regulation described previously for GarA, is not unique to this molecule and seems to play an important role in other FHA domain containing proteins. However, several questions remain.

The hypothesis of a pThr377 driven intra-molecular association and its role in MviN regulation needs to be tested in living cells. Since it has already been shown that deletion of FhaA leads to morphological defects giving rise to shorter and wider cells in mycobacteria, one may speculate that disrupting the intra-molecular association through introduction of a non-phosphorylatable T377A mutant may cause the opposite effects [127]. In addition to cell morphology, the growth rate of such a mutant strain would also be expected to vary.

Overall, roles of both the long linker region and the N-terminal domain remain to be determined. Although a structure has been reported for this domain, its function remains elusive and no interacting partners could be found. The cellular phenotype of a deletion of this domain, or its overexpression, may provide some clues about its cellular activity. No function has been assigned to the long linker rich in glycine, tyrosine and proline residues that separates the two globular FhaA domains. One may test whether this linker is a substrate for the *M. tuberculosis* tyrosine kinase PtkA. This kinase has been shown to phosphorylate a tyrosine residue directly preceded by proline in the mycobacterial protein PtpA [188]. Indeed, such an epitope, and many more solvent accessible tyrosine residues, can be found in the FhaA linker.

From a protein engineering perspective, it would be interesting to see how the equilibrium constant of the intra-molecular association is linked to both the linker length and the apparent affinity for the phospho-epitope. Presumably, shorter distances between the pThr and the FHA domain will lead to higher affinities due to a higher sampling rate. If the linker becomes too short to make the interaction however, a phospho-dependent oligomerisation may be favoured. Possible mechanisms include Chk2 like head to tail dimerization or higher oligomers to be formed by ‘daisy-chain’ assemblies [53]. Similarly, the equilibrium constants of the GarA intra-molecular associations could be determined to allow for comparisons to be made between the two systems. GarA is thought to be able to engage in 3 different intra-molecular

associations. Two mutually exclusive phosphorylation sites have been reported to trigger such a binding event and even the native protein is thought to engage in an intra-molecular association [68, 69, 75].

On a wider 'systems' level, phosphoproteomics studies are expected to provide new information about the connectivity of mycobacterial Ser/Thr kinase signalling. As the majority of mycobacterial STPK have sequence specificities that also favour FHA domain binding, FHA domain containing proteins are likely to play many important roles in mycobacterial signal transduction [101].

8.4.2 PIH1D1

In terms of *in vitro* experiments, the most significant outstanding issue is that of the molecular structure of the protein-phosphopeptide complex that would definitively identify residues that are essential to phospho-ligand binding affinity and specificity, informing further investigation by mutagenesis studies in vitro and in vivo. Two further putative PIH domain containing proteins, PIH1D2 and kintoun, are encoded in the human genome but their phospho-binding capabilities remain to be determined. As well as other human PIH domain containing proteins, the PIH1D1 orthologues from other species could be characterised biochemically and structurally to determine which functions of the protein are conserved. The consensus sequence of PIH1D1 may also be refined further. Interestingly, the OPAL array suggested a preference for isoleucine, leucine and valine in the +3 position, similar to many FHA domains. A phospho-peptide incorporating all PIH1D1 preferences seen from the OPAL array should be tested for PIH1D1 binding to determine the affinity for such an optimal peptide. In term, these could be used in co-crystallisation experiments. The phospho-independent binding surface of PIH1D1 also needs to be better characterised. Since the *S. cerevisiae* orthologue of PIH1D1 does not contain the C-terminal domain, it is likely that at least one phospho-independent binding surface is located on the N-terminal domain.

Finally, it remains to be seen whether phospho-dependent PIH1D1 interactions are essential just for SMG1 and mTOR maturation, or also for the stabilisation of other proteins. The PIH1D1 consensus sequence of D-pS-D-D agrees well with the substrate specificity of CK2 which selects for acidic residues both N-terminal and C-terminal to the phosphorylated residue [189]. There are several other phospho-reader proteins that have evolved to bind to CK2 sites. The FHA and BRCT domain containing protein Nbs1 and the aprataxin FHA domain for example have specificity for pS-D-pT-D motifs

[56, 190]. The sequence specificity of PIH1D1 suggests that this protein may interact with a different subset of CK2 substrates potentially recruiting these to the R2TP-Hsp90-prefoldin complex for maturation. These proteins may be recruited through a phosphorylated adaptor protein such as Tel2, which is a PIKK co-factor, or they may be phosphorylated directly. It has been shown that some proteins are protected from caspases by CK2 mediated phosphorylation near caspase recognition sites [191]. Similarly, CK2 phosphorylation protects XRCC1 from ubiquitination and subsequent proteasomal degradation [192]. The recruitment of CK2 substrates to a central chaperone complex via the PIH1D1 phospho-reader domain may be another way in which CK2 manipulates cellular concentrations of its target proteins.

9.0 Bibliography

1. Krebs, E.G. and J.A. Beavo, *Phosphorylation-dephosphorylation of enzymes.* Annual review of biochemistry, 1979. **48**: p. 923-59.
2. Manning, G., et al., *The protein kinase complement of the human genome.* Science, 2002. **298**(5600): p. 1912-34.
3. Veech, R.L., et al., *Cytosolic phosphorylation potential.* The Journal of biological chemistry, 1979. **254**(14): p. 6538-47.
4. Hutter, M.C. and V. Helms, *Influence of key residues on the reaction mechanism of the cAMP-dependent protein kinase.* Protein science : a publication of the Protein Society, 1999. **8**(12): p. 2728-33.
5. Seet, B.T., et al., *Reading protein modifications with interaction domains.* Nature reviews Molecular cell biology, 2006. **7**(7): p. 473-83.
6. Manning, G., et al., *The protist, Monosiga brevicollis, has a tyrosine kinase signaling network more elaborate and diverse than found in any known metazoan.* Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(28): p. 9674-9.
7. Matsuoka, S., et al., *ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage.* Science, 2007. **316**(5828): p. 1160-6.
8. Nigg, E.A., *Mitotic kinases as regulators of cell division and its checkpoints.* Nature reviews Molecular cell biology, 2001. **2**(1): p. 21-32.
9. Hunter, T., *Tyrosine phosphorylation: thirty years and counting.* Current opinion in cell biology, 2009. **21**(2): p. 140-6.
10. Lizcano, J.M. and D.R. Alessi, *The insulin signalling pathway.* Current biology : CB, 2002. **12**(7): p. R236-8.
11. Wetlaufer, D.B., *Nucleation, rapid folding, and globular intrachain regions in proteins.* Proceedings of the National Academy of Sciences of the United States of America, 1973. **70**(3): p. 697-701.
12. Kühne, W., *Über das Verhalten verschiedener organisirter und sog. ungeformter Fermente.* Verhandlungen des Naturhistorisch-medizinischen Vereins zu Heidelberg, 1877. **1**(3): p. 190.
13. Blake, C.C., et al., *Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Angstrom resolution.* Nature, 1965. **206**(4986): p. 757-61.
14. Moran, M.F., et al., *Src homology region 2 domains direct protein-protein interactions in signal transduction.* Proceedings of the National Academy of Sciences of the United States of America, 1990. **87**(21): p. 8622-6.
15. Lim, W.A. and T. Pawson, *Phosphotyrosine signaling: evolving a new cellular communication system.* Cell, 2010. **142**(5): p. 661-7.
16. Gibson, T.J., *Cell regulation: determined to signal discrete cooperation.* Trends in biochemical sciences, 2009. **34**(10): p. 471-82.
17. Greenstein, A.E., et al., *Structure/function studies of Ser/Thr and Tyr protein phosphorylation in Mycobacterium tuberculosis.* Journal of molecular microbiology and biotechnology, 2005. **9**(3-4): p. 167-81.
18. Stock, A.M., V.L. Robinson, and P.N. Goudreau, *Two-component signal transduction.* Annual review of biochemistry, 2000. **69**: p. 183-215.
19. Forst, S., J. Delgado, and M. Inouye, *Phosphorylation of OmpR by the osmosensor EnvZ modulates expression of the ompF and ompC genes in Escherichia coli.* Proceedings of the National Academy of Sciences of the United States of America, 1989. **86**(16): p. 6052-6.

20. Johnson, L.N., M.E. Noble, and D.J. Owen, *Active and inactive protein kinases: structural basis for regulation.* Cell, 1996. **85**(2): p. 149-58.
21. Szarek, P., et al., *Physical nature of intermolecular interactions within cAMP-dependent protein kinase active site: differential transition state stabilization in phosphoryl transfer reaction.* The journal of physical chemistry B, 2008. **112**(37): p. 11819-26.
22. Valiev, M., et al., *The role of the putative catalytic base in the phosphoryl transfer reaction in a protein kinase: first-principles calculations.* Journal of the American Chemical Society, 2003. **125**(33): p. 9926-7.
23. Hubbard, S.R., *Crystal structure of the activated insulin receptor tyrosine kinase in complex with peptide substrate and ATP analog.* The EMBO journal, 1997. **16**(18): p. 5572-81.
24. Xu, W., et al., *Crystal structures of c-Src reveal features of its autoinhibitory mechanism.* Molecular cell, 1999. **3**(5): p. 629-38.
25. Kanaseki, T., et al., *Structural features of Ca2+/calmodulin-dependent protein kinase II revealed by electron microscopy.* The Journal of cell biology, 1991. **115**(4): p. 1049-60.
26. Miller, S.G. and M.B. Kennedy, *Regulation of brain type II Ca2+/calmodulin-dependent protein kinase by autophosphorylation: a Ca2+-triggered molecular switch.* Cell, 1986. **44**(6): p. 861-70.
27. Ziegler, S.F., et al., *Novel protein-tyrosine kinase gene (hck) preferentially expressed in cells of hematopoietic origin.* Molecular and cellular biology, 1987. **7**(6): p. 2276-85.
28. Won, A.P., J.E. Garbarino, and W.A. Lim, *Recruitment interactions can override catalytic interactions in determining the functional identity of a protein kinase.* Proceedings of the National Academy of Sciences of the United States of America, 2011. **108**(24): p. 9809-14.
29. Alexander, J., et al., *Spatial exclusivity combined with positive and negative selection of phosphorylation motifs is the basis for context-dependent mitotic signaling.* Science signaling, 2011. **4**(179): p. ra42.
30. Joseph, R.E., et al., *A remote substrate docking mechanism for the tec family tyrosine kinases.* Biochemistry, 2007. **46**(18): p. 5595-603.
31. Krebs, E.G., A.B. Kent, and E.H. Fischer, *The muscle phosphorylase b kinase reaction.* The Journal of biological chemistry, 1958. **231**(1): p. 73-83.
32. Owen, D.J., et al., *The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase gcn5p.* The EMBO journal, 2000. **19**(22): p. 6141-9.
33. Brahms, H., et al., *Symmetrical dimethylation of arginine residues in spliceosomal Sm protein B/B' and the Sm-like protein LSm4, and their interaction with the SMN protein.* RNA, 2001. **7**(11): p. 1531-42.
34. Koch, C.A., et al., *The common src homology region 2 domain of cytoplasmic signaling proteins is a positive effector of v-fps tyrosine kinase function.* Molecular and cellular biology, 1989. **9**(10): p. 4131-40.
35. Felder, S., et al., *SH2 domains exhibit high-affinity binding to tyrosine-phosphorylated peptides yet also exhibit rapid dissociation and exchange.* Molecular and cellular biology, 1993. **13**(3): p. 1449-55.
36. Ladbury, J.E., et al., *Measurement of the binding of tyrosyl phosphopeptides to SH2 domains: a reappraisal.* Proceedings of the National Academy of Sciences of the United States of America, 1995. **92**(8): p. 3199-203.
37. Elia, A.E.H., L.C. Cantley, and M.B. Yaffe, *Proteomic screen finds pSer/pThr-binding domain localizing Plk1 to mitotic substrates.* Science, 2003. **299**(5610): p. 1228-31.
38. Manke, I.A., et al., *BRCT repeats as phosphopeptide-binding modules involved in protein targeting.* Science, 2003. **302**(5645): p. 636-9.
39. Yaffe, M.B., et al., *The structural basis for 14-3-3:phosphopeptide binding specificity.* Cell, 1997. **91**(7): p. 961-71.

40. Elia, A.E.H., et al., *The molecular basis for phosphodependent substrate targeting and regulation of Plks by the Polo-box domain.* Cell, 2003. **115**(1): p. 83-95.
41. Yaffe, M.B. and S.J. Smerdon, *The use of in vitro peptide-library screens in the analysis of phosphoserine/threonine-binding domain structure and function.* Annu Rev Biophys Biomol Struct, 2004. **33**: p. 225-44.
42. Li, J., et al., *Structural and functional versatility of the FHA domain in DNA-damage signaling by the tumor suppressor kinase Chk2.* Molecular cell, 2002. **9**(5): p. 1045-54.
43. Jackson, S.P. and J. Bartek, *The DNA-damage response in human biology and disease.* Nature, 2009. **461**(7267): p. 1071-8.
44. Polo, S.E. and S.P. Jackson, *Dynamics of DNA damage response proteins at DNA breaks: a focus on protein modifications.* Genes & development, 2011. **25**(5): p. 409-33.
45. Bork, P., et al., *A superfamily of conserved domains in DNA damage-responsive cell cycle checkpoint proteins.* The FASEB journal : official publication of the Federation of American Societies for Experimental Biology, 1997. **11**(1): p. 68-76.
46. Mohammad, D.H. and M.B. Yaffe, *14-3-3 proteins, FHA domains and BRCT domains in the DNA damage response.* DNA Repair, 2009. **8**(9): p. 1009-17.
47. Downs, J.A., N.F. Lowndes, and S.P. Jackson, *A role for Saccharomyces cerevisiae histone H2A in DNA repair.* Nature, 2000. **408**(6815): p. 1001-4.
48. Rogakou, E.P., et al., *DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139.* The Journal of biological chemistry, 1998. **273**(10): p. 5858-68.
49. Stiff, T., et al., *ATM and DNA-PK function redundantly to phosphorylate H2AX after exposure to ionizing radiation.* Cancer research, 2004. **64**(7): p. 2390-6.
50. Stucki, M., et al., *MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks.* Cell, 2005. **123**(7): p. 1213-26.
51. Lou, Z., et al., *MDC1 is coupled to activated CHK2 in mammalian DNA damage response pathways.* Nature, 2003. **421**(6926): p. 957-61.
52. Jungmichel, S., et al., *The molecular basis of ATM-dependent dimerization of the Mdc1 DNA damage checkpoint mediator.* Nucleic Acids Res, 2012.
53. Li, J., et al., *Chk2 oligomerization studied by phosphopeptide ligation: implications for regulation and phosphodependent interactions.* The Journal of biological chemistry, 2008. **283**(51): p. 36019-30.
54. Cai, Z., N.H. Chehab, and N.P. Pavletich, *Structure and activation mechanism of the CHK2 DNA damage checkpoint kinase.* Molecular cell, 2009. **35**(6): p. 818-29.
55. Chapman, J.R. and S.P. Jackson, *Phospho-dependent interactions between NBS1 and MDC1 mediate chromatin retention of the MRN complex at sites of DNA damage.* EMBO reports, 2008. **9**(8): p. 795-801.
56. Lloyd, J., et al., *A supramodular FHA/BRCT-repeat architecture mediates Nbs1 adaptor function in response to DNA damage.* Cell, 2009. **139**(1): p. 100-11.
57. Matsuzaki, K., A. Shinohara, and M. Shinohara, *Forkhead-associated domain of yeast Xrs2, a homolog of human Nbs1, promotes nonhomologous end joining through interaction with a ligase IV partner protein, Lif1.* Genetics, 2008. **179**(1): p. 213-25.
58. Hofmann, K. and P. Bucher, *The FHA domain: a putative nuclear signalling domain found in protein kinases and transcription factors.* Trends in biochemical sciences, 1995. **20**(9): p. 347-9.
59. Durocher, D., et al., *The FHA domain is a modular phosphopeptide recognition motif.* Molecular cell, 1999. **4**(3): p. 387-94.
60. Mahajan, A., et al., *Structure and function of the phosphothreonine-specific FHA domain.* Science signaling, 2008. **1**(51): p. re12.
61. Pennell, S., et al., *Structural and functional analysis of phosphothreonine-dependent FHA domain interactions.* Structure, 2010. **18**(12): p. 1587-95.

62. Spivey, V.L., et al., *Forkhead-associated (FHA) Domain Containing ABC Transporter Rv1747 Is Positively Regulated by Ser/Thr Phosphorylation in Mycobacterium tuberculosis.* The Journal of biological chemistry, 2011. **286**(29): p. 26198-209.
63. Liao, H., et al., *Structure of the FHA1 domain of yeast Rad53 and identification of binding sites for both FHA1 and its target protein Rad9.* Journal of molecular biology, 2000. **304**(5): p. 941-51.
64. Liao, H., I.J. Byeon, and M.D. Tsai, *Structure and function of a new phosphopeptide-binding domain containing the FHA2 of Rad53.* Journal of molecular biology, 1999. **294**(4): p. 1041-9.
65. Wang, P., et al., *II. Structure and specificity of the interaction between the FHA2 domain of Rad53 and phosphotyrosyl peptides.* Journal of molecular biology, 2000. **302**(4): p. 927-40.
66. Takagi, M., et al., *A novel nucleolar protein, NIFK, interacts with the forkhead associated domain of Ki-67 antigen in mitosis.* The Journal of biological chemistry, 2001. **276**(27): p. 25386-91.
67. Byeon, I.-J.L., et al., *Sequential phosphorylation and multisite interactions characterize specific target recognition by the FHA domain of Ki67.* Nature structural & molecular biology, 2005. **12**(11): p. 987-93.
68. Nott, T.J., et al., *An intramolecular switch regulates phosphoindependent FHA domain interactions in Mycobacterium tuberculosis.* Science signaling, 2009. **2**(63): p. ra12.
69. Villarino, A., et al., *Proteomic identification of M. tuberculosis protein kinase substrates: PknB recruits GarA, a FHA domain-containing protein, through activation loop-mediated interactions.* Journal of molecular biology, 2005. **350**(5): p. 953-63.
70. O'Hare, H.M., et al., *Regulation of glutamate metabolism by protein kinases in mycobacteria.* Molecular microbiology, 2008. **70**(6): p. 1408-23.
71. Pike, B.L., et al., *Diverse but overlapping functions of the two forkhead-associated (FHA) domains in Rad53 checkpoint kinase activation.* The Journal of biological chemistry, 2003. **278**(33): p. 30421-4.
72. Sweeney, F.D., et al., *Saccharomyces cerevisiae Rad9 acts as a Mec1 adaptor to allow Rad53 activation.* Current biology : CB, 2005. **15**(15): p. 1364-75.
73. Sun, Z., et al., *Rad53 FHA domain associated with phosphorylated Rad9 in the DNA damage checkpoint.* Science, 1998. **281**(5374): p. 272-4.
74. Jia-Lin Ma, N. and D.F. Stern, *Regulation of the Rad53 protein kinase in signal amplification by oligomer assembly and disassembly.* Cell cycle (Georgetown, Tex), 2008. **7**(6): p. 808-17.
75. Barthe, P., et al., *Dynamic and structural characterization of a bacterial FHA protein reveals a new autoinhibition mechanism.* Structure, 2009. **17**(4): p. 568-78.
76. Wang, B., et al., *Archaeal eukaryote-like serine/threonine protein kinase interacts with and phosphorylates a forkhead-associated-domain-containing protein.* Journal of bacteriology, 2010. **192**(7): p. 1956-64.
77. Wehenkel, A., et al., *Mycobacterial Ser/Thr protein kinases and phosphatases: physiological roles and therapeutic potential.* Biochimica et biophysica acta, 2008. **1784**(1): p. 193-202.
78. Di Sabato, G. and W.P. Jencks, *Mechanism and Catalysis of Reactions of Acyl Phosphates I. Nucleophilic Reactions.* J. Am. Chem. Soc., 1961. **83**: p. 4440.
79. WHO, *Global Tuberculosis Control*, W.H. Organisation, Editor. 2011.
80. Division, U.N.S. *National Accounts Main Aggregates Database*. 2011; Available from: http://unstats.un.org/unsd/snaama/dnllist.asp.
81. Di Perri, G., et al., *Nosocomial epidemic of active tuberculosis among HIV-infected patients.* Lancet, 1989. **2**(8678-8679): p. 1502-4.
82. Martinez, E., et al., *Incidence and causes of death in HIV-infected persons receiving highly active antiretroviral therapy compared with estimates for the general*

population of similar age and from the same geographical area. HIV medicine, 2007. **8**(4): p. 251-8.
83. WHO, *Anti-Tuberculosis Drug Resistance in the World*. 2008, World Health Organisation.
84. Hershkovitz, I., et al., *Detection and molecular characterization of 9,000-year-old Mycobacterium tuberculosis from a Neolithic settlement in the Eastern Mediterranean.* PloS one, 2008. **3**(10): p. e3426.
85. Hippocrates, *Aphorisms*. 400 BC.
86. Editor. *History of the MRC*. 2006 02-02-12]; Available from: http://www.mrc.ac.uk/About/History/index.htm.
87. Editor, *Jean Antoine Villemin - Obituary*, in *British Medical Journal*. 1892. p. 1091.
88. Koch, R., *Die Aetiologie der Tuberculose.* Berliner Klinische Wochenschrift, 1882. **15**: p. 221-230.
89. Hart, P.D. and I. Sutherland, *BCG and vole bacillus vaccines in the prevention of tuberculosis in adolescence and early adult life.* British Medical Journal, 1977. **2**(6082): p. 293-5.
90. Waksman, S.A., *Streptomycin and neomycin: an antibiotic approach to tuberculosis.* British Medical Journal, 1950. **2**(4679): p. 595-600.
91. Murray, J.F., *A century of tuberculosis.* American journal of respiratory and critical care medicine, 2004. **169**(11): p. 1181-6.
92. Glickman, M.S. and W.R. Jacobs, Jr., *Microbial pathogenesis of Mycobacterium tuberculosis: dawn of a discipline.* Cell, 2001. **104**(4): p. 477-85.
93. Ernst, J.D., *Macrophage receptors for Mycobacterium tuberculosis.* Infection and immunity, 1998. **66**(4): p. 1277-81.
94. Aderem, A. and D.M. Underhill, *Mechanisms of phagocytosis in macrophages.* Annual review of immunology, 1999. **17**: p. 593-623.
95. Bainton, D.F., *The Discovery of Lysosornes.* The Journal of Cell Biology, 1981. **91**(3): p. 66-76.
96. Vandal, O.H., et al., *A membrane protein preserves intrabacterial pH in intraphagosomal Mycobacterium tuberculosis.* Nature medicine, 2008. **14**(8): p. 849-54.
97. Peyron, P., et al., *Foamy macrophages from tuberculous patients' granulomas constitute a nutrient-rich reservoir for M. tuberculosis persistence.* PLoS pathogens, 2008. **4**(11): p. e1000204.
98. Russell, D.G., *Who puts the tubercle in tuberculosis?* Nature reviews Microbiology, 2007. **5**(1): p. 39-47.
99. Rees, R.J. and P.D. Hart, *Analysis of the host-parasite equilibrium in chronic murine tuberculosis by total and viable bacillary counts.* British journal of experimental pathology, 1961. **42**: p. 83-8.
100. Sharma, K., et al., *Transcriptional control of the mycobacterial embCAB operon by PknH through a regulatory protein, EmbR, in vivo.* Journal of bacteriology, 2006. **188**(8): p. 2936-44.
101. Prisic, S., et al., *Extensive phosphorylation with overlapping specificity by Mycobacterium tuberculosis serine/threonine protein kinases.* Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(16): p. 7521-6.
102. Srivastava, S.K., et al., *NAD+-dependent DNA ligase (Rv3014c) from Mycobacterium tuberculosis: novel structure-function relationship and identification of a specific inhibitor.* Proteins, 2007. **69**(1): p. 97-111.
103. Cole, S.T., et al., *Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.* Nature, 1998. **393**(6685): p. 537-44.

104. Grundner, C., L.M. Gay, and T. Alber, *Mycobacterium tuberculosis serine/threonine kinases PknB, PknD, PknE, and PknF phosphorylate multiple FHA domains.* Protein science : a publication of the Protein Society, 2005. **14**(7): p. 1918-21.
105. Molle, V., et al., *An FHA phosphoprotein recognition domain mediates protein EmbR phosphorylation by PknH, a Ser/Thr protein kinase from Mycobacterium tuberculosis.* Biochemistry, 2003. **42**(51): p. 15300-9.
106. Gupta, M., et al., *Forkhead-associated domain-containing protein Rv0019c and polyketide-associated protein PapA5, from substrates of serine/threonine protein kinase PknB to interacting proteins of Mycobacterium tuberculosis.* The Journal of biological chemistry, 2009. **284**(50): p. 34723-34.
107. Molle, V., et al., *Two FHA domains on an ABC transporter, Rv1747, mediate its phosphorylation by PknF, a Ser/Thr protein kinase from Mycobacterium tuberculosis.* FEMS microbiology letters, 2004. **234**(2): p. 215-23.
108. Belanger, A.E., et al., *The embAB genes of Mycobacterium avium encode an arabinosyl transferase involved in cell wall arabinan biosynthesis that is the target for the antimycobacterial drug ethambutol.* Proceedings of the National Academy of Sciences of the United States of America, 1996. **93**(21): p. 11919-24.
109. Alderwick, L.J., et al., *Molecular structure of EmbR, a response element of Ser/Thr kinase signaling in Mycobacterium tuberculosis.* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(8): p. 2558-63.
110. Molle, V., et al., *EmbR2, a structural homologue of EmbR, inhibits the Mycobacterium tuberculosis kinase/substrate pair PknH/EmbR.* The Biochemical journal, 2008. **410**(2): p. 309-17.
111. Jang, J., et al., *Functional characterization of the Mycobacterium tuberculosis serine/threonine kinase PknJ.* Microbiology, 2010. **156**(Pt 6): p. 1619-31.
112. Jones, P.M. and A.M. George, *The ABC transporter structure and mechanism: perspectives on recent research.* Cellular and molecular life sciences : CMLS, 2004. **61**(6): p. 682-99.
113. Curry, J.M., et al., *An ABC transporter containing a forkhead-associated domain interacts with a serine-threonine protein kinase and is required for growth of Mycobacterium tuberculosis in mice.* Infection and immunity, 2005. **73**(8): p. 4471-7.
114. Belanger, A.E. and G.F. Hatfull, *Exponential-phase glycogen recycling is essential for growth of Mycobacterium smegmatis.* Journal of bacteriology, 1999. **181**(21): p. 6670-8.
115. Weldingh, K., et al., *Two-dimensional electrophoresis for analysis of Mycobacterium tuberculosis culture filtrate and purification and characterization of six novel proteins.* Infection and immunity, 1998. **66**(8): p. 3492-500.
116. Tian, J., et al., *Variant tricarboxylic acid cycle in Mycobacterium tuberculosis: identification of alpha-ketoglutarate decarboxylase.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(30): p. 10670-5.
117. Tian, J., et al., *Mycobacterium tuberculosis appears to lack alpha-ketoglutarate dehydrogenase and encodes pyruvate dehydrogenase in widely separated genes.* Molecular microbiology, 2005. **57**(3): p. 859-68.
118. de Carvalho, L.P.S., et al., *Activity-based metabolomic profiling of enzymatic function: identification of Rv1248c as a mycobacterial 2-hydroxy-3-oxoadipate synthase.* Chemistry & biology, 2010. **17**(4): p. 323-32.
119. Niebisch, A., et al., *Corynebacterial protein kinase G controls 2-oxoglutarate dehydrogenase activity via the phosphorylation status of the OdhI protein.* The Journal of biological chemistry, 2006. **281**(18): p. 12300-7.
120. Cowley, S., et al., *The Mycobacterium tuberculosis protein serine/threonine kinase PknG is linked to cellular glutamate/glutamine levels and is important for growth in vivo.* Molecular microbiology, 2004. **52**(6): p. 1691-702.

121. Patel, D., *Functional Studies of Serine/Threonine Protein Kinase Signalling in Mycobacterium tuberculosis*, in *National Institute for Medical Research*. 2006, University of London: London. p. 164-165.
122. McDowell, M.A., et al., *Structural and functional studies on the N-terminal domain of the Shigella type III secretion protein MxiG.* The Journal of biological chemistry, 2011. **286**(35): p. 30606-14.
123. Tong, Y., et al., *Phosphorylation-independent dual-site binding of the FHA domain of KIF13 mediates phosphoinositide transport via centaurin alpha1.* Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(47): p. 20346-51.
124. Sureka, K., et al., *Novel role of phosphorylation-dependent interaction between FtsZ and FipA in mycobacterial cell division.* PloS one, 2010. **5**(1): p. e8590.
125. Dziadek, J., et al., *Conditional expression of Mycobacterium smegmatis ftsZ, an essential cell division gene.* Microbiology, 2003. **149**(Pt 6): p. 1593-603.
126. Roumestand, C., et al., *Structural Insight into the Mycobacterium tuberculosis Rv0020c Protein and Its Interaction with the PknB Kinase.* Structure, 2011. **19**(10): p. 1525-34.
127. Gee, C.L., et al., *A phosphorylated pseudokinase complex controls cell wall synthesis in mycobacteria.* Science signaling, 2012. **5**(208): p. ra7.
128. Inoue, A., et al., *Involvement of an essential gene, mviN, in murein synthesis in Escherichia coli.* Journal of bacteriology, 2008. **190**(21): p. 7298-301.
129. Mir, M., et al., *The extracytoplasmic domain of the Mycobacterium tuberculosis Ser/Thr kinase PknB binds specific muropeptides and is required for PknB localization.* PLoS pathogens, 2011. **7**(7): p. e1002182.
130. Barthe, P., et al., *The structure of PknB extracellular PASTA domain from mycobacterium tuberculosis suggests a ligand-dependent kinase activation.* Structure, 2010. **18**(5): p. 606-15.
131. McMahon, S.B., et al., *The novel ATM-related protein TRRAP is an essential cofactor for the c-Myc and E2F oncoproteins.* Cell, 1998. **94**(3): p. 363-74.
132. Lovejoy, C.A. and D. Cortez, *Common mechanisms of PIKK regulation.* DNA repair, 2009. **8**(9): p. 1004-8.
133. Gottlieb, T.M. and S.P. Jackson, *The DNA-dependent protein kinase: requirement for DNA ends and association with Ku antigen.* Cell, 1993. **72**(1): p. 131-42.
134. Taccioli, G.E., et al., *Ku80: product of the XRCC5 gene and its role in DNA repair and V(D)J recombination.* Science, 1994. **265**(5177): p. 1442-5.
135. Bakkenist, C.J. and M.B. Kastan, *DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation.* Nature, 2003. **421**(6922): p. 499-506.
136. Cobb, J.A., et al., *DNA polymerase stabilization at stalled replication forks requires Mec1 and the RecQ helicase Sgs1.* The EMBO journal, 2003. **22**(16): p. 4325-36.
137. Dazert, E. and M.N. Hall, *mTOR signaling in disease.* Current opinion in cell biology, 2011. **23**(6): p. 744-55.
138. Yamashita, A., et al., *Human SMG-1, a novel phosphatidylinositol 3-kinase-related protein kinase, associates with components of the mRNA surveillance complex and is involved in the regulation of nonsense-mediated mRNA decay.* Genes & development, 2001. **15**(17): p. 2215-28.
139. Chawla, R. and C.M. Azzalin, *The telomeric transcriptome and SMG proteins at the crossroads.* Cytogenetic and genome research, 2008. **122**(3-4): p. 194-201.
140. McMahon, S.B., M.A. Wood, and M.D. Cole, *The essential cofactor TRRAP recruits the histone acetyltransferase hGCN5 to c-Myc.* Molecular and cellular biology, 2000. **20**(2): p. 556-62.
141. Robert, F., et al., *The transcriptional histone acetyltransferase cofactor TRRAP associates with the MRN repair complex and plays a role in DNA double-strand break repair.* Molecular and cellular biology, 2006. **26**(2): p. 402-12.

142. Takai, H., et al., *Tel2 regulates the stability of PI3K-related protein kinases.* Cell, 2007. **131**(7): p. 1248-59.
143. Gangloff, Y.-G., et al., *Disruption of the mouse mTOR gene leads to early postimplantation lethality and prohibits embryonic stem cell development.* Molecular and cellular biology, 2004. **24**(21): p. 9508-16.
144. Brown, E.J. and D. Baltimore, *ATR disruption leads to chromosomal fragmentation and early embryonic lethality.* Genes & development, 2000. **14**(4): p. 397-402.
145. Herceg, Z., et al., *Disruption of Trrap causes early embryonic lethality and defects in cell cycle progression.* Nature genetics, 2001. **29**(2): p. 206-11.
146. Takai, H., et al., *Tel2 structure and function in the Hsp90-dependent maturation of mTOR and ATR complexes.* Genes & development, 2010. **24**(18): p. 2019-30.
147. Horejsi, Z., et al., *CK2 phospho-dependent binding of R2TP complex to TEL2 is essential for mTOR and SMG1 stability.* Molecular cell, 2010. **39**(6): p. 839-50.
148. Kakihara, Y. and W.A. Houry, *The R2TP complex: Discovery and functions.* Biochimica et biophysica acta, 2012. **1823**(1): p. 101-7.
149. Wiech, H., et al., *Hsp90 chaperones protein folding in vitro.* Nature, 1992. **358**(6382): p. 169-70.
150. Lai, B.T., et al., *Quantitation and intracellular localization of the 85K heat shock protein by using monoclonal and polyclonal antibodies.* Molecular and cellular biology, 1984. **4**(12): p. 2802-10.
151. Imamura, T., et al., *Involvement of heat shock protein 90 in the degradation of mutant insulin receptors by the proteasome.* The Journal of biological chemistry, 1998. **273**(18): p. 11183-8.
152. Boulon, S., et al., *HSP90 and its R2TP/Prefoldin-like cochaperone are involved in the cytoplasmic assembly of RNA polymerase II.* Molecular cell, 2010. **39**(6): p. 912-24.
153. Cloutier, P., et al., *High-resolution mapping of the protein interaction network for the human transcription machinery and affinity purification of RNA polymerase II-associated complexes.* Methods, 2009. **48**(4): p. 381-6.
154. Martin-Benito, J., et al., *Structure of eukaryotic prefoldin and of its complexes with unfolded actin and the cytosolic chaperonin CCT.* The EMBO journal, 2002. **21**(23): p. 6377-86.
155. Huen, J., et al., *Rvb1-Rvb2: essential ATP-dependent helicases for critical complexes.* Biochemistry and cell biology = Biochimie et biologie cellulaire, 2010. **88**(1): p. 29-40.
156. Gribun, A., et al., *Yeast Rvb1 and Rvb2 are ATP-dependent DNA helicases that form a heterohexameric complex.* Journal of molecular biology, 2008. **376**(5): p. 1320-33.
157. Zhao, R., et al., *Navigating the chaperone network: an integrative map of physical and genetic interactions mediated by the hsp90 chaperone.* Cell, 2005. **120**(5): p. 715-27.
158. Smith, D.F., *Tetratricopeptide repeat cochaperones in steroid receptor complexes.* Cell stress & chaperones, 2004. **9**(2): p. 109-21.
159. Jiménez, B., et al., *Structure of minimal tetratricopeptide repeat domain protein tah1 reveals mechanism of its interaction with pih1 and hsp90.* The Journal of biological chemistry, 2012. **papers in press**.
160. Zhao, R., et al., *Molecular chaperone Hsp90 stabilizes Pih1/Nop17 to maintain R2TP complex activity that regulates snoRNA accumulation.* The Journal of cell biology, 2008. **180**(3): p. 563-78.
161. Inoue, M., et al., *PIH1D1, a subunit of R2TP complex, inhibits doxorubicin-induced apoptosis.* Biochemical and biophysical research communications, 2010. **403**(3-4): p. 340-4.
162. Montenarh, M., *Cellular regulators of protein kinase CK2.* Cell and tissue research, 2010. **342**(2): p. 139-46.
163. Wiseman, T., et al., *Rapid measurement of binding constants and heats of binding using a new titration calorimeter.* Analytical biochemistry, 1989. **179**(1): p. 131-7.

164. Indyk, L. and H.F. Fisher, *Theoretical aspects of isothermal titration calorimetry.* Methods in enzymology, 1998. **295**: p. 350-64.
165. IUPAC, *Standards in isothermal microcalorimetry*, I. Wadso, Editor. 2001.
166. Sherwood, D. and J. Cooper, *Crystals, X-rays and Proteins: Comprehensive Protein Crystallography*. 2010: Oxford University Press.
167. Heinrich, B., et al., *PILATUS: A single photon counting pixel detector for X-ray applications.* Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 2009. **607**(1).
168. Matthews, B.W., *Solvent content of protein crystals.* Journal of molecular biology, 1968. **33**(2): p. 491-7.
169. Hendrickson, W.A., J.L. Smith, and S. Sheriff, *Direct phase determination based on anomalous scattering.* Methods in enzymology, 1985. **115**: p. 41-55.
170. Macek, B., et al., *Phosphoproteome analysis of E. coli reveals evolutionary conservation of bacterial Ser/Thr/Tyr phosphorylation.* Molecular & cellular proteomics : MCP, 2008. **7**(2): p. 299-307.
171. Kim, J.H., et al., *Biochemical characterization of human Ecdysoneless reveals a role in transcriptional regulation.* Biological chemistry, 2010. **391**(1): p. 9-19.
172. Liu, Z.R., et al., *Crosslinking of the U5 snRNP-specific 116-kDa protein to RNA hairpins that block step 2 of splicing.* RNA, 1997. **3**(11): p. 1207-19.
173. Close, D., et al., *Crystal structures of the S. cerevisiae Spt6 core and C-terminal tandem SH2 domain.* Journal of molecular biology, 2011. **408**(4): p. 697-713.
174. Leung, C.C.Y., et al., *Molecular basis of BACH1/FANCJ recognition by TopBP1 in DNA replication checkpoint control.* The Journal of biological chemistry, 2011. **286**(6): p. 4292-301.
175. Lougheed, K.E.A., et al., *Effective inhibitors of the essential kinase PknB and their potential as anti-mycobacterial agents.* Tuberculosis, 2011. **91**(4): p. 277-86.
176. Arnvig, K.B., et al., *Sequence-based analysis uncovers an abundance of non-coding RNA in the total transcriptome of Mycobacterium tuberculosis.* PLoS pathogens, 2011. **7**(11): p. e1002342.
177. Huang, C.-C.F., et al., *Intermolecular binding between TIFA-FHA and TIFA-pT mediates tumor necrosis factor alpha stimulation and NF-kappaB activation.* Molecular and cellular biology, 2012. **32**(14): p. 2664-73.
178. Scheffers, D.-J. and M.G. Pinho, *Bacterial cell wall synthesis: new insights from localization studies.* Microbiology and molecular biology reviews : MMBR, 2005. **69**(4): p. 585-607.
179. Kang, C.-M., et al., *The Mycobacterium tuberculosis serine/threonine kinases PknA and PknB: substrate identification and regulation of cell shape.* Genes & development, 2005. **19**(14): p. 1692-704.
180. Collins, M.O., et al., *Phosphoproteomic analysis of the mouse brain cytosol reveals a predominance of protein phosphorylation in regions of intrinsic sequence disorder.* Molecular & cellular proteomics : MCP, 2008. **7**(7): p. 1331-48.
181. King, J.B., et al., *Accurate mass-driven analysis for the characterization of protein phosphorylation. Study of the human Chk2 protein kinase.* Analytical chemistry, 2006. **78**(7): p. 2171-81.
182. Gabant, G., et al., *Autophosphorylated residues involved in the regulation of human chk2 in vitro.* Journal of molecular biology, 2008. **380**(3): p. 489-503.
183. Ali, A.A.E., et al., *Specific recognition of a multiply phosphorylated motif in the DNA repair scaffold XRCC1 by the FHA domain of human PNK.* Nucleic acids research, 2009. **37**(5): p. 1701-12.
184. Li, H., et al., *Structure of human Ki67 FHA domain and its binding to a phosphoprotein fragment from hNIFK reveal unique recognition sites and new views to the structural basis of FHA domain functions.* Journal of molecular biology, 2004. **335**(1): p. 371-81.

185. Yata, K., et al., *Plk1 and CK2 act in concert to regulate Rad51 during DNA double strand break repair.* Molecular cell, 2012. **45**(3): p. 371-83.
186. Shiozaki, E.N., et al., *Structure of the BRCT repeats of BRCA1 bound to a BACH1 phosphopeptide: implications for signaling.* Molecular cell, 2004. **14**(3): p. 405-12.
187. Durocher, D., et al., *The molecular basis of FHA domain:phosphopeptide binding specificity and implications for phospho-dependent signaling mechanisms.* Molecular cell, 2000. **6**(5): p. 1169-82.
188. Bach, H., D. Wong, and Y. Av-Gay, *Mycobacterium tuberculosis PtkA is a novel protein tyrosine kinase whose substrate is PtpA.* The Biochemical journal, 2009. **420**(2): p. 155-60.
189. Songyang, Z., et al., *A structural basis for substrate specificities of protein Ser/Thr kinases: primary sequence preference of casein kinases I and II, NIMA, phosphorylase kinase, calmodulin-dependent kinase II, CDK5, and Erk1.* Molecular and cellular biology, 1996. **16**(11): p. 6486-93.
190. Becherel, O.J., et al., *CK2 phosphorylation-dependent interaction between aprataxin and MDC1 in the DNA damage response.* Nucleic acids research, 2010. **38**(5): p. 1489-503.
191. Duncan, J.S., et al., *A peptide-based target screen implicates the protein kinase CK2 in the global regulation of caspase signaling.* Science signaling, 2011. **4**(172): p. ra30.
192. Parsons, J.L., et al., *XRCC1 phosphorylation by CK2 is required for its stability and efficient DNA repair.* DNA repair, 2010. **9**(7): p. 835-41.