MRC National Institute for Medical Research

Division of Molecular Structure

Structural and Functional Analysis of a Phospho-Dependent Molecular Switch: Rv1827 from *Mycobacterium tuberculosis*

A thesis submitted by

Timothy James Nott

In partial fulfillment of the requirements of
University College London
For the degree of Doctor of Philosophy

March 2009

Declaration

I, Timothy James Nott, 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 Research. I confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the text. Some of the work discussed in Chapter 7 was performed in collaboration with Dr. Helen O'Hare (University of Leicester) and Dr. Jiejin Li (National Institute for Medical Research).

Acknowledgements

I would like to begin these acknowledgements by sincerely thanking Dr. Steve Smerdon for giving me the opportunity of working for my PhD in his lab. His 'I'll never tell you to do anything, but then again I'll never tell you not to do anything' approach afforded me the independence freely to pursue my interests. Nevertheless he has always been on hand to and help me out and show me the bigger picture when I needed a better view of it. I thank Steve also for his open-mindedness and for permitting me to complete what is ostensibly an NMR spectroscopy-based thesis in his X-ray crystallography lab…

I would like to thank Dr. Geoff Kelly for all his assistance and support. His unrelenting patience, good will and enthusiasm has inspired me no end. Three years ago I'd never have predicted becoming as fascinated as I am with NMR. Cheers Geoff.

To the members of the Smerdon lab I would also like to extend my deepest gratitude. Without their daily support, encouragement and understanding I wouldn't have made it this far. Simon, I cannot thank you enough.

Lastly I would like to thank my friends who have propped me up when I found the going tough and kept my feet on the ground when I was getting too high. Lucas, Emily, Andy, Joe and Libby, I'm ready to rejoin society.

Abstract

Forkhead-associated (FHA) domains have gained considerable prominence as ubiquitous phosphothreonine-dependent binding modules; however, their precise roles in Ser/Thr kinase pathways and mechanisms of regulation remain unclear. From experiments with Rv1827, an FHA domain–containing protein from *Mycobacterium tuberculosis*, a complete molecular description of an FHA-mediated Ser/Thr protein kinase signalling process is derived. First, binding of the FHA domain to each of three metabolic enzyme complexes regulates their catalytic activities but does not require priming phosphorylation. However, phosphorylation of a threonine residue within a conserved N-terminal motif of Rv1827 triggers its intramolecular association with the FHA domain of Rv1827, thus blocking its interactions with each of the three enzymes. The nuclear magnetic resonance structure of this inactivated form and further mutagenic studies show how a novel intramolecular phospho-switch blocks the access of the target enzymes to a common FHA interaction surface and how this shared surface accommodates three functionally related, but structurally diverse, binding partners. Thus a remarkable and unsuspected versatility in the FHA domain that allows for the transformation of multiple kinase inputs into various downstream regulatory signals has been revealed.

Contents

List of figures

List of tables

List of abbreviations

Å	angstrom
µl/µM/µs	micro-litre/micro-molar/micro-second
AMP	adenosine monophosphate
ARIA	ambiguous restraints for iterative assignment
ATP	adenosine triphosphate
CD	circular dichroism
Da	dalton
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
FHA	forkhead-associated
GDH	NAD^+-dependent glutamate dehydrogenase
GDP	guanosine diphosphate
GltB	β-subunit of the GS complex
GS	glutamate synthase
GST	glutathione-S-transferase
GTP	guanosine triphosphate
H1	residues 27-34 in Rv1827 pThr 22
HPLC	high performance liquid chromatography
HSQC	heteronuclear single quantum coherence
IPTG	isopropyl-β-D-thiogalactopyranoside
Kd	dissociation constant
KGD	α-ketoglutarate decarboxylase
LB	Luria broth
MALDI-TOF	matrix-assisted laser desorption ionisation - time of flight
MALLS	multi angle laser light scattering
min.	minute
ml/mM/ms	milli-litre/milli-molar/milli-second
MWCO	molecular weight cutoff
NAD	nicotinamide adenine dinucleotide
nl/nM/ns	nano-litre/nano-molar/nano-second
NMR	nuclear magnetic resonance
NO	nitric oxide
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser effect spectroscopy
pH	[the] power of hydrogen ($-\log_{10}$ of the hydrogen ion concentration)
RNA	ribonucleic acid
Rv1827 pThr 22	Rv1827 phosphorylated at Thr 22 by PknB
sec.	second
STPK	serine/threonine protein kinase
TALOS	torsion angle likelihood obtained from shift and sequence similarity
TCA	tricarboxylic acid
TCEP	tris(carboxyethyl)phosphine
TPR	tetratricopeptide
UV	ultra-violet

Anything that happens, happens.

Anything that, in happening, causes something else to happen, causes something else to happen.

Anything that, in happening, causes itself to happen again, happens again.

It doesn't necessarily do it in chronological order, though.

Douglas Adams, *Mostly Harmless*

1 Introduction

1.1 Signal transduction

1.1.1 Overview

Signal transduction is the process by which biological systems transduce, assimilate, integrate and disseminate information at the molecular level. This allows appropriate temporal and spatial biochemical responses to be elicited from a vast array of stimuli. Complex cellular processes such as responses to the extracellular environment, cell cycle progression, regulation of DNA repair, protein translation and monitoring of the metabolic status of the cell all require sensitive and dynamic signal transduction mechanisms in order to maintain stability of the system.

Signals are primarily encoded and transmitted through protein-protein, protein-nucleic acid and protein-small molecule interactions. However, they can also take the form of morphological gradients of molecules and hormones as well as environmental changes in pH, salinity and temperature. Over the last forty years understanding of the underlying mechanisms and principles of molecular signal transduction has been and continues to be a principle aim of biological research.

1.1.2 Historical perspective

The first instances of the term 'signal transduction' as a way of describing cellular responses to stimuli began to appear in the literature in the early 1970s (Rensing, 1972). Arguably it was the discovery and initial characterization of G-proteins and G-protein coupled receptors by Martin Rodbell in the early 1980s that demonstrated a molecular

basis for signal transduction for the first time (Rodbell, 1980). The ensuing revolution in molecular biology of the 1980s and early 1990s began to identify molecular mechanisms in a multitude of cellular contexts and lead to an explosion of interest in signal transduction pathways as evidenced by the increasing number of publications on the subject (Figure 1.1). Since then it has been adopted as a concept used to describe the molecular mechanisms underpinning dynamic biological behavior.

The relatively recent advances in the fields of systems biology and computational biology suggest that molecular signal transduction may in fact be more broadly understood and modelled in terms of the general principles of biochemical oscillators (Novak and Tyson, 2008). Biochemical oscillatory behaviour is another way of describing how the components of macro-molecular biological networks interact with each other. Instead of representing the flow of signals linearly in an A to B to C fashion, higher order properties and tendencies of the whole network are considered. By considering signalling processes in this way it is becoming possible to understand how signal transduction pathways that control different aspects of cellular behaviour interact with each other. After all, molecular signalling controlling one type of cellular event cannot be completely understood except in the context of all the signalling processes in operation. This is clearly a formidable task. Therefore by representing and modelling the properties of local reaction networks as simplified oscillatory motifs and modules, hypotheses can be made on how biological systems operate in ways that may have been obscured when considering one signalling process at a time.

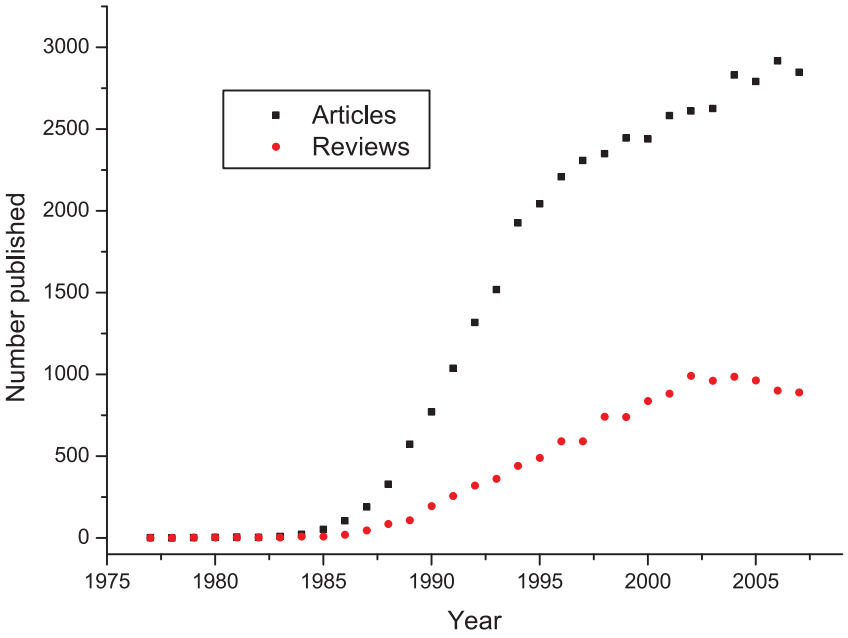

Figure 1.1. Frequency of the term 'signal transduction' in the titles and abstracts of scientific papers. Searches were performed at http://www.ncbi.nlm.nih.gov/sites/entrez?db=pubmed .

1.1.3 Underlying principles of signal transduction – biochemical oscillators

Interestingly, the concept of biological oscillators predates that of signal transduction and was first recognized in the oscillatory behavior of metabolic processes such as glycolysis (Hess and Boiteux, 1971; Pye and Chance, 1966), the horseradish peroxidase reaction (Olsen and Degn, 1977; Olsen and Degn, 1978) and cyclic AMP production (Gerisch et al., 1975). Since then biochemical oscillations have been observed and described as the mechanisms at the root of circadian rhythms (Dunlap, 1999), gene regulatory networks and protein interaction networks (Momiji and Monk, 2008; Monk, 2003), metabolic networks (Higgins, 1964; Sel'kov, 1968) and cell cycle progression (Gerhart et al., 1984; Novak and Tyson, 2008). More recently, synthetic gene regulatory networks with oscillatory behavior have been constructed in *Escherichia coli* based on a negative feedback loop composed of three repressor genes and their corresponding promoters. Here, green fluorescent protein (GFP) expression was observed to oscillate autonomously on a time scale of hours. This proof-of-principle approach demonstrates that considering signaling and interaction networks as having fundamental oscillatory behavior offers instructive new insight and biotechnological possibilities (Elowitz and Leibler, 2000).

Currently, aspects of biological behavior are described from the atomic scale to that of populations of organisms, and at every level in between. That descriptions of biological events at so many scales are made is testimony to how fascinated people are by its diversity and intricacy. However, finding ways to describe how one level is related to another is not straight forward. Indeed, categorizing biological processes into discrete scales at all is a minefield. Therefore attempting to find scale-independent ways of

characterizing biological systems will, in the long run, greatly aid our understanding of how they work.

Biochemical oscillators display scale-independent characteristics such as periodicity, robustness and entrainment. These are properties that transcend those of the individual molecules or reaction partners and involve the full topology of the reaction network (Novak and Tyson, 2008). Description in this way is fundamentally different from the linear representations of reaction networks and signaling pathways that biologists are familiar with today. However, current understanding can be used to build scale-independent descriptions and models that may be applied outside their original context. It is hoped that this will help to unify our understanding of biological processes in general.

Four general requirements for biochemical oscillators have been identified, and can be recognized as key properties of many well characterized signal transduction systems and processes. These are negative feedback, temporal delay (produced by embedded positive feedback or a series of reaction intermediates), sufficient non-linearity in reaction kinetics (as a result of the complexity of the environment in which the reactions occur) and the appropriate balancing of the timescales of opposing chemical reactions.

This combination of positive and negative feedback loops supports bistability, which is the ability of the system to adopt either of two alternative steady states (separated by unstable intermediate states) under identical conditions. The state that the system falls into at any given time depends on hysteresis, which is the property of a bistable system to be switched between either steady state by a transient signal, and switched back again by another transient signal. Therefore the state of the system depends not only on its

present conditions but also on its recent history. Early models of glycolytic oscillations fall into this class of oscillators (Higgins, 1964; Sel'kov, 1968), as well as models determining the lytic versus lysogenic switch of bacteriophage lambda (Atsumi and Little, 2006a; Atsumi and Little, 2006b; Oppenheim et al., 2005).

The reversibility of biochemical reactions in oscillatory systems and cell signaling processes is almost ubiquitous and is the basis of much of their inherent sensitivity and flexibility. For example, proteins are post-translationally modified (by phosphorylation, acetylation, methylation, hydroxylation and ubiquitination etc.) and unmodified, G-proteins cycle between GTP- and GDP-bound forms, second messengers are synthesized and degraded or released and sequestered, proteins are imported and exported across membranes and proteins are degraded and re-synthesized (Ferrell, 2002). Protein phosphorylation is the most widely employed reversible post-translational modification in signal transduction systems and it has been estimated that approximately 30% of all eukaryotic proteins are phosphorylated on at least one residue at some point during their existence (Cohen, 2000).

1.2 Signal transduction via protein phosphorylation

1.2.1 Overview

Protein phosphorylation has been implicated in virtually all basic cellular processes, ranging from metabolic regulation, cellular growth, division and differentiation, motility, intracellular trafficking, membrane translocation, muscle contraction, immunity, learning and memory (Manning et al., 2002a; Manning et al., 2002b; Ubersax and Ferrell, 2007). Protein kinases catalyze the transfer of the γ-phosphate of ATP to specific amino acid side chains within proteins. The free energy of the phosphorylation reaction is -12 kcal

mol^{-1}, making this form of post-translational modification extremely stable and resistant to uncatalyzed hydrolysis. Serine/threonine protein kinases (STPKs) and tyrosine kinases phosphorylate the hydroxyl groups of serine/threonine or tyrosine residues, respectively. Conversely, protein phosphatases dephosphorylate phosphorylated residues and are also classed according to their specificity: serine/threonine phosphatases, tyrosine phosphatases and dual specificity phosphatases that can dephosphorylate phosphorylated serine/threonine or tyrosine residues.

The extent to which protein phosphorylation is used as a means to signal transduction lies with the diversity of effects that it can have on phosphorylated target proteins Most are physical consequences of the addition of two negative charges and the possibility of engaging in up to four additional hydrogen bonds per phosphoryl group. Thus protein phosphorylation may directly induce conformational changes by altering intramolecular electrostatic interactions, act to increase/decrease stability or solubility, create or ablate protein-protein interaction surfaces and activate or inhibit enzymatic activity. The significance of the generation of phosphorylation-dependent binding surfaces will be discussed in more detail in section 1.4.

1.3 The protein kinase domain

Given that phosphorylation can have diverse effects on protein function, both protein kinases and phosphatases require robust mechanisms for specific target selection. Indeed, it has long been recognized that temporally and spatially aberrant eukaryotic protein kinase activity contributes to oncogenic disease states (Eckhart et al., 1979; Levinson et al., 1978). The structures of several hundred protein kinase domains have been deposited in the protein data bank (www.rcsb.org/pdb) to date. Many structures of

kinase domains have now been solved in both active and inactive conformations, as well as in complex with ATP and ATP analogues, regulatory subunits, and protein or peptide substrates. This has lead to increased understanding at the structural level of the features that determine and maintain kinase specificity as well as their catalytic mechanisms.

The canonical catalytic domain of classical protein kinases is approximately 250 amino acids in length and conforms to a conserved bilobal architectural fold. This consists of a small N-terminal lobe (~85 amino acids) comprised of β-sheets and a larger C-terminal lobe (~170 amino acids) comprised mostly of α-helices. ATP binds in a cleft at the interface of the two lobes with the adenosine moiety buried in a hydrophobic pocket. This results in orientation of the negatively charged phosphate backbone towards the solution. Protein substrates bind along the cleft and conserved residues within the catalytic domain transfer the γ-phosphate of ATP onto the hydroxyl groups of serine/threonine or tyrosine residues.

The structures of active kinase domains in the vicinity of the catalytic site are very similar as the conformation of catalytic residues must be constrained such that their orientation is favorable for phosphoryl transfer. However, the conformations of inactive kinase domains vary greatly. Diversity in the mechanisms by which an inactive kinase domain is converted into an active conformation is one way of providing signaling specificity and regulation of different kinase families (Canagarajah et al., 1997; Huse and Kuriyan, 2002; Jeffrey et al., 1995; Yamaguchi and Hendrickson, 1996). Most specific kinase inhibitors exert their effect by binding to and maintaining the inactive conformation of the kinase, preventing conversion to the active state.

1.3.1 Catalysis and activation of the protein kinase domain

There are several conserved architectural features of kinase domains that regulate their catalytic activity through conformational change. One of these is called the phosphate-binding loop (P-loop). In the active kinase conformation the P-loop constrains ATP in the cleft between the N- and C-lobes of the kinase domain and consists of a glycine-rich sequence (GXGXΦG) where X is any amino acid and Φ is typically a tyrosine or phenylalanine residue. The inherent flexibility conferred by the glycine residues facilitate interactions between the backbone of the P-loop and the phosphate groups of the ATP molecule, while the aromatic residue caps the site of phosphate transfer (Huse and Kuriyan, 2002). The substrate peptide binds in an extended conformation along the solvent exposed side of the cleft and is positioned close to the ATP γ-phosphate. Conserved residues in the kinase catalytic loop execute the transfer of the ATP γ-phosphate to the substrate peptide.

As one of the first kinases to be studied in detail, Protein kinase A (PKA) has been used to exemplify the catalytic mechanism of STPKs. In the residue numbering of PKA, the conserved Asp 166 is believed to act as a base for proton abstraction from the hydroxyl group of the substrate target residue resulting in its attack on the ATP γ-phosphate (Figure 1.2). Asn 171 is thought to orientate Asp 166 for catalysis, and in conjunction with Asp 184 binds two divalent cations required for nucleotide recognition (Huse and Kuriyan, 2002). Asp 184 is a component part of the highly conserved Asp-Phe-Gly (DFG) motif, by which the N-terminal end of the activation loop is defined. Furthermore the ATP α- and β-phosphates are coordinated by the side chains of Lys 72, which is itself coordinated by Glu 91. Glu 91 resides in another conserved architectural

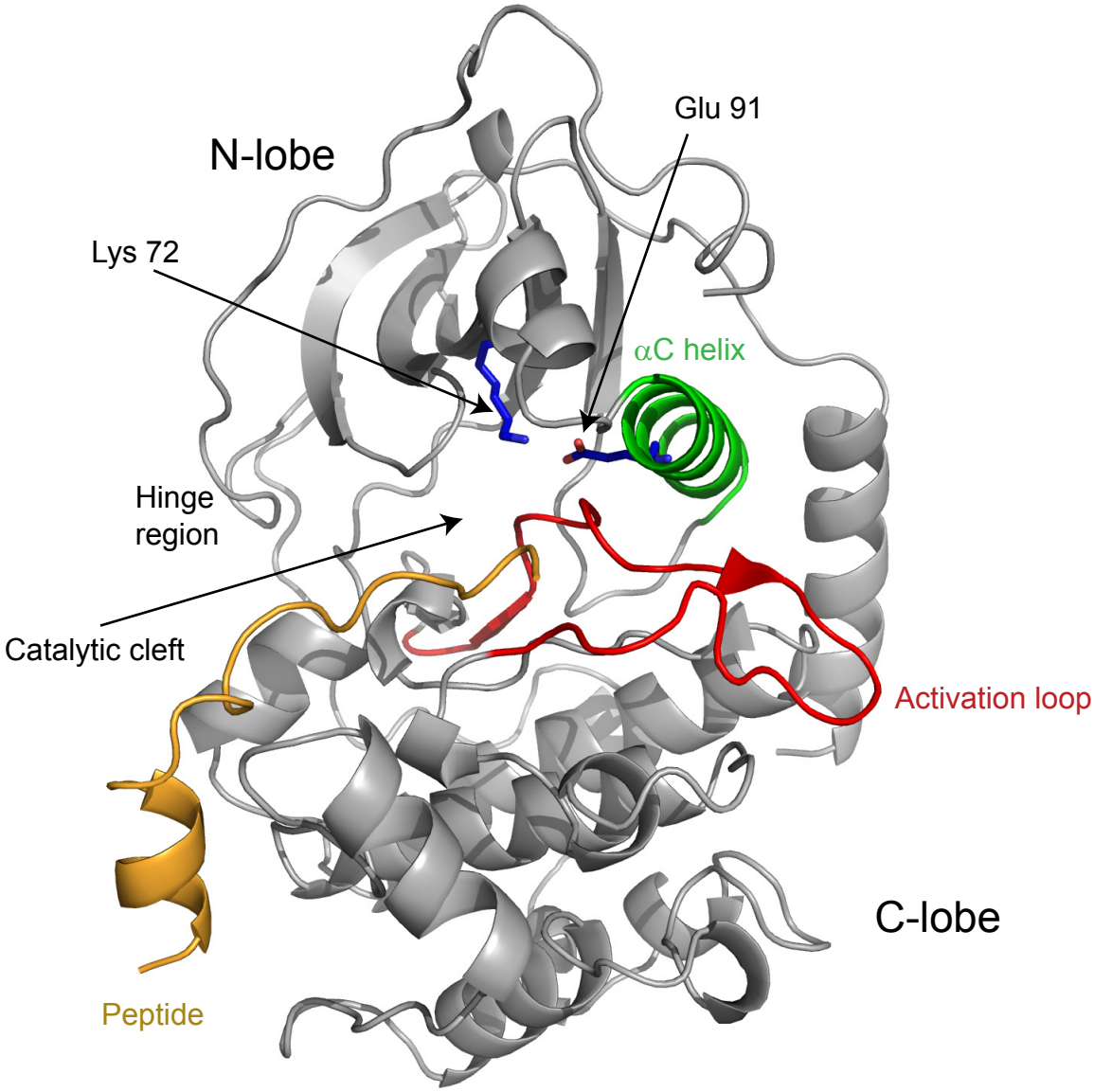

Figure 1.2. Structure of the kinase domain. Cartoon representation of cAMP-Dependent Protein Kinase (Protein kinase A, PKA) in complex with a peptide inhibitor (orange) (pdb accession 2CPK). The αC helix is shown in green and the activation loop is shown in red. Lys 72 and Glu 91 are shown as sticks. ATP and the substrate of the kinase bind at the cleft between the N- and C-lobes.

feature of kinase domains, the αC helix. Appropriate orientation of the αC helix stabilizes residues forming the active catalytic site.

Reorientation of the αC helix is another mechanism by which kinase activity can be regulated. For example the binding of cyclin to the αC helix of cyclin-dependent kinases results in its rotation from an inactive to an active conformation (De Bondt et al., 1993; Jeffrey et al., 1995). On the other hand, binding of the SH3 domain of Src kinases stabilizes an orientation of the αC helix in which Glu 91 does not interact with Lys 72 resulting in poor ATP binding (Gonfloni et al., 2000; Moarefi et al., 1997; Sicheri et al., 1997; Xu et al., 1997).

Thus conversion of a kinase domain from an inactive to an active state can occur through subtle conformational changes brought about by several different mechanisms. In many kinases phosphorylation of the activation loop is the minimum requirement to produce full activity. In the inactive conformation the activation loop occludes the catalytic site by collapsing into it and thereby preventing both cofactor and substrate from binding (Hubbard et al., 1994). However, upon phosphorylation the activation loop adopts an open conformation in which the phosphate group of the modified residue forms an ion pair with an asparagine that directly precedes the catalytic Asp 166 (Hubbard, 1997). The number and position of the phosphorylated residues in the activation loop varies between kinases, but interaction with Asn 165 reinforces the orientation of the DFG motif and promotes catalysis. In some cases the phosphorylated residue may be replaced with either glutamate or aspartate and catalytic activity is still conferred through the resultant similar ion pairing (Huang and Erikson, 1994; Orr and Newton, 1994). In many cases phosphorylation of the activation loop and precise

orientation of αC helix are both required for the kinase to transition from a catalytically inactive to active state (Engh and Bossemeyer, 2001; Huse and Kuriyan, 2002).

1.4 Phospho-protein interaction domains

1.4.1 Overview

Over the past twenty five years modular protein-protein interaction domains have emerged as fundamental components in cellular organization and provide a means of dynamically controlling information in signal transduction pathways (Pawson, 2004) and coupling oscillatory behaviors. Protein interaction domains are often found as modules within the multi-domain architecture of regulatory proteins and generally serve to temporally and spatially regulate the sensory or enzymatic activity of the protein within the context of multi-molecular signaling complexes (Yaffe and Smerdon, 2004). By mediation of specific protein-protein interactions alteration of protein synthesis, degradation and post-translational modifications are regulated. These constitute important ways of conferring temporal control on such systems. Spatial regulation mandates accurate targeting of proteins to sub-cellular compartments or organelles. This is achieved via import/export sequences or through protein interaction domains that reversibly recognize specific peptide sequences, protein surfaces, lipids, nucleic acids and/or post-translationally modified peptides (Figure 1.3) (Pawson and Nash, 2000; Pawson and Nash, 2003; Yaffe and Smerdon, 2004).

The versatility of interaction domain-containing proteins is exemplified by the signaling between transforming growth factor-β (TGF-β) receptor kinases and R-SMAD signal transducer proteins (Wrighton et al., 2009). TGF-β is a signaling hormone, and along with its family members (the nodals, activins, bone morphogenic proteins (BMPs),

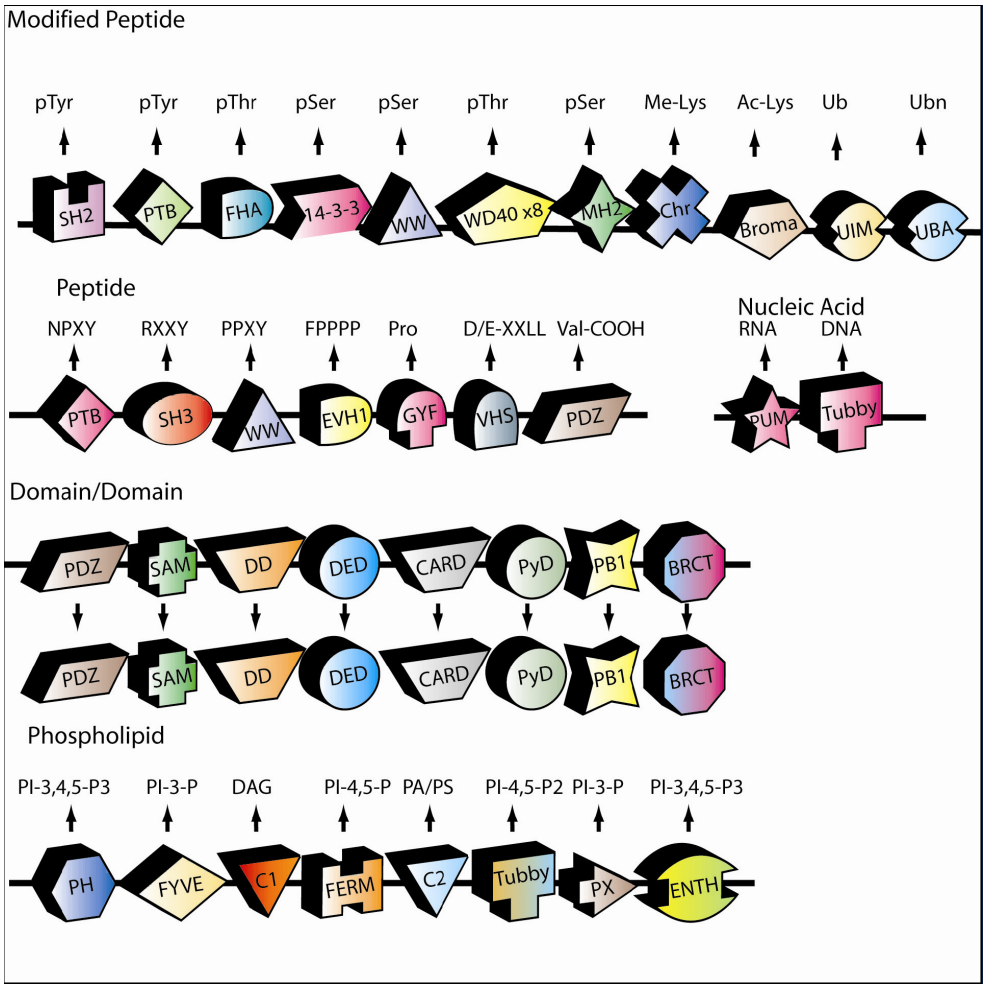

Figure 1.3. Modular interaction domains used in signal transduction processes. A selection of protein interaction domains and their ligands are shown. The full names are given in the SMART (Simple Modular Architecture Research Tool) database (http://smart.embl-heidelberg.de/). Adapted from (Pawson and Nash, 2003).

myostatin, anti-Muellerian hormone (AMH) and others) controls eukaryotic cell division, differentiation, migration, adhesion, organization and programmed cell death (Massague and Gomis, 2006). Upon binding of TGF-β to the extracellular domains of transmembrane TGF-β receptor kinases, auto-phosphorylation of exposed serine residues in a Gly/Ser-rich sequence in an intracellular juxta-membrane region occurs. The C-terminal MH2 (mothers against decapentaplegic homologue 2) domain of R-SMAD then binds to the phosphorylated Gly/Ser-rich sequence. After recruiting the scaffold protein SARA, the MH2 domain is subsequently phosphorylated by the TGF-β receptor kinase at a C-terminal serine. Now in a phosphorylated form, the MH2 domain associates with two other phosphorylated MH2 domains, each binding to the C-terminal phosphorylation site of the adjacent molecule. This trimer then dissociates from the TGF-β receptor kinase and is translocated to the nucleus where the N-terminal MH1 (Mad homology 1) domains of the R-SMAD proteins bind to DNA, whilst the MH2 domains associate with various transcriptional components altering gene expression (Miyazono et al., 2004; Wu et al., 2001). This example is a demonstration of the multi-level complexity in molecular signaling that can be readily achieved by linking protein interaction domains together in the same polypeptide chain.

The Src homology (SH2) domain was the first phospho-interaction domain to be studied. It was identified as a constituent part of the cytoplasmic tyrosine kinases Fps, Src and Abl (Sadowski et al., 1986). The SH2 domains of these proteins were shown to confer specificity in tyrosine kinase signaling by mediating protein-pTyr interactions through a "dedicated non-catalytic domain" (DeClue et al., 1987; Koch et al., 1989; Pawson, 1988; Pawson et al., 1988; Sadowski et al., 1988). Detailed studies have since been undertaken on the growing number of phosphoprotein-interaction domains and include pTyr-interaction domains such as SH2 and PTB-domains and pSer/pThr specific

interaction domains such as WW domains, 14-3-3 domains, the WD40 repeats in F-Box containing proteins, the polo-box domains of Polo-like kinases and FHA domains (Pawson, 2004; Yaffe and Smerdon, 2004).

The atomic structures of isolated phospho-dependent interaction domains show that in general they fold in a manner that generates a ligand-binding surface with low internal mobility that is used to recognize a short phosphopeptide motif. In the next section this and other features of phospho-interaction domains will be discussed using the FHA domain as a primary example.

1.5 The Forkhead-associated domain

1.5.1 Overview

Forkhead-associated domains (FHA domains) were initially identified as a domain contained in a subset of members of the Forkhead transcription factor family (Hofmann and Bucher, 1995). To date, FHA domains have been identified in more than 2000 proteins (Pfam database) in prokaryotes and eukaryotes (Mahajan et al., 2008). Since their discovery FHA domains have been implicated to play important signaling roles in DNA damage responses, cellular growth and cell cycle regulation (Mahajan et al., 2008). Their presence is highly correlated with that of STPKs, highlighting a functional linkage between these two classes of proteins.

FHA domains are unique among phospho-interaction signaling domains as they are the only known domain to specifically recognize only pThr residues. Unlike FHA domains, WW and 14-3-3 domains do not discriminate between pThr- and pSer-containing ligands. *In vitro* orientated peptide library screening has identified that specificity of FHA

interactions is usually determined by residues -4 to +3 relative to the pThr of the ligand (Yaffe and Smerdon, 2004). FHA domains can be broadly categorized into those that bind to pTXXD motifs (where pT is phospho-threonine and X is any amino acid) and those that bind to pTXXΨ (where Ψ is a bulky non-polar residue such as isoleucine, leucine or valine). However, recent research shows that as well as selectivity for the pThr + 3 residue, several FHA domains employ more sophisticated means of selecting binding partners allowing greater functional versatility.

1.5.2 Structure of FHA domains

The structures of 16 unique FHA domains have been solved at atomic resolution by either nuclear magnetic spectroscopy (NMR) or x-ray crystallography in their free form, complexed with ligands or both (www.rcsb.org/pdb). Sequence alignments predict that the core FHA domain homology region consists of 55-75 amino acids, forming 5 or 6 β-strands that in turn fold into two β-sheets, which generates a twisted β-sandwich (Figure 1.4 top left). Flanking β-strands stacking antiparallel with the core homology region extend the surfaces of the β-sheets producing the 11 or 12 β-strands that are usually observed in FHA domains. Buried hydrophobic sidechains exclude solvent from the centre of the β-sandwich and stabilize the tertiary structure. The modular nature of FHA domains is highlighted by the way that the first and last β-strands are adjacent and run in an anti-parallel direction to each other. Thus hydrogen bonding between the backbone of the N-terminal residue of the first β-strand and the C-terminal residue of the last β-strand usually defines the domain limits.

Defined in this way, the minimal FHA domain constructs reported so far range from 95 (in the case of Ki67-FHA) to 121 amino acids in length (in the case of Rad53-FHA2).

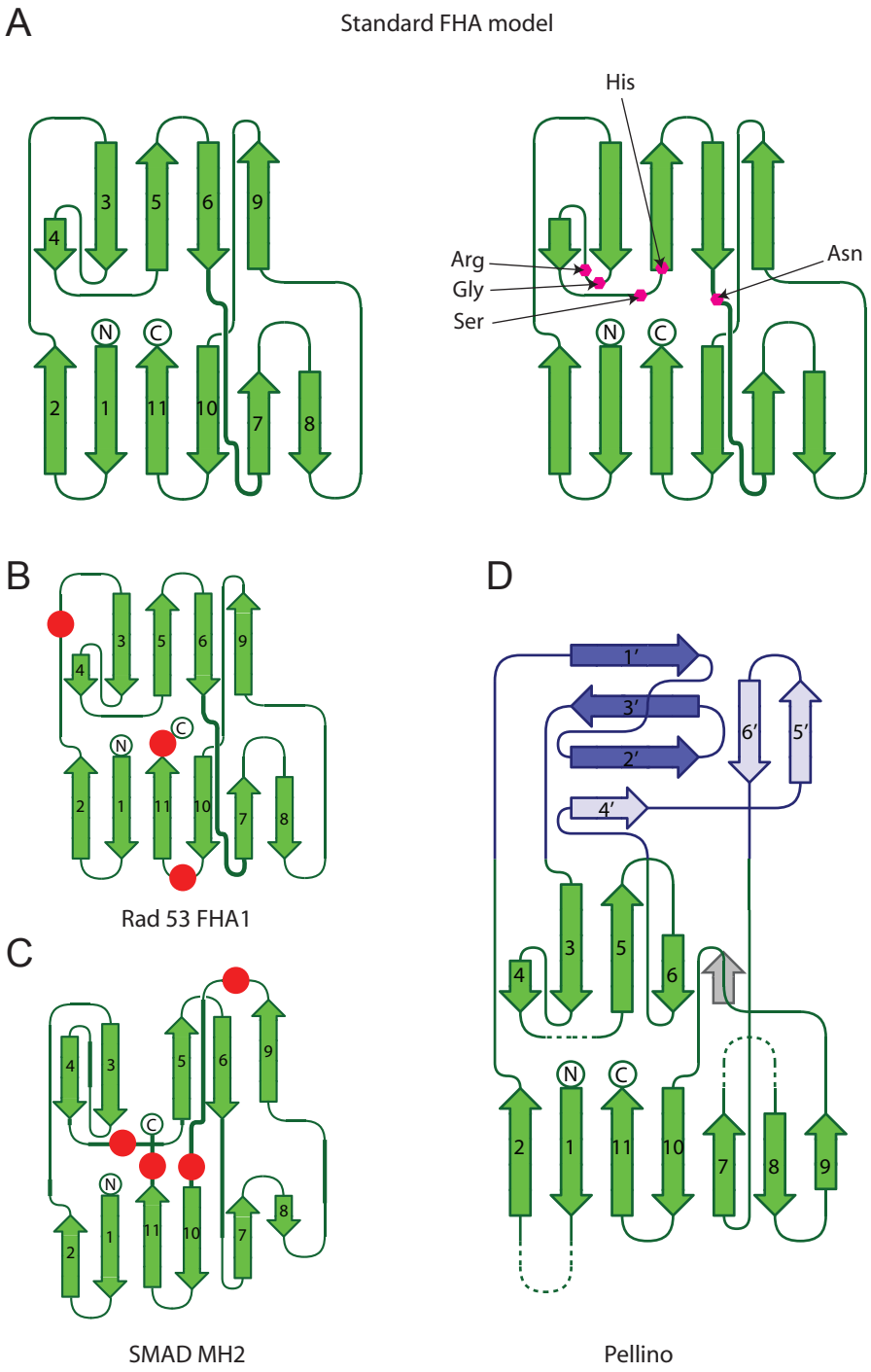

Figure 1.4. Comparison of the domain topology and overall β-sandwich structure of FHA domains. (A) The standard FHA domain model. left FHA domain β-strands are represented by green arrows. right The topological location of the five most highly conserved FHA domain residues are indicated with pink filled circles. (B) FHA domain topology of Rad53 FHA1. The location of helical insertions are indicated with red filled circles. (C) Topology of the SMAD MH2 domain. The β-sandwich architecture is similar to FHA domains, but MH2 domains lack the conserved FHA domain residues shown in A right. (D) Pellino protein domain topology. As well as sharing core FHA domain β-sandwich topology, Pellino proteins contain an extra 'wing' insertion (light and dark blue) of β-strands. In each schematic the N- and C-termini of the domains are indicated. β-strand numbering starts at the N-terminal strand.

This variability arises as a result of insertions of varying length in the loops between β-strands and is shown for Rad53 FHA1 in Figure 1.4. Red circles show the location of helical insertions into Rad53 FHA1 domains and the SMAD MH2 domain. The MH2 domain shares core FHA β-sandwich topology but differs in the locations of helical insertions, conservation of pThr-binding residues and biological function. The similarity in core topology between FHA domains and MH2 domains can be viewed as a case of convergent evolution supporting differential biological function. An interesting new class of FHA domain variants has recently been discovered in Pellino proteins (Figure 1.4), in which two substantial sequence insertions between β2/β3 and β6/β7 extend from the FHA core and form a "wing" that juxtaposes the pThr binding surface (Lin et al., 2008; Pennell and Smerdon, 2008). The effect of the wing insertion on binding properties is currently not known.

All functionally active FHA domains studied to date contain five highly conserved amino acids that are required for correct folding of the tertiary structure and pThr-binding surface formation (Figure 1.4 top right). FHA domains bind phosphorylated peptides through direct contacts with the pThr-peptide backbone and sidechains and via hydrogen bonds with the pThr phosphate group. The binding surface is formed from the loops between β-strands β3/β4, β4/β5, β6/β7 and β10/β11. The structure of Rad53 FHA1 in complex with a pThr-peptide demonstrates the importance of the conserved phospho-binding FHA domains residues (Arg 70, Ser 85, and Asn 107) (Figure 1.5). The side chains of Arg 70 and Ser 85 directly contact the phosphate through a salt bridge and hydrogen bonds, respectively. The hydroxyl of Ser 85 also forms a hydrogen bond with a backbone amide of the β6/β7 loop helping to stabilize the binding site. Asn 107 coordinates the pThr peptide backbone C-terminal to the phosphorylation site and in concert with other residues from the β6/β7 and β10/β11 loops helps to determine

binding specificity. The conserved Gly 69 and His 88 provide additional architectural stabilization around the phospho-binding site.

1.5.3 Binding specificity and functional versatility of FHA domains

The development of orientated phosphopeptide library screens in the early 1990s has since facilitated the determination of ligand specificity of a variety of phosphopeptide binding domains including PTB, WW, WD40 repeats, 14-3-3 and FHA domains (Yaffe and Smerdon, 2004). Selection specificity is deduced by applying a degenerate library of phosphorylated peptides to immobilized interaction domains and after incubation, unbound phosphopeptides are removed (Songyang and Liu, 2001). Bound peptides are eluted and sequenced by automated Edman degradation. Raw selection ratios are determined by comparison of the relative molar percentage of each amino acid at each Edman degradation sequencing cycle in the recovered peptides to that of each amino acid in the original peptide library mixture at the same position. Thus if the original peptide library contains a central phosphorylated residue flanked by a degenerate series of residues at positions –3, -2, -1, +1, +2 and +3 relative to the fixed phosphorylated amino acid at position 0, binding preferences of phosphopeptide binding domains can be elucidated.

A trend that emerged from orientated phosphopeptide library screening of FHA domains is their selection for the amino acid three residues C-terminal of the pThr (pThr + 3). FHA domains with specificity for the pThr + 3 residue fall broadly into those that select for pTXXD and those that select for pTXX(I/L/V). However, specificity can be additionally determined by the sequence of residues in the loops of the binding surface. For example Rad53 FHA1 shows a clear preference for Asp at the pThr + 3 position as a

result of the presence of Arg 83 in the β4/β5 loop, with which it forms a salt bridge (Figure 1.5 and Figure 1.6). However, Rad53 FHA2 binds preferentially to Ile or Leu at the pThr + 3 position even though it appears to have an Arg (Arg 617) in a structurally equivalent position (Figure 1.6) (Mahajan et al., 2008). The FHA2 structure reveals that the Arg 617 guanidino group makes an intramolecular salt bridge with Asp 683, rendering it unavailable for intermolecular recognition and changing the pThr + 3 preference from a charged residue to a non-polar residue (Byeon et al., 2001; Durocher and Jackson, 2002; Mahajan et al., 2008; Wang et al., 2000).

FHA domains have also been identified that recognize residues N-terminal to the pThr of the ligand. The structure of murine PNK-FHA complexed with a phosphopeptide derived from its biologically relevant binding partner XRCC4 reveals that residues N-terminal to the pThr are cradled between the β2/β3 and β4/β5 loops and are essential for ligand binding (Figure 1.6) (Bernstein et al., 2005). This may be explained by the complimentarity of basic residues on the FHA domain and acidic residues N-terminal to the pThr of the peptide.

Dun1-FHA domain contains two pThr binding sites and has been proposed to act as a detector of multiply phosphorylated SQ-TQ clusters, such as those that precede each of the two FHA domains of Rad53 (Figure 1.6). Rad53 FHA1 is preceded by four phosphorylatable TQ repeats at positions 5, 8, 12 and 15 which are important for Rad53 dimerisation and activation (Lee et al., 2003b). The structure of Dun1-FHA in complex with a Rad53-derived peptide phosphorylated at Thr 5 and Thr 8 determined by nuclear magnetic spectroscopy (NMR) shows that the N-terminal phosphorylation site binds primarily to the conserved Arg 60, aided by the non-conserved Arg 62 in the β3/β4 loop. The second binding site appears to be formed by the non-conserved Arg 102, assisted

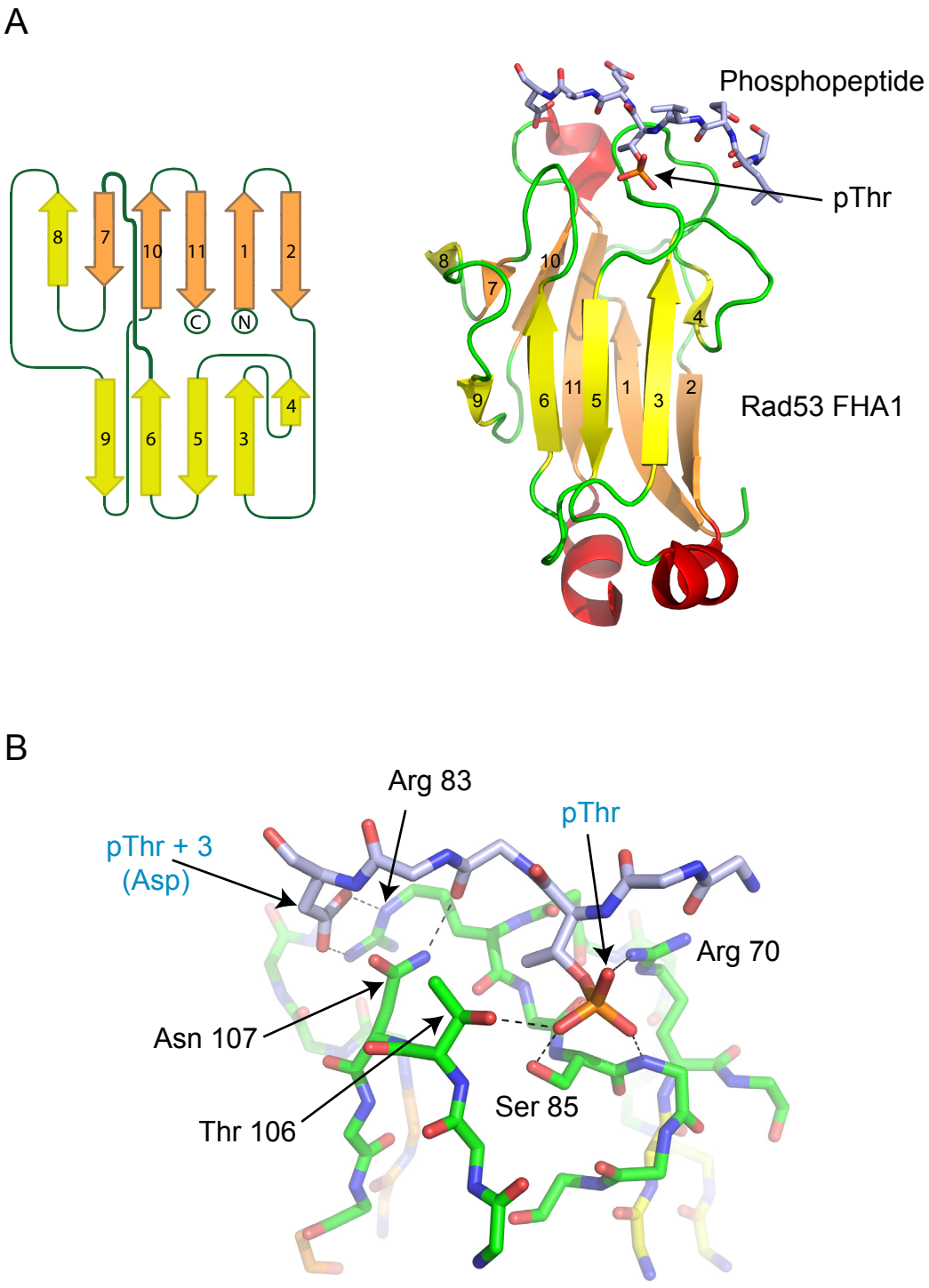

Figure 1.5. Structure Rad53 FHA1-phosphopeptide complex (1G6G). (A) The protein fold of the Rad53 FHA1 domain from Saccharomyces cerevisiae in complex with a synthetic phosphopeptide. The phospho-binding surface is formed by residues from the loops connecting the β3/β4, β4/β5, β6/β7 and β10/β11 strands. (B) Key FHA domain residues make direct contacts with the phosphothreonine (pThr) containing ligand. The pThr-containing ligand and key FHA domain residues are shown in stick representation and hydrogen bonds are indicated with dashed lines.

possibly by Lys 100 located in the β6/β7 loop. Interestingly the pThr + 6 residue (Ser 11) has also been implicated in phosphopeptide binding, as evidenced by strong nuclear Overhauser effects (NOEs) between it and Lys 129 of the β10/β11 loop (Mahajan et al., 2008).

The N-terminal FHA domain of Ki67 antigen protein represents a unique case among FHA domains in that it has been shown to bind to multiply phosphorylated peptides using an extended binding surface (Figure 1.6). The NMR structure of Ki67 complexed with a 43 amino acid peptide from hNIFK (residues 226-269) phosphorylated at Ser 230, Thr 234 and Thr 238 displays several novel features. The phosphate group of pThr 238 is coordinated by Arg 31 and Ser 45 and also Asn 67 of the β6/β7 loop which appears to play a role in binding residues at the pThr + 2 position. However, interactions of residues in the β10/β11 and β1/β2 loops with the phosphopeptide stabilize a 14 residue α-helix in the region pThr + 4 to pThr + 18. Residues in the region pThr + 27 to pThr + 29 stack antiparallel with β4 thus extending the β-sheet towards β2 (Byeon et al., 2005; Li et al., 2004). Thus despite being specific for pThr, FHA domains have evolved to use many other types of selection criteria to confer selectivity of binding of pThr-containing ligands.

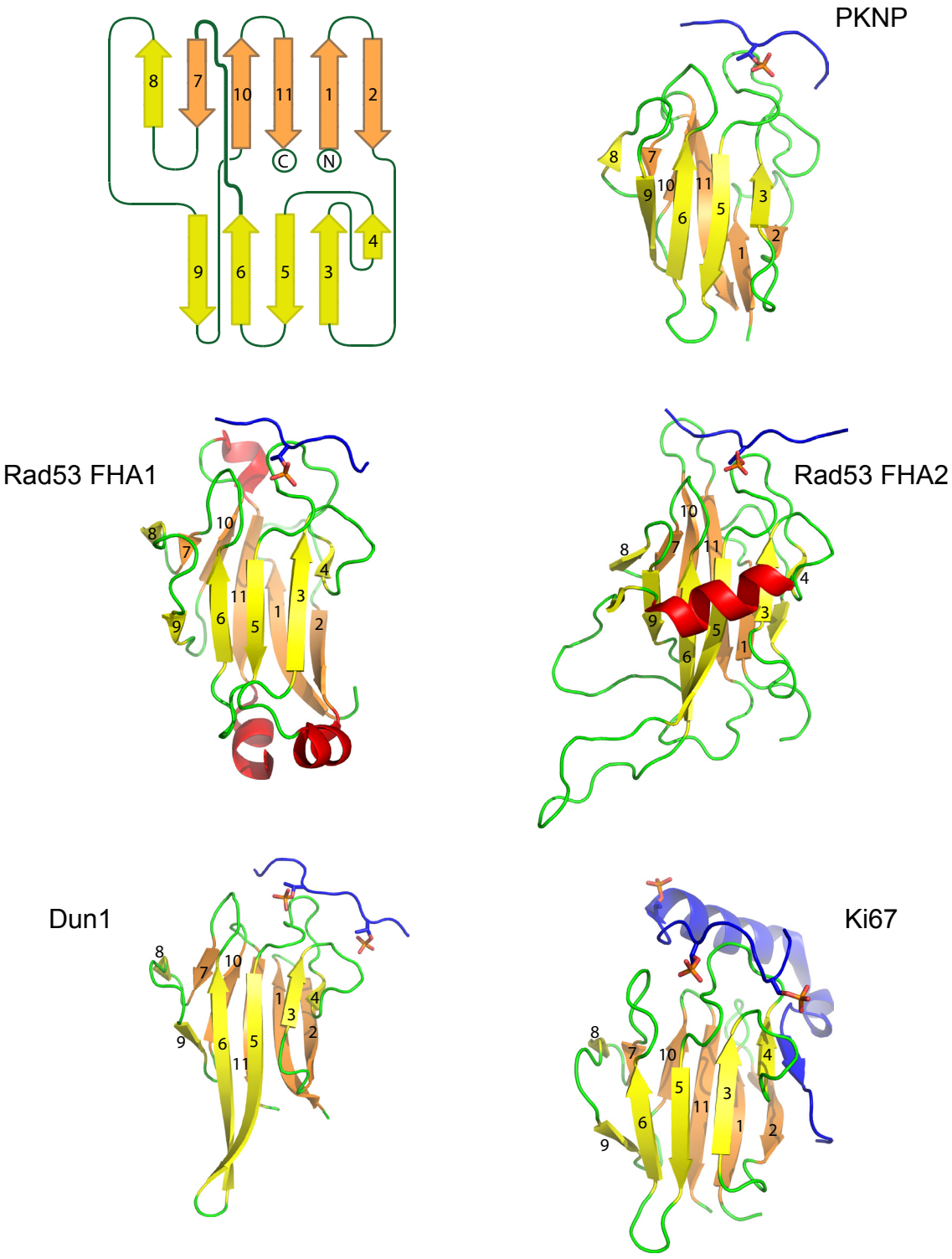

Figure 1.6. Binding specificity and functional versatility of FHA domains. Cartoon representation of the structures of Rad53 FHA1 (1G6G), Rad53 FHA2 (1K2N), PNKP (2W3O), Dun1 (2JQL) and Ki67 (2AFF) are shown (in green (loops), yellow/orange (β-strands) and red (α-helices)) in complex with their respective phosphopeptide ligands (blue). pThr residues in the phosphopeptide ligands are shown as sticks.

1.5.4 The possibility of phospho-independent FHA interactions?

The scope of FHA binding specificity and function has recently been expanded through the bioinformatics approach of evolutionary trace analysis. This technique is based on two hypotheses. The first is that functional sites (catalytic sites, binding sites etc.) on proteins evolve through variations of a conserved architecture. In this way architecture-defining residues remain mostly invariant, but residues responsible for functional specificity could undergo many substitutions, each of which may be associated with a functional variation and evolutionary divergence (Lichtarge and Sowa, 2002). The second hypothesis is that sequence homology trees approximate functional classification hierarchies. Lichtarge and co-workers justify this second hypothesis with the reasonably intuitive assertion that proteins sharing greater sequence identity have diverged more recently than sequences with less similarity and as a result have had less time to diverge functionally (Lichtarge et al., 1996; Lichtarge and Sowa, 2002).

When applied to FHA domains, evolutionary trace analysis predicts several discrete functional surfaces in addition to the canonical phospho-interaction surface (Lee et al., 2003a). Consistent with the structure of Ki67 FHA complexed with the hNIFK (residues 226-269) tri-phosphopeptide outlined above, β4 was predicted as a functional surface (Figure 1.7, C). Another predicted FHA domain recognition site incorporating residues from β9 and the β5/β6 loop is the location of the Ile 157 mutation to Thr of Chk2 FHA in Li-Fraumeni cancer patients (Figure 1.7, B) (Lee et al., 2003a; Liang and Van Doren, 2008). Mutation at this site, which is on the opposite side of the domain to the phospho-interaction surface, disrupts binding of BRCA1, CDC25A and p53, all of which have been implicated in human cancers(Li et al., 2002).

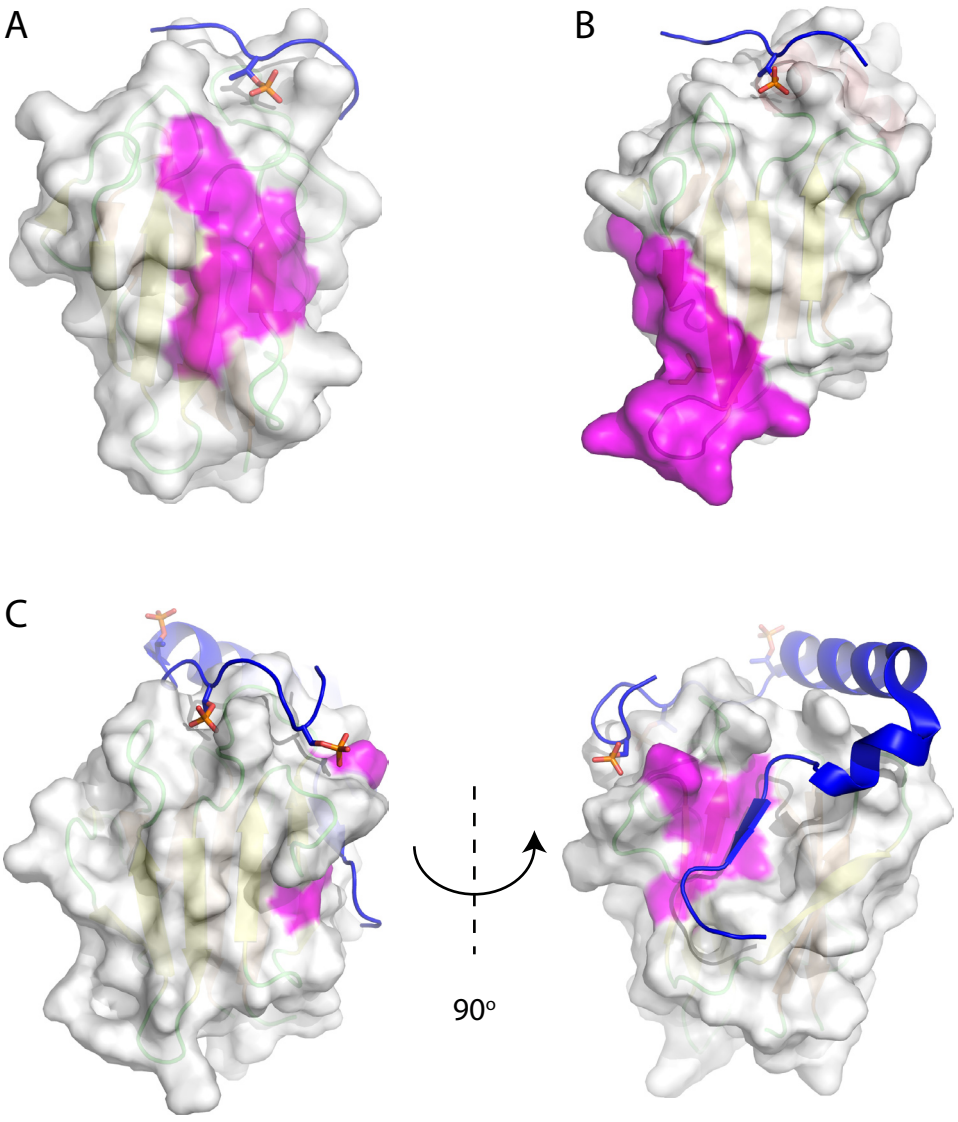

Figure 1.7. Bioinformatically-predicted conserved FHA domain surfaces. Conserved surfaces of the FHA domains are show in pink. Phosphopeptide ligands are shown as in Figure 1.6. (A) PNKP FHA domain-phosphopeptide complex (2W3O) shown in the same orientation as in Figure 1.6, highlighting the location of the predicted surface formed from residues in β7 and β10. (B) Chk2 FHA domain in complex with a synthetic phosphopeptide (1GXC). A predicted conserved surface is formed by residues from the loop connecting β5/β6 and from β9. (C) left Ki67 FHA domain-phosphoepeptide complex (2AFF) in the same orientation as in Figure 1.6. right Ki67 FHA domain rotated by 90° showing the location of the conserved surface formed by β4, which stacks anti-parallel with the C-terminal end of the phosphopeptide ligand.

Another functional surface predicted by bioinformatics lies around the adjacent β7 and β10 strands (illustrated using PNK FHA domain in Figure 1.7, A). To date, there have been no reports of FHA domains employing this surface to engage in functionally relevant phospho-independent interactions. However, the crystal structure of human CHFr displays forms a chain-swapped dimer at this site such that the expected β7/β8 hairpin straightens out into a single strand and hydrogen bonds with the equivalent strand of the other molecule of the dimer (Stavridi et al., 2002). This swaps β9-β11 into a β-sandwich with strands β1-β7 from the other molecule. The extent to which this chain swapped dimer feature is biologically relevant remains to be seen. It may be that it is a crystallographic artifact. The next section will address the phosphorylation signaling potential of *M. tuberculosis*, focusing on the kinases that have been studied in greatest detail.

1.6 The phosphorylation-signaling potential of M. tuberculosis

1.6.1 Overview

Like many prokaryotes, *M. tuberculosis* possesses the capability of conducting signal transduction by means of reversible phosphorylation signaling. Averaging between 1-2% of the genetic content across all bacteria, two-component systems are the most widely employed mechanism of phosphorylation signaling in prokaryotes (Wehenkel et al., 2008). In their simplest form, two-component systems consist of a sensor histidine kinase and a response regulator, although signaling cascades involving multiple histidine kinases and/or response regulators may function as part of complex histidyl-aspartyl phospho-relays. As well as possessing 11 complete two-component systems, *M. tuberculosis* also has at its disposal 11 eukaryotic-like STPK genes (PknA to L, except C),

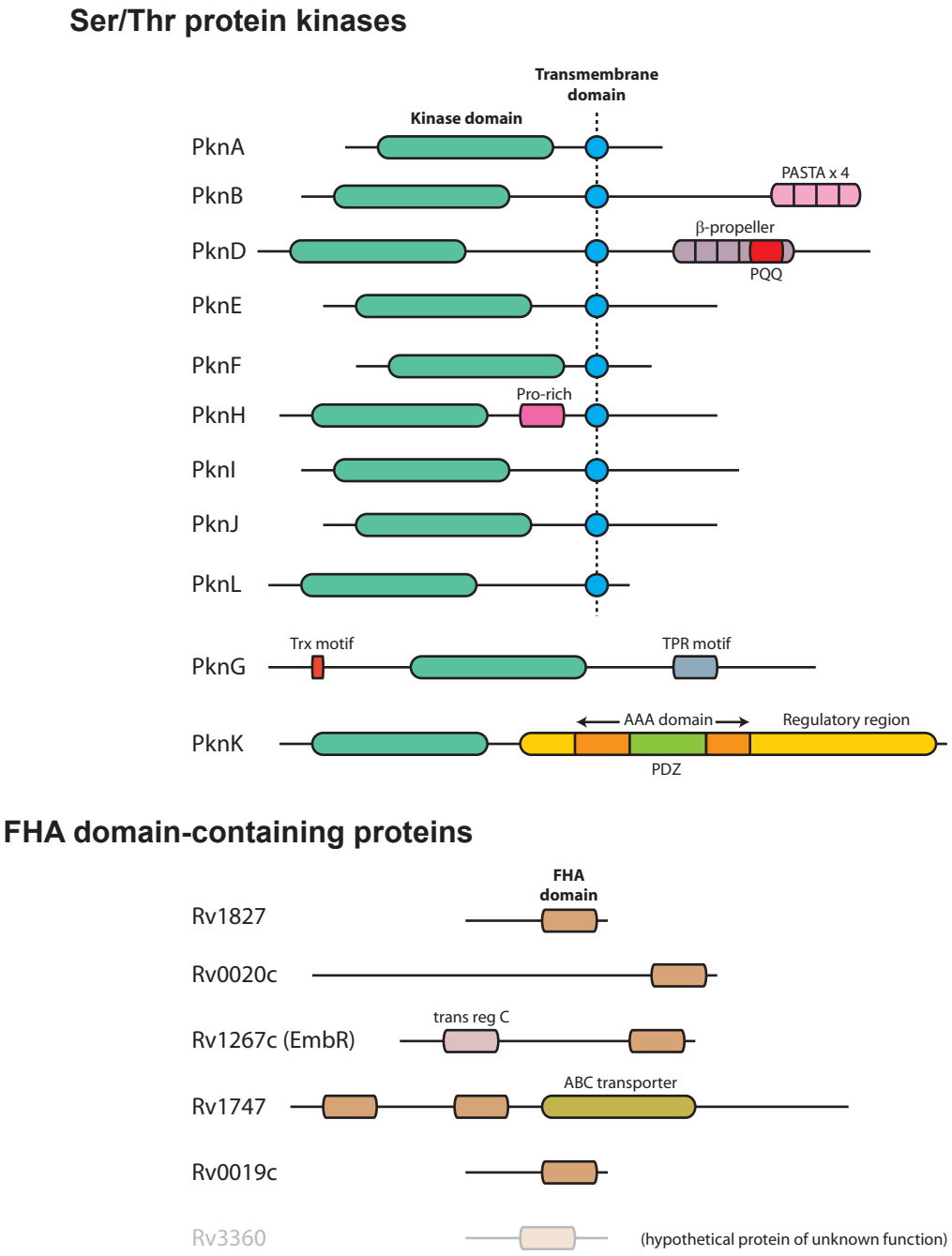

Figure 1.8. The phosphorylation signaling potential of M. tuberculosis. Schematic representation of the domain architecture of the Ser/Thr protein kinases, FHA domain-containing proteins and Ser/Thr protein phosphatase of M. tuberculosis. PQQ, pyrroloquinoline quinine; TRP, tetratricopeptide; Trx, thioredoxin. Figure adapted from (Av-Gay and Everett, 2000; Pallen et al., 2002).

three phospho-protein phosphatases and 5 FHA domain-containing proteins (Figure 1.8) (Av-Gay and Everett, 2000; Pallen et al., 2002; Wehenkel et al., 2008).

Little is known about the physiological roles and signaling mechanisms of the mycobacterial STPKs but they have been implicated in the regulation of growth, development, metabolism and interactions with host signaling processes. Their low sequence identity of approximately 30% with human kinases makes them attractive therapeutic targets and, to date, the kinase inhibitors erbstatin and staurosporine have been shown to prevent phagocytosis of *M. leprae* by mouse macrophages (Prabhakaran et al., 2000). Additionally, the protein kinase inhibitor H7 was found to inhibit the growth of *M. bovis* Bacille Calmette Guerin (BCG) and *M. smegmatis* mc^2 (Drews et al., 2001)

1.6.2 Protein kinase characterization in Mycobacterium tuberculosis

Of the 11 eukaryotic-like STPKs of *M. tuberculosis,* 9 are predicted to function as transmembrane receptor kinases, with an N-terminal kinase domain linked through a single transmembrane helix to one or more extracellular C-terminal domains thought to serve as signal sensors. PknK and PknG possess no predicted transmembrane domain and are thought to be cytosolic kinases.

1.6.3 PknB

The structure of the PknB kinase domain was the first prokaryotic STPK to be solved by x-ray crystallography and confirms high architectural conservation of the overall fold and the catalytic site to eukaryotic kinase domains (Ortiz-Lombardia et al., 2003; Young et al., 2003). Again, like many eukaryotic kinase domains, phosphorylation of the activation loop of PknB has been found to regulate its catalytic activity. At least two activation

loop threonines (Thr 171 and Thr 173) are fully phosphorylated in the active state. Dephosphorylation of Thr 171 and/or Thr 173 by their cognate phosphatase, PstP, or their substitution for alanine significantly decreases PknB kinase activity (Boitel et al., 2003). However, the available crystal structures of the PknB kinase domain show poor electron density in the region of the activation loop, suggesting that PknB employs an induced fit mechanism of binding acceptor substrates or other factors that modulate its activity (Boitel et al., 2003; Young et al., 2003).

The C-terminal extracellular sensor domain of PknB consists of four repeated PASTA (Penicillin-binding protein And Ser/Thr kinase Associated) domains, frequently found in bacterial STPKs and high molecular weight penicillin-binding proteins (Boitel et al., 2003; Wehenkel et al., 2008; Yeats et al., 2002; Young et al., 2003). The two PASTA domain repeats of *Streptococcus pneumoniae* Pbp2x (penicillin-binding protein 2x) were found to bind the β-lactam antibiotic cefuroxime. This lead to the hypothesis that PASTA domains play a role in sensing stem peptides from unlinked peptidoglycans in the cell wall (Gordon et al., 2000; Jones and Dyson, 2006; Yeats et al., 2002). As the specificity of PASTA domains for binding peptidoglycans is yet to be fully characterized, hypotheses about the actual functions of PknB remain speculative. However, it has been linked with the regulation of cellular growth and morphology. Over-expression of PknA or PknB in *M. tuberculosis* resulted in a broad, bulgy cell morphology while partial depletion of PknA or PknB using RNAi caused a narrow elongated morphology (Kang et al., 2005).

1.6.4 PknD

The crystal structure of PknE and three independent crystal forms of PknB show that a conserved surface, on the opposite side of the N-lobe of the respective kinase domains relative to their catalytic sites, enables dimerisation (Boitel et al., 2003; Gay et al., 2006; Ortiz-Lombardia et al., 2003; Wehenkel et al., 2006; Young et al., 2003). It has been suggested that back-to-back dimer formation in *M. tuberculosis* kinases promotes their activation in an analogous manner to the RNA-dependent antiviral protein PKR (Dar et al., 2005; Dey et al., 2005). In this model, ligand binding to the extracellular receptor domains of PknB induces back-to-back dimer formation of the intracellular kinase domains. This would result in reorientation of the αC helix, the C-terminal end of which contributes to the dimer interface and in so doing stabilizes the active conformation (Wehenkel et al., 2008). Indeed, allosteric activation of PknD by dimerisation has recently been demonstrated (Greenstein et al., 2007). In this study the authors constructed chimeric PknD kinase domain-fusion constructs that dimerised upon addition of rapamycin. Phosphorylation activity was found to be stimulated upon dimerisation. The previously unphosphorylated PknD kinase domains auto-phosphorylated and were then able to phosphorylate substrates *in vitro*. Activation and substrate specificity were shown to be attenuated by mutations at the dimer interface, although catalytically inactive mutants were shown to activate wild-type PknD kinase domain solely through dimerisation. Once auto-phosphorylated, the PknD kinase domain does not require dimerisation to maintain its activity, supporting the model of kinase activation through ligand-induced dimer formation.

Little is known about the function of PknD but genomic and expression data implicate it in sensing levels of extracellular phosphate (Av-Gay and Everett, 2000). The

extracellular sensor domain of PknD was solved by x-ray crystallography and was found to form a highly symmetric six-bladed β-propeller. This is tethered to the transmembrane domain via a flexible linker (Good et al., 2004). The structure shows a concave surface that is the site for ligand/protein-binding for many other β-propeller structures. The authors suggested that the quaternary structure of the intracellular kinase domain could be altered through the binding of multivalent ligands. However, the structure fails to provide clues as to how a signal could be transmitted to the intracellular kinase domain. Close inspection suggests that there would be little conformational alteration upon ligand binding and the flexibility of the transmembrane linker would appear to limit the propagation of any conformational change.

1.6.5 PknG

Of the two soluble cytosolic *M. tuberculosis* kinases PknG is the best studied. PknG is part of a putative operon containing proteins involved in glutamate uptake that is conserved within the mycobacteria and closely related actinomycetes. The PknG kinase domain is situated between an N-terminal rubredoxin-like domain and a C-terminal tetratricopeptide (TPR) domain. The TPR domain appears to be responsible for dimerisation of PknG in the crystal structure of a truncated construct of PknG lacking the N-terminal rubredoxin-like domain and in complex with the ATP-competitive inhibitor AX20017 (Scherr et al., 2007). This structure shows that AX20017 binds deep into a cleft shaped by a unique set of amino acids distinct from those found in human kinases, making PknG an attractive target for the design of anti-mycobacterial drugs. PknG has also been reported to be secreted within macrophage phagosomes where it inhibited phagosome-lysosome fusion, thereby mediating intracellular survival of mycobacteria (Walburger et al., 2004).

Determination of the auto-catalytic activity of PknG and its sites of auto-phosphorylation remains an active field of research. *In vitro* kinase assays performed by Cowley and co-workers reported that PknG auto-phosphorylated within its kinase domain as well as on the C-terminal TPR domain (Cowley et al., 2004). However, O'Hare and colleagues have reported that PknG has no phosphorylated residues in its catalytic domain or C-terminal TPR domain. They found phosphorylation sites solely in the N-terminal region of the protein at Thr 23, Ser 31/Thr 32 and within the Thr 63/Thr 64/Ser 65 motif. Additionally a construct in which the N-terminal 73 residues had been deleted failed to incorporate radiolabeled ATP, yet was still catalytically active against substrate peptides (O'Hare et al., 2008). Consistent with this is the finding that no phosphorylated residues were found in the PknG crystal structure which lacked the N-terminal 73 residues (Scherr et al., 2007). From these data it seems that PknG auto-phosphorylation sites lie within the N-terminal 73 residues and that phosphorylation of the catalytic domain is not required for activity. This contrasts with PknB which, like many eukaryotic kinase domains, requires activation loop phosphorylation for catalytic activity against substrates.

M. tuberculosis PknG and its homologue in the related actinomycete *C. glutamicum* share approximately 45% sequence identity and have been implicated in the regulation of intracellular glutamate levels. *C. glutamicum* PknG has been shown to exert regulation of glutamate metabolism through phosphorylation of the small FHA domain-containing protein OdhI. In its unphosphorylated state, OdhI binds to and inhibits the putative E1 subunit (OdhA) of the α-ketoglutarate dehydrogenase complex (Niebisch et al., 2006). Recent work has demonstrated similar findings in *M. tuberculosis* and *M. smegmatis*, showing that through phosphorylation of the mycobacterial homologue of OdhI

(Rv1827, also referred to as GarA) inhibition of the homologue of OdhA (α-ketoglutarate decarboxylase; KGD; Rv1248c) is relieved (O'Hare et al., 2008). Despite sharing approximately 58% sequence identity with OdhA, KGD has been found to function as an α-ketoglutarate decarboxylase in a variant tricarboxylic cycle that appears to have been evolutionarily optimized for mycobacterial persistence in infected macrophages (Tian et al., 2005a).

The phosphorylation-mediated regulation of core metabolic activity within *M. tuberculosis* is currently poorly understood. However, recent advances in this active field of research will be discussed more detail within the context of the results presented in subsequent chapters.

1.7 Research objectives

Since its identification as an FHA domain-containing culture filtrate protein, Rv1827 has been shown to be a substrate for at least two mycobacterial Ser/Thr protein kinases (Villarino et al., 2005). Furthermore, phosphorylation of Rv1827 has been implicated in the regulation of core metabolic processes in mycobacteria and the closely related actinomycete *C. glutamicum* (Niebisch et al., 2006; O'Hare et al., 2008). However, a molecular mechanism that links phosphorylation signaling with metabolic regulation through the phosphorylation state of Rv1827 has not been forthcoming.

Therefore, the principal objective of this investigation was structural and functional characterization of Rv1827. This began with the identification of Rv1827 interaction partners in *M. tuberculosis* cell-free extracts, and confirmed the metabolic context of Rv1827 activity to be regulation of glutamate metabolism. Then, through a combination

of biochemical, biophysical and structural analyses we determined that abrogation of Rv1827 interactions with its metabolic binding partners was caused by a phosphorylation-induced intramolecular association of the N-terminus of Rv1827 with its FHA domain. This lead us to propose a mechanistic description of Rv1827 functionality that explains how phosphorylation signaling is coupled to the regulation of a branching point in mycobacterial central metabolism.

2 Materials and Methods

2.1 Bioinformatics

2.1.1 Protein and DNA information

Sequence information on selected *M. tuberculosis* genes, their flanking regions and the proteins encoded therein was obtained from the Tuberculist Web Server (http://genolist.pasteur.fr/TubercuList/). This resource is provided by the Institut Pasteur and is an online *M. tuberculosis* genome database.

The properties of proteins that can be calculated from their amino acid sequences (in the presence and absence of N-terminal tags) such as theoretical isoelectric point (pI), molar extinction coefficient, proportion of each amino acid and molecular weight are useful to know when attempting purification or biophysical/biochemical analysis. The protein sequences of target genes were analysed using DNASTAR Lasergene 8 and the website calculator PROTParam, available from the ExPASy Proteomics Server (http://www.expasy.ch/tools/protparam.html).

Other ExPASy proteomics tools (http://www.expasy.ch/tools/) were also used for analysis of nucleotide and gene sequences. Protein sequence alignments were generated using the ClustalW2 general purpose multiple sequence alignment program (http://www.ebi.ac.uk/Tools/clustalw2/index.html).

2.1.2 Sequence homology searches

Protein sequence homologues were identified using the NCBI BLAST (Basic Local Alignment Search Tool) or PSI-BLAST (Position-Specific Iterated BLAST) servers

(http://www.ncbi.nlm.nih.gov/BLAST/) with filtering of low complexity regions. These are regions of sequence which have an amino acid distribution bias and can affect the accuracy of the BLAST results. Searches for sequence homologues were made against the non-redundant database, which comprises all non-redundant GenBank CDS translations plus the PDB (Protein Data Bank; http://www.rcsb.org/pdb/), SwissProt, PIR and PRF.

2.2 Molecular biology

2.2.1 Plasmid construction

pGEX 6P1 vector (GE Healthcare Life Sciences) containing a 486bp BamHI/HindIII fragment encoding the full length *M. tuberculosis* Rv1827 protein (162 residues) was provided by Dr. D. Patel (NIMR). This was used as the template from which truncated constructs were generated and on which site-directed mutagenesis was performed. pGEX 6P1 vector containing a 2265bp EcoRI/XhoI fragment encoding full length *M. tuberculosis* PknG protein was provided by Dr. K. Lougheed (NIMR). pGEX 6P1 vector with a 996bp LIC (ligation-independent cloning) fragment encoding the intracellular segment of *M. tuberculosis* PknB protein (332 residues) and pGEX 6P1 vector with a 837bp LIC fragment encoding the kinase domain of PknB protein (279 residues) were provided by Dr. D. Patel (NIMR).

2.2.2 Bacterial strains

E. coli NovaBlue and TOP10 Singles Competent Cells (Novagen) were used for initial cloning and DNA purification procedures (see section 2.2.7 for transformation protocols). *E. coli* Rosetta 2 (DE3) pLysS and *E. coli* BL21 (DE3) competent cells

(Novagen) were used in the expression of all protein constructs (Table 2.1). The DE3 lysogen contains the T7 polymerase gene under the control of the *lacUV5* promoter. Addition of IPTG (isopropyl-β-D-thiogalactopyranoside) (Peirce) induces constitutive expression of T7 polymerase. For expression vectors under the control of a T7 promoter, the polymerase transcribes the mRNA at high copy number resulting in over-expression of the desired protein.

E. coli strains	Genotype
NovaBlue	*endA1 hsdR17*(r_{K12}^{-} m_{K12}^{+}) *supE44 thi-1 recA1 gyrA96 relA1 lac* F'[*proA*$^{+}$*B*$^{+}$ *lacI*^{q}ZΔ*M15*::Tn*10*] (Tet^{R})
TOP10	F- *mcr*A Δ(*mrr-hsd* RMS-*mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*araleu*) 7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G (Invitrogen)
BL21 (DE3)	*B, F, hsdSB (rB-, mB-), gal, dcm, ompT, λ(DE3* (Novagen)
Rosetta 2 (DE3) pLysS	F^{-} *ompT hsdS*$_{B}$(r_{B}^{-} m_{B}^{-}) *gal dcm* (DE3) pLysS pRARE2[3] (Cam^{R})

Table 2.1. Genotypes of *E. coli* strains used for cloning and protein expression.

2.2.3 DNA manipulation and analysis

Agarose gels were prepared and run according to standard procedures (Sambrook, 1989). All plasmid DNA was purified from overnight cultures of *E. coli* NovaBlue Singles competent cells (Novagen) using either QIAprep Spin Miniprep or Midiprep Kits from Qiagen, according to the manufacturer's instructions. Final DNA concentrations were determined by UV spectrophotometry at 260nm.

2.2.4 Polymerase chain reaction (PCR)

PCR reactions were carried out using an MJ Research DNA Engine Dyad Peltier thermal cycler and the KOD Hot Start DNA Polymerase kit (Novagen). All reactions were carried out in thin-wall 200 µl PCR tubes (Abgene) at a final volume of 50 µl using a heated lid at a constant temperature of 100°C. The standard procedure used is a modified version of the GC-rich protocol provided by the manufacturer. The addition of DMSO (Sigma) (which reduces primer secondary structure) and the use of a 68°C extension temperature yields a higher and more consistent PCR success rate.

Reagent	Volume (µl)
dH_2O	31.5
KOD Reaction Buffer (10X)	5
$MgSO_4$ (25 mM)	2
DMSO (100%)	2.5
dNTP Mix (2 mM each)	5
Template DNA (5 ng/µl)	1
Oligonucleotide Primer 1 (15 µM)	1
Oligonucleotide Primer 2 (15 µM)	1
KOD HotStart DNA polymerase (1 n rxn^{-1})	1
Total	50

Table 2.2. Typical PCR reaction. 5% DMSO (final) in the reaction aids reduction in primer and template secondary structure that can interfere with PCR. The final concentration of each oligonucleotide is 0.2 mM.

Step	Temperature (°C)	Time
1	94 (initial heating)	2 min.
2	94 (denaturation of primers)	15 secs.
3	55 (annealing)	30 secs.
4	68 (extension)	1 min. kb^{-1}
5	(cycle to 2nd step 28 times)	
6	68 (final extension)	1 min. kb^{-1}.
7	4 (end of PCR program)	5 min.

Table 2.3. Typical thermocycler program for PCR reaction. An initial 2 min melt ensures full priming in the first cycle. In most cases 55°C was found to be sufficient regardless of primer T_m. The use of 29 cycles typically yielded approximately 1 µg/50 µl reaction.

All PCR products were purified using the QIAquick PCR Purification kit (Qiagen) according to the manufacturer's instructions. DNA from each reaction was eluted in 50µl H_2O.

2.2.5 Restriction enzyme digestion reactions

In all cases when double restriction enzyme digests were performed, buffer conditions compatible with both enzymes were chosen (according to the double digest table in the Roche catalogue). All digests were carried out at 10U restriction enzyme/1µg target DNA at 37°C for two hours. Reactions were halted where possible by heating according to the manufacturer's instructions (Roche).

Reagent	Volume (µl)
dH_2O	14
Reaction Buffer (10X)	2
BSA (20X)	1
DNA (1 µg)	1
Restriction enzyme 1 (10 u μl^{-1})	1
Restriction enzyme 2 (10 u μl^{-1})	1
Total	20

Table 2.4. Typical restriction enzyme digest for vector and PCR products. BSA was used in accordance with the manufacturer's instructions (Roche). The reaction volume was chosen as the lowest volume without exceeding the final glycerol concentration of 5% in order to avoid enzyme star activity (non-specific DNA cleavage).

Digested DNA products were purified by electrophoresis using a standard 1% agarose (Bio-Rad) gel. Gels were stained with 0.5 µg ml^{-1} ethidium bromide (Bio-Rad) and visualized on a 258 nm UV-transilluminator. The desired vector or PCR product was excised from the gel using a sterile scalpel blade and purified using the QIAquick Gel Extraction Kit (Qiagen). Each reaction was eluted in 50 µl H_2O.

2.2.6 Ligation reactions

Ligation reactions were incubated for one hour at room temperature using T4 DNA ligase according to the manufacturer's instructions (NEB).

Reagent	Volume (µl)
dH_2O	3
Reaction Buffer (10X)	2
Linearised vector (33 fmol)	8
Digested PCR insert (100 fmol)	6
T4 DNA ligase (400 u μl^{-1})	1
Total	20

Table 2.5. Typical ligation reaction. A ratio of 3:1 (PCR insert:vector) was used where possible.

2.2.7 Transformations

All transformations of bacterial cells were carried out according to the instructions supplied by the manufacturer of the competent cells (Invitrogen). A typical transformation was performed as follows: 1 µl of DNA solution (at a stock concentration of 50 ng μl^{-1}) was added to the appropriate competent cell aliquot (50 µl), incubated on ice for 10 minutes, heat-shocked at 42°C for 30 seconds, incubated on ice for 2 minutes, incubated at 37°C for 30 min. after the addition of 200 µl LB and plated on the appropriate media. 1 ng and 50 ng total DNA were used for cloning and expression vectors, respectively. 5 µl of the total ligation reaction mixture (Table 2.5) were used to transform competent cells. All plates were incubated overnight at 37°C.

2.2.8 Site-directed mutagenesis and DNA sequence verification

All site-directed mutagenesis experiments were carried out using the QuikChange kit (Stratagene) according to the manufacturer's instructions. All mutagenic primers were

HPLC purified (Sigma) to remove compounds that might reduce the efficiency of site-directed mutagenesis reactions. 1 μl of the mutated plasmid DNA was transformed and purified as above.

Reagent	Volume (μl)
dH_2O	40
Reaction Buffer (10X)	5
dNTP Mix	1
Template DNA (10 ng)	1
Oligonucleotide Primer 1 (15 μM)	1
Oligonucleotide Primer 2 (15 μM)	1
Pfu DNA Polymerase (2.5 u μl^{-1})	1
Total	50

Table 2.6. Typical site-directed mutagenesis reaction. Where possible primers were designed to conform to the manufacturer's parameters: T_m 75-85°C, GC content 40-60%, length 25-45 bases, length of regions flanking the mutations 11-21 bases, G or C at the end of the primer.

Step	Temperature (°C)	Time
1	95	30 secs.
2	95	30 secs.
3	55	1 min.
4	68	1 min. kb^{-1}
5	(cycle to 2nd step 12 times)	
6	4	5 min.

Table 2.7. Typical thermocycler program for a QuikChange (Stratagene) reaction.

All DNA clones and mutations were sequenced by Cogenics using standard fluorescent dideoxy terminator techniques to confirm construct identity. Sequences were aligned and analyzed using SeqMan (DNASTAR Lasergene 8).

2.3 Protein purification techniques

2.3.1 Protein concentration determination

Protein concentration was determined directly by spectroscopic analysis using a Cary 50 Bio (Varian) spectrophotometer. A spectrum (200 nm to 350 nm) was recorded for each sample. Concentration was determined according to:

$$A_{280} = \varepsilon_{280} \cdot c \cdot l$$

where c is the molar concentration, A_{280} is the absorbance at 280 nm, ε_{280} is the molar extinction coefficient at 280 nm (calculated using the DNASTAR Protean software) and l is the path-length of the cuvette.

2.3.2 Protein concentration and storage

Protein concentration was carried out using Vivaspin ultrafiltration concentrators (VivaScience). The capacity (ml) and membrane molecular weight cutoff (MWCO) varied depending upon sample volume and molecular weight. The most commonly used concentrators were 20 ml Vivaspins of 3, 5, 10 or 30 kDa MWCO. Most samples were concentrated in Vivaspins by centrifugation at 4000*g*. The time necessary for concentration varied greatly with initial sample volume, final concentration required, and the nature of the protein being analyzed. In most cases, final protein preparations were snap frozen in 10-200 µl aliquots (in eppendorf tubes) in the appropriate buffer by immersion in liquid nitrogen and then stored at -80°C.

2.3.3 SDS-PAGE

All SDS-PAGE experiments were performed using precast Novex 1.0 mm Bis-Tris gels (Invitrogen) according to the manufacturer's instructions. Gels were run at 200 V for 38 min. in an XCell SureLock mini-cell (Invitrogen) on a Bio-Rad PowerPac 300. For each gel, broad range protein standard (Bio-Rad) molecular marker was used. Each protein sample was mixed with NuPAGE 4x LDS sample buffer (Invitrogen) and NuPAGE sample reducing agent (Invitrogen) according to the manufacturer's instructions and heated to 100°C for 10 min. prior to loading. The typical final volume loaded was between 10 and 30 µl per well. Gels were developed by staining with InstantBlue Coomassie stain (Expedeon) according to the manufacturer's instructions. Gels were imaged on a Kodak Digital Science Image Station 440CF using transmitted ambient light with a yellow (523 nm) high-pass filter.

2.3.4 Protein expression

All expression studies used either *E. coli* Rosetta 2 (DE3) pLysS or *E. coli* BL21 (DE3) competent cells (Novagen), which were transformed according to the manufacturer's instructions. One colony from an overnight plate was used to inoculate 2 ml of LB medium containing the appropriate antibiotics. The culture was incubated for 8 hours at 37°C and 1 ml was used to inoculate 250 ml of LB + antibiotics. The 250 ml culture was grown overnight at 37°C and was subsequently used as a starter culture. 1L of LB + antibiotics was inoculated with 10 ml starter culture and grown to an optical density at 600 nm of 0.4 at 37°C. The temperature was then reduced to 30°C and when an optical density at 600 nm of 0.6-0.8 was achieved (mid-log phase) IPTG was added to a final concentration of 1 mM to induce protein expression. At three and a half hours post induction cells were harvested by centrifugation at 6,000*g* for 10 min. The pellet was

washed by re-suspension in phosphate-buffered saline (PBS) (Sigma) and re-harvested as before. All cell pellets were stored in 50 ml Falcon tubes (Corning) at -20°C.

2.3.5 Bacterial lysis

All bacterial lysis was carried out at 4°C. Bacterial pellets were fully resuspended in 10 ml g^{-1} (lysis buffer/pellet) according to the individual purification protocol and supplemented with 30 kU g^{-1} rLysozyme and 25 U ml^{-1} Benzonase (Novagen) to aid in lysis and reduce lysate viscosity. Lysis buffer was augmented with one EDTA-free Protease Inhibitor Cocktail Tablet (Roche) 50 ml^{-1}. Cells were lysed by sonication (Branson Sonifier 450) on ice at 40% power with six bursts of ~20 secs. 60 ml^{-1} sonicate volume. The lysate was clarified by centrifugation at 45,000*g* for one hour. Aliquots of the whole sonicate, clarified lysate, and insoluble pellet (dissolved in ~10 M urea) were removed for SDS-PAGE analysis.

2.3.6 GST-purification

All proteins were expressed as 3C-protease cleavable GST-fusions and unless otherwise stated were purified as follows. Cell pellets were resuspended and lysed (as above) in GST-lysis buffer (50 mM TrisCl pH 8.0, 300 mM NaCl, 0.5 mM tris-(carboxyethyl) phosphine (TCEP) (Pierce)). The clarified lysate was applied to an appropriate amount of glutathione Sepharose 4B resin (typically 2-5 ml) (Amersham) pre-equilibrated in GST-lysis buffer. The slurry was incubated overnight at 4°C with gentle mixing. Resin was recovered by centrifugation at 500g for 5 min. An aliquot was taken of the supernatant for SDS-PAGE analysis and the rest discarded. The resin was diluted with an equal volume of lysis buffer and carefully removed into an empty 25 ml glass Econo-column (Bio-Rad) pre-equilibrated in GST-lysis buffer and allowed to settle under

gravity. The resin was washed with 50 column volumes (i.e. total resin volume = 1 column volume) of wash buffer (50 mM TrisCl pH 8.0, 300 mM NaCl, 0.5 mM TCEP) under gravity flow.

If an un-cleaved GST-fusion protein was required, the target protein was eluted at this stage using 3 x ½ column volumes of wash buffer supplemented with 10 mM reduced glutathione (pH 8.0). Aliquots of each elution fraction were taken for SDS-PAGE analysis. Alternatively if a 3C-cleaved target protein was required, the washed resin was equilibrated with 3C-cleavage buffer (20 mM TrisCl pH 8.0, 200 mM NaCl, 1 mM EDTA (Sigma), 0.5 mM TCEP) and diluted with 1 column volume to produce a 50% resin slurry. Non-cleavable GST-tagged 3C protease, supplied by J. Lloyd and Dr. Jiejin Li (NIMR), was added to the resin according to their recommendations and the 50% slurry was incubated overnight with gentle mixing. The target protein was eluted with 3C-cleavage buffer in 3 x ½ column volume fractions. Again, aliquots of each elution fraction were taken for SDS-PAGE analysis.

2.3.7 Ion-exchange chromatography

In the case of PknB purification, an ion-exchange step was required following 3C-cleavage. Following cleavage the NaCl concentration of the buffer was adjusted to 50 mM by dilution of the protein sample in preparation for ion-exchange. The protein sample was applied to a Q-sepharose column (by cycling the column flow-through back through the column) equilibrated with 50 mM NaCl, 50 mM TrisCl pH 8.0. Bound proteins were eluted using a NaCl gradient of 0-300 mM, 50 mM TrisCl pH 8.0. Fractions containing PknB were identified using SDS-PAGE, pooled and the protein was

concentrated (to 2-5 mg ml^{-1}) using a Vivaspin centrifugal concentrator (10 kDa cutoff) and subjected to size-exclusion chromatography.

2.3.8 Size-exclusion chromatography

The eluent from the previous column was concentrated to 3-5 ml and loaded onto a pre-packed HiLoad Superdex prep grade column (Amersham) using an AKTA Prime system. Either Superdex 75 or 200 (S75 or S200) was used depending on the molecular weight of the target protein. The column was pre-equilibrated in gel filtration buffer (typically 10 mM TrisCl pH 8.0, 50 mM NaCl) and run at 0.3 ml min^{-1} overnight collecting in 3 ml fractions. Fractions containing the appropriate peaks were pooled, concentrated, and stored at -80°C.

2.3.9 NMR sample preparation

Recombinant uniformly ^{15}N- and ^{13}C-labelled full-length Rv1827 was expressed using the pGEX 6P1 vector system (GE Healthcare) in Rosetta 2(DE3) pLysS *E. coli* (Novagen), grown in minimal media. 1 L ^{15}N-^{13}C- 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) (99at.% ^{15}N min. from GOSS Scientific Instruments Ltd.), D-glucose (2 g) (U-$^{13}C6$, 99% from Cambridge Isotope Laboratories Inc.) and ampicillin (100 μg ml^{-1}). All minimal media was 0.2 μm filter-sterilised and induced expression cultures (see above) were grown at 18°C overnight and harvested as above.

NMR samples contained 0.75-1.0 mM uniformly ^{15}N- and ^{13}C-labelled Rv1827 (1-162) in 90% H_2O/10% ^2H_2O containing 20 mM sodium acetate pH 5.8 and 50 mM NaCl.

Phosphorylation of uniformly ^{15}N- and ^{13}C-labelled Rv1827 was carried out as previously described prior to dialysis into NMR buffer using Slide-A-Lyzer MINI Dialysis Units (Pierce) with a 3500 Da MWCO.

2.4 Biochemical and biophysical techniques

2.4.1 Phosphorylation studies

Stoichiometric phosphorylation of full length Rv1827 (162 residues) at Thr 22 was achieved by diluting the protein to between 0.25 and 1.0 mM in 50 mM TrisCl pH 8.0, 150 mM NaCl, 5 mM $MgCl_2$, supplemented with ATP at twice the final Rv1827 molar concentration and adding PknB kinase domain (residues 1-279) to produce a final molar ratio of substrate:kinase of 250:1. PknB kinase domain was purified using GST, ion exchange and size exclusion chromatography. Stoichiometric phosphorylation of Rv1827 at Thr 22 was typically accomplished in 20 min. at room temperature. Phosphorylation of Rv1827 at Thr 21 by PknG was achieved under the same conditions in ~20 hours. PknG was purified by GST and size exclusion chromatography. For some applications removal of the kinase from the phosphorylated Rv1827 samples was necessary. This was achieved by size-exclusion chromatography (outlined in section 2.3.8).

2.4.2 Limited proteolysis

Sequencing grade modified trypsin (Promega) was used in all the limited proteolysis experiments at a ratio of target protein:trypsin of 500:1 (w/w) at room temperature. For SDS-PAGE analysis, aliquots of the reaction mixture were taken at defined time points (e.g. at 0, 1, 5, 10, 20 etc. mins after addition of trypsin) and immediately mixed with LDS

sample loading buffer (Invitrogen) and NuPAGE® Sample Reducing Agent (Invitrogen) at 100°C for 2 min. to stop the reaction.

2.4.3 Preparation of M. tuberculosis cell-free lysate and Pull-down assays

*M. tuberculosi*s cell-free extract (lysate) was prepared (by Debbie Hunt, NIMR) from *M. tuberculosi*s H37Rv grown in 100 ml Dubos broth containing 0.05% Tween 80 (v/v) supplemented with 0.2% glycerol (v/v) and 4% Dubos medium albumin (v/v) in 1 L polycarbonate culture bottles (Nalgene) in a Bellco roll-in incubator (at 2 rpm) at 37°C. When cultures reached an optical density (λ=600 nm) of 0.8-1.0 bacteria were harvested by centrifugation (10000g, 20 min.). Cell pellets were washed with 20 ml 10 mM TrisCl pH 7.5 before resuspension in 1.5 ml 10 mM TrisCl pH 7.5 containing one EDTA-free protease inhibitor cocktail tablet (Roche) 50 ml^{-1}. Bacteria were placed on ice and lysed in the presence of glass beads (150-212 µm, Sigma) in a Ribolyser (Hybaid) and placed on ice for 5 min. Glass beads and intact bacteria were removed by centrifugation twice at 13000g for 15 min. at 4°C. Cell-free extracts were further filtered through a low binding Durapore 0.22 µm membrane filter (Ultrafree-MC, Millipore). Aliquots of 10 µl glutathione Sepharose 4B resin (Amersham) pre-equilibrated in 50 mM TrisCl pH 8.0, 300 mM NaCl and 0.5 mM TCEP were saturated with gel filtration purified GST and GST-Rv1827 constructs. Unbound protein was removed by five cycles of centrifugation at ~500g for 3 min., aspiration of the supernatant and resuspension of the sedimented resin in 0.5 ml of the same buffer. Each aliquot of saturated resin was mixed with 100 µl clarified *M. tuberculosi*s cell-free lysate and incubated at 4°C for 3 hours with mixing. Unbound protein was removed as described. LDS sample loading buffer (Invitrogen) and NuPAGE® Sample Reducing Agent (Invitrogen) were added directly to the

recovered resin according to the manufacturer's instructions. Protein content was analysed by SDS-PAGE.

2.4.4 Dynamic light scattering

Dynamic light scattering was carried out on a DynaPro-801 dynamic light scattering instrument (DynaPro). Protein samples were diluted according to the manufacturer's instructions (usually 4-5 mg ml^{-1}). Each sample was passed through a 0.22 μm filter and a minimum of fifty measurements were taken at 20°C.

2.4.5 Multi-angle laser light scattering

Samples of purified protein were applied at a flow rate of 0.5 ml min^{-1} to a Superdex 200 (or 75) 10/300 GL column pre-equilibrated in 50 mM TrisCl pH 8.0, 300 mM NaCl mounted on a Jasco HPLC. A DAWN-HELEOS multi-angle 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 refractive index change (dn/dc = 0.186) using 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 using ASTRA v5.1 (Wyatt Technology Corp., Santa Barbara, CA).

2.4.6 Circular dichroism spectroscopy

Rv1827 samples were diluted to 210 μg ml^{-1} with 50mM phosphate buffer (pH 6.85) for far-UV CD and to 3.75 mg ml^{-1} in the same buffer for near-UV CD. Spectra were

acquired on a JASCO-J715 spectropolarimeter. Spectral collection of far-UV (260-190 nm) and near-UV (250-320 nm) CD were carried out in 1 mm and 10 mm path-length quartz cuvettes, respectively. Spectra were taken with a 100 nm $min.^{-1}$ scan rate, 0.25 secs. time constant and 2 nm spectral bandwidth. Final spectra are the average of at least 20 scans, and data analysis was carried out using Specpro (Dr. S. Martin, NIMR). CD signal was measured in millidegrees and converted to $\Delta\varepsilon_{mrw}$ according to:

$$\Delta\varepsilon_{mrw} = \frac{S \cdot (mrw)}{3290 \cdot c \cdot l}$$

where S is the CD signal in millidegrees, mrw is the mean residue weight of the protein, c is the protein concentration (mg ml^{-1}) and *l* is the cell path length (cm)

Thermal denaturations were carried out in a 2 mm path length quartz cuvet from 2-95°C (2°C min^{-1}; 4 secs. time constant) and were monitored at 202 nm.

2.4.7 Mass spectrometric techniques

2.4.7.1 Electrospray mass spectrometry

Mass spectrometry was utilised in this study primarily to verify the identity and chemical homogeneity of purified protein samples. It was also used to determine the phosphorylation state of putative substrates for the kinases that have been investigated. Protein molecular weight was determined using a stand-alone syringe pump (Perkin Elmer, Foster City CA) coupled to a Platform electrospray mass spectrometer (Micromass, Manchester, UK). Samples were desalted on-line using a 2 mm x 10 mm guard column (Upchurch Scientific, Oak Harbor WA) packed with 50 μm Poros RII

resin (Perseptive Biosystems, Framingham) inserted in place of the sample loop on a rheodyne 7125 valve. Proteins were injected onto the column in 10% acetonitrile, 0.10% formic acid, washed with the same solvent and then step eluted into the mass spectrometer in 70% acetonitrile, 0.1% formic acid at a flow rate of 10 µl $min.^{-1}$. The mass spectrometer was calibrated using myoglobin and standard samples comprised of 100-300 pmol protein at a minimum concentration of 1 µM.

2.4.7.2 MALDI-TOF mass spectrometry

For protein identification, bands were excised and subjected to tryptic in-gel digestion in accordance with a protocol supplied by Dr. S. Howell (NIMR). Each band of interest was excised with a sterile scalpel and placed in a 0.5 ml tube. SDS-PAGE buffer components and Coomassie stain were removed by two washes with extraction buffer (200 mM ammonium bicarbonate, 50% acetonitrile) for 30 min. at 37°C. The gel sample was reduced with extraction buffer supplemented with 20 mM dithiothreitol (DTT) (Sigma) for 1 hour at 37°C. The DTT was removed with three washes of extraction buffer for 30 min. at 37°C. The cysteines in the gel sample were alkylated (in the dark) by extraction buffer supplemented with 5 mM iodoacetamide for 20 min. at room temperature. The iodoacetamide was removed with three washes of 20 mM ammonium bicarbonate, 50% acetonitrile buffer for 30 min. at room temperature. All buffer was removed and 0.5 ml of 100% acetonitrile was added to the gel sample and incubated for 15 min. at room temperature. The acetonitrile was removed and the gel sample was dried in a desiccator for 1 hour. The fully desiccated gel sample was stored at room temperature until proteolytic digestion was performed. Digestion was performed by adding 100 ng (in 50 µl of 5 mM ammonium bicarbonate) of modified sequencing-grade

trypsin (Promega) to the desiccated gel sample. The digestion was performed at 37°C overnight. MALDI-TOF analysis was performed using 0.3 µl of the tryptic digest.

A Reflex III MALDI time-of-flight mass spectrometer (Bruker Daltonik GmbH, Bremen, Germany) equipped with a nitrogen laser and a Scout-384 probe was used to obtain positive ion mass spectra of digested protein with pulsed ion extraction in reflectron mode. An accelerating voltage of 26 kV was used with detector bias gating set to 2 kV and a mass cut-off of m/z 650. Thin layer matrix surfaces of α-cyano-4-hydroxycinnamic acid mixed with nitrocellulose were prepared as described by Shevchenko et al., 1996. A 0.3 µl aliquot of acidified digestion supernatant was deposited onto the thin layer and allowed to dry prior to rinsing with water. Analysis of the resulting spectra was carried out by Mascot (Matrix Science Ltd.), which uses mass spectrometry data to identify proteins from primary sequence databases.

3 Nuclear magnetic resonance spectroscopy: materials and methodology

3.1 Overview

There are two techniques by which protein structures are routinely determined at atomic resolution: nuclear magnetic resonance (NMR) spectroscopy and x-ray crystallography. Structures solved by X-ray crystallography originate directly from measurements of the electron density in a crystal lattice, whereas NMR is an inherently indirect technique. NMR solution structures are modeled using data derived from experimentally determined chemical shifts. Unlike X-ray crystallography, NMR spectroscopy can also be used to analyze time-dependent chemical phenomena such as reaction kinetics and intramolecular dynamics, and as such is a powerful technique for investigating the properties of proteins in solution.

Both X-ray crystallography and NMR spectroscopy have their limitations. X-ray crystallography is dependent on the generation of diffracting protein crystals, which is often rate limiting and not always possible for a given macromolecule. Structure determination by NMR spectroscopy is limited primarily by the size of the macromolecule under investigation, as the complexity of NMR spectra increase dramatically with increased molecular weight, and the stability of samples over long periods of time at room temperature. Structure determination by NMR spectroscopy can also be a lengthy procedure and it is difficult to eliminate human error in the analysis of NMR spectra. Nonetheless, NMR solution structures make up a significant percentage of those deposited in the Protein Data Bank (~14% in March 2009).

3.2 Nuclear magnetic resonance spectroscopy

3.2.1 Spectroscopy

NMR spectra were acquired at 25°C on Varian Inova 800 and 600 and Bruker Avance 700 and 600 MHz spectrometers, all equipped for $^{1}H/^{15}N/^{13}C$ triple-resonance experiments. The Bruker spectrometers were controlled by TOPSPIN software (http://www.bruker-biospin.com/topspin.html) and the Varian spectrometers by VNMR software (http://www.varian.com). In most experiments water suppression was achieved using WATERGATE (Piotto et al., 1992). NMR experiments and acquisition of the spectra that were used for the structural and hydrodynamic investigation of Rv1827 and the PknB-phosphorylated form were performed by Dr. Geoff Kelly (NIMR).

^{1}H-^{15}N Heteronuclear Single Quantum Coherence (HSQC) spectra were acquired for ^{15}N singly or ^{15}N- and ^{13}C-doubly isotopically labeled protein and used for various purposes such as evaluation of sample quality and consistency, titrations and assignment purposes, which resulted in the employment of various acquisition parameters (1024 x 64 to 2048 x 1024 data points).

Shown in Table 3.1 are the two (^{1}H-^{15}N HSQC) and three dimensional (all other) NMR experiments that were performed for backbone assignment, sidechain assignment and generation of distance restraints for Rv1827 and the PknB-phosphorylated form. The buffer conditions and sample concentrations used in each experiment are described in section 2.3.9.

Experiment	Purpose of experiment
^{1}H-^{15}N HSQC	General
HNCACB	Backbone assignment
CBCA(CO)NH	
HN(CA)CO	
HNCO	
HNCA	
(H)C(CCO)NH	Sidechain assignment
H(CCCO)NH	
HCCH-TOCSY	
^{15}N-edited NOESY	Distance restraints
^{13}C-edited NOESY	

Table 3.1. NMR spectroscopy experiments and the reasons for performing them.

3.2.2 Spectral processing

All NMR spectra were processed using nmrPipe (Delaglio et al., 1995). Typically, the acquisition dimension was multiplied by a Gaussian window function and a 90^{o} shifted sine-bell function in the indirect dimensions to enhance spectral apodisation (resolution). Polynomial solvent subtraction was used on the acquisition dimension and linear prediction was used on the indirect dimension to improve and increase the spectral resolution if required. All dimensions were zero-filled to the next appropriate power of two.

3.2.3 Spectral analysis and resonance assignment

NMR spectra were analysed using CARA (Computer-Aided Resonance Assignment) version 1.5.5/NEASY (http://www.nmr.ch/). ^{1}H, ^{15}N and ^{13}C resonance assignment of the backbone was achieved by analysis of HNCACB, CBCA(CO)NH, HN(CA)CO, HNCO and HNCA triple resonance experiments (Bax and Grzesiek, 1993) (Table 3.1 and Figure 3.1 (top)). Assignment of ^{1}H, ^{15}N and ^{13}C sidechain resonances was

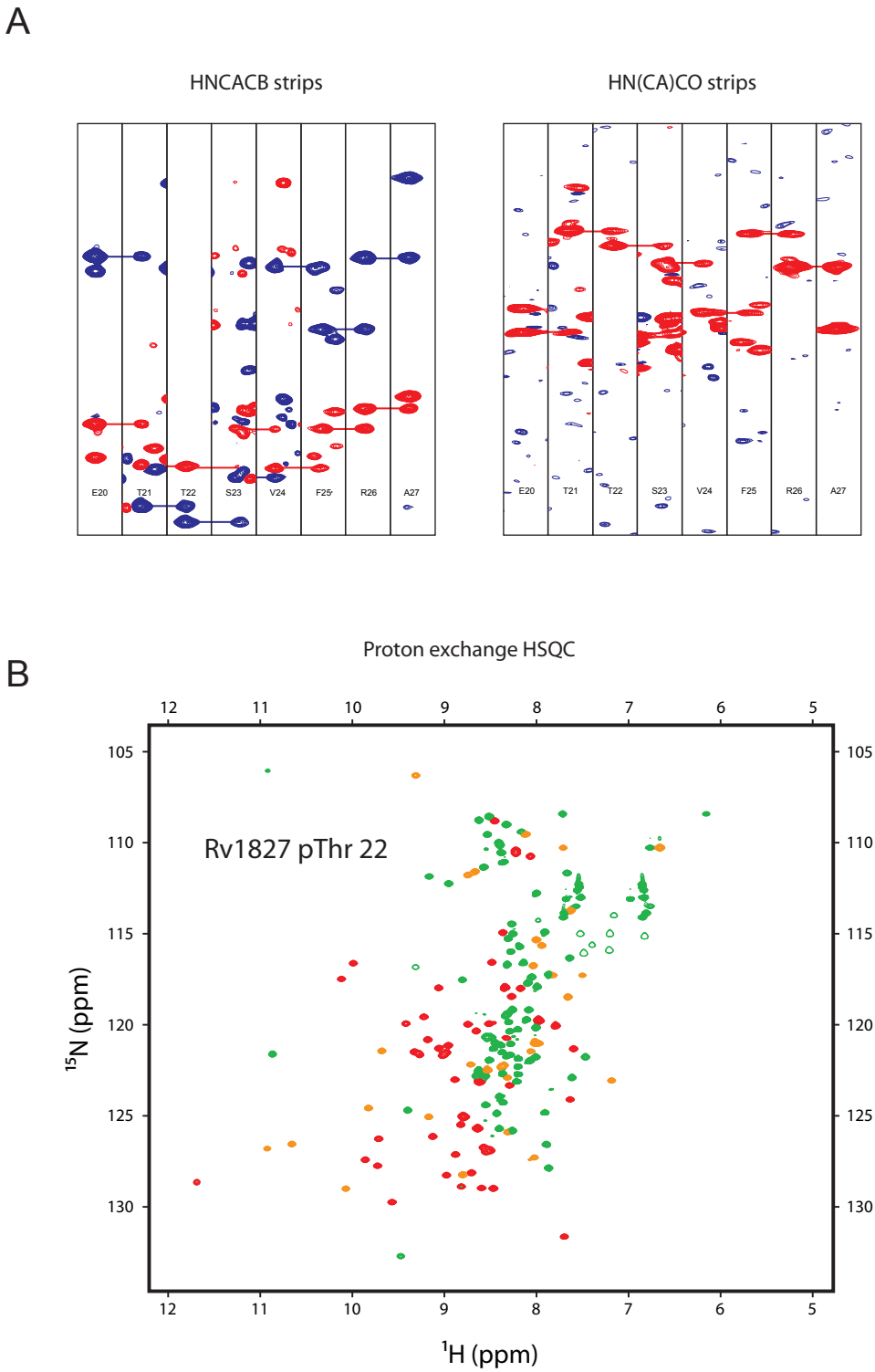

Figure 3.1. Resonance assignment using HNCACB and HN(CA)CO spectra (A) and ^{1}H-^{15}N HSQC spectrum of Rv1827 pThr 22 coloured according to the results of the ^{2}H exchange experiment (B). Peaks that exchanged within 2 minutes are coloured green. Peaks that exchanged within 17 hours are coloured orange and those that did not exchange within 17 hours are coloured red. Peaks folded in the ^{15}N-dimension are shown as hollow contours. See Figure 6.2 for annotated ^{1}H-^{15}N HSQC spectrum.

completed using (H)C(CCO)NH, H(CCCO)NH, three-dimensional HCCH-TOCSY and ^{15}N-edited and ^{13}C-edited NOE spectra.

3.3 Hydrodynamic analysis by NMR

3.3.1 Relaxation experiments

The dipolar nuclear magnetic spin relaxation of protonated heteronuclei (such as ^{15}N) is mediated by the overall tumbling of the molecule in solution and by the internal motions of ^{15}N-^{1}H bond vectors. Therefore by measurement of ^{15}N relaxation parameters, such as the longitudinal or transverse relaxation rate constants and the steady state ^{1}H-^{15}N NOE, useful information can be gained about the intramolecular dynamics of a protein.

^{15}N T_1, T_2 and ^{1}H-^{15}N NOE enhancement relaxation experiments were performed at 600 MHz (at 25°C), using standard acquisition methods (Kay et al., 1989). Buffer conditions and sample concentrations are described in section 2.3.9. Tau values used in the T_1 and T_2 relaxation experiments are shown in Table 9.2.

3.3.2 Longitudinal and transverse relaxation

The relaxation of an amide ^{15}N nucleus at high field frequency is dominated by the dipolar interaction with the directly attached proton spin (Kay et al., 1989). Longitudinal (spin-lattice or T_1) and transverse (spin-spin or T_2) relaxation experiments can be used to monitor the reorientation of amide ^{15}N-^{1}H bond vectors in the backbone of proteins over a wide range of timescales. T_1 relaxation values are sensitive to high frequency ^{15}N-^{1}H bond vector motions (on the nano- to pico-second time scale) and T_2 relaxation values are sensitive to slower dynamic processes.

The ^{15}N-^{1}H bond vectors of residues in unstructured regions of proteins reorient rapidly in solution and have similar T_1 and T_2 relaxation times (Figure 3.2). As proteins become larger, more structured and tumble more slowly in solution, their bond vectors reorient more slowly and their ^{15}N T_2 relaxation rate gets proportionally faster. Unlike the T_2 mechanism, T_1 relaxation is dependent on the transfer of the energy of excited nuclei to their surroundings (the lattice). This transfer of energy (and thereby the rate of T_1 relaxation) will be most efficient if the ^{15}N-^{1}H bond vector reorients at the Larmor frequency. If its motion is faster or slower, the nuclei will not be able to transfer energy to their surroundings as efficiently, resulting in a slower rate of T_1 relaxation (Figure 3.2). Therefore, for a large macromolecule that is tumbling slowly in solution, the rate of T_2 relaxation of its amide ^{15}N nuclei will be fast, and the rate of T_1 relaxation will be slow.

The ratio between the average T_1 and T_2 relaxation values for a given molecule can be used to determine its correlation time (τ_c). This may be imagined as the time it takes for the molecule in solution to rotate through an angle of one radian. In the context of a protein, which has one ^{15}N-^{1}H bond vector per residue (except for proline residues), global and local correlation times can also be estimated. As the T_2 relaxation of the amide ^{15}N nuclei in well-folded globular proteins usually occurs 5 to 10 times more rapidly than their T_1 relaxation, estimation of the correlation time is possible.

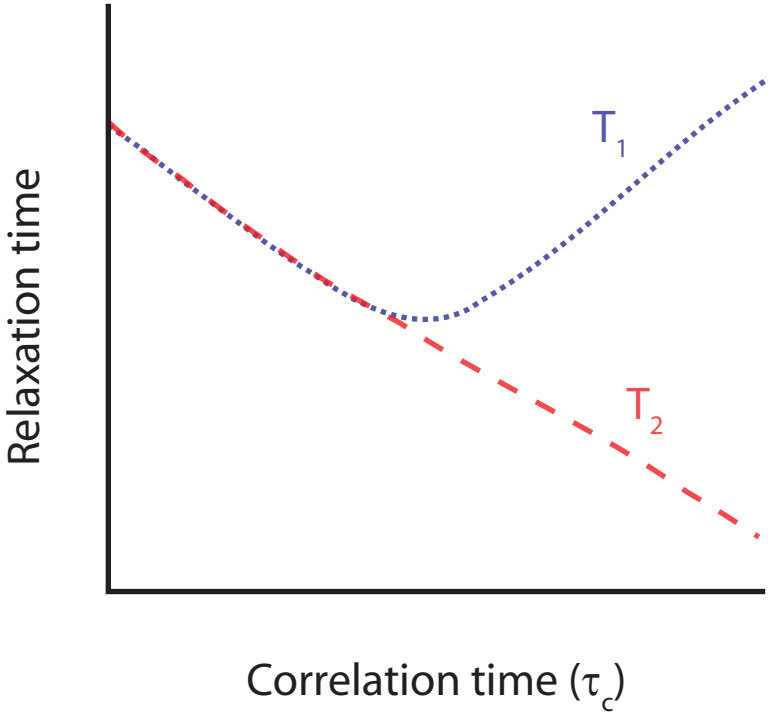

Figure 3.2. Schematic representation of T_1 (blue) and T_2 (red) relaxation times as a function of correlation time (τ_c).

3.3.3 The nuclear Overhauser effect

The nuclear Overhauser effect (NOE) arises as a result of the fact that dipolar-coupled spins do not relax independently, and is one of the most important effects in NMR spectroscopy. The NOE is characterized by the cross-relaxation rate constant, which is defined by the steady-state NOE enhancement. The steady-state NOE experiment is most simply understood by using a dipolar-coupled two-spin (A, X) system as an example. If the X spin system is irradiated by a weak radio frequency field (that does not disturb the A spin) for a period of time then the average populations across the transitions are equalized and the A spin magnetization evolves to a steady-state value. Under these conditions the X spins are said to be saturated. The value of the longitudinal magnetization (population difference) for the A spin is altered by saturation of the X spins. If the steady-state NOE enhancement is positive then the longitudinal magnetization across the A spin transitions are increased by reducing the longitudinal magnetization across the X spin transitions.

To measure the value of the NOE enhancement a steady-state NOE difference experiment is performed in which two spectra are recorded. In the first spectrum the X spin is saturated for a period of time sufficient to establish the NOE enhancement of the A spin. Then a 90° pulse followed by a reverse INEPT sequence is applied and the free-induction decay is recorded. In the second spectrum the X spin is not saturated. Instead, a 90° pulse is applied to the system at equilibrium and the free-induction decay is recorded. Thus the NOE enhancement is calculated as the fractional change in the intensity (I) of A on saturating X

$$NOE = \frac{I_{saturated}}{I_{unsaturated}}$$

Typical NOE enhancement values of approximately 0.8 are expected for the ^{15}N nuclei of large molecules with well defined tertiary structure tumbling slowly (on the nano-second timescale) in solution and values significantly lower are expected for nuclei of small molecules tumbling rapidly (on the sub-nanomolar timescale) in solution.

3.4 Dihedral angles and hydrogen bonds

3.4.1 Dihedral angles

Dihedral torsion angle (φ and ψ) restraints were calculated using the program TALOS (Cornilescu et al., 1999), which compares chemical shift assignments with a database of known protein assignments with known secondary structural elements. The implementation of TALOS-derived dihedral angle restraints in structure calculations is described in more detail in section 6.2.1. The dihedral angle restraints used in calculation of Rv1827 pThr 22 solution structures are listed in section 9.3.

3.4.2 Hydrogen bonds

Amide proton exchange was monitored by acquiring successive ^{15}N-HSQC spectra following dissolution of lyophilized protein in 100% $^{2}H_2O$ containing 20 mM sodium acetate pH 5.8 and 50 mM NaCl. The first ^{15}N-HSQC spectrum was acquired 5 minutes after dissolution of the lyophilized protein and showed the presence of 79 backbone amide protons that had not yet exchanged (Figure 3.1). After 17 hours only 49 backbone amide protons had not exchanged from which it was inferred that these amides were

engaged in hydrogen-bonds, confirmed by appropriate H_α-H_α correlations in the ^{13}C-edited NOE spectra. The implementation of hydrogen bond restraints in structure calculations is discussed in more detail in section 6.2.2. The hydrogen bond restraints used in calculation of Rv1827 pThr 22 solution structures are listed in section 9.4.

3.5 Structure determination by NMR

3.5.1 NOESY spectra

All NOESY spectra were acquired, processed and analyzed as previously outlined. Distance constraint assignment was performed following analysis of three-dimensional ^{1}H-^{15}N-NOESY-HSQC with a mixing time of 120 ms and mixing times of 100 and 120 ms for ^{1}H-^{13}C-NOESY-HSQC spectra acquired in 100% $^{2}H_2O$ and 90% H_2O/10% $^{2}H_2O$, respectively.

The ^{1}H-^{13}C-NOESY-HSQC spectra were analyzed and annotated using an assigned HCCH-TOCSY peaklist as a template, effectively representing intra-residue NOEs from which all other observable inter-residue NOESY peaks were picked. The ^{1}H-^{15}N-NOESY-HSQC spectra were analyzed and annotated using an assigned H(CCCO)NH peaklist as a template, which was used to represent intra-residue NOEs, from which other observable inter-residue NOESY peaks were identified and analysed.

The secondary structure of FHA domains is almost entirely composed of β-sheets, consequently NOESY spectral assignment began with identification of H_α-H_α NOEs that defined the register of β-sheet-forming residues. From these short and medium range assignments, the appropriate hydrogen bonds were inferred from the amide proton

exchange data. This procedure is described in more detail in section 6.2.2. Long-range NOEs with unique assignment possibility were also identified at this stage.

3.5.2 Distance calibration

Distance calibration involves converting peak intensities or volumes into upper-distance limits between atoms. As both peak intensity and volume are directly proportional to one another (assuming that the line-widths of all peaks are identical), either may be used for peak calibration.

Peak intensities were directly measured in XEASY and converted into upper distance constraints using the awk script 'distance_calibration.csh' (section 9.2), which categorized peak intensities into strong/medium/weak, in accordance with known distances in secondary structural elements (Wuthrich, 1986). Peak overlap was resolved on an individual basis usually through comparison of the equivalent regions of the 90% H_2O/10% ^2H_2O and 100% ^2H_2O spectra. Pseudoatom correction was added to all cases where there was chemical shift redundancy for stereo-pairs that were not stereo-specifically assigned.

3.5.3 ARIA 1.2

ARIA 1.2 was used to calculate the structure of Rv1827 pThr 22 and the model of the unphosphorylated form (Linge et al., 2003). Distance restraints from the ^1H-^{15}N-NOESY-HSQC and ^1H-^{13}C-NOESY-HSQC were included for structure calculation. Also incorporated were hydrogen bond restraints (section 9.4) from the ^2H exchange experiment and dihedral angles derived from TALOS (section 9.3). Crystallography and NMR System (CNS) was used to perform simulated annealing combined with molecular

dynamics protocols (in torsion angle space for structural calculations and Cartesian space for water refinement) (Brunger et al., 1998). The parameters used for the simulated annealing and water refinement protocols are shown in section 9.6.

3.6 NMR titrations

The NMR titration was performed by recording a ^{1}H-^{15}N HSQC spectrum after each addition of phosphopeptide. The results of the phosphopeptide titration are discussed in section 5.4.5. The concentrations of the phosphopeptide and Rv1827 throughout the titration are summarized in Table 3.2. Spectra were recorded at 25^{o}C using Rv1827 dissolved in the buffer described in section 2.3.9. The phosphopeptide was dissolved in 20 mM sodium acetate (pH 5.8) and 50 mM NaCl. The sequence of the phosphopeptide was:

E-V-T-V-E-T-pT-S-V-F-R-A-D-[Y-K-K]

In the above sequence, the phosphothreonine residue is denoted by pT and the C-terminal Y-K-K tag was added to increase the solubility of the phosphopeptide and so that its concentration could be determined by UV spectrophotometry.

Titration number	Peptide (μM)	Rv1827 (μM)	molar equivalent (peptide:Rv1827)	Total vol. (μl)
0	0.0	251.0	0.00	500
1	7.9	248.0	0.03	505
2	23.4	244.0	0.10	515
3	45.5	237.0	0.19	530
4	72.9	228.0	0.32	550
5	104.6	218.0	0.48	575
6	145.0	206.0	0.70	610
7	185.3	193.0	0.96	650
8	249.1	173.0	1.44	725
9	343.2	143.0	2.40	875
10	422.4	120.0	3.52	1050

Table 3.2. NMR titration of pThr 22 peptide into unphosphorylated Rv1827.

4 A role for Rv1827 in the regulation of glutamate metabolism

4.1 M. tuberculosis represents pathogencity on a global scale

In 1993 the World Health Organisation (WHO) declared tuberculosis as a global emergency. It is currently recognized as a worldwide pandemic with approximately 2 billion people infected with *M. tuberculosis,* the causative agent of tuberculosis. Tuberculosis is also a leading killer among HIV-infected people with weakened immune systems: approximately 200,000 people living with HIV/AIDS die from tuberculosis every year. With the emergence of multiply and extensively drug-resistant tuberculosis strains in all countries surveyed by the WHO, the identification of novel drug targets and therapies has never been more important.

Tuberculosis in humans is a complex disease caused by bacterial populations located in discrete lesions, or micro-environments (Boshoff and Barry, 2005). Some types of micro-environments are conducive to bacterial replication, whereas others restrict their growth without sterilising the infecting population. Within pulmonary macrophages and human granulomas *M. tuberculosis* is exposed to a number of host-generated chemical stresses including diminished oxygen (hypoxia) and increased nitric oxide (NO) concentrations (Gomez and McKinney, 2004).

4.2 Survival of M. tuberculosis within the human macrophage

M. tuberculosis is able to respond to the microaerophilic environment of the human macrophage through the *dos*RST system (Sardiwal et al., 2005; Shiloh et al., 2008). This is a classical two-component regulatory system comprising of a transmembrane sensor histidine kinase (DosS/T) and a response regulator (DosR). In response to hypoxia and

environments with elevated NO, DosS/T undergoes autophosphorylation and hence activation. In the activated state DosS phosphorylates an Asp residue of the response regulator DosR and in doing so causes a signal about the extracellular environment to be transmitted into the cell (Sousa et al., 2007).

DosR/T are transcription factors that once activated induce the transcription of approximately 50 genes, known as the dormancy regulon (Shiloh et al., 2008). The function of many of the transcribed genes remains unknown. Those that have been characterized include an electron transporter (ferredoxin, *fdxA*) (Ricagno et al., 2007), a heat shock protein (α-crystallin, *hspX* or *acr*) (Hu et al., 2006), and an operon associated with nitrite and nitrate reduction (*narK2X*) (Shiloh et al., 2008; Sohaskey and Wayne, 2003). Research has shown that *in vitro* stimulation via hypoxia and NO initiates a metabolic shift within cultured bacilli to a state that recapitulates the latent phase of mycobacterial infection (Wayne and Sohaskey, 2001). Understanding the nature and regulation of metabolic changes associated with the shift from actively replicating to dormant populations of *M. tuberculosis* may in turn lead to the identification of anti-mycobacterial therapeutic targets.

4.2.1 Survival-promoting adaptations of M. tuberculosis

Adaptations resulting in such metabolic flexibility are thought to involve the *M. tuberculosis* core intermediary metabolism and specifically the tricarboxylic acid (TCA) cycle. The TCA cycle is an essential biochemical pathway within cellular metabolism, yielding reducing equivalents for energy generation and biosynthetic reactions as well as precursors for lipids, amino acids and heme.

Many microorganisms use unique or alternative TCA cycle reactions that appear to be required to maintain TCA cycle functionality where expected enzyme reactions (as defined by the presence or absence of the corresponding genes or demonstrable enzymatic activity) are "missing" (Cordwell, 1999). *M. tuberculosis* is no exception and lacks an α-ketoglutarate dehydrogenase (KDH) activity. Ordinarily this would generate succinyl-CoA and consequently succinate from α-ketoglutarate (Tian et al., 2005a; Tian et al., 2005b).

M. tuberculosis has substituted KDH activity for an α-ketoglutarate decarboxylase (KGD) activity that in concert with the succinic semi-aldehyde dehydrogenase capacity of Rv0234c/Rv1731 (GabD1/GabD2) can generate succinate oxidatively. The literature on the biochemical pathways in the *M. tuberculosis* TCA cycle also suggests the ability to generate succinate reductively via the glyoxylate shunt. Furthermore a functional γ-aminobutyrate shunt may facilitate the inter-conversion of α-ketoglutarate and γ-aminobutyrate with succinic semi-aldehyde and glutamate (Bott, 2007; O'Hare et al., 2008). Thus the *M. tuberculosis* TCA cycle appears to be able to operate in half cyclic modes, characteristic of organisms adapted to microaerophilic conditions, to generate necessary metabolic intermediates (Tian et al., 2005a).

4.3 Rv1827 and phosphorylation-mediated metabolic regulation

Rv1827 (known also as GarA (glycogen-accumulation regulator A) or CFP17) was originally identified as a *M. tuberculosis* culture filtrate protein (CFP) found as a 162 residue (17.6 kDa) full length protein and an N-terminally truncated form comprising residues 31-162 (Weldingh et al., 1998). Both forms were found to induce protective immunity in animal models (Weldingh and Andersen, 1999).

Rv1827 is highly conserved across the actinomycetes which include the mycobacteria and corynebacteria, but a eukaryotic homologue is absent. In *C. glutamicum,* phosphorylation of its Rv1827 homologue (OdhI) by PknG has been found to regulate intracellular glutamate levels (Bott, 2007; Niebisch et al., 2006; Schultz et al., 2007). It was found that OdhI phosphorylation by PknG resulted in significantly less glutamate catabolism, and hence a high degree of glutamate accumulation within the cell. Deletion of the PknG gene (Δ*pknG*) was found to be sufficient to cause the glutamate utilization defect, as was mutation of the PknG-phosphorylation site on OdhI when PknG was present (Niebisch et al., 2006). However, a Δ*pknG* Δ*odhI* double deletion mutation abolished the glutamate utilization defect and glutamate appeared to be catabolised normally. This suggested that kinase-mediated regulation of intracellular glutamate concentrations is exerted through regulation of OdhI activity. Indeed, the activity of *C. glutamicum* 2-oxoglutarate dehydrogenase (OdhA) was later found to be regulated by phosphorylation of OdhI. OdhA catalyses the dehydrogenation of α-ketoglutarate and leads to the generation of succinate via a succinyl-CoA intermediate. Binding of OdhI to OdhA inhibits the enzyme, and this inhibitory effect can be relieved through OdhI phosphorylation by PknG (Asakura et al., 2007; Schultz et al., 2007).

Sequence alignments predict that Rv1827 harbors a C-terminal FHA domain and an N-terminal extension of approximately 50 amino acids of no apparent predicted structure or function. Full length Rv1827 was found to be the optimal substrate for PknB in mycobacterial cell-free extracts and both Rv1827 and OdhI were found to be phosphorylated by their respective PknG kinases *in vivo* (Niebisch et al., 2006; O'Hare et al., 2008; Villarino et al., 2005). In each case, single phosphorylation sites were mapped to an N-terminal region of Rv1827. PknB specifically phosphorylates Thr 22 and PknG

specifically phosphorylates Thr 21. Phosphorylation of Rv1827 by PknB and PknG were found to be mutually exclusive events (England et al., 2008; Niebisch et al., 2006; O'Hare et al., 2008; Villarino et al., 2005). Interestingly both phosphorylation sites conform to the Rv1827 preferred phospho-binding motif as determined by oriented peptide library screening (Yaffe and Smerdon, 2004). Each phosphorylation site maintains a hydrophobic residue at the pThr + 3 position, suggesting a common mechanism by which signals from two different kinases can be transmitted through a common structural response of Rv1827.

Rv1827 can be used as an *in vitro* substrate for PknD, PknE and PknF, although no functional significance has been attributed to these events (Villarino et al., 2005). It has recently been reported that Rv1827 FHA domain binds to a phosphopeptide encompassing the PknB phosphorylation site of its own N-terminus, with an apparent k_d of 0.385 (±0.003) μM (England et al., 2008). This prompted the authors to suggest that N-terminal phosphorylation of Rv1827 induces an intramolecular interaction through tight recognition of the FHA domain for its phosphorylated N-terminus.

In summary, it seems that OdhI is able to control glutamate metabolism in *C. glutamicum* and Rv1827 presumably functions in an analogous manner in *M. tuberculosis*. An additional level of control is exerted by N-terminal phosphorylation of Rv1827/OdhI. However, it is unclear how Rv1827 or OdhI interact with their respective enzyme targets. This leads to the question of whether the interaction is dependent on the phosphorylation state of the enzymes or can Rv1827/OdhI interact with them phospho-independently. Moreover it is perplexing how phosphorylation of Rv1827 within its N-terminal region abrogates these interactions. As a first step towards addressing these

issues it was important to find out whether Rv1827 interacts only with KGD, or if it has multiple binding partners.

4.4 Identification of Rv1827 binding partners by pull-down assays

In an attempt to characterize Rv1827 binding partners in *M. tuberculosis* cell-free extracts, a GST-Rv1827 fusion protein construct was utilized to perform pull-down experiments. In these experiments Rv1827 was found to bind specifically and reproducibly to three high molecular weight proteins that were identified by MALDI-TOF mass spectrometry (Figure 4.1 left, lane 3) (S. L. Westcott, PhD Thesis 2004). Two of these, Rv1248c (α-ketoglutarate decarboxylase; KGD) and Rv2476c (NAD^+-specific glutamate dehydrogenase 4; GDH-4) have recently been reported as inhibitory targets of Rv1827 (Niebisch et al., 2006; O'Hare et al., 2008). KGD has a mass of 134 kDa and GDH has a mass of 177 kDa. The N-terminus of KGD is homologous to the acyltransferase domain of the E2 subunit of the α-ketoglutarate dehydrogenase complex, while the C-terminus is homologous to the E1 subunit of the same complex. Rv3859c (α-subunit (GltB) of the glutamate synthase complex, GS) was additionally identified as a novel binding partner (S. L. Westcott, PhD Thesis 2004). This protein has a mass of 166 kDa. Intriguingly all three enzymes utilize α-ketoglutarate as substrate, placing Rv1827 activity squarely in the context of core metabolism, as hinted by the work in *C. glutamicum*.

Additional pull-down assays were carried out using PknB-phosphorylated Rv1827, the minimal FHA domain (residues 55-149) and a mutant protein containing an Ala substitution of Ser 95 (S95A). Ser 95 is a highly conserved FHA domain residue whose mutation is widely reported to abrogate pThr-motif binding. Whilst interactions with all three enzymes are apparently unaffected by removal of the N-terminal 54-residues, they

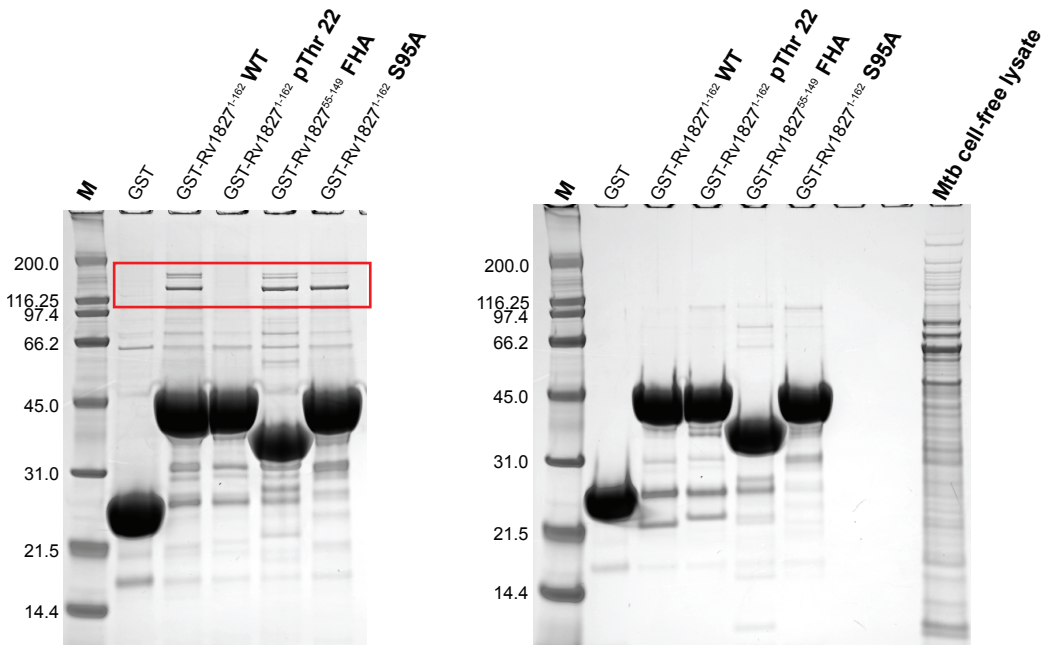

Figure 4.1. Identification of Rv1827 binding partners. left **Pull-down gel showing that Rv1827 wild-type (WT) binds to three high molecular weight proteins (lane 3). Phosphorylation of Rv1827 at Thr 22 by PknB (pThr 22) abrogates Rv1827 interactions (lane 4). The Rv1827 FHA domain is sufficient for binding to the high molecular weight proteins (lane 5). Mutation of the conserved Rv1827 FHA domain Ser 95 to Ala (S95A) causes loss of interactions with GltB, reduces interactions with GDH and maintains interactions with KGD at wild-type levels.** right **Reference gel showing the purified GST-fusion constructs used in the pull-down experiments. Gel loading is at approximately the same level as in the pull-down experiments. The location of the three metabolic enzymes is indicated with a red box.**

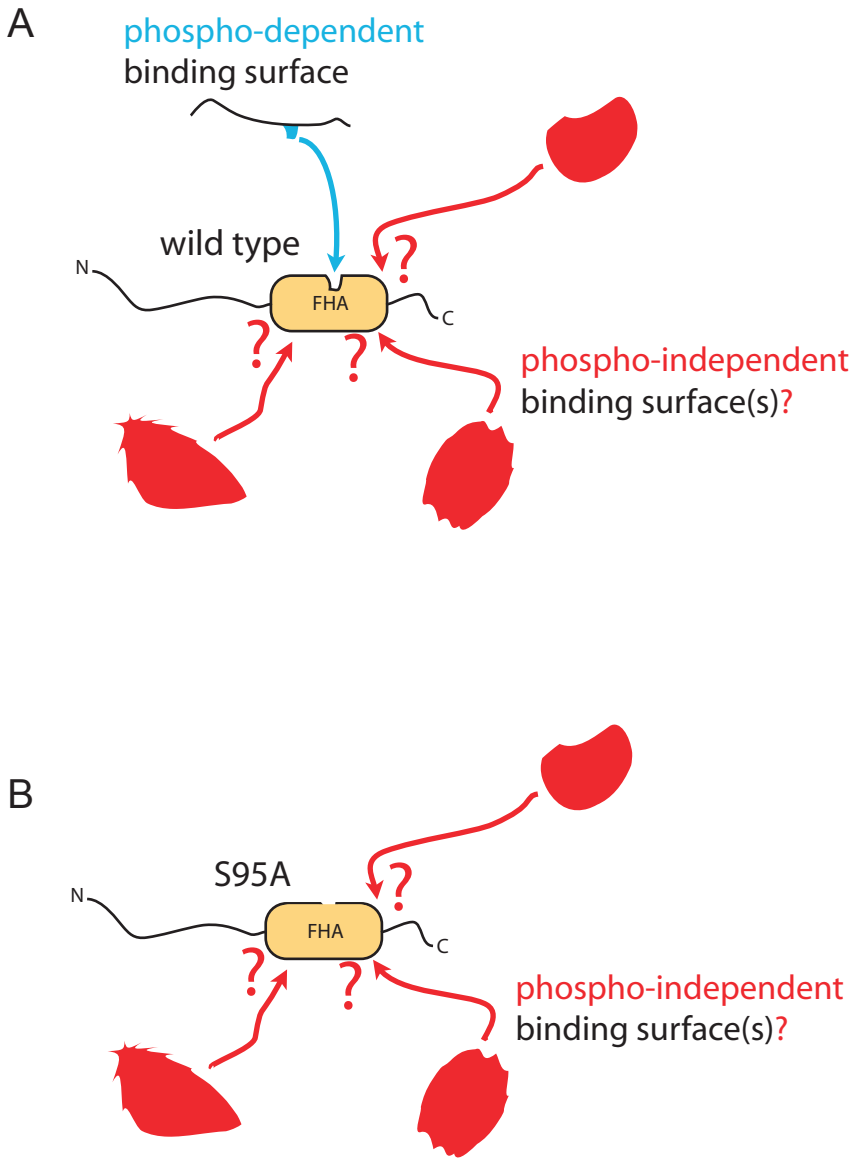

Figure 4.2. Schematic representation of possible phospho-dependent and phospho-independent Rv1827 FHA domain interactions. (A) The canonical phosphopepitde-binding surface is represented with a 'notch' in the FHA domain. The phospho-dependent ligand is represented in black and blue, and phospho-independent binding partners are represented in red. (B) The S95A mutant can no longer bind phospho-dependently, but can still potentially use other FHA domain surfaces for phospho-independent interactions.

are abolished by Thr 22 phosphorylation (Figure 4.1 left, lanes 5 and 4). Thus Rv1827 FHA domain alone is sufficient for binding its metabolic enzyme targets and phosphorylation of the N-terminal region completely blocks these interactions. Furthermore, the phospho-binding deficient S95A mutant eliminates interaction with GltB, whilst association with GDH is reduced and binding to KGD is unaffected (Figure 4.1 left, lane 6). The reduced ability of Rv1827 S95A to bind GDH and GltB is suggestive of classical pThr-dependent interactions, yet phosphorylation sites on KGD, GDH and GltB have not been found (Nott et al., 2009). These experiments show that PknB-phosphorylation of Rv1827 regulates interactions of its FHA domain with three functionally related TCA cycle enzymes, and that association with at least two, and most likely all three, involves a phospho-independent FHA binding function. This is represented in the bottom panel of Figure 4.2 showing that potential Rv1827 phospho-independent binding surfaces may be distinct from the canonical pThr binding site or partially coincident with it.

4.5 Rv1827 modulates glutamate and a-ketoglutamate metabolism and ammonia assimilation

O'Hare and co-workers have recently tried to characterize the effect that Rv1827 has on KGD and GDH activity. They used recombinantly expressed KGD and GDH in an *in vitro* coupled-assay reaction system that used ferricyanide reduction as the final readout for the activity of the metabolic enzymes. In this way they made measurements of the activity of the enzymes in the presence of increasing concentrations of Rv1827. Their *in vitro* ferricyanide reductase assays showed that Rv1827 acts as a potent inhibitor of each enzyme. Recombinant KGD purified from *E. coli* was found to have a maximal ferricyanide reductase activity of 1.8 s^{-1} and a K_m for α-ketoglutarate of 480 μM, which is

similar to that observed for the recombinant *M. smegmatis* homologue MSMEG_4699 (1.6 s^{-1} and a K_m for α-ketoglutarate of 540 μM). Concentration-dependent inhibition of KGD was observed upon addition of Rv1827, but not PknB-phosphorylated Rv1827 (O'Hare et al., 2008; Tian et al., 2005b).

Interestingly Rv1827 was found to affect the affinity of GDH for its substrate and thus the reaction rates of both its anabolic and catabolic reactions. Thus whether Rv1827 acts as an inhibitor or an activator *in vivo* depends on substrate concentration. If it is assumed that α-ketoglutarate concentrations are similar to those estimated for *E. coli* (0.1-0.9 mM) and *C. glutamicum* (1 mM) and that the concentration of glutamate is below 600 mM, then Rv1827 would inhibit GDH activity and reduce the rate of the low-affinity route of ammonia assimilation from α-ketoglutarate (O'Hare et al., 2008). Recent data show that Rv1827 is a strong activator of glutamate synthase (GS) activity in *M. smegmatis* and *M. Bovis* cell-free lysates (Nott et al., 2009). The combined activity of the GS complex and glutamine synthetase (GlnA) forms the high-affinity pathway of ammonia assimilation from α-ketoglutarate.

Since KGD, GDH and the GS complex all utilize α-ketoglutarate as a substrate, these combined data suggest that Rv1827 exerts control of carbon flux and nitrogen assimilation at an important branch of central metabolism. Either α-ketoglutarate can be oxidized in the TCA cycle by KGD or be converted into glutamate. This may be achieved either by GDH (under conditions of nitrogen abundance) or the combined activities of the GS complex and GlnA under conditions of nitrogen limitation. In all cases, Rv1827 regulatory activity is lost upon its phosphorylation by PknG or PknB (Nott et al., 2009; O'Hare et al., 2008).

4.5.1 Interactions are dependent on Rv1827 phosphorylation status

Regardless of the differential effects Rv1827 has on the three α-ketoglutarate utilizing enzymes, it is remarkable that a single phosphorylation event can regulate such complex core metabolic activity. In order to understand the differences between Rv1827 and the PknB-phosphorylated form through biophysical and biochemical analytical methods it was necessary to be able to generate them both in large quantities and at high purity. In this respect the stoichiometric phosphorylation Rv1827 at Thr 22 by PknB was extremely fortunate. This meant that large quantities of Rv1827 pThr 22 could be generated and, under the appropriate reaction conditions, none of the unphosphorylated form of Rv1827 would be left in the sample. Thus a batch of unphosphorylated Rv1827 could be split and half of it treated with PknB so that accurate comparisons could be made between the two forms.

Figure 4.3 shows that upon phosphorylation by PknB the mass of Rv1827 increases from 17629.82 ± 0.53 Da to 17710.46 ± 0.48 Da. The increase in mass of 80.64 Da corresponds to the addition of one phosphate per molecule. Importantly, none of the unphosphorylated species is present in the sample of Rv1827 pThr 22.

4.6 Conclusions

M. tuberculosis possesses a flexible core metabolism that appears to be able to function oxidatively or reductively depending on environmental pressures. Furthermore, signaling through Ser/Thr protein kinases is able to direct carbon flux in the TCA cycle and pathways of nitrogen assimilation to meet cellular needs for energy and biosynthetic precursors. The small FHA domain-containing protein Rv1827 differentially regulates

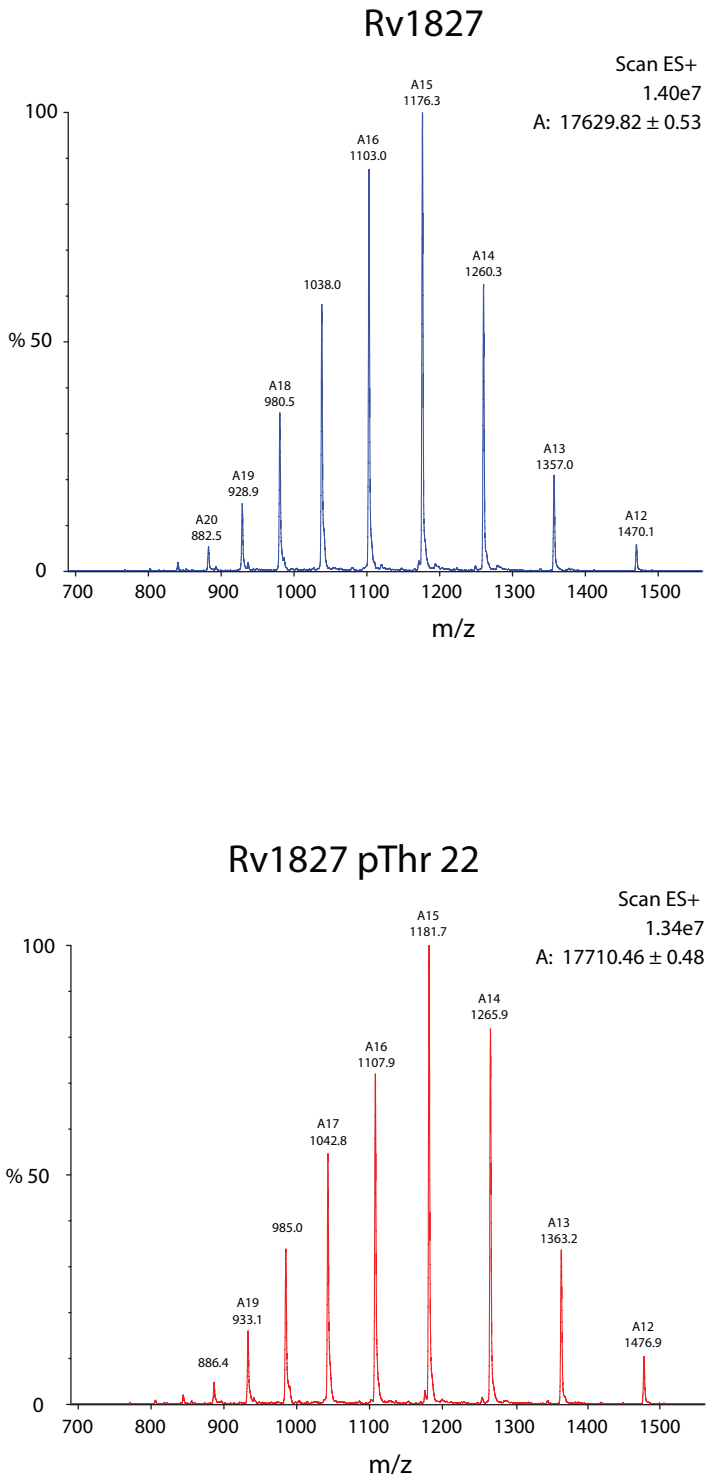

Figure 4.3. Rv1827 is stoichiometrically phosphorylated on one residue by PknB. Electro-spray mass spectra of Rv1827 (top, blue) and Rv1827 pThr 22 (bottom, red). Upon phosphorylation by PknB the mass of Rv8127 increases from 17629.82 ± 0.53 to 17710.46 ± 0.48 Da. An increase in mass of 80.64 Da is correlated with the addition of a single phosphate group per molecule.

the activities of three α-ketoglutarate utilizing metabolic enzymes. The isolated FHA domain suffices for inhibition of KGD and GDH and activation of the GS complex. Interestingly, a mutation (Ser 95 to Ala) that has been widely used to demonstrate FHA domain exclusivity for interaction with phosphopeptides does not affect Rv1827 interactions with at least one of its target enzymes. This strongly suggests that the FHA domain possesses phospho-independent binding capability.

Remarkably, phosphorylation of one of two adjacent threonines in the N-terminus of Rv1827 by PknG or PknB abrogates binding of the FHA domain to its metabolic targets. Phosphorylation by PknG (at Thr 21) and PknB (at Thr 22) are mutually exclusive events. Therefore, phosphorylation of Rv1827 at either threonine induces a conformational change that renders the adjacent one inaccessible to the other kinase. Furthermore the observation that N-terminal phosphorylation disrupts Rv1827 FHA domain-mediated interactions with its target metabolic enzymes suggest that an intra-molecular interaction is triggered that occludes the interaction surface.

Rv1827 that is stiochiometrically phosphorylated at Thr 22 by PknB can be generated with high purity and at high yield. This permitted rigorous biochemical, biophysical and structural investigation of the differences of these two forms of Rv1827 with the aim of understanding the mechanism by which N-terminal phosphorylation controls modulation of carbon flux in the *M. tuberculosis* TCA cycle.

5 Phosphorylation-induced conformational change of Rv1827

5.1 Overview

In seeking a molecular explanation for the observation that N-terminal phosphorylation prevents Rv1827 FHA domain interactions with its metabolic target enzymes a biochemical, biophysical and hydrodynamic study was conducted on unphosphorylated Rv1827 and the PknB-phosphorylated form. The experiments aimed to characterize the postulated phosphorylation-induced conformational change prior to structural analysis.

5.2 Phosphorylation confers stability

5.2.1 Limited proteolysis

The first method that was used investigate structural stability was limited tryptic proteolysis. Trypsin is a protease that can cleave the peptide backbone of proteins C-terminal of Arg or Lys residues. Historically, it has been used as a means of identifying domain boundaries in proteins as it cannot cleave the peptide backbone in regions that are engaged in well defined secondary or tertiary structure. It therefore targets flexible regions. Following limited typtic proteolysis, cleavage sites can be identified from the masses of regions resistant to cleavage or by N-terminal sequencing by Edman degradation.

Simulated tryptic proteolysis of Rv1827 performed using the online ExPASY PeptideCutter tool (http://www.expasy.org/tools/peptidecutter/) identified 12 possible cleavage sites from the amino acid sequence. 9 sites lie within the predicted FHA domain (residues 55-149) and are likely to be protected from tryptic cleavage as a result of the tight, ordered packing of residues typically found in FHA domains. One lies at the

extreme C-terminus of the FHA domain (Lys 151) and two potential tryptic sites are N-terminal of the FHA domain at Lys 10 and Arg 26.

When subjected to limited trypsin proteolysis unphosphorylated Rv1827 is efficiently cleaved at Arg 26, which is four/five residues C-terminal of the PknB/G phosphorylation sites (Figure 5.1, left). As expected the FHA domain is resistant to cleavage by trypsin. For Rv1827 pThr 22 the Arg 26 tryptic site is almost completely protected, indicating that upon phosphorylation a conformational change occurs in this region of the molecule and results in its stabilization. (Figure 5.1, right). However, even in the PknB-phosphorylated form, trypsin is able to cleave after Lys 151. This site lies two residues downstream of the last residue in the most C-terminal β-strand of the FHA domain.

Interestingly, when Rv1827 pThr 21 (phosphorylated by PknG) is subjected to limited trypsin proteolysis the Arg 26 tryptic site is protected, albeit to a lesser degree than for Rv1827 pThr 22 (Figure 5.1, middle). A possible explanation for this difference may be that in the pThr 21 form Arg 26 is effectively shifted away from the site of phosphate binding by approximately 3.4 Å (the distance between peptide bonds in a polypeptide chain), leaving it slightly less restrained and consequently more accessible to trypsin proteolysis. Cleavage at Lys 151 is observed in Rv1827 pThr 21 at a similar efficiency as in Rv1827 pThr 22, demonstrating that differential N-terminal phosphorylation has no effect of the conformation of the protein C-terminal of the FHA domain.

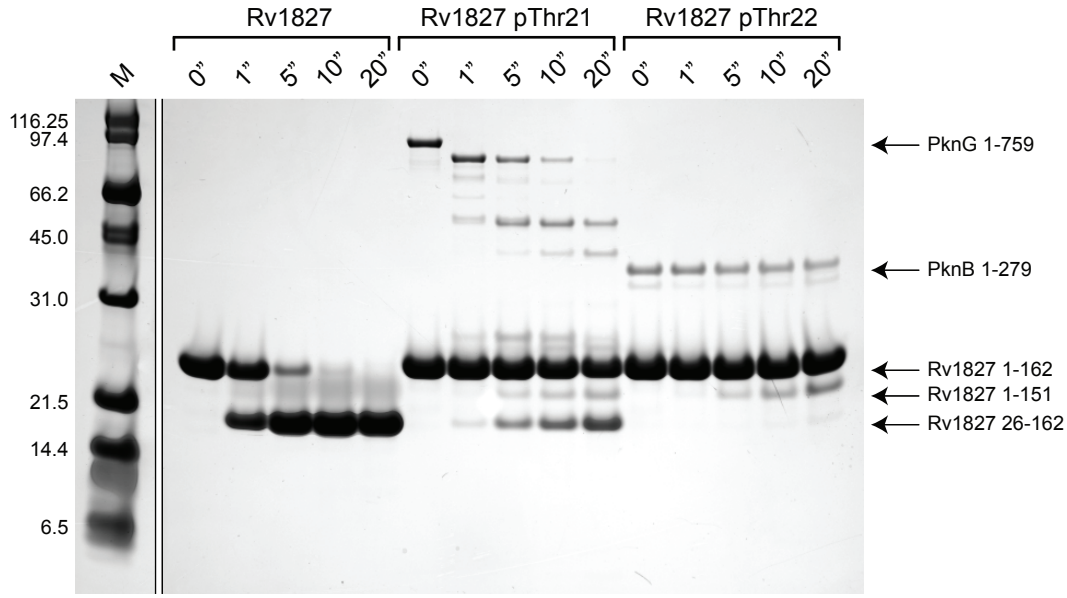

Figure 5.1. Limited tryptic proteolysis of unphosphorylated and PknG- and PknB-phosphorylated Rv1827. Molecular weight markers (M) are displayed in kDa. Unphosphorylated Rv1827 is cleaved by trypsin at Arg 26 (determined by mass spectrometry) within approximately 10 min. The rate of trypsin cleavage at Arg 26 is reduced in the PknG-phosphorylated form (Rv1827 pThr 21), and when Rv1827 is phosphorylated by PnkB (Rv1827 pThr 22), Arg 26 is protected from tryptic proteolysis. The limited tryptic cleavage profiles of PknG and PknB can be seen as higher molecular weight bands on the gel.

5.2.2 Circular dichroism

Another way of probing the stability of a protein is through thermal (or chemical) denaturation of the secondary structure. This process can be followed using circular dichroism (CD) spectroscopy. Amino acids have a chiral centre about their respective C_α atoms (except Gly) that differentially transmits left- and right-circularly polarized light if the amino acid is in a region of β-strand, α-helix or random coil. Therefore the heating-induced transition between the folded protein (with secondary structure) and unfolded polypeptide chain can yield information about the thermal stability of the protein.

By following the CD thermal unfolding profiles at 202 nm it was found that PknB-phosphorylation of Rv1827 at Thr 22 confers a significant increase in thermal stability compared with the unmodified form (Figure 5.2 A). Whereas unphosphorylated Rv1827 was found to be completely unfolded at 60°C, Rv1827 pThr 22 had to be heated to 80°C before the secondary structure was completely denatured. Thus phosphorylation of the N-terminal region affects stability across the entire protein. A likely explanation for this is that the N-terminal region interacts tightly with the FHA domain, as recently suggested by England and co-workers (England et al., 2008).

Rv1827 produces a far-UV CD spectrum characteristic of a protein rich in β-strands. However, comparison of the far-UV CD spectra of Rv1827 and Rv1827 pThr 22 shows that phosphorylation induces a small but significant change in CD signal at approximately 202 nm (Figure 5.2 B, left). This shift in signal may be correlated with the formation and stabilization of a short region of secondary structure with α-helical propensity that is not present in the unmodified protein. However, as peaks in the near-

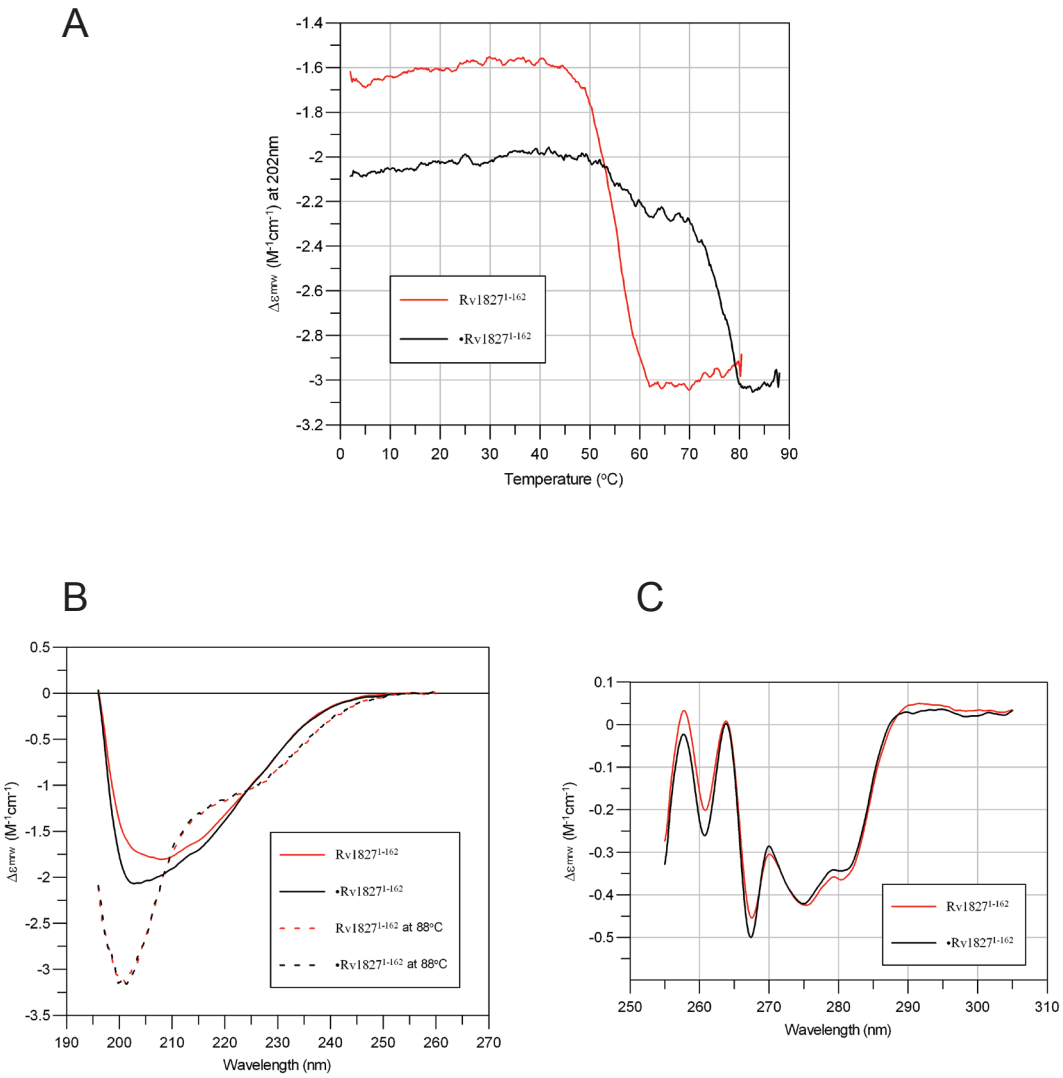

Figure 5.2. Biophysical comparison of Rv1827 and Rv1827 pThr 22. (A) Thermal denaturation of Rv1827 (red lines) and Rv1827 pThr 22 (black lines) measured using circular dichroism (CD) at 202 nm. Whereas unphosphorylated Rv1827 is completely denatured by 60°C, Rv1827 maintains some secondary structure up to 80°C at which temperature it is completely denatured. (B) left Far-UV CD spectra of Rv1827 (red) and Rv1827 pThr 22 (black) at room temperature (solid lines) and at 88°C (dashed lines). The coincidence of the spectra of the two forms of Rv1827 at 88°C indicates that measurements were taken on samples of the same concentration. right Near-UV CD spectra of Rv1827 and Rv1827 pThr 22. The coincidence of peaks in the spectra of the respective samples indicates that residues contributing to the near-UV CD spectra (Phe at 261 nm; Tyr at 275 and 282nm) were unaffected by phosphorylation-induced conformational change

UV CD spectra are coincidental for both phosphorylated and unphosphorylated forms of Rv1827 we concluded that residues contributing to the near-UV CD spectra (Phe at 261 nm; Tyr at 275 and 282nm) were unaffected by phosphorylation-induced conformational change (Figure 5.2 B, right).

5.3 Phosphorylation induces a conformation compaction

In order to understand the stabilizing effect that phosphorylation confers on Rv1827 in more detail, hydrodynamic parameters of the unphosphorylated and PknB-phosphorylated forms were measured. It was initially hypothesized that N-terminal phosphorylation would induce 'head to tail' dimerisation whereby the phospho-threonine of one molecule would bind to the FHA domain of another, as seen for the Chk2 FHA domain. This mode of FHA binding has been shown to play important roles in the activation of kinases possessing FHA domains. In this model dimerisation through FHA domain-mediated interactions brings the kinase domains into close proximity promoting auto-phosphorylation *in trans*.

5.3.1 Multi-angle laser light scattering

Multi-angle laser light scattering (MALLS) is a technique that provides a direct measure of absolute molecular mass by coupling size exclusion chromatography eluents directly to static light-scattering and concentration detectors. Since the signal from light-scattering detectors is directly proportional to the molecular mass of the protein times the concentration (mg ml^{-1}), combining this signal with that from a concentration detector (refractive index or absorbance), determination of the molecular mass of each peak eluting from the size exclusion chromatography column is possible.

The mean weight-averaged molecular weight of Rv1827 and Rv1827 pThr 22 were determined by size exclusion chromatography coupled to MALLS to be 17700 g mol^{-1} (±1%) and 17500 g mol^{-1} (±1%), respectively (Figure 5.3 A). Therefore both Rv1827 and Rv1827 pThr 22 (which have formula masses of 17629 and 17709 Da, respectively) are monomeric in solution and phosphorylation does not promote Chk2-like dimerisation. Interestingly, the chromatographic peak of Rv1827 pThr 22 is retarded by approximately thirty seconds compared with the unphosphorylated protein, suggesting that phosphorylation induces a conformational compaction. However, it cannot be discounted that the addition of a phosphate group caused Rv1827 pThr 22 to interact with the matrix of the chromatography column, thus retarding its retention time.

5.3.2 Dynamic light scattering

The second hydrodynamic technique that was employed was dynamic light scattering (DLS). This technique can show whether a protein sample is homogeneous or heterogeneous by measuring its polydispersion. Both Rv1827 and Rv1827 pThr 22 were found to be monodisperse in solution, indicating their respective purity and homogeneity. DLS can also be used to measure the hydrodynamic radius of a protein and thereby its translational diffusion coefficient (D_T). The hydrodynamic radius of a protein is effectively a measure of the distance along its longest axis, and is therefore related to how quickly it can diffuse through the solution. A small, globular protein will have a small effective hydrodynamic radius and therefore a high rate of diffusion. Conversely a large or elongated protein will have a high effective hydrodynamic radius and will be able to diffuse less quickly through the solution.

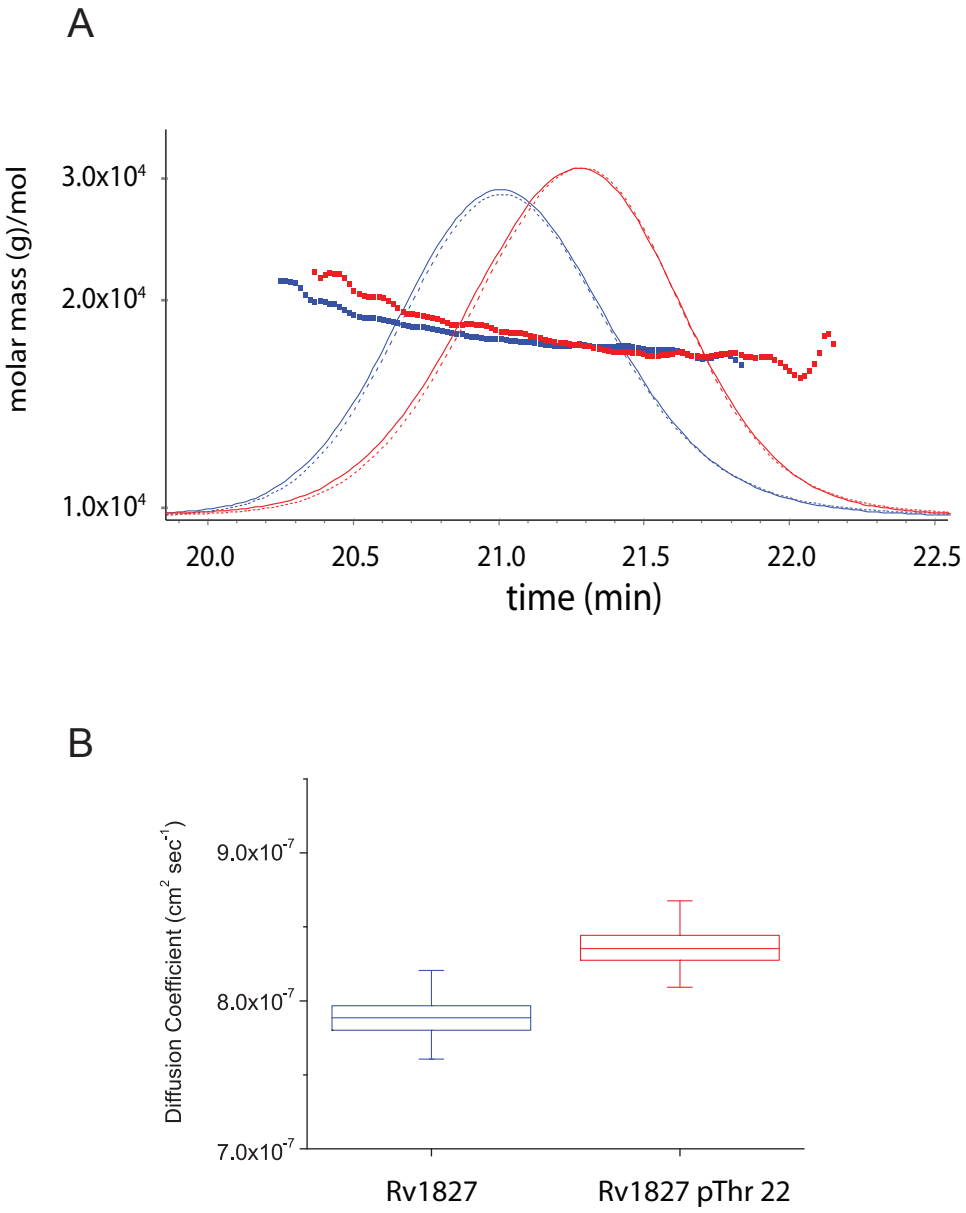

Figure 5.3. Hydrodynamic analysis of Rv1827 and Rv1827 pThr 22. (A) Comparison of the chromatograms (measured at 280 nm) of Rv1827 (blue) and Rv1827 pThr 22 (red) by size-exclusion chromatography coupled to multi-angle laser light scattering (SEC-MALLS). The weight-average molecular weight of material in the chromatographic peaks of the respective forms of Rv1827 are shown as square dots. The differential refractive index (dashed lines) and scatter at 90° (solid lines) closely coincide for each sample, indicating their respective homogeneity. The mean weight-average molecular weight of Rv1827 and Rv1827 pThr 22 were determined to be 17700 g mol^{-1} (± 1%) and 17500 g mol^{-1} (± 1%), respectively, indicating that both unphosphorylated and phosphorylated forms are monomeric in solution. (B) Comparison of the translational diffusion coefficients (D_T) of Rv1827 and Rv1827 pThr 22, obtained by dynamic light scattering (DLS). Both samples were at 250 µM. Upon phosphorylation, the D_T of Rv1827 increased from 7.90 x 10^{-7} cm^{-1} sec^{-1} to 8.4 x 10^{-7} cm^{-1} sec^{-1} implying that phosphorylation induces an intramolecular conformational compaction of the molecule.

Comparison of the translational diffusion coefficients of the two forms of Rv1827 shows a small but significant increase for Rv1827 pThr 22 (8.4 x 10^{-7} cm^{-1} sec^{-1}) compared with the unphosphorylated form (7.90 x 10^{-7} cm^{-1} sec^{-1}) (Figure 5.3 B). Again this implies that phosphorylation induces an intramolecular conformational compaction of the molecule. It was therefore reasoned that upon phosphorylation of Rv1827 at Thr 22 by PknB (and by inference Thr 21 phosphorylation by PknG) an intramolecular association is promoted whereby residues of the N-terminus become tightly associated with the FHA domain conferring proteolytic protection, increased thermal stability, and a more compact global conformation

5.4 NMR-based hydrodynamic analysis of unphosphorylated and phosphorylated Rv1827

Knowing that both the unphosphorylated and PknB-phosphorylated states of Rv1827 were monomeric and monodisperse in solution meant that NMR could be used to further investigate their respective hydrodynamic properties. Whereas the techniques described in the previous section are informative and relatively labor unintensive, they are inherently of low resolution. They could not provide unambiguous explanations for the differences in biochemical or biophysical properties that phosphorylation induced in Rv1827. However, NMR can be used to analyze the hydrodynamic properties of a protein at atomic resolution and therefore address these questions directly.

5.4.1 ^{1}H-^{15}N HSQC assignment and shift-mapping

Backbone amide resonance assignment was initially used as a means of determining regions of the molecule over which changes to the local chemical environment were

induced following phosphorylation. Assignment of the respective ^{1}H-^{15}N HSQC spectra of Rv1827 and Rv1827 pThr 22 was achieved as described in section 3.2.3 and are shown in Figure 5.4 and Figure 5.5. Both forms of Rv1827 yielded well dispersed ^{1}H-^{15}N HSQC spectra, so their assignment was relatively straight forward.

The Rv1827 pThr 22 ^{1}H-^{15}N HSQC spectrum contained 195 peaks in total from the 167 residues in the protein construct, which has 5 more residues than the gene sequence. The 5 extra residues (Gly-Pro-Leu-Gly-Ser) are present at the N-terminus of the protein and are a consequence of the GST purification protocol (outlined in section 2.3.6). The construct contains 10 proline residues, including the residual N-terminal tag which was left following 3C-cleavage. Out of a possible 157 assignments for backbone amide resonances 156 were unambiguously made. 3 (out of a possible 5) Gln sidechain pairs and 6 (out of a possible 8) Asn sidechain pairs were assigned. Arg HN_{ε} and His HN_{δ} were folded in the ^{15}N-dimension, and are represented as hollow contours in the ^{1}H-^{15}N HSQC spectra. 8 out of a possible 9 Arg HN_{ε} resonances were assigned and both of the His HN_{δ} were assigned. 3 peaks in the Rv1827 pThr 22 ^{1}H-^{15}N HSQC spectrum were unassigned.

From the ^{1}H-^{15}N HSQC spectrum of unphosphorylated Rv1827, 155 out of a possible 157 backbone amide resonances were assigned. However, no sidechain amide group resonance assignments were possible as ^{15}N-NOESY spectra were not analyzed for the unphosphorylated from of Rv1827.

Analysis of chemical shift perturbations upon phosphorylation of Rv1827 was used an initial method of identifying regions of the molecule in which conformational changes

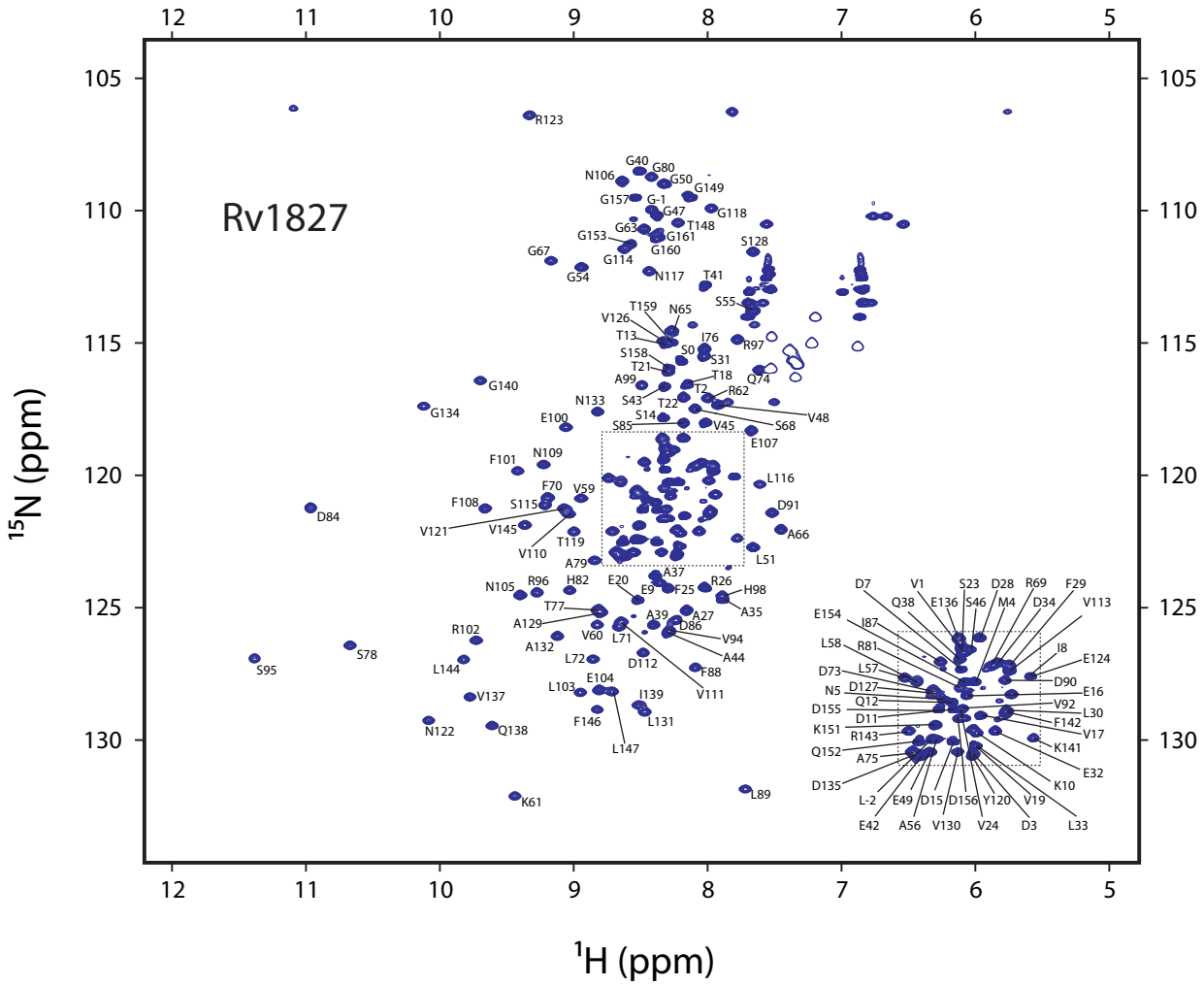

Figure 5.4. ^{1}H-^{15}N HSQC spectrum of Rv1827. The spectrum was acquired at 600 MHz at 25°C in 20 mM sodium acetate (pH 5.8), 50 mM NaCl and 10% $^{2}H_2O$. Peaks folded in the ^{15}N-dimension are shown as hollow contours. The inset at the lower right of the panel corresponds to the crowded region in the center of the spectrum. Unassigned peaks are not annotated.

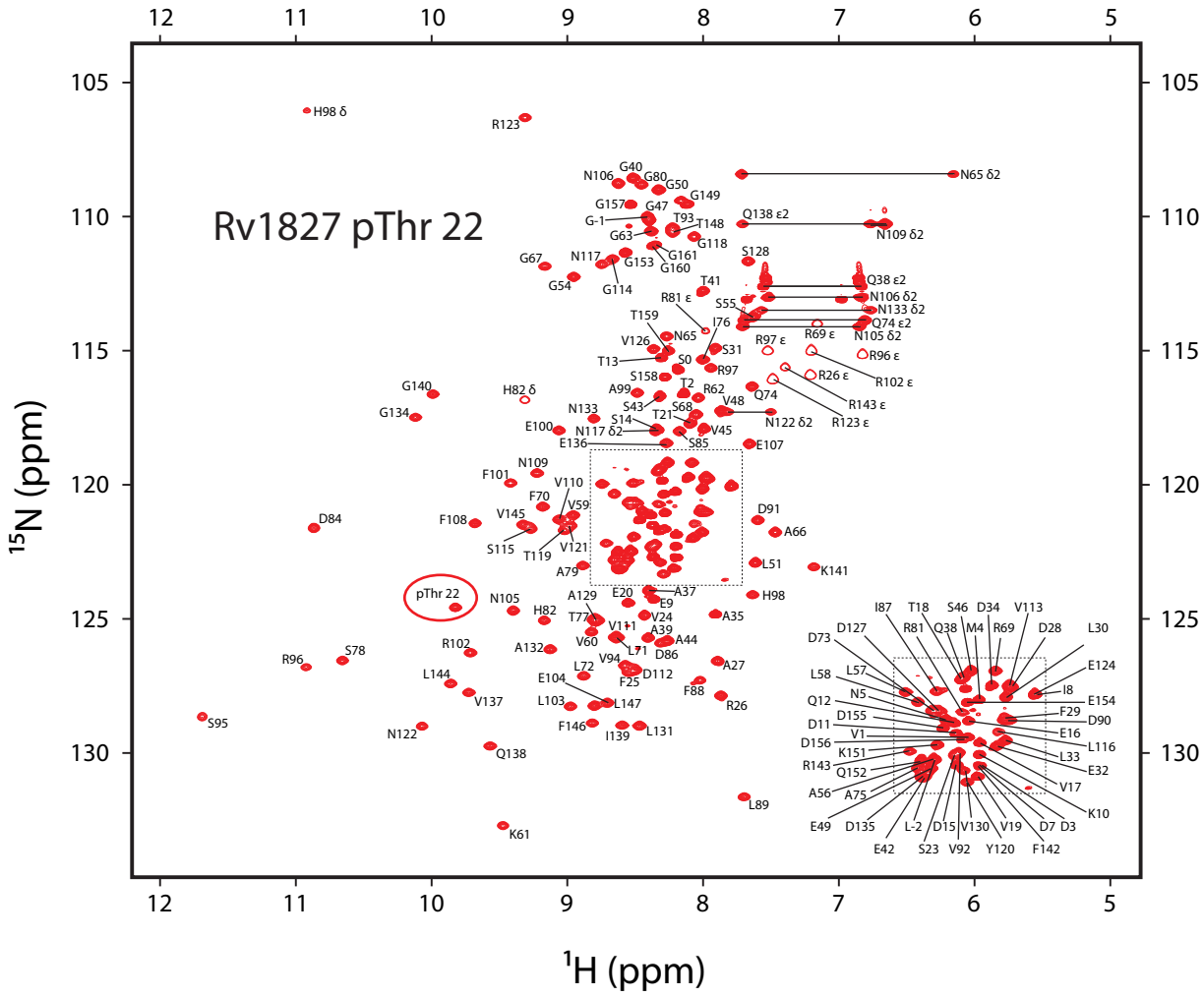

Figure 5.5. ^{1}H-^{15}N HSQC spectrum Rv1827 pThr 22. The spectrum was acquired at 600 MHz at 25°C in 20 mM sodium acetate (pH 5.8), 50 mM NaCl and 10% $^{2}H_2O$. Peaks folded in the ^{15}N-dimension are shown as hollow contours. Assigned δ-NH_2 (asparagine) and ε-NH_2 (glutamine) pairs are linked with horizontal bars. The inset at the lower right of the panel corresponds to the crowded region in the center of the spectrum. pThr 22 is highlighted with a red ellipse. Unassigned peaks are not annotated.

were induced. Figure 5.6 (top panel) shows ^{1}H-^{15}N chemical shift perturbations induced in Rv1827 following phosphorylation of Thr 22, and was calculated according to:

$$\Delta\delta = \sqrt{\left(\left(\Delta\delta^{1}H\right)^{2} + \left(\Delta\delta^{15}N/6\right)^{2}\right)} \qquad \textbf{Equation 5.1}$$

where Δδ is the combined chemical shift perturbation and ^{1}H and ^{15}N are the proton and nitrogen chemical shifts in parts per million (ppm) for the backbone amide of each residue.

It is clear that significant perturbations are induced over four regions (Figure 5.6, top panel). The chemical environment of residues between Thr 21 and Leu 33 are significantly perturbed following phosphorylation, implying that an extended change in conformation occurs over this region. The greatest chemical shift difference is found at pThr 22 and may be attributed to a combination of binding-induced ordering and the proximity of the phosphate group to the backbone amide, significantly down-shifting both ^{1}H and ^{15}N resonances.

The chemical shift perturbations of three other short regions (centered on Ser 95, Asn 117 and Lys 141, respectively) were hypothesized to correspond to the phospho-binding surface. In common with other functional FHA domains this is formed by three loops each connecting two anti-parallel β-strands, as shown in Figure 1.4. Interestingly Arg 81 does not show a large chemical shift perturbation upon phosphorylation, although it is predicted to be critical for phosphate binding. This may be explained by the fact that in available FHA-phosphopeptide structures it is the Arg 81 NH_2 groups at the end of the side chain that contact the phosphate. The backbone amide makes only structural

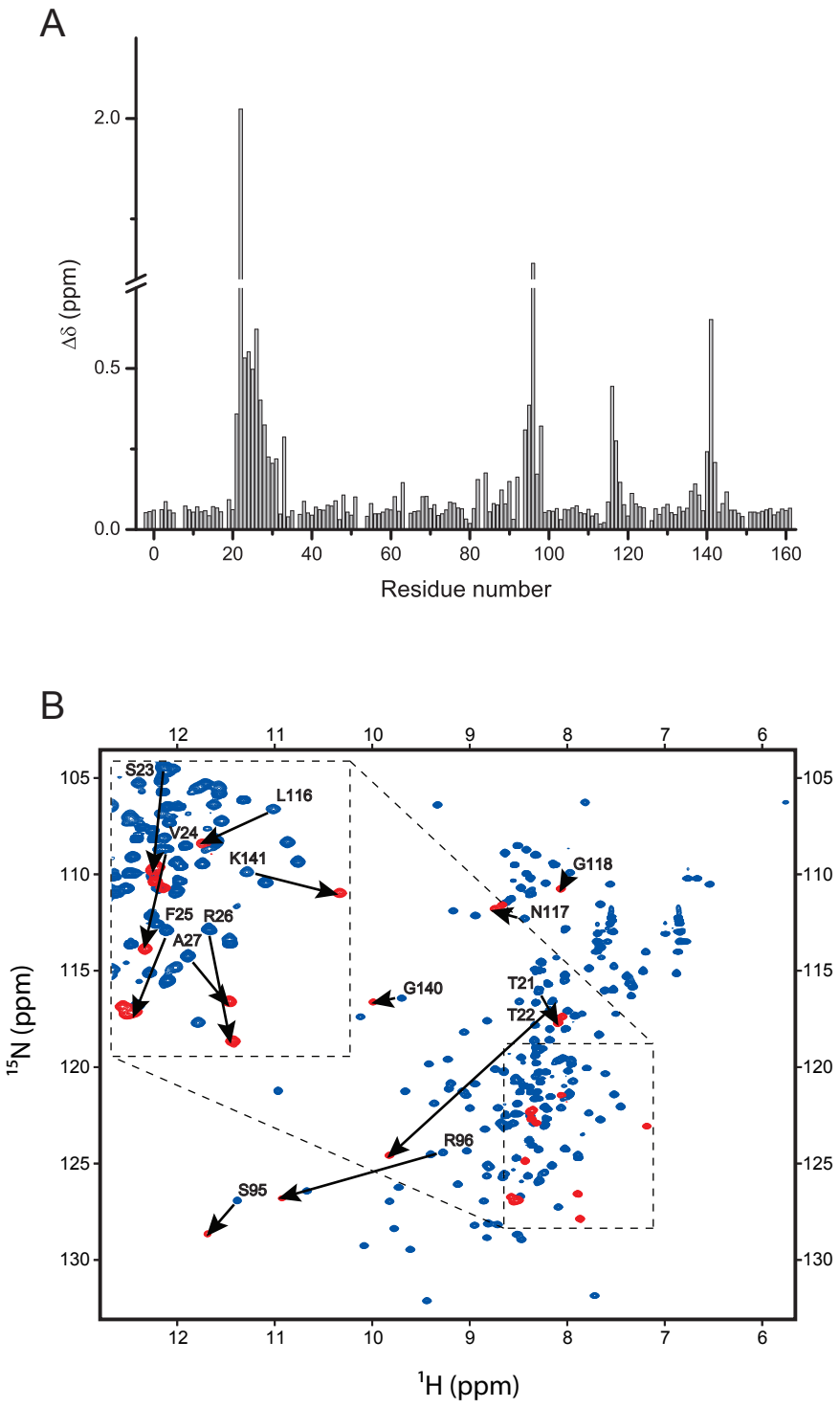

Figure 5.6. ^{1}H-^{15}N HSQC spectrum shift mapping of Rv1827 and Rv1827 pThr 22. (A) Chemical shift perturbation profile of Rv1827 upon its phosphorylation by PknB. Chemical shift perturbation values were calculated according to Equation 5.1. (B) Overlaid ^{1}H-^{15}N HSQC spectra of Rv1827 (blue) and selected residues of Rv1827 pThr 22 (red). The change in chemical shift of residues that is induced by phosphorylation is indicated with arrows.

contacts with the FHA domain and is therefore a significant distance from the site of phosphate binding. This also supports the idea that FHA domains have a pre-formed phospho-interaction surface and do not employ an induced-fit mode of binding.

Figure (Figure 5.6 B) shows the ^{1}H-^{15}N HSQC spectrum of Rv1827 (blue) with a selection of residues most perturbed by phosphorylation annotated and their new positions superimposed (red). Threonine 22 is shown to shift by the greatest amount, closely followed by Arg 96, which in other FHA-phosphopeptide structures makes direct contact with the phosphate group with its backbone amide.

5.4.2 ^{1}H-^{15}N relaxation experiments

The protein backbone dynamics of unphosphorylated and PknB-phosphorylated Rv1827 were investigated using ^{15}N relaxation measurements. T_1, T_2 and ^{1}H-^{15}N-NOE relaxation experiments were performed as each reports on internal molecular mobility by different mechanisms. The principles behind the respective relaxation experiments are outlined in section 3.3.1.

5.4.3 Backbone dynamics of Rv1827 pThr 22

The T_1, T_2, and ^{1}H-^{15}N-NOE relaxation profiles for Rv1827 pThr 22 and Rv1827 are shown in Figure 5.7 and display interesting similarities and differences for these two forms of the protein. In each relaxation profile, missing residues correspond to prolines, overlapped peaks or peaks with extremely weak intensity.

The FHA domain is defined in the T_2 and ^{1}H-^{15}N-NOE relaxation profiles of Rv1827 pThr 22 as being formed by residues 55-149 [agreeing with the ^{2}H exchange data described in section 3.4.2 and Figure 3.1]. Over this region of the protein backbone the average T_2 relaxation time was found to be 68.00 ± 10.14 ms for Rv1827 pThr 22 (Table 5.1). The average ^{1}H-^{15}N-NOE enhancement value was found to be 0.73 ± 0.10. A T_2 relaxation time of ~68 ms and ^{1}H-^{15}N-NOE enhancement of ~0.73 characterizes what may be thought of as the rigid core of the molecule. T_2 relaxation times and ^{1}H-^{15}N-NOE enhancements of this order are observed for residues 22-25 in addition to within the FHA domain. Therefore upon phosphorylation by PknB, pThr 22 and several residues C-terminal to it associate intramolecularly with the FHA domain and become as rigidly ordered as the core of the molecule. This results in an increase in thermal stability as discussed in section 5.2.2 and shown in Figure 5.1, B.

	T_1 (ms)	T_2 (ms)	^{1}H-^{15}N NOE
Rv1827	845.16 ± 53.15	65.06 ± 10.55	0.74 ± 0.10
Rv1827 pThr22	888.27 ± 66.32	68.00 ± 10.14	0.73 ± 0.10

Table 5.1. Summary of backbone dynamics of Rv1827 and Rv1827 pThr 22 across FHA domains residues (55-149).

This trend is also observed by T_1 relaxation, although the data are more variable and the FHA domain is apparently less well defined. It is possible that anisotropy of the protein structure contributes to the observed T_1 variability. The N-terminal 19 residues and residues C-terminal of Lys 151 show extremely high internal motion. From this it is concluded that the extreme N- and C-terminal residues are completely unrestrained. This explains the ability of trypsin to cleave the peptide backbone of Rv1827 C-terminal of Lys 151.

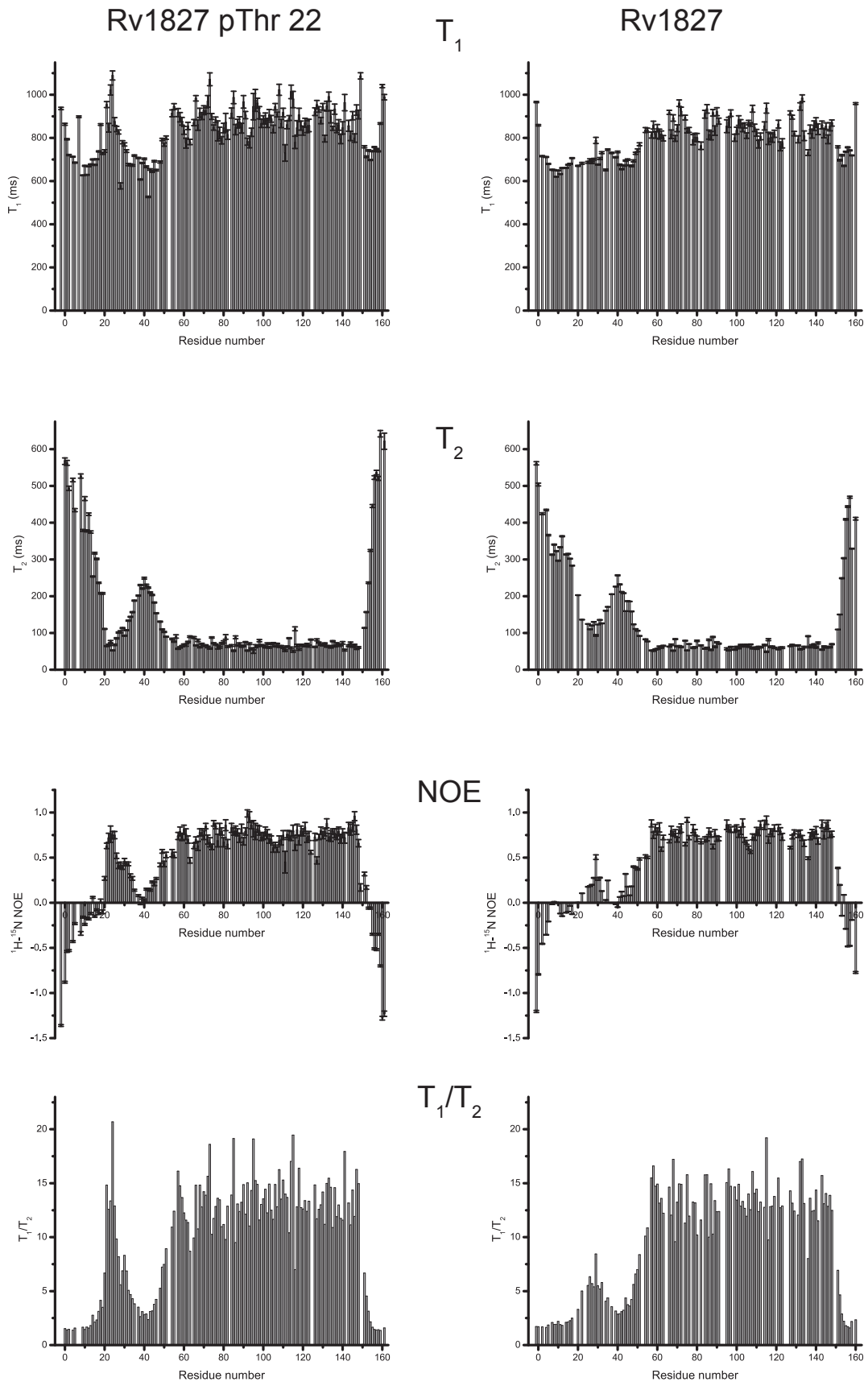

Figure 5.7. T_1, T_2 and NOE relaxation values (acquired at 600 MHz), and T_1/T_2 ratios for Rv1827 (right) and Rv1827 pThr 22 (left).

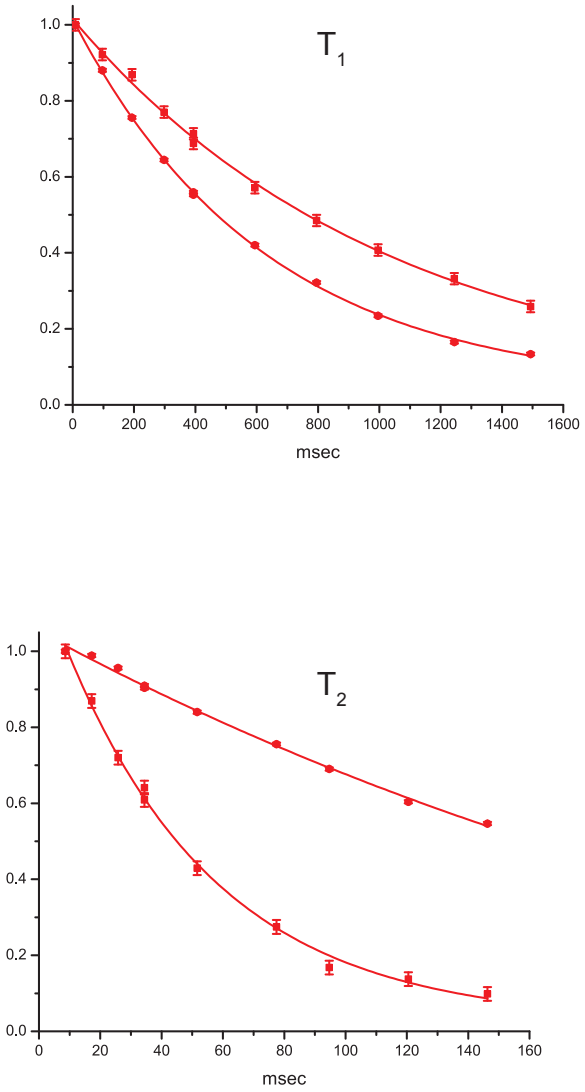

Figure 5.8. T_1 and T_2 decays for Val 24 (square dots) and Ala 39 (round dots) in Rv1827 pThr 22.

An interesting feature of Rv1827 pThr 22 is the peak of T_2 relaxation time centered about Gly 40, which on both the N- and C-terminal sides falls back to that of the core values. This suggests that between the site of phosphorylation and the N-terminus of the FHA domain there exists a 'linker' of increasing and then decreasing internal mobility. This pattern is mirrored in the ^{1}H-^{15}N-NOE relaxation profile, but is less well defined by T_1 across that region. Examples of the raw T_1 and T_2 relaxation data for Ala 39 (linker region) and Val 24 (core region) are shown in Figure 5.8.

5.4.4 Backbone dynamics of unphosphorylated Rv1827

As in the PknB-phosphorylated form, the FHA domain is clearly defined by T_2 relaxation and ^{1}H-^{15}N-NOE enhancement measurements (Figure 5.7). The average T_2 relaxation time and ^{1}H-^{15}N-NOE value over residues 55-149 are 65.06 ± 10.55 ms and 0.74 ± 0.10, respectively (Table 5.1). These are highly similar to those observed for Rv1827 pThr 22 and it is concluded therefore that phosphorylation of Rv1827 by PknB does not change the average backbone dynamics of the FHA domain. Comparison of the T_1 relaxation profiles for the phosphorylated and unphosphorylated forms of the protein over the region N-terminal to the FHA domain highlights a significant difference in internal mobility at the T_1 timescale. In the unphosphorylated form, the N-terminus of the protein exhibits high internal mobility in the N-terminal 50 residues leading to recovery of the longitudinal magnetization of the backbone amide bond vectors significantly more quickly than that of the core.

Despite the fast internal motion seen in T_1 relaxation, a trough in the T_2 relaxation centered on Ala 27 (mirrored by dipolar-coupled cross-relaxation processes in the same region) suggests that the mobility of this region is nonetheless partially constrained. A

combination of transient structural propensity centered on Ala 27 and conformational exchange in which this region is associated with and dissociated from the FHA domain is proposed to explain the dynamic properties of the N-terminus of unphosphorylated Rv1827.

The T_1/T_2 ratios (per residue) for the unphosphorylated and PknB-phosphorylated forms of Rv1827 are shown in Figure 5.7 (D). The average T_1/T_2 ratios for the FHA domains (residues 55-149) of the phosphorylated and unphosphorylated forms are very similar, as are the extreme N- and C-termini (Table 5.2). This suggests that phosphorylation has no significant effect on the apparent mobilities of these regions. However, phosphorylation substantially decreases the internal mobility of residues 20-33 (Figure 5.7, D). Average T_1/T_2 ratios over this region increase from 5.62 ± 1.27 to 9.75 ± 4.57 upon phosphorylation, implying that a change in conformation is induced that extends significantly further than those most commonly observed for FHA-phosphopeptide interactions. It seems that under the conditions of intra-molecular phosphopeptide binding a novel FHA-binding mode may be operational. Residues up to pThr + 10 may contribute to the overall affinity of the intramolecular interaction.

	Average region-specific T_1/T_2 ratios		
	FHA domain (residues 55-149)	Residues 20-33	Termini (residues 1-19 and 150-162)
Rv1827	13.51 ± 1.96	5.62 ± 1.27	2.32 ± 1.16
Rv1827 pThr22	13.40 ± 2.27	9.75 ± 4.57	2.26 ± 1.27

Table 5.2. Comparison of the average region-specific T_1/T_2 ratios for Rv1827 and Rv1827 pThr 22.

5.4.5 Unphosphorylated Rv1827 can exist in a 'pre-bound' conformation

To investigate the possibility that the dynamic behavior of the N-terminal residues is the product of transient ordering within an otherwise unrestrained region, pThr 22 phosphopeptide was titrated against non-phospho Rv1827 and ^{1}H-^{15}N chemical shift perturbations were monitored by NMR. Upon phosphopeptide addition the chemical shifts of residues involved in phosphate binding recapitulate those of the Rv1827 pThr 22 form. However, residues 21-30 move to new locations in the ^{1}H-^{15}N-HSQC spectrum distinct from those in either the un-modified or modified protein spectra (Figure 5.9). This suggests that phosphopeptide interactions with the FHA domain induce changes in the chemical environment of these residues, presumably by displacement. It is therefore proposed that, rather than acting to tether a disordered N-terminal region, Thr 22-phosphorylation appears to stabilise a pre-existing but low occupancy association of residues 21-30 with the FHA domain (Figure 5.9). This explains a small but consistent decrease in apparent affinity observed for binding of pThr 22 and pThr 21 phosphopeptides to full-length un-phosphorylated Rv1827 as compared to the N-terminally truncated FHA domain alone (Nott et al., 2009). It additionally provides a rationale for the increased KGD and GDH inhibition also apparent in enzyme assays following deletion of residues 1-54 (Nott et al., 2009).

5.5 Conclusions

A combination of biochemical, biophysical and hydrodynamic analyses have shown that, upon phosphorylation by PknB at Thr 22, Rv1827 undergoes a novel intra-molecular conformational compaction. This yields significant proteolytic protection at a site four (or in the case of PknG-phosphorylation five) residues C-terminal of the

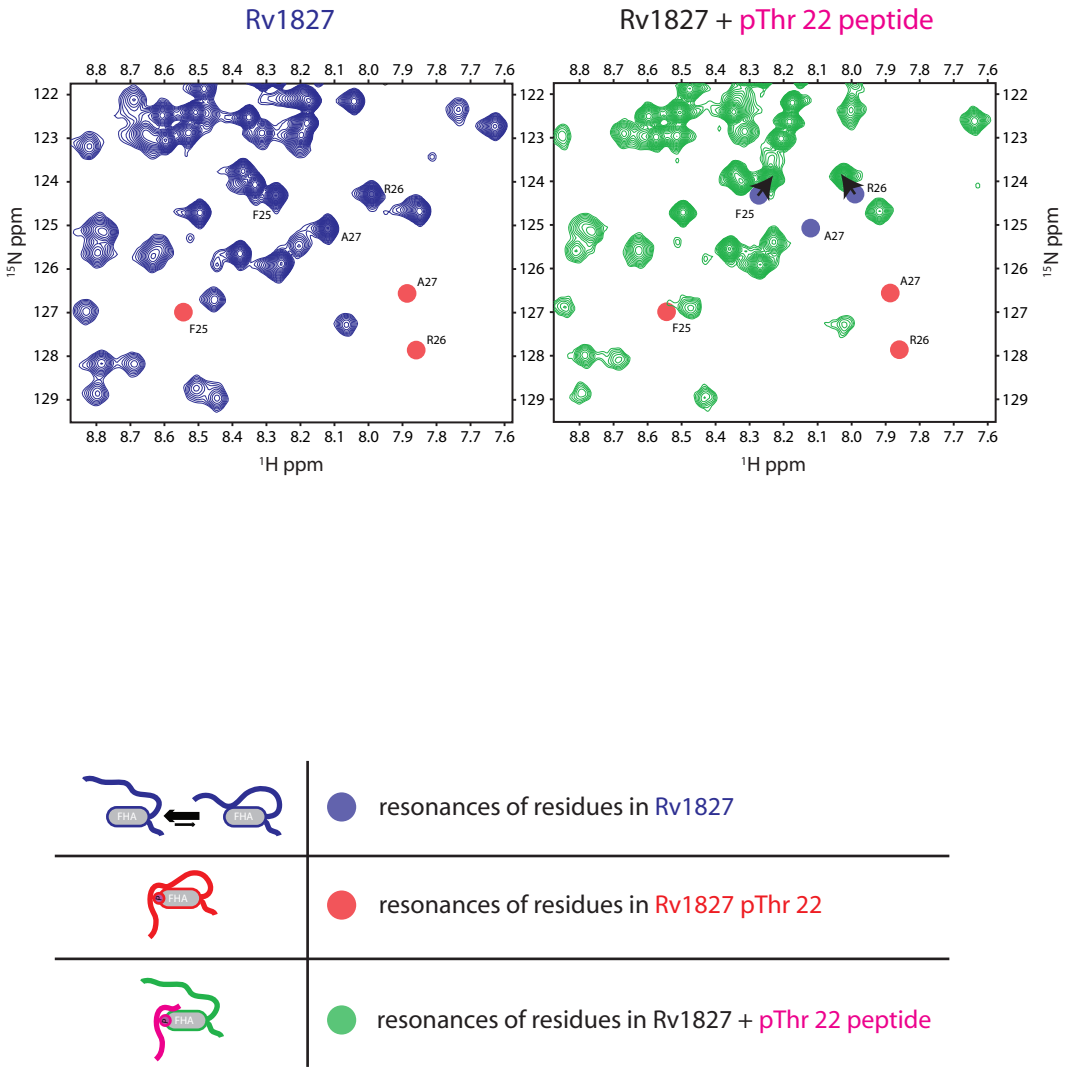

Figure 5.9. Unphosphorylated Rv1827 can exist in a 'pre-bound' conformation. top **Comparison of an equivalent region of the ^{1}H-^{15}N HSQCs of Rv1827 (blue, left) and Rv1827 after addition of 2.5 molar equivalents of pThr 22 peptide (green, right). Resonances of F25, R26 and A27 in the unphosphorylated (blue filled circles) and pThr 22 (red filled circles) forms are labeled. The new resonances of F25 and R26 upon addition of phosphopeptide are indicated with arrows. Addition of phosphopeptide displaces A27 to a new, unknown location.** bottom **Cartoon representing Rv1827 in the unphosphorylated, pThr 22 and pThr 22-peptide bound conformations (blue, red and green with pink, respectively).**

phosphorylation site. The resultant intramolecular binding substantially increases the thermal stability of the molecule. It appears also that a small region of α-helix is stabilized as evidenced by the change in CD signal at ~202 nm.

NMR chemical shift perturbation mapping suggests that the phosphate binding site is formed by the residues most highly conserved in other studied FHA domains, namely Arg 81, Ser 95, Asn 117 and Lys 141. T_1, T_2 and ^{1}H-^{15}N NOE relaxation measurements show that regions of high mobility are apparent at the extreme N- and C-termini and the 34-54 region, while residues around pThr 22 display dynamic behavior identical to that of the core FHA domain. Remarkably, restricted (though still significant) internal motion is also observed for the N-terminal region in the non-phosphorylated form.

We conclude that interactions in the 'prebound' form resemble those observed in the final 'closed' conformation, and contribute to the overall stability of intra-molecular binding in the phosphorylated protein (Figure 5.10). Through these conformational effects, the phosphorylated N-terminal region appears to function as a molecular 'off' switch enabling robust competition for FHA domain interactions with target enzymes as shown in Figure 4.1. This suggests that the surface used by the FHA domain to interact with its metabolic targets may be partially coincidental with the extended surface used for intramolecular phosphate binding.

The next chapter will describe the methodology that was used for solving the solution structure of Rv1827 pThr 22 and present a description of the final ensemble of structures.

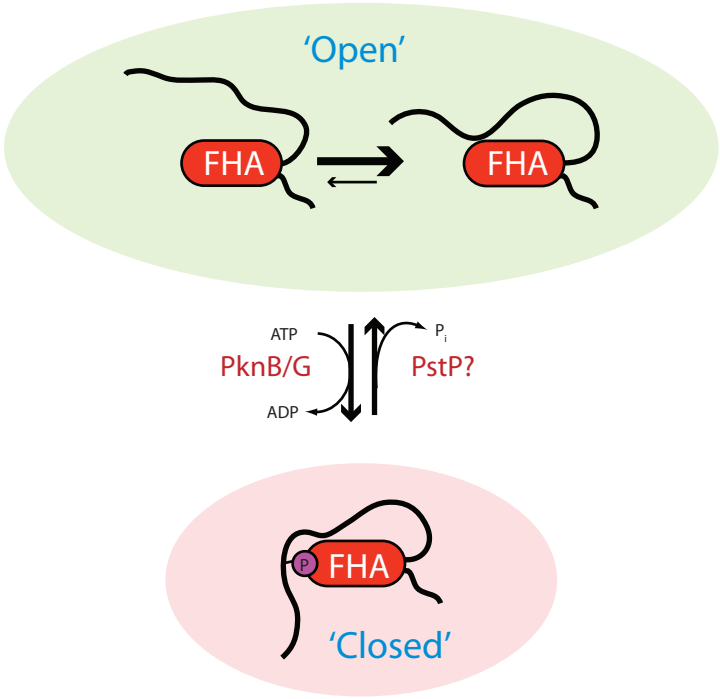

Figure 5.10. Model of conformation exchange upon phosphorylation. In the unphosphorylated state, Rv1827 appears to conformationally exchange between a 'pre-bound' form in which the N-terminus is associated with the FHA domain and an 'un-bound' form in which the N-terminus is unrestrained. When Rv1827 is phosphorylated by either PknG or PknB, a tight intramolecular association of the N-terminus with the FHA domain is induced. This intramolecular association abrogates interactions of the FHA domain with its metabolic binding partners.

6 Solution structure of Rv1827 pThr 22

6.1 Overview

In order to understand the effect of phosphorylation of Rv1827 at the molecular level, the solution structure of Rv1827 pThr 22 was solved by NMR spectroscopy and a model of the unphosphorylated form was generated for comparison. A combination of manual and automatic procedures were employed for the structure determination of Rv1827 pThr 22. Manual structure calculations were performed using constraints derived from ^{1}H-^{13}C- and ^{1}H-^{15}N-NOESY spectra, hydrogen bonds that were inferred from ^{2}H-exchange experiments and backbone torsion angle restraints derived from TALOS (Torsion Angle Likelihood Obtained from Shift and sequence similarity). ARIA 1.2 was used for automatic structure calculation using unassigned peak lists, chemical shift assignments and hydrogen bond information defined in the manual calculations. ^{1}H-^{15}N relaxation data were additionally used in a final step of structure refinement. The quality of calculated structures was checked using PROCHECK-NMR and assessed manually against known structures of other FHA domains.

6.2 Structure determination methodology

6.2.1 Prediction of secondary structure

The initial structural characterization of Rv1827 pThr 22 began with prediction of the secondary structure. NMR was used to predict secondary structure as the chemical shift of backbone atoms is sensitive to the torsion angles either side of the C_α atom. For each amino acid the torsion angle between the backbone amide (NH) and the C_α atom is the phi (Φ) angle and the torsion angle between the C_α atom and the backbone carbonyl (CO) is the psi (Ψ) angle. Residues in regions of α-helical secondary structure have

characteristically different Φ and Ψ torsion angles to residues in parallel and anti-parallel β-sheets. For example, the C_α chemical shifts of residues in α-helices have typically up-field shifted resonances compared to random coil values and those in β-sheets have typically down-field shifted resonances.

The secondary structure of a protein can be predicted by trying to match the backbone chemical shifts of amino acid triplets against those of known secondary structure. The program TALOS does this by trying to find statistically significant matches in a database of ~24000 triplets. If significant matches are found, TALOS predicts the Φ and Ψ torsion angles of the amino acid at the centre of each amino acid triplet.

TALOS incorporates the assignments of $^{15}N(^{1}H)$, $^{1}H(^{15}N)$, $^{13}C_\alpha$, $^{13}C_\beta$, $^{1}H_\alpha$ and $^{13}C(O)$ chemical shifts of each amino acid for secondary structural prediction. Consequently greater coverage of assignment results in more accurate predictions. Table 6.1 shows a comparison of the proportion of complete backbone assignments (of $^{15}N(^{1}H)$, $^{1}H(^{15}N)$, $^{13}C_\alpha$, $^{1}H_\alpha$ and $^{13}C(O)$) of unphosphorylated Rv1827 and the PknB-phosphorylated form. These assignments were obtained as described in section 3.2.3. The highest proportion of assignment was made over the most ordered regions of the two forms of Rv1827.

		Rv1827		Rv1827 pThr 22	
	Total atoms	Number unassigned	% assignment	Number unassigned	% assignment
Full length (-4-162)	843	119	85.9	45	94.7
Residues 20-149 (structure)	656	61	90.7	13	98.0
Core residues (20-25, 55-149)	510	34	93.3	7	98.6

Table 6.1. Backbone assignment of Rv1827 and Rv1827 pThr 22. Backbone atoms are defined as $^{15}N(^{1}H)$, $^{1}H(^{15}N)$, $^{13}C_\alpha$, $^{1}H_\alpha$ and $^{13}C(O)$.

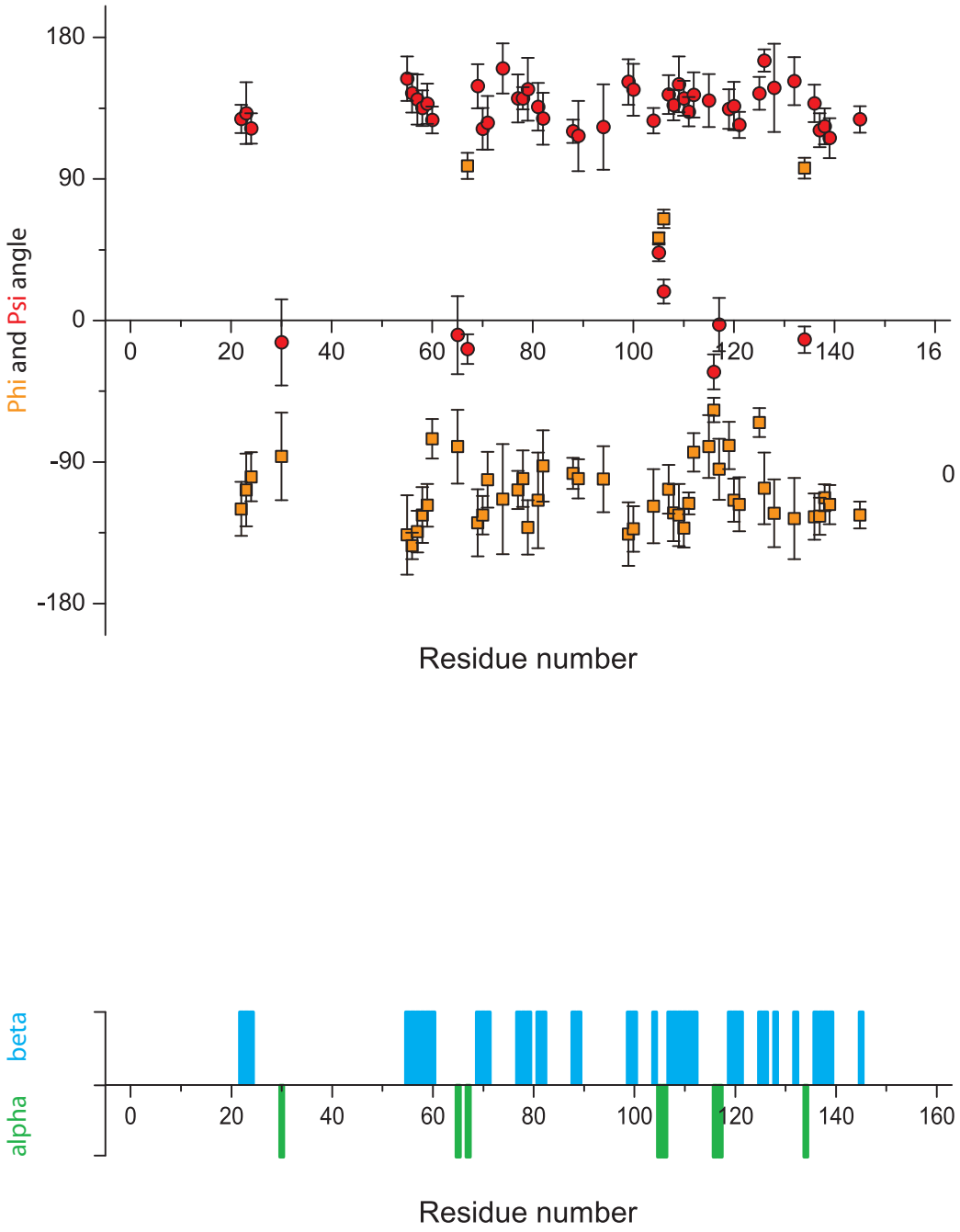

Figure 6.1. Secondary structural prediction of Rv1827 pThr 22 by the program TALOS. top **Residues for which significant matches of Phi (orange) and Psi (red) torsion angles were made using TALOS.** bottom **Secondary structure predicted by Phi and Psi torsion angles. Regions of β-strands are shown in blue and regions of α-helix/turns are shown in green. The FHA domain of Rv1827 is largely predicted to consist of β-strands.**

Figure 6.1 (A) shows the TALOS predictions of Φ and Ψ torsion angles of Rv1827 pThr 22 for residues for which at least nine matches were found. Figure 6.1 (B) shows the secondary structural predictions for matched residues, and have been catagorised into predicted β-strands and α-helices. Residues 55 to 145 were predicted to predominantly form β-strands. The short stretches of predicted α-helices were hypothesized to form loops between β-strands. Interestingly, a short region of β-strand is predicted between pThr 22 and Val 24, suggesting that residues around the phosphorylation site form an extended conformation similar to those seen in other FHA-phosphopeptide complexes (Byeon et al., 2005; Li et al., 2004). A residue at the pThr + 8 (Leu 30) is predicted to reside in a region of α-helix, although statistically significant matches were not found for residues N- and C-terminal to it. In support of the ^{15}N relaxation data discussed in section 5.4.3, no secondary structure was predicted for the termini of the protein or the region between Leu 30 and the Ser 55 due to the high internal mobility of these regions.

For the structural calculations presented in the following sections TALOS-derived Φ and Ψ torsion angle restraints were incorporated with a TALOS error of 10° and a TALOS factor of 2. It was decided that TALOS predictions for pThr 22, Ser 23 and Val 24 should not be included in calculations as the presence of the phosphate group on pThr 22 could potentially cause chemical shift perturbations of the backbone atoms of these residues not related to secondary structure. Another motivation for not including these residues was that modified residues are not represented in the TALOS database of amino acid triplets.

6.2.2 Hydrogen bond restraints

The introduction of hydrogen bond restraints to structure calculations can be useful. They help to define the secondary structure of the protein at an early stage in the calculations so that possible ambiguity in other restraints does not become a limiting or biasing factor in structure refinement. Hydrogen bond restraints were included in all but fully automatic structure calculations. 37 hydrogen bonds were inferred from a combination of a ^{2}H exchange experiment and $^{1}H_{\alpha}$-$^{1}H_{\alpha}$ NOEs identified in ^{1}H-^{13}C-NOESY spectra. The results of the ^{2}H exchange experiment are represented in Figure 6.2, which shows the Rv1827 pThr 22 ^{1}H-^{15}N-HSQC spectrum coloured according to how quickly backbone amide protons exchanged with ^{2}H from the solution upon dissolution of the lyophilized protein in 100% $^{2}H_{2}O$. Amide protons that exchanged within 2 min. are coloured green and those that exchanged between 2min. and 17 hours are coloured orange. Those that had not exchanged after 17 hours in 100% $^{2}H_{2}O$ are coloured red and were assumed to be engaged in long-lived hydrogen bonds.

In the context of a secondary structure composed mostly of anti-parallel β-strands the unambiguous identification of inter-strand $^{1}H_{\alpha}$-$^{1}H_{\alpha}$ NOEs was used to define the register of β-strands in the β-sheets. From these data, candidate inter-strand hydrogen bonds were inferred. If a non-exchanging amide proton was identified from the ^{2}H exchange experiment along with the corresponding unambiguous reciprocal $^{1}H_{\alpha}$-$^{1}H_{\alpha}$ NOE relationship for the $^{1}H_{\alpha}$ of the same residue, a hydrogen bond restraint between that amide group and the backbone carbonyl of the adjacent residue of the adjacent strand was included in calculations. In order to maintain the co-linearity of the hydrogen bonds the distance between the carbonyl oxygen and the amide proton was set to 2.1 ± 0.5 Å

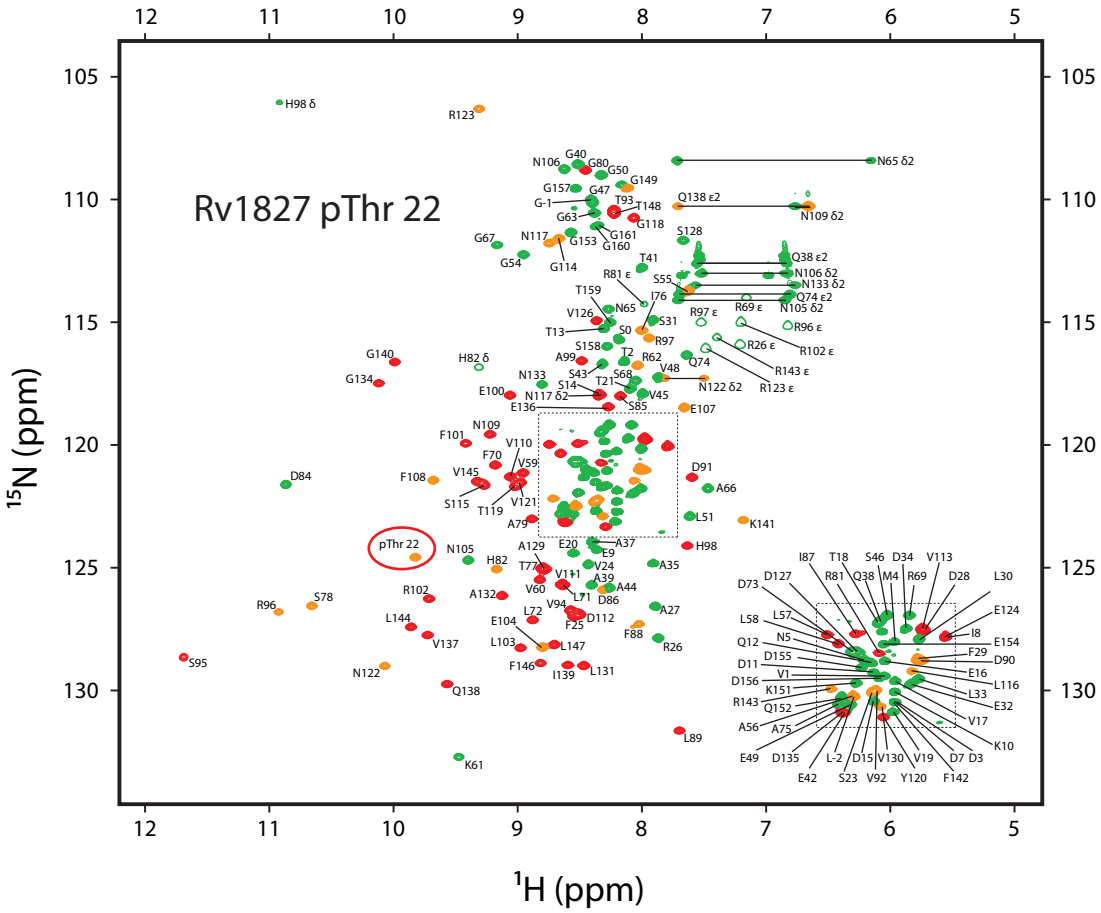

Figure 6.2. Annotated ^{1}H-^{15}N HSQC spectrum of Rv1827 pThr 22 coloured according to the results of the ^{2}H exchange experiment. Peaks that exchanged within 2 minutes are coloured green. Peaks that exchanged within 17 hours are coloured orange and those that did not exchange within 17 hours are coloured red. Peaks folded in the ^{15}N-dimension are shown as hollow contours. Assigned δ-NH_2 (asparagine) and ε-NH_2 (glutamine) pairs are linked with horizontal bars. The inset at the lower right of the spectrum corresponds to the crowded regions in the center. pThr 22 is highlighted with a red ellipse. Unassigned peaks are not annotated.

and the distance between the carbonyl oxygen and the amide nitrogen was set to 3.0 ± 0.5 Å. Consequently 74 hydrogen bond restraints for 37 hydrogen bonds between β-strands were included in total. Two hydrogen bond restraints of the same distance were included between the H_ε and N_ε of Arg 81 and the carbonyl of Val 94 as a hydrogen bond at this position has been observed in all FHA-phosphopeptide complex crystal structures solved to date.

Three hydrogen bond restraints were included between the phosphorus atom of pThr 22 and the hydroxyl group of Ser 95, the backbone amide of Arg 96 and the terminal NH_2 group of Arg 81. These are three conserved FHA domain-phosphopeptide hydrogen bonds and are shown in detail in Figure 1.5 for the Rad53 FHA1-phosphopeptide complex crystal structure. The phosphorus atom was chosen instead of its individual phosphate oxygens so as not to influence the sidechain dihedral angles of the pThr residue in structure calculations. The distance between the phosphorus atom and the respective hydrogen bonding protons was set as 3.1 ± 0.5 Å. The distance between the hydroxyl oxygen of Ser 95 and the phosphorus atom was set as 4.0 ± 0.5 Å, as were the distances between the nitrogen of the Arg 96 backbone amide and the nitrogen of the Arg 81 terminal NH_2 group.

In the available crystal structures of FHA domain-phosphopeptide complexes these three FHA domain residues engage in conserved hydrogen bonds with the oxygen atoms of the phosphate group in the pThr ligand. However, as no stereochemical information was obtainable about the phosphate oxygens in pThr 22, hydrogen bond restraints between the phosphorus atom and FHA domain residues were included instead. This left free rotation around the chi2 and chi3 bonds of pThr 22.

6.2.3 Resonance assignment of Rv1827 pThr 22

Through analysis of the various triple resonance spectra outlined in sections 3.2.3 and 3.5.1, the assignment of Rv1827 pThr 22 backbone and side chain resonances was achieved. In total, Rv1827 pThr 22 contains 1962 potentially assignable atoms. Assignment of several subsets of atoms such as the sidechain NH_2 groups of Arg residues, Lys NH_3^+ groups and in many cases the hydroxyl protons of Ser and Thr residues, assignment was not possible. Additionally, the protons of methylene groups were not stereo-specifically assigned in any cases. If it was not possible to stereo-specifically assign the resonances of sidechain proton pairs due to spectral resolution (or overlapping resonances), the same ^{13}C and ^{1}H chemical shifts were assigned to each (in a form of pseudoatom correction).

Shown in Table 6.2 is the proportion of assignments made for different regions of Rv1827 pThr 22 across all assignable atoms. No pseudoatom groups are included so the total number of atoms is non-redundant.

	Rv1827 pThr 22		
	Total atoms	unassigned	% assignment
Full length (-4-162)	1962	318	83.8
Residues 20-149 (structure)	1553	188	87.9
Core residues (20-25, 55-149)	1242	155	87.5

Table 6.2. Non-redundant assignment statistics for Rv1827 pThr 22.

6.2.4 Automatic versus manual structure determination

For all the ensembles of structures presented in the following sections the N-terminal 19 residues and the C-terminal 12 residues of Rv1827 pThr 22 have been omitted for clarity.

Figure 6.3 (A) shows the scheme with which the ensembles of structures in this section are coloured.

Initial solution structure calculations of Rv1827 pThr 22 were performed automatically using ARIA 1.2 (Linge et al., 2003). These calculations incorporated the complete set of resonance assignments for Rv1827 pThr 22, TALOS-derived Φ and Ψ torsion angle restraints and complete peak lists from the ^{1}H-^{15}N-NOESY spectrum (recorded in 10% $^{2}H_2O$) and ^{1}H-^{13}C-NOESY spectra (one recorded in 10% and the other recorded in 100% $^{2}H_2O$). Distance calibration of peak intensities for the respective spectra was carried out as described in section 3.5.2. Figure 6.3 (B) shows the 20 lowest energy structures calculated automatically out of 50 water-refined structures. A poor degree of convergence is observed for the FHA domain (residues 55-149, black). The 35 residues N-terminal to it are un-converged and appear bunched up against the N-terminal end of the domain with no discernable secondary or tertiary structure. The number of unambiguous and ambiguous restraints (categorized by type and distance) used in these calculations are shown in Table 6.3.

		intra.	seq.	med.	long	Total
automatic	unambiguous	1227	656	218	696	2797
	ambiguous	369	248	142	501	1260
	all	1596	904	360	1197	4057
manual	all	1039	69	73	893	2074
automatic and manual	unambiguous	1190	610	218	824	2842
	ambiguous	416	267	154	554	1391
	all	1606	877	372	1378	4233

Table 6.3. Comparison of the number of unambiguous and ambiguous restraints incorporated in structure calculations automatically and manually when used in isolation (*top* and *middle*) and combined (*bottom*). intra. (intra-residue), seq. (sequential, $i \rightarrow i+1$), med. (medium, $\leq i \rightarrow i+4$) and long ($\geq i \rightarrow i+5$).

For comparison, Figure 6.3 (C) shows the 20 lowest energy structures calculated from manually derived NOE distance restraints (summarized in Table 6.3), hydrogen bond restraints and TALOS-derived Φ and Ψ torsion angle restraints. Again, residues comprising the FHA domain are poorly converged. The region N-terminal to the FHA domain is now distributed over a greater area as a result of the addition of hydrogen bonds constraining the phosphorus atom of pThr 22 to Ser 95, Arg 96 and Arg 81. However, no discernable secondary or tertiary structure is present around the site of phosphorylation.

Significantly higher convergence over the FHA domain was achieved by combining the distance restraints manually-derived from the NOESY spectra, hydrogen bonds and TALOS-derived Φ and Ψ torsion angles with complete resonance assignments and unassigned peak lists (Figure 6.3, D). A higher degree of convergence was also observed for pThr 22, Ser 23 and Val 24. Under these circumstances the residues between the phosphorylation site and the FHA domain were more constrained as they were effectively 'tethered' at the N- and C-terminal ends. Despite low convergence, residues in this linking region followed a path predominantly across one face of the FHA domain by doubling back over the phosphorylation site. The Φ and Ψ torsion angle restraints for Leu 30 are partly responsible for this U-turn in direction along with an intense and overlapped ^{1}H-^{13}C NOESY peak between Ala 27 and Arg 97 that was assigned unambiguously (yet erroneously) by ARIA.

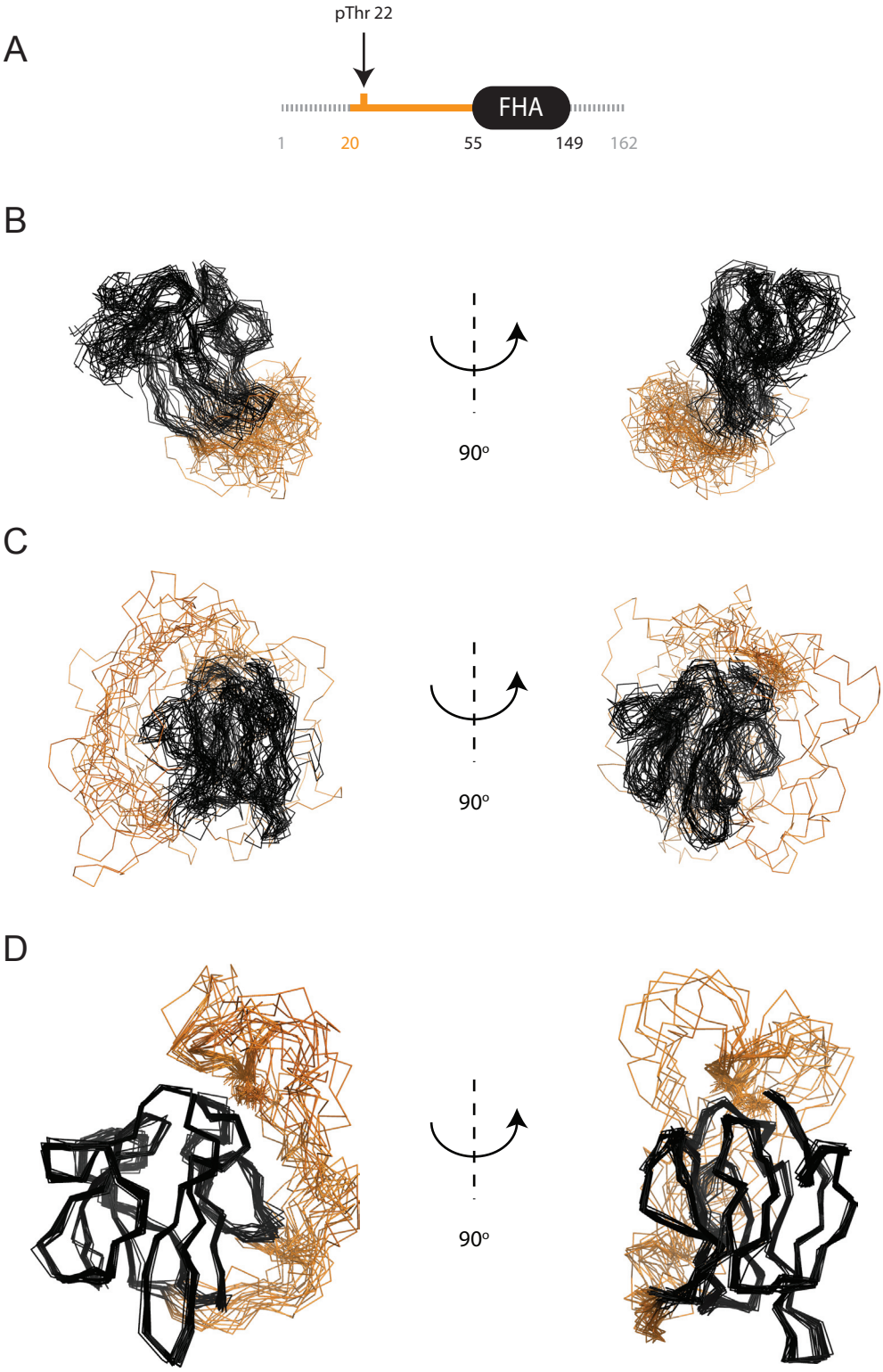

Figure 6.3. Manual versus automatic structure calculation of Rv1827 pThr 22. (A) Schematic showing the colouring used in the images below. The extreme N- (1-19) and C-terminal (150-162) residues (grey) have not been included in the structure pictures for clarity. (B) Ensemble of Rv1827 solution structures calculated automatically by ARIA 1.2. (C) Ensemble of Rv1827 structures calculated using manually-derived distance restraints and hydrogen bonds. (D) Ensemble of Rv1827 structures calculated using both manually-derived and automatic procedures.

6.3 Structural refinement

6.3.1 Identification of unambiguous NOESY-derived distance restraints

An iterative process of structure calculation and manual evaluation of distance restraints derived from NOESY peak lists facilitated refinement of Rv1827 pThr 22 solution structures. A set of manually assigned unambiguous NOESY-derived restraints was iteratively used in structure calculations incorporating full NOESY peak lists to automatically extract more NOESY-derived restraints that were then evaluated manually for ambiguity. If judged unambiguous these new restraints were added to the initial list of restraints and the process was repeated until convergence was reached in which no more NOESY peaks could be unambiguously assigned as distance restraints (Table 6.4).

		Intra.	seq.	med.	long	Total
refined automatic and manual	unambiguous	1156	561	216	844	2777
	ambiguous	444	286	157	657	1544
	all	1600	846	373	1501	4320

Table 6.4. Comparison of the proportion of the total ambiguous and unambiguous distance restraints of different types derived from NOESY spectra. intra. (intra-residue), seq. (sequential, $i \rightarrow i+1$), med. (medium, $\leq i \rightarrow i+4$) and long ($\geq i \rightarrow i+5$).

Application of this iterative procedure revealed the ensemble of 20 lowest energy structures shown in Figure 6.4. A high degree of convergence is evident across the FHA domain and in the pThr 22 to +3 position. Significant convergence is also present in the linker region, and shows it following a narrow path across one face of the FHA domain. Such high convergence in what is demonstrably a region of high internal mobility was reasoned to be an inappropriate *in silico* over-interpretation. The implications of this are discussed in detail in the next section.

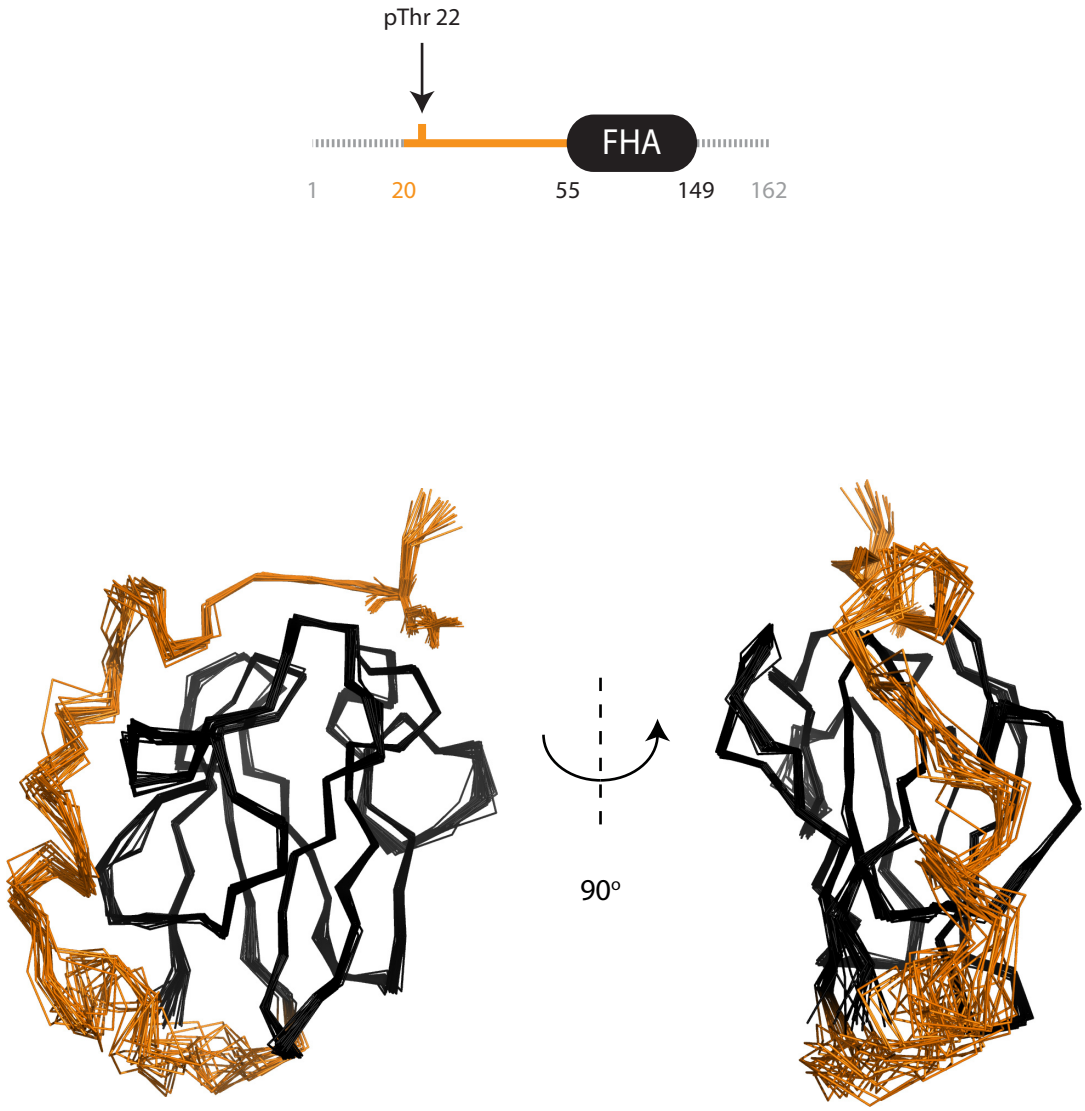

Figure 6.4. Ensemble of solution structures of Rv1827 pThr 22 calculated using the final set of distance restraints and automatic procedures. top **Schematic showing the colouring of the ensemble of structures (same as in Figure 6.3).** bottom **Ensemble of Rv1827 pThr 22 solution structures, showing a high degree of convergence across the FHA domain (black) and the linker region (orange).**

6.3.2 Restraint evaluation and thresholding using ^{15}N relaxation profiles

NMR-based relaxation measurements give a powerful insight into the internal molecular motion of the protein backbone. As discussed in section 5.4.3 and shown in Figure 5.7, the T_2 and ^{1}H-^{15}N NOE relaxation profiles of Rv1827 pThr 22 show that the site of phosphorylation binds to the FHA domain and becomes as relatively inflexible as it. However, the N- and C-termini are highly mobile and the linker region (residues 25-54) has a peak of internal mobility centered on Gly 40. As the extreme termini have apparently no intrinsic structure, correlated with the identification of no medium- to long-range NOEs, they were not incorporated into further calculations.

Dealing with the linker region presented an interesting challenge. It was reasoned that distance restraints for residues with a long T_2 relaxation time (in excess of 150 ms, corresponding to ^{1}H-^{15}N-NOE enhancement values lower than 0.3) ought to be treated in the same way as the termini, and discarded. This corresponded to the 12 most highly mobile residues of the linker region (residues 35-46). As with the termini of the protein only intra-residue and sequential NOEs were identified in this region.

The explanation for the over-interpretation of data in automatic structure calculation procedures (resulting in the ensembles shown in Figure 6.4) lies in the nature of NOESY experiments. The cross-relaxation of dipolar-coupled spins becomes less efficient with increased distance between the spin systems according to r^{-6}. NOEs of atoms in highly mobile regions of a molecule arise as a result of the cross-relaxation of dipolar-coupled spins that are constitutively close in space, but without containing structural information. These are observed predominately as intra-residue and sequential NOEs which, as these

spin systems are necessarily close in space, produce intense resonances due to a high degree of conformational averaging. In distance calibration of the spectra, these in turn are attributed the tightest distance restraints in structure calculations. So far so good: intra-residue NOEs have no long-range structural significance and so will not bias structure calculations. Not quite. What about the overlap in resonances in highly mobile regions? Here, the chemical shifts of atoms approach their random coil values and the resonances of subgroups of atoms of residues of the same type are consequently considerably more overlapped than in ordered regions of the molecule. This poses a problem. Overlapped intra-residue and sequential NOEs can be equally well interpreted in automatic structure calculation as highly weighted medium- and long-range structural constraints between residues of the same or similar types.

If these peaks are removed from peak lists, yet the chemical shifts of atoms they correspond to remain in the assignments table, automatic assignment and structure calculation will continually try to find new assignments for them. If the peaks remain in the peak lists and the atoms are removed from the chemical shift table, they may be assigned incorrectly to other atoms.

Both the peaks and the corresponding atom chemical shifts must be omitted from automatic structure calculations in order to mitigate the potentially misleading effects of automatic erroneous assignment of these types of resonances. Conceptually this is simple to understand, but practically untenable due to the possibility of human error in the evaluation of NOESY spectra that may contain thousands of peaks. In other words asking a person to *completely* assign a NOESY spectrum and then selectively remove resonances that have been judged by another empirical method to be too highly mobile to be relevant for structure calculations is very error prone.

Another way to deal with this issue is evaluation at the restraint level, not the peak and chemical shift lists level. In an idealized situation, NOESY spectra used for structure calculation would not contain overlapped resonances and consequently each atom of a molecule would have a unique chemical shift. If this were the case, automatic resonance assignment and structure calculation could unambiguously derive a set of non-violating restraints to describe the molecule. Therefore in an ideal world there would be equivalence between the peak lists plus the chemical shifts and the set of restraints that satisfy them.

As Rv1827 pThr 22 has been demonstrated to exhibit high internal mobility at the termini and in the region between ~Leu 30 and Ser 55, automatic *in silico* procedures for structure calculation encountered many of the problems outlined above. Therefore at the stage where more unambiguous restraint information could not be extracted from the respective NOESY spectra, NOE-derived distance restraints were manually evaluated in light of the relaxation data described previously, and subsequent calculations were carried out in the absence of peak lists. By this method a molecular model of the protein that satisfied all the available empirical data could be generated without at every step reintroducing ambiguity and bias from the mobile regions of the molecule.

Several unambiguous medium- and long-range NOEs were identified for residues 27-34 and 48-54. These two regions lie at the N- and C-terminal ends of the linker. The T_2 and ^{1}H-^{15}N NOE relaxation characteristics over these regions appear to be neither as ordered as the core of the molecule, nor as disordered as the center of the linker or the termini. Interestingly, the amino acid sequences of these regions are highly conserved in Rv1827 homologues in the actinomycetes. We reasoned therefore that identifying local

secondary structure in these regions would aid our understanding of how to incorporate information about them into subsequent structural refinement. The strategy that was employed involved characterizing intra-regional distance restraints defining secondary structure and then long-range distance restraints between the respective regions and the FHA domain. These regions are represented in cyan in Figure 6.5.

Residues between Ala 27 and Asp 34 were characterized to form a region approximating a short stretch of helix. Unambiguous $i \rightarrow i$+4 NOEs were identified between Phe 29 and Leu 33 and suggested that a canonical α-helix was present. However, $i \rightarrow i$+6 NOEs between Arg 26 and Ala 32 and $i \rightarrow i$+7 NOEs between Arg 26 and Leu 33 made it clear that this region did not form an α-helix, but a region with less well defined helical propensity. This region will now be referred to as H1. These findings agreed with the nature of the TALOS Φ and Ψ torsion angle predictions for Leu 30, which is at the centre of this region. Despite the high sequence conservation for the 7 residues immediately N-terminal to the FHA domain, no intra-regional NOE relationships were identified that stabilized any discernable secondary structure in this region.

Next, distance restraints between H1 and residues on the FHA domain were evaluated for ambiguity. A set of 16 unambiguous distance restraints were identified that orientated H1 with respect to the FHA domain. Several FHA domain residues were identified that made multiple contacts with H1. NOE relationships between Gln 138 and Ala 27, Phe 29 and Leu 33 suggested that this residue was important not only for the H1-FHA interaction, but also for stabilizing intra-regional interactions within H1 itself. Long-range NOEs between Ala 27 and Lys 141 and also Glu 32 and Arg 143 suggested that basic residues on the FHA domain were also important for maintaining the H1-FHA

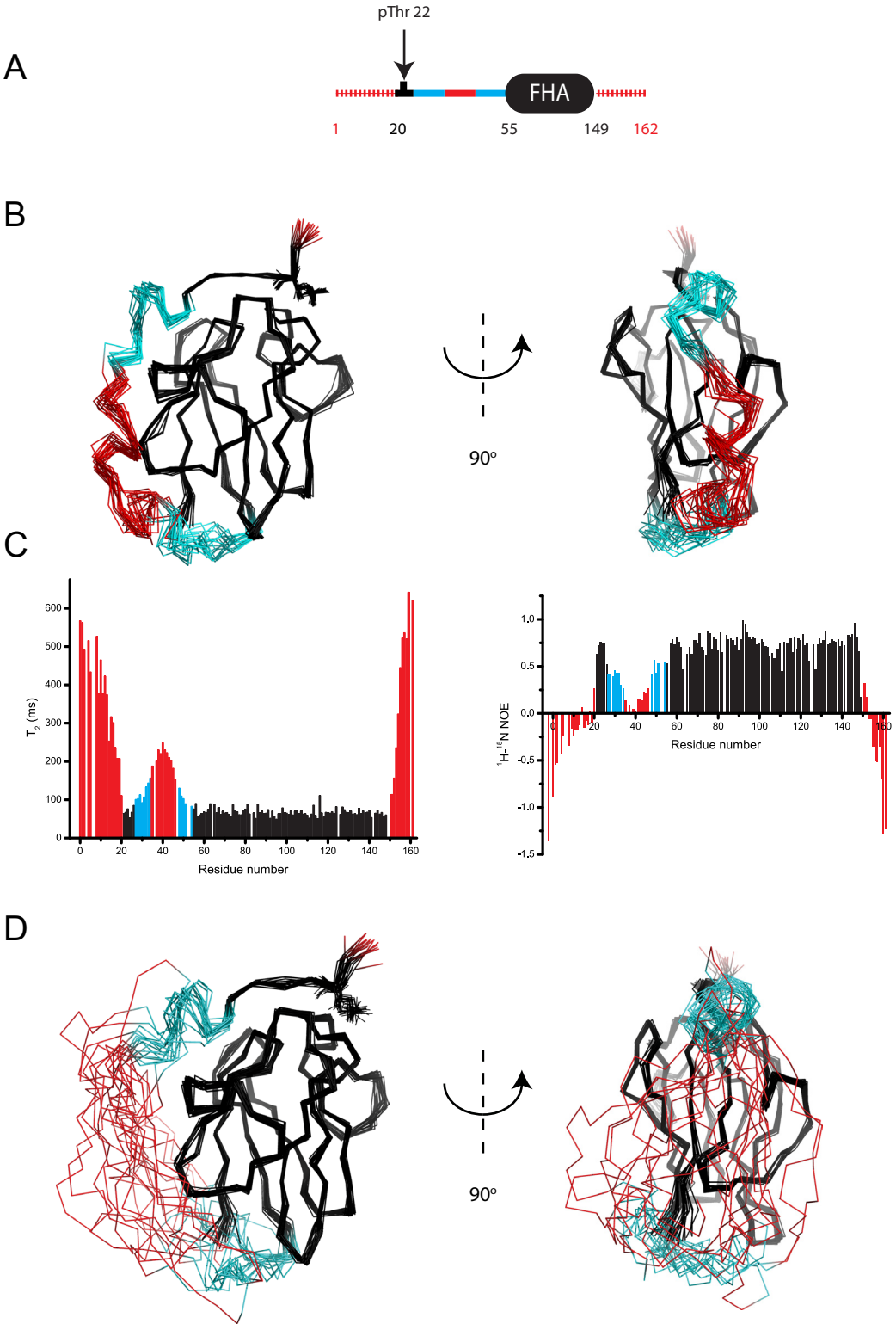

Figure 6.5. Restraint thresholding of Rv1827 pThr 22 solution structure calculations using T_2 and NOE relaxation profiles. (A) Schematic showing the regions of Rv1827 in which different thresholding procedures were employed (red: no distance restraints, cyan: intra-regional distance restraints, black: all distance restraints). (B) Ensemble of solution structures from Figure 6.4, coloured according to the schematic in (A). (C) Rv1827 pThr 22 T_2 (left) and NOE (right) relaxation profiles coloured according to the schematic in (A). (D) Ensemble of Rv1827 pThr 22 solution structures produced by applying the distance restraint thresholding procedure, coloured according to (A). By being un/less-restrained, residues in the linker region explore more conformational space, reflecting their high internal mobility.

domain association. NOEs between the aromatic ring of Phe 29 and the $^{1}H_{\beta}$ protons of Asn 117 were also unambiguously assigned. This is an interesting observation as Asn 117 is one of the most highly conserved FHA domain residues, and annotated as a residue critical for phosphopeptide-binding. In the available structures of FHA domains in complex with phosphopeptide ligands Asn 117 is engaged in hydrogen bonds with the pThr +2 or +3 residue. However, in the intramolecularly bound form of Rv1827 pThr 22, Asn 117 makes contacts with the aromatic ring of the pThr + 7 residue as well as side chain interactions with Val 24 and Phe 25.

At the C-terminal end of the linker region NOE relationships were identified between residues 51 to 54 and residues 148 and 149, which form the C-terminal end of the final β-strand of the FHA domain. However, identification of these long-range NOEs is unsurprising since Ser 55 is the first residue of the N-terminal β-strand of the FHA domain, and is adjacent and anti-parallel with Thr 148 and Gly 149. Therefore the C-terminal end of the linker is necessarily close in space to these residues of the FHA domain and the NOE relationships appeared to be of no structural significance.

After removal of all ambiguous distance restraints and the restraints involving regions coloured red in Figure 6.5 (A, B and C), the final proportions of distance restraints used to calculate the final ensembles of Rv1827 pThr 22 structures are summarized in Table 6.5. Removal of restraints according to the scheme in Figure 6.5 (C) yielded the ensemble of structures shown Figure 6.5 (D). Both sets of ensembles in Figure 6.5 are coloured with the same scheme to highlight the effect of completely removing restraints involving residues 35-46 on the convergence of the linker.

	intra	seq.	med.	long	Total
Final restraints	659	385	179	775	1998

Table 6.5. Summary of final NOE-derived distance restraints for Rv1827 pThr 22 after restraint thresholding. intra. (intra-residue), seq. (sequential, $i \rightarrow i+1$), med. (medium, $\leq i \rightarrow i+4$) and long ($\geq i \rightarrow i+5$).

6.4 Rv1827 pThr 22 solution structure evaluation

Statistical analyses of the ensemble of the 20 lowest energy solution structures of Rv1827 pThr 22 are described in section 9.8, Table 9.1 and Figure 9.1 of the appendix section. Presented in this section is a more qualitative analysis, focused on a molecular model of Rv1827 pThr 22 that satisfies all the available empirical data. As discussed in sections 5.4.3 and 5.4.4, T_1, T_2 and ^{1}H-^{15}N NOE relaxation profiles have provided significant insight into the internal molecular mobility of the unphosphorylated and PknB-phosphorylated forms of Rv1827. Additionally they have provided a rational basis on which to selectively manipulate NOESY-derived distance restraints (section 6.3.2) so that the calculated ensembles of conformations reflect both core structure and highly dynamic regions of the molecule. However, as non-standard methodology was employed, careful evaluation of the molecular model must follow.

6.4.1 Qualitative analysis and structure validation

A caveat of x-ray crystallography as an experimental technique for describing molecular structure is that a crystal structure does not display dynamic information about the molecule, and necessarily so. A high resolution crystal structure therefore represents a snap-shot in time of the molecule under investigation. On the other hand NMR is amenable to the study of protein dynamics at atomic resolution, but a problem is still

encountered when viewing an ensemble of solution structures one molecule at a time. One molecule from an ensemble of solution structures contains no intrinsic information about the other molecules of the ensemble. This is especially problematic when a protein is known to have high internal mobility in specific regions. In this case it is important to keep in mind that any one molecule is unrepresentative of the conformational space capable of being explored by its highly dynamic regions. Therefore viewing ensembles of solution structures can be far more informative in giving the viewer a description of the regions with high and low internal motility. This description derives from the level of convergence of regions of the molecules in an ensemble. In ensembles of solution structures a higher degree of convergence is correlated with higher proportions of non-violating, unambiguous distance restraints used in calculations. It follows that higher convergence effectively represents there being an increasingly restricted set of conformations capable of satisfying the set of distance restraints. Conversely, the fewer the restraints there are, the greater the conformational space capable of being explored in simulated annealing calculations.

Both of these effects have been incorporated into the calculation of solution structure ensembles of Rv1827 pThr 22. A high degree of convergence is observed for the FHA domain (residues 55-149) and the pThr 22 to + 3 region (residues 22-25). The pairwise r.m.s. deviation from the mean structure calculated among the 20 lowest energy structures out of 50 refined structures over these core regions for backbone atoms is 0.64 ± 0.09 Å. However, necessarily poor convergence is present in the linker region (residues 26-54). As discussed in sections 5.4.3 and 6.3.2 the peak of internal mobility of the linker is centered on Gly 40, with the T_2 relaxation profile clearly showing that internal mobility decreases on either side back to core levels over approximately 13

residues (Figure 5.7, B and Figure 6.5, B). Consequently this trend ought to be recapitulated in the ensemble of structures representing Rv1827 pThr 22.

Figure 6.6 (A, left) shows the pairwise r.m.s. deviation from the mean structure calculated among the 100 lowest energy structures out of 500 refined structures for backbone atoms. By increasing the number of structures generated, a greater amount of conformational space can be sampled by the linker region in the ensemble. However, in doing this the pairwise r.m.s. deviation from the mean structure over core regions is driven down to potentially unrealistic accuracy (0.33 ± 0.20 Å for residues 22-25 and 55-149). Nevertheless, the highest pairwise r.m.s. deviation from the mean structure for the linker region is centered approximately on Gly 40, recapitulating the T_2 relaxation profile. Figure 6.6 (A, right) shows the ensemble of Rv1827 pThr 22 solution structures coloured according to the pairwise r.m.s. deviation from the mean structure. Colour ranges of ≤0.1 Å r.m.s. deviation (minimum, blue) and ≥7.5 Å r.m.s. deviation (maximum, red) were used to colour the residues of the molecules in the ensemble.

An easier way to understand the colour mapping is that the most highly converged residue on the core is blue and residues with a pairwise r.m.s. deviation from the mean structure equal to or greater than the 5 most highly disordered residues of the linker are in red. Intermediate levels of convergence are colour-graded in a rainbow spectrum (from blue, turquoise, green, yellow and orange to red. Figure 6.6 (B, left and C, left) show the T_2 and ^{1}H-^{15}N NOE relaxation profiles of Rv1827 pThr 22 coloured according to the same scheme as the pairwise r.m.s. deviation from the mean structure. Residues with T_2 relaxation rate constants ≤48 ms are blue and those ≥230 ms are red. Residues with ^{1}H-^{15}N NOE enhancements ≤0.05 are the reddest and those ≥0.85 are the bluest.

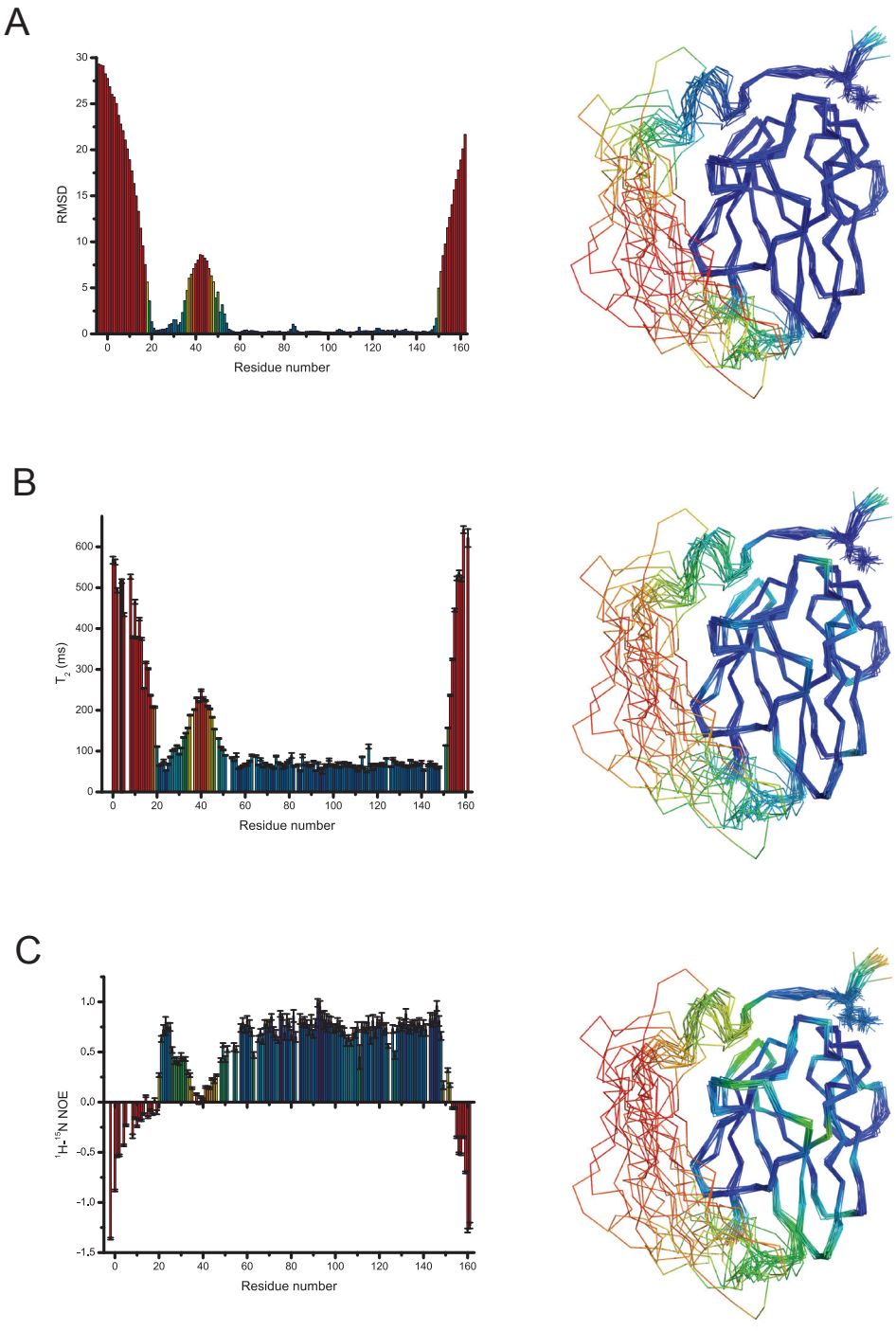

Figure 6.6. Qualitative evaluation of the Rv1827 pThr 22 ensemble of 20 lowest energy structures. Structures are the same as those in Figure 6.5 (D). (A) r.m.s. deviation calculated over backbone atoms for the ensemble of solution structures (left) and the ensemble of structures coloured according to the r.m.s. deviation (≤ 0.1 Å blue, ≥ 7.5 Å red) (right). (B) T_2 relaxation profile of Rv1827 pThr 22 and the ensemble of structures coloured according to T_2 relaxation (≤ 45 ms blue, ≥ 230 ms red). (C) NOE relaxation profile of Rv1827 pThr 22 and the ensemble of structures coloured according to NOE enhancement (≤ 0.05 blue, ≥ 0.85 red).

Again, the colours have been transposed onto the structures to the right of the respective relaxation profiles. This makes it easier to see how well the pairwise r.m.s. deviation from the mean structure of the ensemble reflects the relaxation data.

The FHA domain and the N-terminal phosphorylation site clearly show the highest convergence (blue) and the centre of the linker region clearly shows the lowest convergence (red). When coloured according to the relaxation data, the increased internal mobility of the H1 region with respect to the core is highlighted (by cyan to yellow colouring), although not absolutely reflected by the pairwise r.m.s. deviation from the mean structure, where it is more blue. This difference highlights the difficulty of generating a molecular model that satisfies all facets of the empirical information about a biological molecule.

6.4.2 Qualitative description of the solution structure of Rv1827 pThr 22

The FHA domain of Rv1827 is formed by residues 55-149. They adopt a classical FHA domain fold comprising a compact 11-stranded β-sandwich topology. This is shown in Figure 6.7. Residues forming β-strands are coloured red in the ribbons representation in the top panel and are shown in red cartoon form in the middle panel. No α-helical insertions are present in the loops between β-strands. Therefore the short regions of TALOS-predicted α-helix within the FHA domain correspond to tight turns between the β-strands.

The location and orientation of the most highly conserved N-terminal residues are well defined by a total of 65 long-range NOEs. Overall, residues 21-33 adopt an extended conformation encompassing a short region of helical propensity (H1, blue). This region

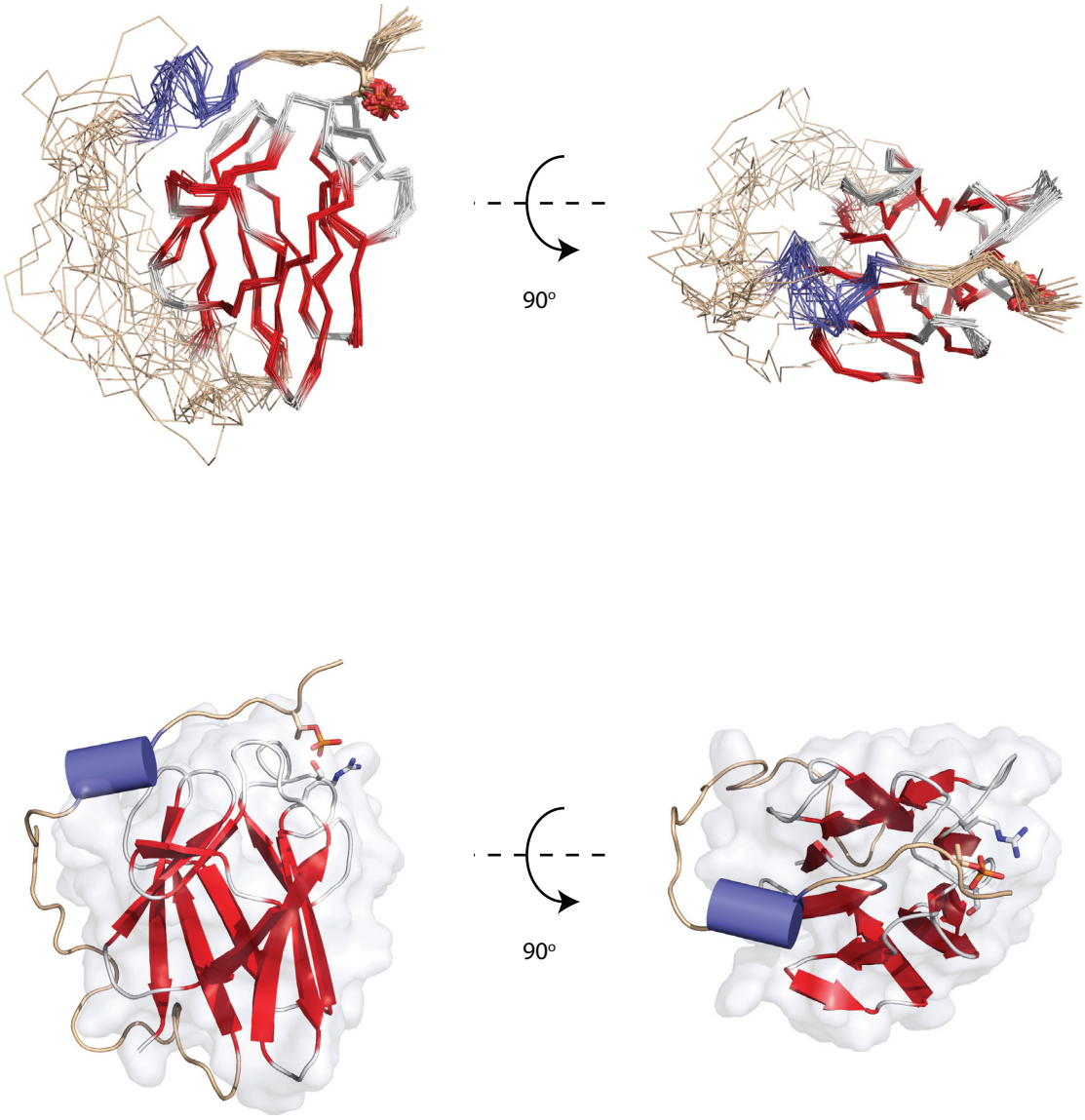

Figure 6.7. Rv1827 pThr 22 final solution structures. top ribbons representation of Rv1827 pThr 22. β-strands are coloured red and loops between strands of the FHA domain are coloured white. The H1 region is coloured blue and pThr 22 is shown as sticks. bottom Cartoon and transparent surface representation of a representative structure of Rv1827 pThr 22. β-strands are shown as red arrows and the H1 region as a blue cylinder. Shown as sticks are pThr 22, Arg 81 and Ser 95.

follows a path across the FHA surface distinct from that previously observed in FHA-phosphopeptide complex structures. This difference is highlighted in Figure 6.8, which shows a comparison between the Rv1827 FHA domain interaction with its phosphorylated N-terminus and the Chk2-phosphopeptide complex (Li et al., 2002). The phosphorylated N-terminus of Rv1827 is orientated almost orthogonal with respect to the β-sandwich faces of the FHA domain compared to the phosphopeptide-FHA domain interaction in Chk2, which is approximately parallel. It is difficult to predict whether the isolated Rv1827 FHA domain would bind a phosphopeptide inter-molecularly in the same orientation as intra-molecularly. It may be that association of the H1 region with a specific auxiliary surface on the FHA domain plays a significant part in orientating the extended conformation C-terminal to the phosphorylation site. Nevertheless, when phosphorylated by PknB the conserved residues of the N-terminus ultimately direct the subsequent ensemble of linker conformations along one side of the FHA β-sandwich.

As previously mentioned, FHA domain selectivity for the residue pThr + 3 is a theme that has emerged from both the identification of *in vivo* phosphopeptide binding partners and *in vitro* phosphopeptide library screens. Rv1827 is an interesting case in that two adjacent N-terminal threonines are specifically phosphorylated by two different kinases. When either threonine is phosphorylated, the pThr + 3 residue conforms to the preferred Rv1827 FHA domain selectivity criterion. If Thr 21 is phosphorylated by PknG, Val 24 is the pThr + 3 residue and if Thr 22 is phosphorylated by PknB, Phe 25 is the pThr + 3 residue. Both are relatively large, hydrophobic amino acids and it seems plausible that the binding orientation of the N-terminus would look extremely similar in either case. Indeed this may explain the increased ability of trypsin to cleave Arg 26

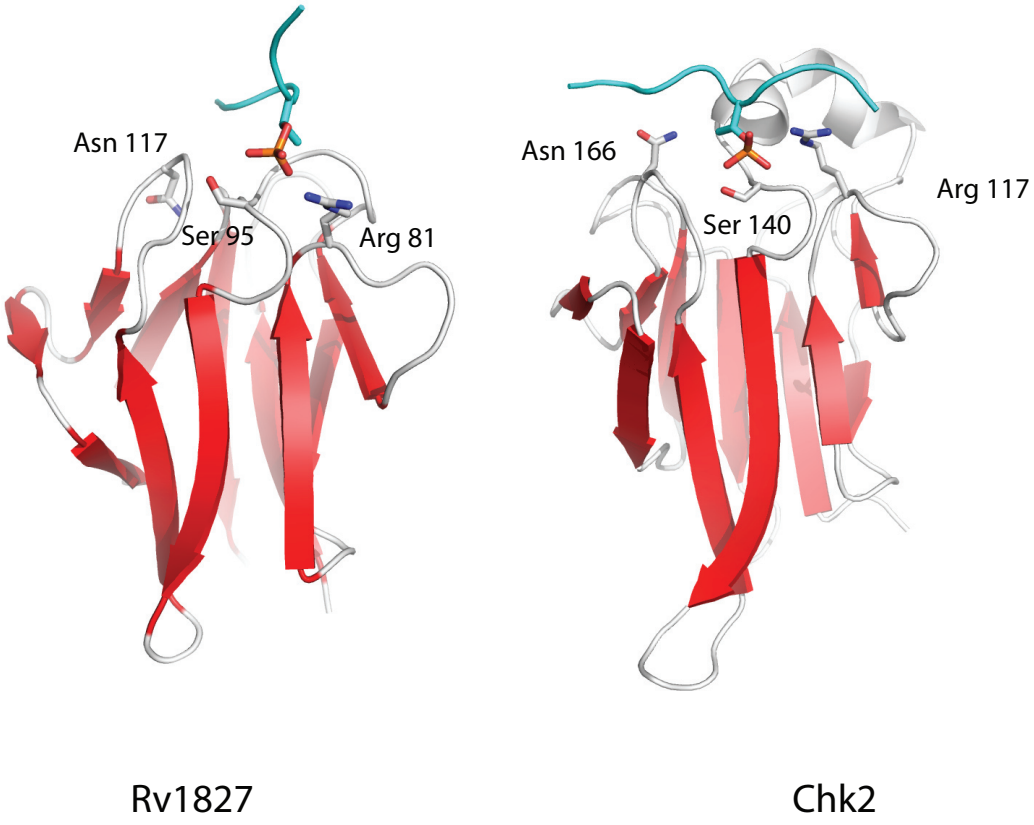

Figure 6.8. Comparison of the phosphopeptide-binding orientations of Rv1827 and Chk2. The FHA domains of Rv1827 and Chk2 are shown in cartoon representation with white loops connecting red β-strands. The respective phosphopeptide ligands are coloured cyan. FHA domain residues important for phosphopeptide binding are shown as sticks, as is the pThr residue of the ligands. Chk2 FHA domain binds phosphopeptides approximately parallel with the planarity of the β-sheets, whereas Rv1827 FHA domain binds its intramolecular phosphopeptide ligand almost orthogonal to the planarity of the β-sheets.

when Rv1827 is phosphorylated at Thr 21, by shifting the register of the extended conformation such that Arg 26 is less well constrained.

6.5 A model of the 'pre-bound' conformation of unphosphorylated Rv1827

It was suggested in section 5.4.1 that in the unphosphorylated state, the region of high sequence conservation within the N-terminus of Rv1827 associates with the FHA domain. The occupancy of this 'pre-bound' state has yet to be characterized, but it is clear that the exchange regime between the completely unbound and pre-bound conformations does not substantially impede the high affinity association of the FHA domain with its target enzymes. It was further proposed that conformations in the unphosphorylated, pre-bound state resemble those in the phosphorylated form, whereby interactions between H1 and the FHA domain stabilize the pThr 22 interaction with the phosphate binding site (Figure 5.10). From this it seemed reasonable to imagine that the conformation of the H1 region in both the unphosphorylated pre-bound and phosphorylated forms of Rv1827 might be very similar.

Based on this assumption, a model of the unphosphorylated, pre-bound conformation of Rv1827 was generated in order to help visualize this form of the protein. For these calculations, distance restraints were based on those used to calculate the structures of Rv1827 pThr 22 presented in the previous sections. However, all restraints were removed for residues N-terminal of Arg 26 (NOE-derived distance restraints and pThr-FHA domain hydrogen bonds). Additionally no statistically significant TALOS predictions were made for residues N-terminal of the FHA domain, supporting the lack of defined secondary structure. The intra-regional restraints for the H1 region (shown as cyan in Figure 6.5 and described in section 6.3.2) were unchanged, as were the long-range

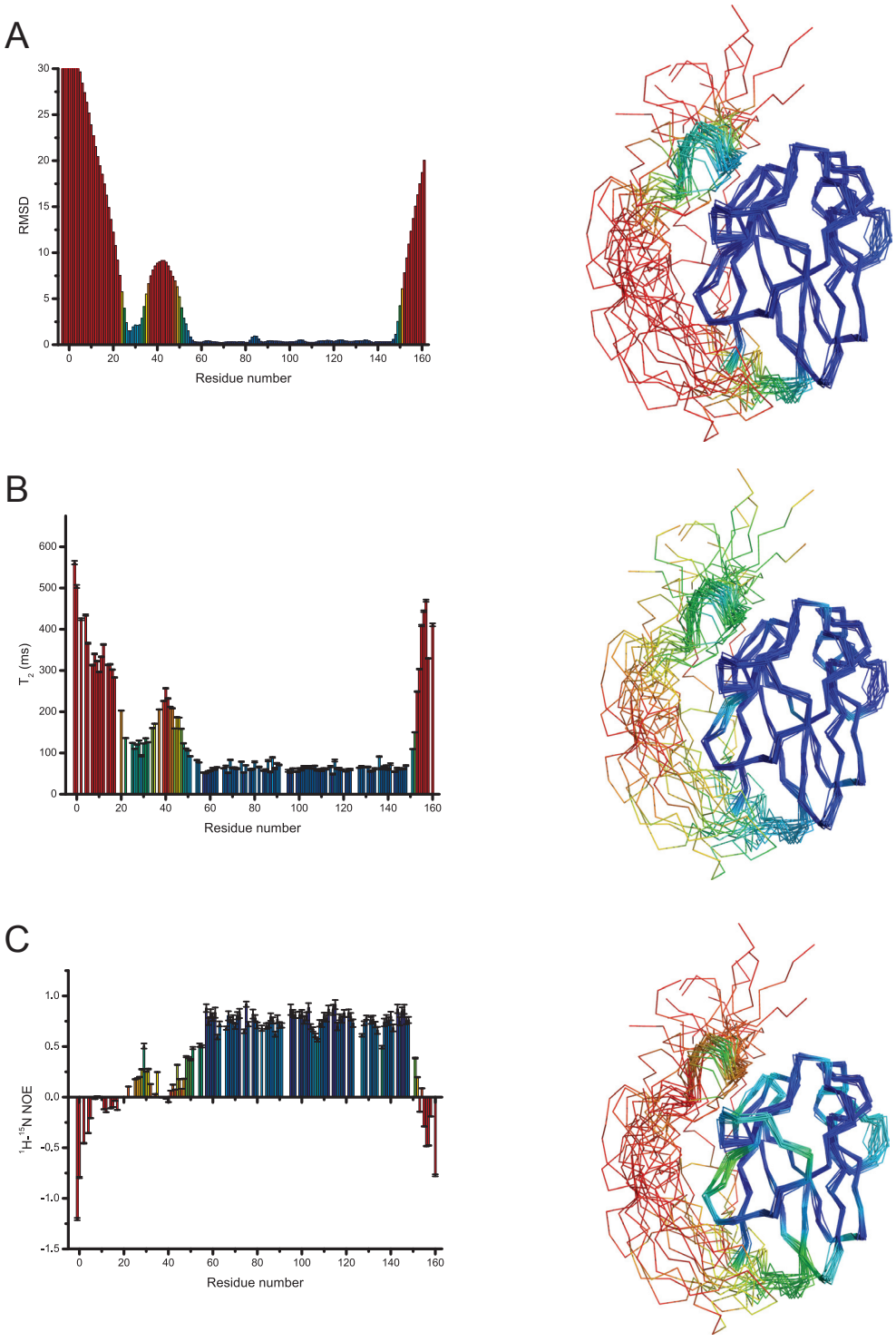

Figure 6.9. Qualitative evaluation of the model of the 'pre-bound' form of unphosphorylated Rv1827. (A) r.m.s. deviation calculated over backbone atoms for the ensemble of solution structures (left) and the ensemble of structures coloured according to the r.m.s. deviation (≤ 0.1 Å blue, ≥ 7.5 Å red) (right). (B) T_2 relaxation profile of Rv1827 pThr 22 and the ensemble of structures coloured according to T_2 relaxation (≤ 45 ms blue, ≥ 230 ms red). (C) NOE relaxation profile of Rv1827 pThr 22 and the ensemble of structures coloured according to NOE enhancement (≤ 0.05 blue, ≥ 0.85 red).

distance restraints between H1 and the FHA domain. The model generated is shown in Figure 6.9, and is coloured according to the same scheme and is in the same format as Figure 6.6. The structures of the pre-bound form of unphosphorylated Rv1827 are aligned over residues 55-149 only. This necessarily produced greater divergence in the H1 region and illustrates its weaker association with the FHA domain than in the phosphorylated form of the protein.

As with the solution structures of Rv1827 pThr 22, colouring the ensemble of models of unphosphorylated Rv1827 according to the pairwise r.m.s. deviation from the mean molecule, T_2 and ^{1}H-^{15}N NOE relaxation profiles, the differential internal mobility over various regions is easier to appreciate. The FHA domain has extremely low internal mobility, and the peak of internal mobility of the linker region is centered on Gly 40. The H1 region adopts approximately the same intra-regional conformation as in the phosphorylated state. This leads to an interesting hypothesis about the mechanism of phosphorylation of Rv1827 by PknB.

We know that PknB activation loop phosphorylation is required for full kinase activity, and that the FHA domain of Rv1827 needs to bind to the phosphorylated activation loop to be phosphorylated N-terminally (Villarino et al., 2005). The approximate distance between the activation loop phosphorylation sites and the catalytic site of the kinase domain is somewhere between 15-25 Å. It is not possible to know for sure as the electron density in the crystal structures of PknB kinase domain is very poor in this region, suggesting high internal mobility. This order of distance is remarkably similar to distances between the site of phosphate binding on the FHA domain and the location of Thr 22 in the unphosphorylated but pre-bound model of Rv1827. It is suggested therefore that as the FHA domain docks onto the phosphorylated activation loop of the

PknB kinase domain, the H1 region effectively increases the local concentration of the phosphorylation target site (Thr 22) at the kinase domain catalytic site. This would lead to more efficient phosphorylation of Thr 22 by PknB. If the N-terminal 54 residues of Rv1287 were totally unrestrained in the unphosphorylated form, the FHA domain might be able to bind and dissociate from the PknB activation loop before the phosphorylation target site had found its way into the catalytic site of the kinase. Under these circumstances, it is imagined that N-terminal phosphorylation of Rv1287 would be significantly less efficient.

So, as well as providing extra stability once N-terminal phosphorylation has been achieved, it may be that latent association of the H1 region with the FHA domain may also promote efficient PknB-mediated phosphorylation.

6.6 Conclusions

The solution structure of Rv1827 pThr 22 demonstrates that N-terminal phosphorylation induces a novel intramolecular conformational compaction in Rv1827 whereby the N-terminal phosphorylation site binds back to surface associated with canonical FHA-phosphopeptide interactions. The N- and C-terminal β-strands of the FHA domain run anti-parallel to each other such that the domain boundary is on the opposite side of the FHA domain from the phosphate-binding surface. This means that in order for pThr 22 to associate with the residues that hydrogen bond to the phosphate group (Arg 81, Ser 95 and Arg 96), the linker region is directed across one of the β-sheets of the domain. This results in an orientation of intramolecular phophopeptide binding that is almost orthogonal to that usually observed in FHA domain-phosphopeptide complexes.

Furthermore, interactions between the H1 region and the FHA domain help to stabilize the intramolecular association and go some way towards explaining the biochemical and biophysical observations outlined in chapter 3. The N-terminal region of high sequence conservation forms an extended conformation that makes contacts with the FHA domain as far as the pThr + 11 position. Thus upon phosphorylation the stabilization of Arg 26 (the pThr + 4 position) may explain its resistance to tryptic proteolysis and the stabilization of the H1 region may explain the small increase in CD signal at 202 nm that suggested the induction of a small region of helix.

It is apparent that there are many caveats and pitfalls when calculating the solution structures of a molecule that contains regions of demonstrably high internal mobility, as evidenced in this case by ^{1}H-^{15}N-relaxation measurements. Relying solely on automatic structure calculation procedures produced ensembles of structures that did not accurately reflect their internal mobility. It was therefore deemed appropriate to apply restraint thresholding procedures manually such that the final ensemble of solution structures satisfied all of the available empirical data.

The initial aim of this work was to understand the mechanism by which N-terminal phosphorylation of Rv1827 abrogates interactions of its FHA domain with its target metabolic enzymes. The solution structure of the phosphorylated form and the model of the unphosphorylated form cannot achieve this, but importantly they can provide a rational for further investigation and experimentation. The identification of the surfaces on the FHA domain of Rv1827 that are used to interact with the metabolic enzymes will be explored in the next chapter.

7 Identification of Rv1827 intermolecular interaction surfaces

7.1 Overview

Chapter 4 began with the identification of Rv1827 interaction partners in *M. tuberculosis* cell-free extracts and the observation that phosphorylation of Rv1827 at Thr 22 by PknB abrogated these interactions. It was subsequently found that in its unphosphorylated form, Rv1827 is a potent inhibitor of the enzymes KGD and GDH, and it is an activator of the GS complex (Chapter 5) (Nott et al., 2009). The phosphorylation state of Rv1827 therefore has a significant impact on core biochemical processes surrounding α-ketoglutarate metabolism and nitrogen assimilation in *M. tuberculosis* and the closely related actinomycetes.

In order to understand the conformational changes that Rv8127 undergoes following N-terminal phosphorylation, a detailed biochemical, biophysical, hydrodynamic and structural investigation of the unphosphorylated and phosphorylated forms of the protein was performed. However, despite having several novel and interesting features, the solution structures could not fully explain how in the unphosphorylated state the FHA domain of Rv1827 interacts with the metabolic enzymes. To address this issue a rationally designed mutagenic study on Rv1827 was performed, guided by the solution structures of Rv1827 pThr 22 and the unphosphorylated model.

7.2 Rational design of Rv1827 mutants

Knowing that conformational changes associated with N-terminal phosphorylation abrogated Rv1827 interactions with its metabolic targets, it was hypothesized that the surface used by the FHA domain to bind intermolecularly must in some way be changed

or occluded by the intramolecular association of the phosphorylation site (pThr 21 or 22) and the FHA domain.

The crystal and solution structures of other FHA domains both free and in complex with their respective phosphopeptide ligands show that few, it any, conformational changes are associated with intermolecular FHA domain interactions. Thus FHA domains appear not to change conformation upon ligand binding. It was therefore reasoned that a mutational analysis focused on surfaces of the Rv1827 FHA domain that were occluded through the intramolecular association of the phosphorylated N-terminus would be more fruitful in identifying intermolecular interaction surfaces. From the solution structures of Rv1827 pThr 22 we saw that two partially overlapping surfaces were occluded by intramolecular binding.

The first of these is the canonical FHA-phosphopeptide interaction surface, which is formed by the loops joining β-strands β3/β4, β4/β5, β6/β7 and β10/β11. These loops contain the conserved residues required for FHA-phosphopeptide interactions (Arg 81, Ser 95, Leu 116, Asn 117 and Lys 141). The second surface comprises residues from β7, β10 and β11 and forms a positively charged cleft within which the linker and H1 regions reside (Figure 7.1, A). Residues from this FHA surface were observed to make contacts with, and help to stabilize, the H1 region in the N-terminally phosphorylated form of Rv1827. The most significant contributions to this stabilization were made by the FHA domain residues Lys 141, Arg 143 and Gln 138 (see section 6.3.2 for more details). Thus as well as being required for intramolecular phosphopeptide interactions, Lys 141 also forms part of this secondary intramolecular interaction surface.

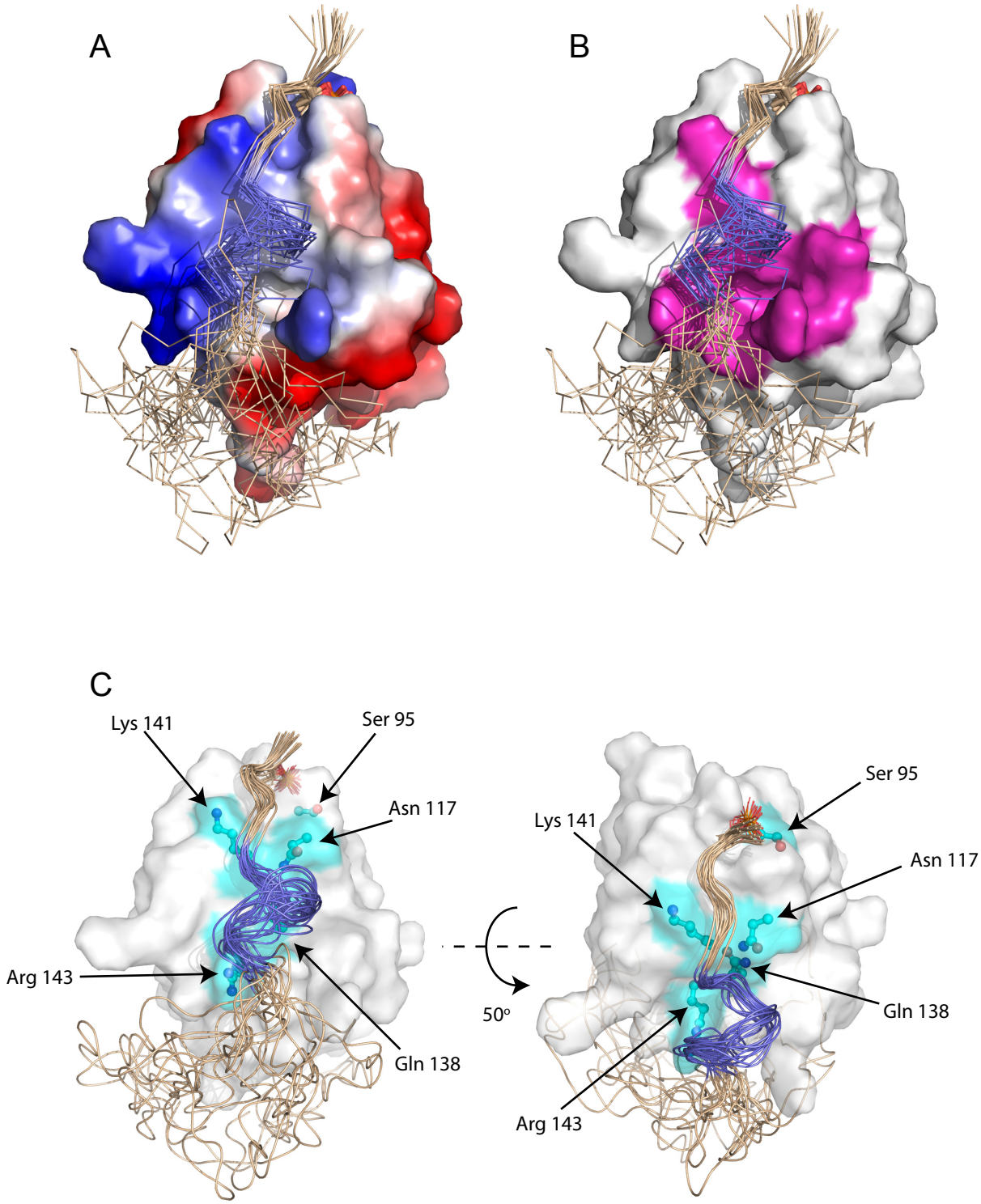

Figure 7.1. Rational design of Rv1827 mutants. All structures in this figure are of the same ensemble (as in Figure 6.7) and are in the same relative orientation. (A) Surface charge of Rv1827 pThr 22 FHA domain (blue: positive, red: negative) with the N-terminal region shown in ribbons representation (H1 region is light blue). (B) Rv1827 FHA domain surface (white) with the location of the conserved van Doren surface highlighted in pink. (C) The location of residues targeted in site directed mutatgenesis (stick and ball representation with cyan surface) shown relative to the N-terminal phosphorylation site (pThr 22 shown in lines representation), the H1 region (blue) and the linker region both in cartoon representation. The location of residues targeted in for mutagenesis largely coincide with the predicted van Doren surface.

The FHA domain region formed by β7, β10 and β11 was identified by bioinformatics as a conserved surface (the van Doren surface) and potentially therefore is of conserved function (see section 1.5.4 and Figure 7.1, B). This and its stabilizing contacts with the H1 region made it an attractive target for mutagenesis and hinted at a possible mechanism by which N-terminal phosphorylation abrogated intermolecular interactions. Could it be that in Rv1827, the van Doren surface is employed for intermolecular interactions and that N-terminal phosphorylation induces stabilization of the H1 region such that this surface is occluded?

Based on the net positive charge of the β-sheet that is occluded by H1 in the solution structures of Rv1827 pThr 22, it was possible that electrostatic interactions were involved in the Rv1827 FHA-enzyme interactions, so we targeted charged residues of this region. We created mutant Rv1827 constructs that contained individual Lys 141 and Arg 143 substitutions for Glu (K141E and R143E), and a double mutant that contained both mutations (K141E/R143E). We also created a mutation of Gln 138 to Arg (Q138R) in order to investigate whether adding a basically charged residue to the region would have any effect on intermolecular interactions. The design of mutagenic primers and the site-directed mutagenesis reactions were performed by Lasse Stach (NIMR).

We saw previously that mutation of the conserved Rv1827 FHA domain phosphate-binding Ser 95 to Ala (S95A) mutation reduced interactions with GDH and eliminated interactions with the β-subunit (GltB) of the GS complex (see section 4.4). We wondered if mutation of another conserved residue would have the same effect. We therefore mutated Asn 117 to Ala (N117A), which has been demonstrated in other FHA domains to significantly reduce their ability to bind to phosphopeptides. The locations of all the residues mutated are shown in Figure 7.1 (C).

7.3 Identification of intermolecular Rv1827 interaction surfaces

In an attempt to characterize the effects that the Rv1827 FHA domain mutations had on its intermolecular interactions with KGD, GDH and GltB (the β-subunit of the GS complex) we performed GST-fusion pull-down assays. These experiments took the same form as those described previously (in section 4.4), and the results are shown in Figure 7.2 and summarized in Table 7.1. All of the Rv1827-fusion constructs used in the pull down assays contained the full length form of Rv1827 (162 residues), which was N-terminally fused to GST.

	Wild type	K141E	R143E	K141E/ R143E	Q138R	N117A	S95A
GDH	+	-	-	-	+	+	-
GltB (GS)	+	-	+	-	+	-	-
KGD	+	-	-	-	+	-	+

Table 7.1. Summary of mutant Rv1827 interactions with KGD, GltB (GS complex) and GDH.

7.4 Effects of mutation of conserved FHA domain residues

As seen previously, wild-type, unphosphorylated Rv1827 bound to KGD, GDH and GltB (Figure 4.1 and Figure 7.2). The S95A mutant associated strongly with KGD but showed reduced binding to GDH and eliminated interactions with GltB. The N117A mutation caused interactions with both KGD and GltB to be lost, but association with GDH was maintained at approximately wild-type levels. It seems therefore that two

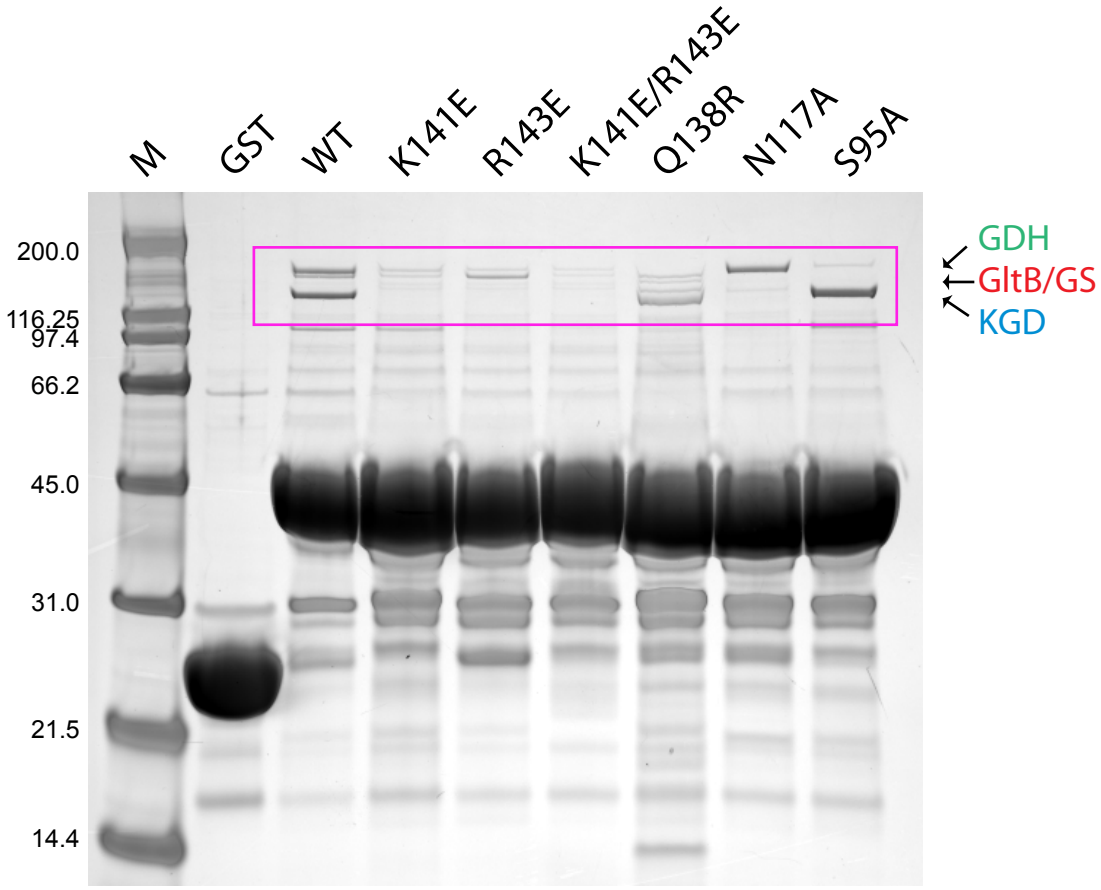

Figure 7.2. Identification of intramolecular and intermolecular Rv1827 interaction surfaces. Pull-down assays with WT Rv1827 and the six indicated mutants of Rv1827.

conserved FHA domain residues that are essential for phosphopeptide interactions are independently required for two different phospho-independent interactions. In Rv1827, Ser 95 is required for association with GltB and GDH (but not KGD) and Asn 117 is required for association with KGD and GltB (but not GDH).

7.5 Effects of mutation of non-conserved FHA domain residues

Interestingly, a differential pattern of intermolecular interactions was also observed for the other Rv1827 mutants. In the case of the K141E and K141E/R143E double mutants, the interactions with all three of the metabolic enzymes were virtually eliminated. However, the R143E mutation maintained interaction with GltB at wild type levels but lost the ability to bind to KGD and GDH. Therefore Arg 143 is required for Rv1827 interactions with KGD and GDH (but not GltB) and Lys 141 is required for effective interactions with all three metabolic targets.

The Q138R mutant maintained interactions with all of the enzymes and, in addition, an interacting protein of ~145 kDa was observed on the pull down gel (Figure 7.2). As yet, identification of this high molecular weight protein has not been attempted. It would appear that changing the surface charge of the FHA domain at the region associated with H1-FHA interactions significantly alters the ability of Rv1827 to interact with its three metabolic targets. Decreasing the positive charge of this area reduces intermolecular interactions and substituting a negatively charged residue for a positively charged one promotes interactions. In the case of the Rv1827 FHA domain, mutational analysis of the conserved van Doren surface has, for the first time, demonstrated that this region forms a functional protein-protein interaction surface.

7.6 Biophysical and kinetic analysis of Rv1827 mutants

To investigate whether the Rv8127 mutants maintained their differential modulatory activity as well as binding to their metabolic targets, a subset of them were tested in both surface plasmon resonance (SPR) experiments (by Dr. Jiejin Li at NIMR) and enzyme activity assays (by Dr. Helen O'Hare at Leicester University). The double mutant containing both K141E/R143E substitutions was not used in these experiments as mutation of Lys 141 was shown by the pull down assays to severely restrict interactions with all three enzymes. Similarly the Q138R mutant was not used as it did not significantly change the interaction pattern with the enzymes compared to the wild type Rv1827 protein. In these experiments, recombinant KGD and GDH (supplied by Dr. Helen O'Hare) were used and the activity of the GS complex was tested in *M. smegmatis* cell free lysate. The results of these assays were normalized to the unphosphorylated Rv1827 wild type values and are shown in Figure 7.3 (A), and summarized in Table 7.2.

	Wild type	K141E	R143E	N117A	S95A
GDH	+	-	-	+	-
GltB (GS)	+	-	+	-	-
KGD	+	-	-	-	+

Table 7.2. Combined summary of the effects on enzyme activity (KGD, GltB (GS complex) and GDH) and SPR binding (to KGD and GDH) of the different Rv1827 FHA domain mutants.

The relative SPR binding results largely recapitulate the trend observed in the pull down assays. Both Lys 141 and Arg 143 are required for association with KGD and GDH,

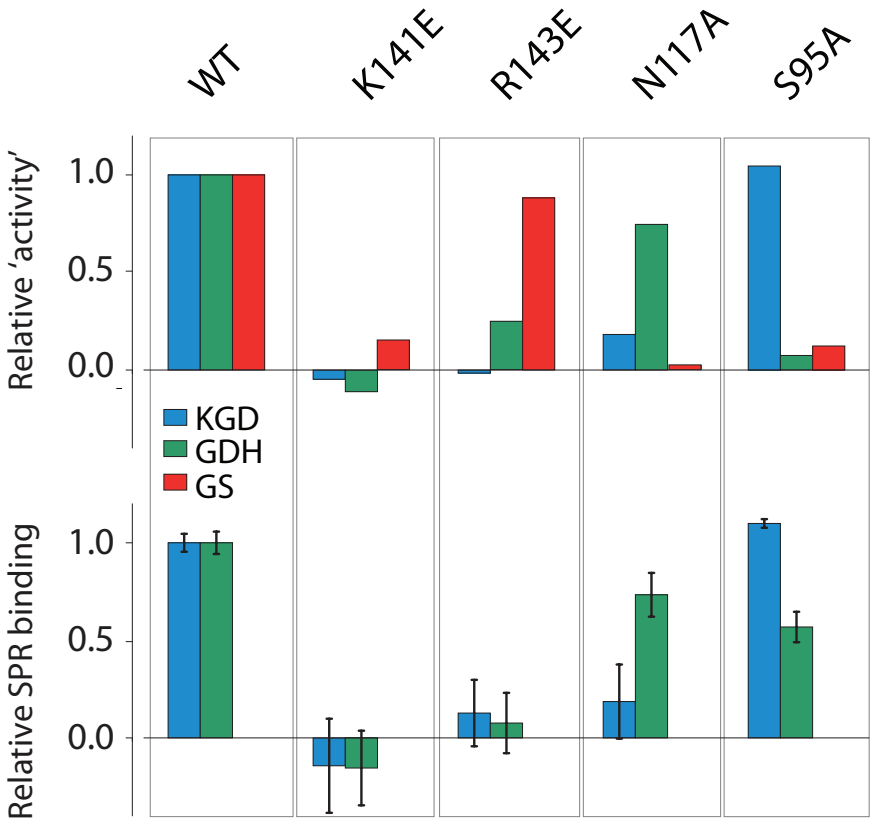

Figure 7.3. Confirmation of intramolecular and intermolecular Rv1827 interaction surfaces. top **The "relative activity" of Rv1827 and its mutants at 4 μM on KGD (blue), GDH (green), and GS (red), normalised to that of the WT protein. Full titrations are shown in (Nott et al., 2009).** lower **Relative SPR binding to immobilised KGD and GDH. Values are the mean of three independent measurements and error bars represent the SEM.**

and Rv1827 interactions with KGD and GDH are relatively insensitive to mutation of Ser 95 and Asn 117, respectively. The Rv1827 K141E mutation is not effective against any of the three enzymes, again demonstrating that it is required for functional intermolecular FHA domain interactions. Reflecting the trend seen in the pull down assay, the activity of the GS complex, GDH and KGD are insensitive to mutation of Arg 143, Asn 117 and Ser 95, respectively.

7.7 Conclusions

These results confirm that the Rv1827 FHA domain surface involved in regulatory intra-molecular interactions is coincident with that employed in the formation of each of the enzyme complexes. The inter-molecular binding events appear to be specific since Rv1827 interactions with KGD, GDH and GS are sensitive to different combinations of three of the four Rv1827 mutations. However, each pairwise interaction is relatively insensitive to mutation at one of the positions (S95A, N117A or R143E), and only K141E has a significant effect on binding to all three enzymes. The differential effects on binding/inhibition/activation of S95A, N117A and R143E preclude the possibility that these mutations are merely acting to either stabilise the 'pre-bound' conformation or globally disrupt FHA folding, since inter-molecular interactions with all three enzymes would be impeded in both scenarios.

Therefore the apparently complex pattern of mutational effects evident in Figure 7.3 are interpreted in terms of a core FHA binding surface that is subtly tailored to accommodate three structurally diverse enzyme targets in a location that is responsive to regulatory Rv1827 phosphorylation (Figure 7.4). These results reveal a clear biological role for phospho-independent interactions in FHA domain function that may

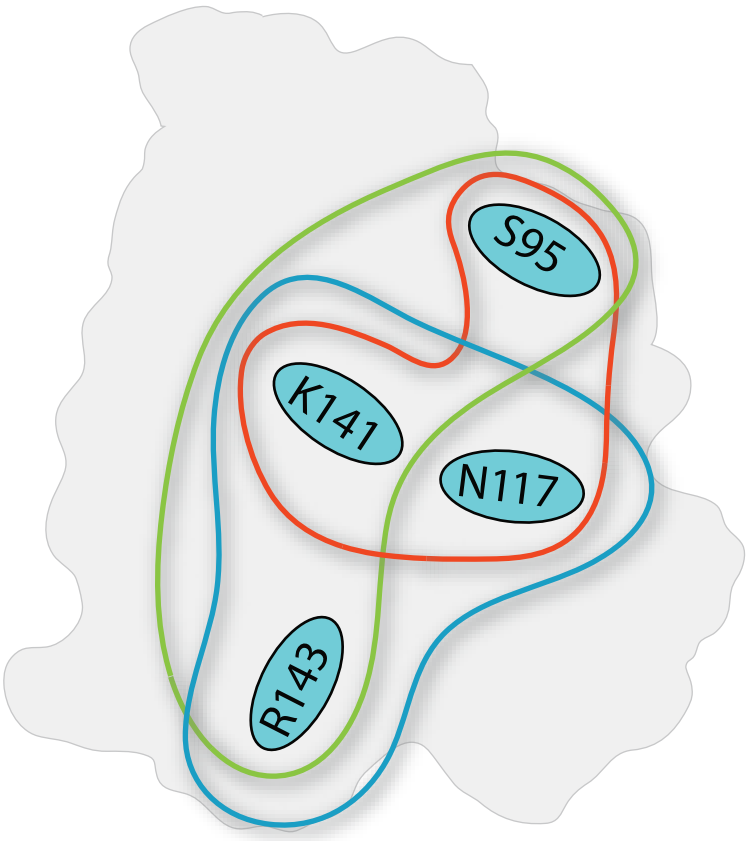

Figure 7.4. Rv1827 FHA domain uses three partially overlapping surfaces to bind intermolecularly. The effects on binding and enzyme inhibition or activation represented as three distinct but substantially overlapping subregions of the FHA surface. KGD, GDH, and GS each use three of the four mutated positions to specifically bind to Rv1827. Lys 141 within the basic FHA region is the only residue whose mutation affects all three interactions.

nonetheless involve residues integral to the canonical pThr binding surface. Consequently, it is now suggested that abrogation of interactions through mutation of highly conserved pThr-binding residues cannot necessarily be taken as *prima facie* evidence for phosphorylation-dependent FHA binding in the absence of supporting data.

8 Discussion

8.1 Rv1827 is a novel FHA domain-containing protein

Modular signaling architectures have evolved to direct varied biological activities through combinatorial linkage of multiple domains with specific function within a single molecule (Pawson, 2007). Many such examples are now known and FHA domains themselves are found in a variety of modular signaling molecules where their apparent phospho-binding function is complemented by other interaction domains and/or enzymatic activities (Li et al., 2002; Oliver et al., 2006; Zhang and Durocher, 2008).

In contrast to such multidomain systems, Rv1827 is a small protein that comprises only a single FHA domain and a seemingly unstructured N-terminal tail region. It nonetheless acts as an activator or inhibitor to regulate three core metabolic enzymes. These differential modulatory effects are summarized in Figure 8.1, where dashed lines represent Rv1827-mediated inhibition of KGD- and GDH-catalyzed reactions (blue and green, respectively) and the bold red line represents activation of GS complex activity. Rv1827 is conserved only across the actinomycetes and, in spite of its deceptively simple architecture, plays an important role in determining carbon flux and nitrogen assimilation.

In chapters 4 and 5 we saw that the phosphorylation of a single site in the N-terminus of Rv1827 abrogates its interactions with its metabolic targets and is associated with substantial changes in its biophysical properties. Whereas unphosphorylated Rv1827 is readily cleaved by trypsin at Arg 26, phosphorylation of Thr 21 (by PknG) and Thr 22 (by PknB) lend significant protection to this tryptic site. Additionally, PknB-mediated phosphorylation confers significant thermal stability to Rv1827, increasing the

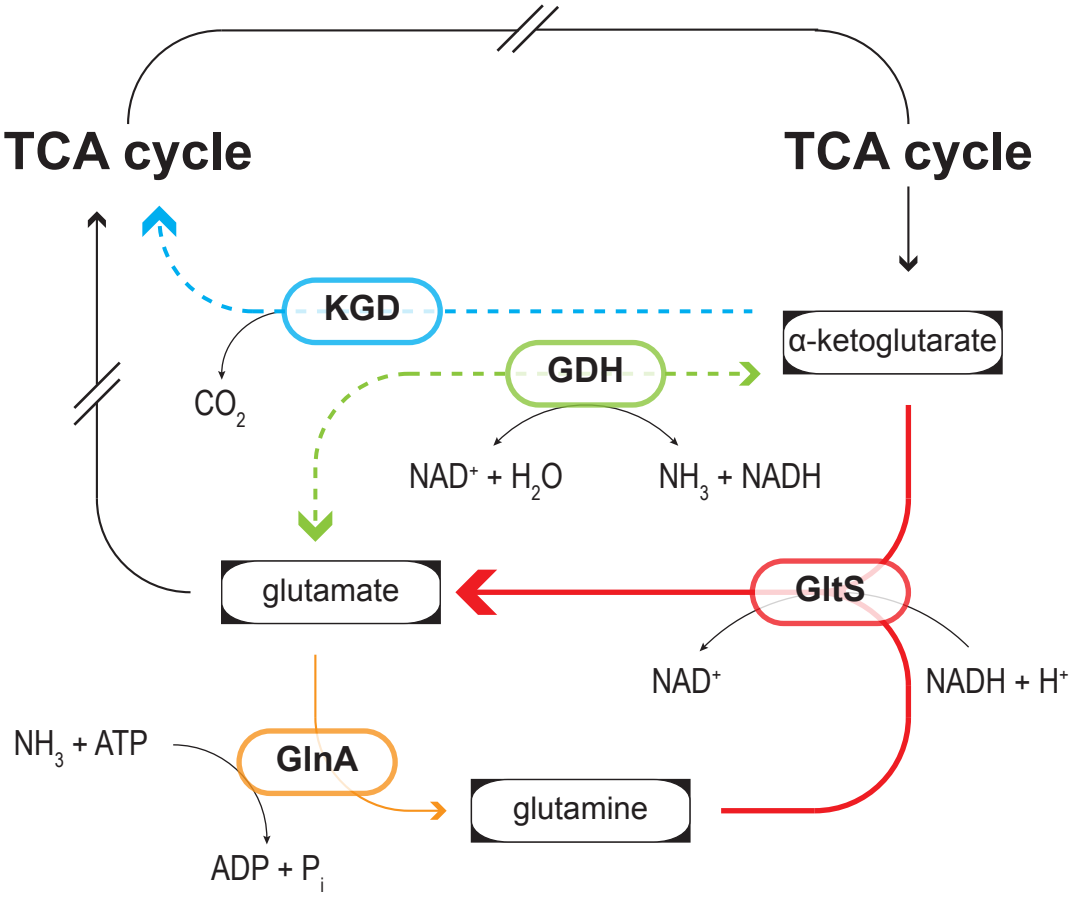

Figure 8.1. Summary of Rv1827 activity on glutamate metabolism and nitrogen assimilation in M. tuberculosis. The three enzymes regulated by Rv1827 lie at the crossroads of the TCA cycle. KGD (blue) decarboxylates α-ketoglutarate. GDH (green) and GS (red)/GlnA (glutamine synthetase, orange) function in the low- and high-affinity pathways of ammonia assimilation, respectively. Inhibition of KGD and GDH (dashed lines) by Rv1827 and its activation of GS (bold red lines) alter the flux of α-ketoglutarate towards glutamate synthesis.

temperature at which its secondary structure is completely denatured from 60 to 80°C. It was found that this increase in stability is brought about by intramolecular association of the N-terminal phosphorylation site with the FHA domain, contrasting the initial hypothesis that phosphorylation would induce Chk2-like dimerisation (Li et al., 2002).

We then investigated the internal mobility and backbone dynamics of unphosphorylated Rv1827 and the conformationally more compact PknB-phosphorylated form using NMR relaxation measurements. When phosphorylated by PknB, the N-terminal pThr 22 site binds back to the FHA domain and becomes highly ordered. Thus the phospho-dependent intramolecular association confers stability across both the phosphorylation site and the FHA domain. Remarkably the ^{1}H-^{15}N relaxation measurements showed that the N-terminal region in unphosphorylated Rv1827 was not as unstructured as we had previously thought. It fact, it adopts a 'pre-bound' conformation, in which there is transient association of residues centered around Ala 27 with those on the FHA domain surface. This surprising result suggested that phosphorylation of Thr 21 (by PknG) or Thr 22 (by PknB) acted to stabilize a pre-existing intramolecular interaction of the N-terminus with the FHA domain. To understand the nature of the intramolecular association of the N-terminal region with the FHA domain, a detailed structural analysis using NMR spectroscopy was subsequently performed.

Rv1827 and the PknB-phosphorylated form were ideally suited to hydrodynamic and structural characterization by NMR spectroscopy due to their relatively small size. Both forms were highly soluble and stable for long periods of time at room temperature. It was also fortunate that homogeneous, mono-disperse samples of each species could be readily generated. However, the calculation of solution structures of Rv1827 pThr 22 and models of the unphosphorylated form presented several challenges. The significant

proportion of disordered/transiently structured residues (~ 40%) meant that standard methodology failed to produce ensembles of structures that reflected the high internal mobility of the region between the N-terminal phosphorylation site and the FHA domain. To address this issue, the T_2 relaxation and NOE enhancement hydrodynamic data were employed to threshold the distance restraints over different regions of the molecule. The implementation of this strategy is discussed in detail in section 6.3.2. The quality of ensembles of solution structures was then assessed by comparing the average pairwise r.m.s. deviation (for backbone atoms) from the average molecule to the T_2 relaxation profiles of Rv1827 and the PknB-phosphorylated form.

The final ensembles of Rv1827 pThr 22 solution structures and models of the unphosphorylated form have several interesting and novel structural features. Whereas the core Rv1827 FHA domain topology is highly similar to those of other FHA domains (an 11 β-stranded twisted β-sandwich), its mode of phosphopeptide binding is significantly different. In the majority of FHA domain-phosphopeptide complex structures solved to date, the phosphopeptide runs along the binding surface parallel to the planarity of the β-sheets. However, despite using the conserved Arg 81 and Ser 95 to co-ordinate the phosphate of pThr 22, the direction of the ensuing peptide chain is skewed towards and over one of the Rv1827 FHA domain β-sheets. A possible explanation for this may be that the contacts between the H1 region and surface formed by β7 and the top portions of β10 and β11 act to extend the canonical phospho-dependent interaction surface into a partially overlapping secondary interaction surface. Thus the intramolecular interaction that is induced upon PknB-mediated phosphorylation is stabilized by both phospho-dependent and phospho-independent FHA domain binding activity.

Remarkably, it appears that the conserved van Doren surface (comprising β7 and the top portions of β10 and β11) has multiple functionalities in Rv1827. Not only is it employed for phospho-independent modulatory interactions of unphosphorylated Rv1827 with its metabolic targets, but is conversely used to stabilize its own kinase-sponsored occlusion. It may even promote efficient PknB-mediated phosphorylation by localizing the kinase target region close to the catalytic site upon interaction of the FHA domain with the phosphorylated activation loop. Furthermore, the H1 region could act like an entropic spring to effectively absorb or dampen the shift in intramolecular phosphopeptide binding register when either Thr 21 or Thr 22 is phosphorylated. Thus virtually identical structural, and therefore functional, responses could be elicited by the signaling activity of either PknG or PknB through phosphorylation of Rv1827.

8.2 A mechanistic insight into the regulation of metabolism

We have derived a molecular description of the mechanism by which Rv1827 activity within core metabolism is regulated through phosphorylation-mediated signaling. Rv1827 is a focal point for signaling cross-talk serving as a substrate for at least two STPKs; PknB, an essential regulator of cell morphology (Fernandez et al., 2006; Kang et al., 2005), and PknG, a virulence factor contributing to *M. tuberculosis* survival in mice and infected macrophages (Cowley et al., 2004; Walburger et al., 2004). Integration of multiple signaling inputs is made possible through a common structural response to PknB or PknG activity that sponsors a regulatory intramolecular interaction between the Rv1827 FHA domain and pThr motifs at its N-terminus. In doing so, intramolecular binding results in occlusion of the FHA domain surfaces used for intermolecular interactions (Figure 8.2). Thus phosphorylation-mediated signaling is coupled to core metabolic activity through an Rv1827 intramolecular "off-switch".

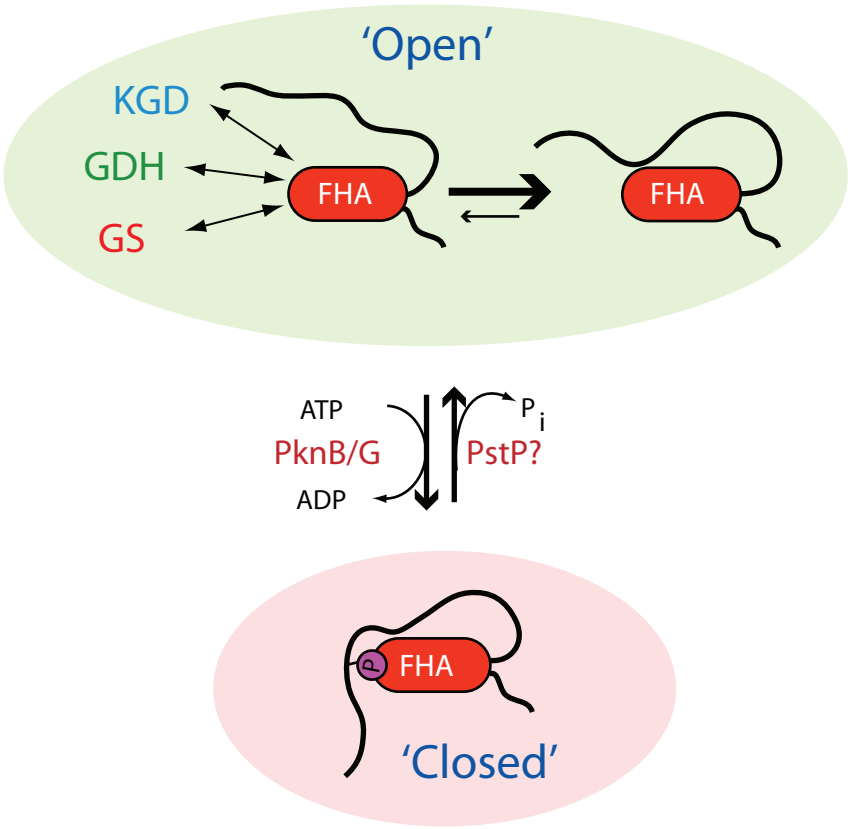

Figure 8.2. Model of the regulation of target enzymes by Rv1827. Rv1827 adopts multiple conformations in the "open" form. Pre-binding of the N-terminal linker is highly stabilized in the "closed", inactive conformation by phosphorylation by PknB (this study) or PknG (O'Hare et al., 2008), disrupting the formation of the enzyme complex. Return to the "open" structure may then arise through action of the MTB Ser/Thr phosphatase, PstP.

Intra-molecular association has been observed previously in modular SH2 and SH3 domain-containing proteins such as Src-family kinases (Boggon and Eck, 2004), Crk (Kobashigawa et al., 2007; Rosen et al., 1995) and Itk (Andreotti et al., 1997) but such a mechanism has not been previously described for any FHA domain protein. Furthermore, in Src, Crk and Itk, intra-molecular binding mainly acts to control accessibility of pTyr (SH2) or PxxP (SH3) motif interaction surfaces on the SH2 or SH3 domains, thereby modulating kinase targeting and activity.

In contrast, it has been shown here how intra-molecular binding can simultaneously modulate FHA domain interactions with both canonical pThr-motif ligands and, more significantly, multiple non-phosphorylated partners, substantially expanding the known repertoire of FHA domain activities and regulatory mechanisms.

Rv1827 is a potent inhibitor/activator of its metabolic targets and uses partially overlapping binding surfaces on its FHA domain to exert such control. Both conserved FHA domain residues and those from a bioinformatically predicted conserved surface are employed for intermolecular Rv1827 binding events. Furthermore these interactions do not require the prior phosphorylation of the Rv1827 binding partners. This demonstrates that as well as being specific for pThr-containing ligands within the context of phosphorylation-mediated signaling, FHA domains can also possess the capability of engaging in phospho-independent protein-protein interactions.

When investigating the activity of KGD in the presence of Rv1827, O'Hare and colleagues found that an Rv1827 concentration of 53 nM gave half maximal KGD enzymatic activity when then the enzyme was at 57 nM. This suggested an apparent affinity for the inhibitory interaction in the mid- to high-nanomolar range, although due

to assay sensitivities the authors found it difficult to accurately determine the affinities of the interactions (O'Hare et al., 2008). The authors also found that the inhibition was apparently non-competitive for α-ketoglutarate. This is suggestive of an allosteric mechanism of inhibition, as Rv1827 reduced the maximal rate of the chemical reaction (V_{max}) without changing the apparent binding affinity of the catalyst for the substrate. Conversely O'Hare and colleagues found that Rv1827 appeared to competitively inhibit GDH by lowing its affinity for its substrate.

It is tempting to speculate about the reasons why two different kinases signal through the same substrate to alter metabolic flux. It may be that *M. tuberculosis* has evolved similar metabolic responses to different environmental factors or stimuli (Figure 8.3). Explanations may lie with the possibility that PknG and PknB use different mechanisms for phosphorylation of Rv1827. From the available data is seems reasonable to infer that in the absence of kinase activity, and if its cytosolic concentration approaches or exceeds that of its interaction partners, Rv1827 will most likely be complexed with the three metabolic enzymes. The function of kinase signaling in this context is metabolic disinhibition/disactivation through Rv1827 phosphorylation. Therefore, the efficiency of the kinases will depend on their ability to phosphorylate Rv1827 in different preformed metabolic complexes (KGD-Rv1827, GDH-Rv1827 and GS-Rv1827).

We know that binding of the FHA domain of Rv1827 to the phosphorylated activation loop of PknB is required for efficient phosphorylation of Rv1827 at Thr 22. The FHA domain-PknB activation loop interaction is mediated by the canonical phospho-binding FHA surface. Therefore if the phospho-binding surface is partially or completely occluded when Rv1827 is bound to GDH or the GS complex (remember that interaction

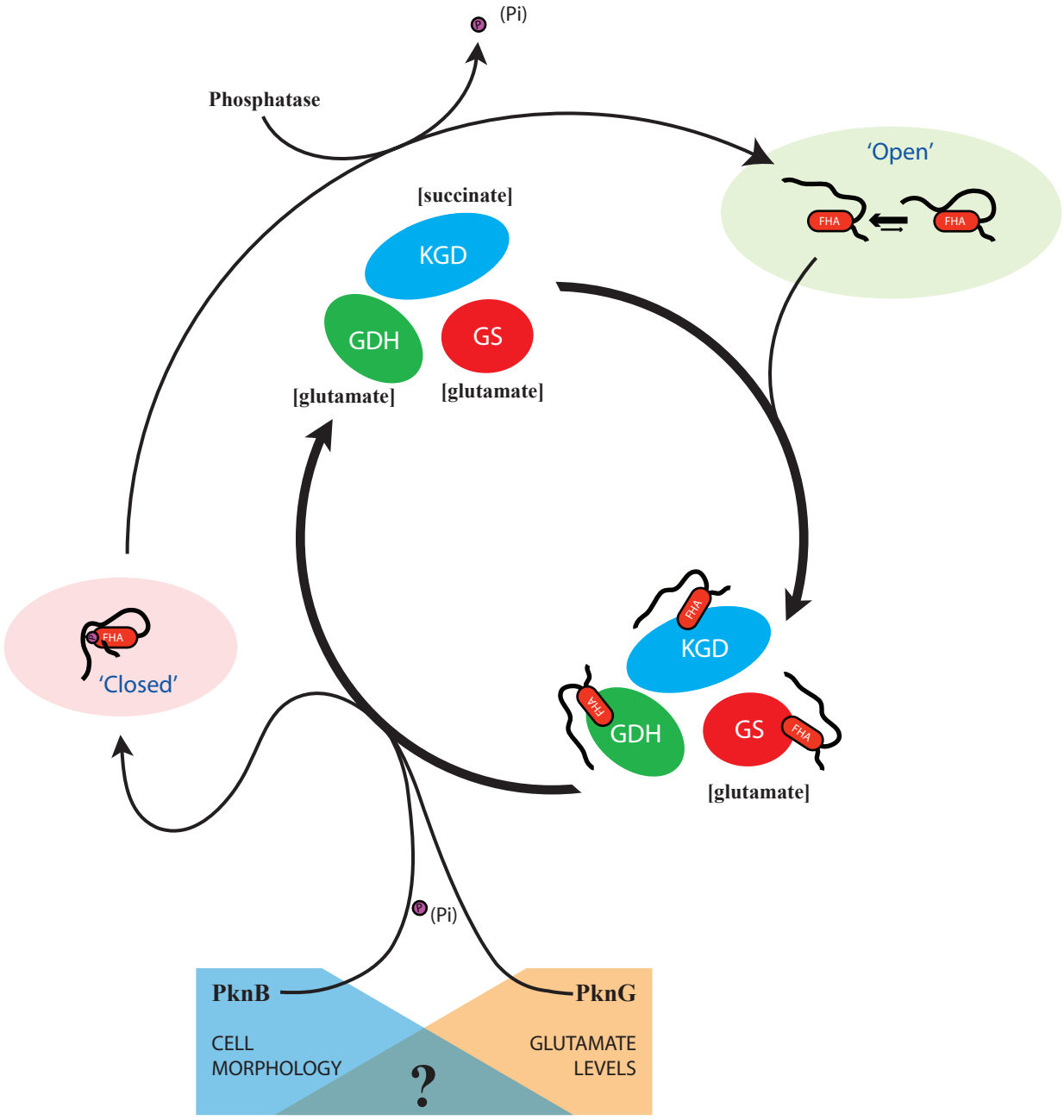

Figure 8.3. Phosphorylation signaling-mediated metabolic regulation. The activity of PknG, PknB, and PstP controls carbon flux and nitrogen assimilation through regulation of the phosphorylation state of Rv1827.

with KGD was insensitive to the S95A mutation), then presumably the efficiency of phosphorylation of Rv1827 by PknB will be greatly reduced. Conversely, PknB-mediated phosphorylation of Rv1827 in complex with KGD may be less attenuated as the phospho-binding surface of the FHA domain may be more exposed and can bind more readily to the PknB activation loop.

As PknG does not contain phosphorylation sites in its activation loop it seems that it may employ a different mechanism. Phosphorylation of the N-terminal region of PknG does not affect its efficiency of phosphorylating a peptide representing its Rv1827 target site, but does increase the efficiency with which PknG can phosphorylate full length Rv1827 (O'Hare et al., 2008). However, the N-terminal phosphorylation sites of PknG that promote Rv1827 phosphorylation are significantly further from the kinase active site than those used by PknB for efficient kinase-substrate docking. Similarly the PknA phosphorylation sites at the C-terminal end of PknG are a significant distance from the catalytic site (Fiuza et al., 2008; O'Hare et al., 2008; Scherr et al., 2007). In this way the PknG target site in the N-terminal region of Rv1827 may take longer to find its way into the active site of the kinase after kinase-substrate docking. Thus dependence on the accessibility of the Rv1827 phospho-binding surface when it is complexed with the different metabolic enzymes may prove to be a mechanism by which these two kinases subtly differ in their regulatory functions. Elucidation of the mechanisms by which PknG and PknB regulate disinhibition/disactivation of metabolic processes through phosphorylation of enzyme-complexed Rv1827 is a natural next step in understanding the phosphorylation signaling mechanisms of *M. tuberculosis*. However, equally important is identification of the stimuli that initially activate the kinases.

8.3 Future perspectives

It has been shown here how phosphorylation signaling can bring about changes in *M. tuberculosis* core metabolic processes, but we still know relatively little about the functional significance of such changes for survival of the organism. Outstanding areas of research include identification of the signals that bring about the change of dormant populations of mycobacteria to actively replicating populations. Recent advances in the implementation of chemostatic systems are beginning to address this issue. Here, mycobacteria can be cultured under tightly controlled conditions that simulate those in specific host environments, such as the human macrophage. Their responses to defined changes in environmental conditions can then be measured by proteome wide mass spectrometry (Rao et al., 2008; Shui et al., 2009), transcriptomics and metabolomics (Halouska et al., 2007).

Further characterization at the structural level of the inhibition of KGD and GDH by Rv1827 and the mechanism of its activation of the GS complex might provide useful hypotheses or predictions regarding metabolic changes associated with dormant/actively replicating transitions. As the three metabolic enzymes with which Rv1827 interacts are likely to form extremely large, highly symmetric multi-protein complexes, cryo-electron microscopy may be a useful tool for this task.

Currently, little is known about the functionality of other Ser/Thr protein kinase networks in *M. tuberculosis*, although we have become aware that characterization of the phospho-proteome by mass spectrometry has recently commenced and, surprisingly, the GroEl chaperone has been found to be a substrate of the *M. tuberculosis* Ser/Thr protein kinases. Such research may in turn lead to testable hypotheses regarding the activity of

other FHA domain-containing proteins, their cognate kinases and interaction partners. If Rv1827 is a representative example of their mechanistic and functional diversity, this will hopefully increase our understanding of the properties of signaling networks in *M. tuberculosis* and lead to the identification of novel therapeutic targets.

9 Appendix

9.1 Sequence alignment of Rv1827 across the actinomycetes

Residues in this alignment are colored according to type.

```
                                        25                        50
                                        .                         .
M. tuberculosis  MTDMNPDIEKDQTSDEVTVETTSVFRADFLSELDAPAQAGTESAVSGVEG 50
M. leprae        MTDMDSGRQEDQTSDEVTVETTSVFRADFLNELDAPAQAGAESVVSGVEG 50
M. marinum       MTDMDAARDRDQTSDEVTVETTSVFRADFLNELDAPTQSGTESSVSGVEG 50
M. ulcerans      MTDMDAARDRDQTSDEVTVETTSVFRADFLNELDAPTQSGTESSVSGVEG 50
M. abscessus     ---MVTDNDKDDTSGEVTAETTSVFRADFATELEAPAQAGAD--VSGVEG 45
M. smegmatis     MTDKDSNLGADQ-SEDVTVETTSVFRADFLNELDAPAAAGTEGAVSGVEG 49
C. amycolatum    ---MTS----NSGQPEASVETTSVFRADLLKEMES--SSNTDSTVSGVEG 41
C. glutamicum    ---MSD----NNGTPEPQVETTSVFRADLLKEMESS--TGTAPASTGAEN 41
C. efficiens     ---MSD----NTGTPEPQVETTSVFRADLLKEMSG---AGSAPAATGADN 40
C. diphtheriae   ---MSD----NTGAPDVQVETTSVFRADLLKEMESG--AGAATASGSDVT 41
                          :    :  .*********: .*:..   :.:     .

                                        75                        100
                                        .                         .
M. tuberculosis  LPPGSALLVVKRGPNAGSRFLLDQAITSAGRHPDSDIFLDDVTVSRRHAE 100
M. leprae        LLAGSALLVVKRGPNAGSRFLLDQAITSAGRHPDSDIFLDDVTVSRRHAE 100
M. marinum       LPAGSALLVVKRGPNAGSRFLLDQPITSAGRHPDSDIFLDDVTVSRRHAE 100
M. ulcerans      LPAGSALLVVKRGPNAGSRFLLDQPITSAGRHPDSDIFLDDVTVSRRHAE 100
M. abscessus     LPAGSALLVVKRGPNAGSRFLLDQATTSAGRHPDSDIFLDDVTVSRRHAE 95
M. smegmatis     LPSGSALLVVKRGPNAGSRFLLDQPTTSAGRHPDSDIFLDDVTVSRRHAE 99
C. amycolatum    LPAGSALLVVKRGPNAGSRFLLDQAVTTAGRHPDSDIFLDDVTVSRRHAE 91
C. glutamicum    LPAGSALLVVKRGPNAGARFLLDQPTTTAGRHPESDIFLDDVTVSRRHAE 91
C. efficiens     LPAGSALLVVKRGPNAGARFLLDQPTTTAGRHPESDIFLDDVTVSRRHAE 90
C. diphtheriae   PPAGAGMLVVKRGPNAGARFLLDRPTTTAGRHPESDIFLDDVTVSRRHAE 91
                   .*:.:**********:*****:. *:*****:****************

                                        125                       150
                                        .                         .
M. tuberculosis  FRLENNEFNVVDVGSLNGTYVNREPVDSAVLANGDEVQIGKFRLVFLTGP
M. leprae        FRLEGNEFHVVDVGSLNGTYVNREPVDSAVLANGDEVQIGKFRLVFLIGS
M. marinum       FRLEGNEFSVVDVGSLNGTYVNREPVDSAVLANGDEVQIGKFRLVFLTGP
M. ulcerans      FRLEGNEFSVVDVGSLNGTYVNRGPVDSAVLANGDEVQIGKFRLVFLTGP
M. abscessus     FRLDSAEFQVVDVGSLNGTYVNREPVDSAVLANGDEVQIGKFRLVFLTGP
M. smegmatis     FRLEGGEFQVVDVGSLNGTYVNREPVDSAVLANGDEVQIGKFRLVFLTGP
C. amycolatum    FRKNGDSYEVVDVGSLNGTYVNREPKNSAVLANGDEVQVGKFRLVFLIGN
C. glutamicum    FRINEGEFEVVDVGSLNGTYVNREPRNAQVMQTGDEIQIGKFRLVFLAGP
C. efficiens     FRINEGEFEVVDVGSLNGTYVNREPRNSQVLQTGDEIQIGKFRLVFLAGP
C. diphtheriae   FRRQDGSFEVVDVGSLNGTYVNREPRNSEVLSSGDEVQIGKFRLVFIEGP
                 ** :  .: ************** * :: *: .***:*:*******: *

                         160
                         .
M. tuberculosis  KQGEDDGSTGGP 162
M. leprae        KQDDDTGRTAGK 162
M. marinum       KQGDDDGGTGGQ 162
M. ulcerans      KQGDDDGGTGGQ 162
M. abscessus     KSAGDDSASGGQ 157
M. smegmatis     KS-DDSGSNA-- 158
C. amycolatum    ------------ 141
C. glutamicum    AE---------- 143
C. efficiens     AA---------- 142
C. diphtheriae   RA---------- 143
```

9.2 NMR NOESY spectra distance calibration (distance_calibration.csh)

The following awk script was used to calibrate the three NOESY spectra used for structure calculations of Rv1827 pThr 22.

```
grep -B1 "# #" 080819_C13_D2O_NOESY_FINAL_VOL.peaks | grep -v "\-\-" > C13_NOE_D2O_peaks.peaks
grep U C13_NOE_D2O_peaks.peaks | cut -dU -f2 | cut -d" " -f11 > 1Cd.txt
#
grep "# #" C13_NOE_D2O_peaks.peaks > 2Cd.txt
#
paste 2Cd.txt 1Cd.txt > 3Cd.txt
#-------------------------------------------------------------------------------------------------------
#
#-------------------------------------------------------------------------------------------------------#!/bin/csh -f
awk '
BEGIN {n=0;}
{$1=="#";}
{ if ($1=="#" && $7 <7.5e+3)
{printf("assign (segid \"    \" and resid %5s and name %5s) (segid \"    \" and resid %5s and name %5s) 2.3 0.5 3.7  !
%5s\n",$3, $4, $5, $6, $7)}
else
 if ($1=="#" && $7 >=7.5e+3 && $7 <=2.5e+4)
{printf("assign (segid \"    \" and resid %5s and name %5s) (segid \"    \" and resid %5s and name %5s) 2.3 0.5 1.7  !
%5s\n",$3, $4, $5, $6, $7)}
else
 if ($1=="#" && $7 >2.5e+4)
{printf("assign (segid \"    \" and resid %5s and name %5s) (segid \"    \" and resid %5s and name %5s) 2.3 0.5 0.7  !
%5s\n",$3, $4, $5, $6, $7)}

      n++;
} ' 3Cd.txt > 13C_finalnoe_D2O.tbl
#-------------------------------------------------------------------------------------------------------#rm 1Cd.txt 2Cd.txt 3Cd.txt
#-------------------------------------------------------------------------------------------------------
#
#
#
#-------------------------------------------------------------------------------------------------------
grep -B1 "# #" 080819_C13_H2O_NOESY_FINAL_VOL.peaks | grep -v "\-\-" > C13_NOE_H2O_peaks.peaks
grep U C13_NOE_H2O_peaks.peaks | cut -dU -f2 | cut -d" " -f11 > 1Ch.txt
#
grep "# #" C13_NOE_H2O_peaks.peaks > 2Ch.txt
#
paste 2Ch.txt 1Ch.txt > 3Ch.txt
#-------------------------------------------------------------------------------------------------------
#
#-------------------------------------------------------------------------------------------------------
#!/bin/csh -f
awk '
BEGIN {n=0;}
{$1=="#";}
{ if ($1=="#" && $7 <7.5e+4)
{printf("assign (segid \"    \" and resid %5s and name %5s) (segid \"    \" and resid %5s and name %5s) 2.3 0.5 3.7  !
%5s\n",$3, $4, $5, $6, $7)}
else
 if ($1=="#" && $7 >=7.5e+4 && $7 <=3.0e+5)
{printf("assign (segid \"    \" and resid %5s and name %5s) (segid \"    \" and resid %5s and name %5s) 2.3 0.5 3.7  !
%5s\n",$3, $4, $5, $6, $7)}
else
 if ($1=="#" && $7 >3.0e+5)
{printf("assign (segid \"    \" and resid %5s and name %5s) (segid \"    \" and resid %5s and name %5s) 2.3 0.5 3.7  !
%5s\n",$3, $4, $5, $6, $7)}
```

```
      n++;
} ' 3Ch.txt > 13C_finalnoe_H2O.tbl
#-------------------------------------------------------------------------------------------------
#rm 1Ch.txt 2Ch.txt 3Ch.txt
#-------------------------------------------------------------------------------------------------
#
#-------------------------------------------------------------------------------------------------
grep -B1 "# #" 080819_N15_NOESY_FINAL_VOL.peaks | grep -v "\-\-" > 15N_NOE_peaks.peaks
grep U 15N_NOE_peaks.peaks | cut -dU -f2 | cut -d" " -f11 > 1N.txt
#
grep "# #" 15N_NOE_peaks.peaks > 2N.txt
#
paste 2N.txt 1N.txt > 3N.txt
#-------------------------------------------------------------------------------------------------
#
#-------------------------------------------------------------------------------------------------
#!/bin/csh -f
awk '
BEGIN {n=0;}
{$1=="#";}
{ if ($1=="#" && $7 <1.0e+5)
{printf("assign (segid \"    \" and resid %5s and name %5s) (segid \"    \" and resid %5s and name %5s) 2.3 0.5 3.7  !
%5s\n",$3, $4, $5, $6, $7)}
else
 if ($1=="#" && $7 >=1.0e+5 && $7 <=4.0e+5)
{printf("assign (segid \"    \" and resid %5s and name %5s) (segid \"    \" and resid %5s and name %5s) 2.3 0.5 1.7  !
%5s\n",$3, $4, $5, $6, $7)}
else
 if ($1=="#" && $7 >4.0e+5)
{printf("assign (segid \"    \" and resid %5s and name %5s) (segid \"    \" and resid %5s and name %5s) 2.3 0.5 0.7  !
%5s\n",$3, $4, $5, $6, $7)}

      n++;
} ' 3N.txt > 15N_finalnoe.tbl
#-------------------------------------------------------------------------------------------------
rm 1N.txt 2N.txt 3N.txt
#-------------------------------------------------------------------------------------------------
#
#-------------------------------------------------------------------------------------------------
cat 13C_finalnoe_H2O.tbl 15N_finalnoe.tbl 13C_finalnoe_D2O.tbl > UNAMBIG_ASSI_080819_three_TEST.tbl
#
#rm 13C_finalnoe_D2O.tbl 13C_finalnoe_H2O.tbl 15N_finalnoe.tbl
#
#rm C13_NOE_D2O_peaks.peaks C13_NOE_H2O_peaks.peaks 15N_NOE_peaks.peaks
```

9.3 Dihedral angle restraints for Rv1827 pThr 22

Amino acid numbering is +5 with respect to the Rv1827 gene numbering due to the addition of 5 residues at the N-terminus of the protein from the pGEX 6P1 expression vector.

VARS RESID RESNAME PHI PSI DPHI DPSI DIST COUNT CLASS
FORMAT %4d %s %8.3f %8.3f %8.3f %8.3f %8.3f %2d %s

35 L -86.440 -13.930 27.810 27.360 5.750 9 Good
60 S -136.330 153.710 25.310 14.120 21.780 10 Good
61 A -143.290 144.640 8.460 12.340 17.370 10 Good
62 L -134.290 140.590 13.240 15.920 15.900 10 Good
63 L -123.730 135.040 17.720 11.260 14.100 10 Good
64 V -117.430 137.850 13.560 12.840 12.140 10 Good
65 V -75.170 127.460 12.540 8.520 15.330 10 Good
70 N -80.240 -9.250 23.540 24.790 31.950 9 Good
72 G 98.240 -18.230 8.330 9.430 21.190 10 Good
74 R -128.640 148.990 21.260 14.070 17.980 10 Good
75 F -123.790 121.950 12.280 13.200 11.020 10 Good
76 L -101.270 125.710 17.780 16.960 11.230 10 Good
79 Q -113.620 160.270 34.980 15.880 15.500 10 Good
82 T -107.820 141.210 12.120 15.350 20.210 10 Good
83 S -100.510 141.250 17.900 7.070 19.020 10 Good
84 A -131.560 147.070 17.350 19.950 27.970 10 Good
86 R -114.290 135.860 30.420 15.200 23.270 9 Good
87 H -92.440 128.330 22.630 16.490 57.900 10 Good
93 F -97.150 120.310 9.840 7.440 29.450 10 Good
94 L -100.600 117.310 12.440 22.410 28.800 10 Good
99 V -100.880 123.100 21.030 27.160 34.860 10 Good
104 A -135.800 151.660 20.250 14.430 24.190 10 Good
105 E -132.500 146.690 14.460 16.390 16.370 10 Good
109 E -118.040 127.030 23.490 8.160 15.810 9 Good
110 N 52.540 43.070 3.380 5.300 19.630 9 Good
111 N 64.620 18.370 5.770 7.700 14.700 10 Good
112 E -107.370 143.670 15.520 12.290 21.740 10 Good
113 F -122.260 136.910 18.060 9.570 14.150 9 Good
114 N -123.740 150.250 19.770 17.730 17.920 10 Good
115 V -132.040 141.280 12.320 11.090 17.730 10 Good
116 V -116.320 132.720 6.910 9.360 18.470 10 Good
117 D -83.780 143.460 12.240 14.400 21.860 9 Good
120 S -80.120 139.930 19.940 16.850 66.410 9 Good
121 L -56.920 -32.680 7.720 11.070 53.360 10 Good
122 N -94.530 -2.620 19.400 16.900 39.350 10 Good
124 T -79.530 134.470 15.000 12.530 21.870 10 Good
125 Y -114.200 136.340 13.670 15.390 16.150 10 Good
126 V -116.860 124.330 17.240 8.340 16.340 10 Good
130 P -64.910 144.370 9.130 10.440 37.520 10 Good
131 V -106.720 165.250 22.740 7.040 34.700 9 Good
133 S -122.510 147.930 21.590 27.970 14.520 9 Good
137 A -125.910 152.180 25.810 15.290 17.220 10 Good
139 G 96.860 -12.220 6.700 8.400 22.730 10 Good
141 E -124.760 138.010 14.740 11.830 16.700 10 Good
142 V -124.450 120.970 11.720 10.840 13.270 10 Good
143 Q -112.730 123.470 8.510 11.380 15.210 10 Good
144 I -117.040 115.950 12.420 12.630 21.550 10 Good
150 V -123.660 127.860 8.540 8.480 13.270 10 Good

9.4 Hydrogen bond restraints for Rv1827 pThr 22

Amino acid numbering is +5 with respect to the Rv1827 gene numbering due to the addition of 5 residues at the N-terminus of the protein from the pGEX 6P1 expression vector.

assign (resid 100 and name HG) (resid 27 and name P) 3.1 0.5 0.5
assign (resid 100 and name OG) (resid 27 and name P) 4.0 0.5 0.5
assign (resid 86 and name HE) (resid 99 and name O) 2.1 0.5 0.5
assign (resid 86 and name NE) (resid 99 and name O) 3.0 0.5 0.5
assign (resid 86 and name QH2) (resid 27 and name P) 3.1 0.5 0.5
assign (resid 86 and name NH2) (resid 27 and name P) 4.0 0.5 0.5
assign (resid 101 and name HN) (resid 27 and name P) 3.1 0.5 0.5
assign (resid 101 and name N) (resid 27 and name P) 4.0 0.5 0.5
assign (resid 61 and name O) (resid 77 and name HN) 2.1 0.5 0.5
assign (resid 61 and name O) (resid 77 and name N) 3.0 0.5 0.5
assign (resid 75 and name O) (resid 63 and name HN) 2.1 0.5 0.5
assign (resid 75 and name O) (resid 63 and name N) 3.0 0.5 0.5
assign (resid 63 and name O) (resid 75 and name HN) 2.1 0.5 0.5
assign (resid 63 and name O) (resid 75 and name N) 3.0 0.5 0.5
assign (resid 73 and name O) (resid 65 and name HN) 2.1 0.5 0.5
assign (resid 73 and name O) (resid 65 and name N) 3.0 0.5 0.5
assign (resid 65 and name O) (resid 73 and name HN) 2.1 0.5 0.5
assign (resid 65 and name O) (resid 73 and name N) 3.0 0.5 0.5
assign (resid 152 and name O) (resid 62 and name HN) 2.1 0.5 0.5
assign (resid 152 and name O) (resid 62 and name N) 3.0 0.5 0.5
assign (resid 62 and name O) (resid 152 and name HN) 2.1 0.5 0.5
assign (resid 62 and name O) (resid 152 and name N) 3.0 0.5 0.5
assign (resid 150 and name O) (resid 64 and name HN) 2.1 0.5 0.5
assign (resid 150 and name O) (resid 64 and name N) 3.0 0.5 0.5
assign (resid 64 and name O) (resid 150 and name HN) 2.1 0.5 0.5
assign (resid 64 and name O) (resid 150 and name N) 3.0 0.5 0.5
assign (resid 148 and name O) (resid 66 and name HN) 2.1 0.5 0.5
assign (resid 148 and name O) (resid 66 and name N) 3.0 0.5 0.5
assign (resid 151 and name O) (resid 139 and name HN) 2.1 0.5 0.5
assign (resid 151 and name O) (resid 139 and name N) 3.0 0.5 0.5
assign (resid 140 and name O) (resid 151 and name HN) 2.1 0.5 0.5
assign (resid 140 and name O) (resid 151 and name N) 3.0 0.5 0.5
assign (resid 149 and name O) (resid 142 and name HN) 2.1 0.5 0.5
assign (resid 149 and name O) (resid 142 and name N) 3.0 0.5 0.5
assign (resid 142 and name O) (resid 149 and name HN) 2.1 0.5 0.5
assign (resid 142 and name O) (resid 149 and name N) 3.0 0.5 0.5
assign (resid 147 and name O) (resid 144 and name HN) 2.1 0.5 0.5
assign (resid 147 and name O) (resid 144 and name N) 3.0 0.5 0.5
assign (resid 141 and name O) (resid 127 and name HN) 2.1 0.5 0.5
assign (resid 141 and name O) (resid 127 and name N) 3.0 0.5 0.5
assign (resid 125 and name O) (resid 143 and name HN) 2.1 0.5 0.5
assign (resid 125 and name O) (resid 143 and name N) 3.0 0.5 0.5
assign (resid 143 and name O) (resid 125 and name HN) 2.1 0.5 0.5
assign (resid 143 and name O) (resid 125 and name N) 3.0 0.5 0.5
assign (resid 123 and name O) (resid 145 and name HN) 2.1 0.5 0.5
assign (resid 123 and name O) (resid 145 and name N) 3.0 0.5 0.5
assign (resid 134 and name O) (resid 115 and name HN) 2.1 0.5 0.5
assign (resid 134 and name O) (resid 115 and name N) 3.0 0.5 0.5
assign (resid 115 and name O) (resid 134 and name HN) 2.1 0.5 0.5
assign (resid 115 and name O) (resid 134 and name N) 3.0 0.5 0.5
assign (resid 132 and name O) (resid 117 and name HN) 2.1 0.5 0.5
assign (resid 132 and name O) (resid 117 and name N) 3.0 0.5 0.5
assign (resid 109 and name O) (resid 112 and name HN) 2.1 0.5 0.5

assign (resid 109 and name O) (resid 112 and name N) 3.0 0.5 0.5
assign (resid 112 and name O) (resid 109 and name HN) 2.1 0.5 0.5
assign (resid 112 and name O) (resid 109 and name N) 3.0 0.5 0.5
assign (resid 107 and name O) (resid 114 and name HN) 2.1 0.5 0.5
assign (resid 107 and name O) (resid 114 and name N) 3.0 0.5 0.5
assign (resid 114 and name O) (resid 107 and name HN) 2.1 0.5 0.5
assign (resid 114 and name O) (resid 107 and name N) 3.0 0.5 0.5
assign (resid 105 and name O) (resid 116 and name HN) 2.1 0.5 0.5
assign (resid 105 and name O) (resid 116 and name N) 3.0 0.5 0.5
assign (resid 116 and name O) (resid 105 and name HN) 2.1 0.5 0.5
assign (resid 116 and name O) (resid 105 and name N) 3.0 0.5 0.5
assign (resid 103 and name O) (resid 118 and name HN) 2.1 0.5 0.5
assign (resid 103 and name O) (resid 118 and name N) 3.0 0.5 0.5
assign (resid 80 and name O) (resid 108 and name HN) 2.1 0.5 0.5
assign (resid 80 and name O) (resid 108 and name N) 3.0 0.5 0.5
assign (resid 106 and name O) (resid 82 and name HN) 2.1 0.5 0.5
assign (resid 106 and name O) (resid 82 and name N) 3.0 0.5 0.5
assign (resid 82 and name O) (resid 106 and name HN) 2.1 0.5 0.5
assign (resid 82 and name O) (resid 106 and name N) 3.0 0.5 0.5
assign (resid 104 and name O) (resid 84 and name HN) 2.1 0.5 0.5
assign (resid 104 and name O) (resid 84 and name N) 3.0 0.5 0.5
assign (resid 84 and name O) (resid 104 and name HN) 2.1 0.5 0.5
assign (resid 84 and name O) (resid 104 and name N) 3.0 0.5 0.5
assign (resid 83 and name O) (resid 92 and name HN) 2.1 0.5 0.5
assign (resid 83 and name O) (resid 92 and name N) 3.0 0.5 0.5
assign (resid 92 and name O) (resid 85 and name HN) 2.1 0.5 0.5
assign (resid 92 and name O) (resid 85 and name N) 3.0 0.5 0.5
assign (resid 85 and name O) (resid 94 and name HN) 2.1 0.5 0.5
assign (resid 85 and name O) (resid 94 and name N) 3.0 0.5 0.5

9.5 ARIA 1.2 new.html

ARIA 1.2 run setup script for structure calculation in the absence of peaklists.

```
<html>
<head>
<title>ARIA - start</title>
</head>
<body bgcolor=#ffffff>
<h2>Parameters for the start:</h2>
<BR>
<h4><!-- ARIA -->
ARIA_DIR=/dms/prog/bin/aria1.2<BR>
ASPECTRUM_1=15N<BR>
CSI_ERROR=30.0<BR>
HBOND_FILE=/dms/home/tnott/ARIA/FHA_PHT_LOOP_NOE/H_BONDS_080727.tbl<BR>
HET1_1=1<BR>
HET1_2=1<BR>
HET1_3=N<BR>
HET1_4=N<BR>
HET1_5=N<BR>
HET2_1=N<BR>
HET2_2=N<BR>
HET2_3=N<BR>
HET2_4=N<BR>
HET2_5=N<BR>
PPMDHET1_1=0.5<BR>
PPMDPRO1_1=0.07<BR>
PPMDPRO2_1=0.05<BR>
PRO1_1=3<BR>
PRO1_2=2<BR>
PRO1_3=N<BR>
PRO1_4=N<BR>
PRO1_5=N<BR>
PRO2_1=2<BR>
PRO2_2=3<BR>
PRO2_3=N<BR>
PRO2_4=N<BR>
PRO2_5=N<BR>
PROJECT_DIR=/dms/home/tnott/ARIA/FHA_PHT_LOOP_NOE<BR>
RUN_NUMBER=557<BR>
SEQ_PDB_FILE=/dms/home/tnott/ARIA/RV1827p22.seq<BR>
TALOS_ERROR=10.0<BR>
TALOS_FACTOR=2.0<BR>
TALOS_FILE=/dms/home/tnott/ARIA/FHA_PHT_LOOP_NOE/geoff_pred2.tab<BR>
UNAMBIG_TBL=/dms/home/tnott/ARIA/FHA_PHT_LOOP_NOE/UNAMBIG_run232_D_40_51_no_OR.tbl
<BR>
WHICH_PEAKS_1=all<BR>
WHICH_PEAKS_2=all<BR>
WHICH_PEAKS_3=all<BR>
WHICH_PEAKS_4=all<BR>
WHICH_PEAKS_5=all<BR>
WHICH_SETUP=XEASY<BR>
XPEAKS_1=/dms/home/tnott/ARIA/FHA_PHT_LOOP_NOE/null.peaks<BR>
XPROT_1=/dms/home/tnott/ARIA/FHA_PHT_LOOP_NOE/080912_RV1827_MASTER.prot<BR>
submit_save=Save updated parameters<BR>
</h4><!-- ARIA -->
</body>
</html>
```

9.6 ARIA 1.2 simulated annealing (SA) and water refinement protocols from run.cns

```
{================== SA protocol ===================}
{* type of molecular dynamics *}
{+ choice: "torsion" "Cartesian" +}
{===>} md_type="torsion";
{* initial seed for random number generator *}
{* change to get different initial velocities *}
{===>} iniseed=89764443;
{* initial temperature for TAD *}
{===>} tadinit_t=20000;
{* initial temperature for TAD and Cartesian dynamics *}
{===>} carinit_t=2000;
{* final temperature after first cooling step *}
{===>} final1_t=1000;
{* finale temperature after second cooling step *}
{===>} final2_t=50;
{* Cartesian time step *}
{===>} timestep=0.003;
{* factor for timestep and number of steps in TAD *}
{===>} tadfactor=9;
{* initial number of MD steps *}
{===>} initiosteps=20000;
{* number of MD steps for refinement *}
{===>} refinesteps=8000;
{* number of MD steps during first cooling stage *}
{===>} cool1_steps=10000;
{* number of MD steps during second cooling stage *}
{===>} cool2_steps=8000;
{* first iteration for Cartesian refinement (after TAD) *}
{===>} cart_firstit=0;

{* Sort structures accordingly to total energy or sum of restraints energies? *}
{+ choice: "totener" "restener" +}
{===>} filesort="totener";

{================== final water refinement  ===================}
{* Do you want water refinement for the last iteration? *}
{+ choice: "yes" "no" +}
{===>} firstwater="yes";
{* Which solvent do you want to use? *}
{+ choice: "water" "dmso" +}
{===>} solvent="water";
{* number of structures for the water refinement *}
{* refine the n best structures regarding energy  *}
{===>} waterrefine=50;
```

9.7 Molmol macro for aligning Rv1827 pThr 22 and unphosphorylated ensembles

Molmol macro for aligning Rv1827 pThr 22 ensemble of 50 structures:

```
InitAll yes
ReadListPdb ./file.nam
CalcBond 1 0 1
SelectMol 'mol.num=50'
FirstMol 'mol.num=2'
SelectMol ''
SelectBond 'bb and res.num=26..153'
DefPropAtom 'displayed' 'selected'
DefPropBond 'displayed' 'selected'
DefPropDist 'displayed' 'selected'
DefPropPrim 'displayed' 'selected'
DefPropMol 'displayed' 'selected'
SelectAtom 'bb and res.num=26..30,60..153'
Fit to_first 'selected'
```

Molmol macro for aligning unphosphorylated Rv1827 ensemble of 50 structures:

```
InitAll yes
ReadListPdb ./file.nam
CalcBond 1 0 1
SelectMol 'mol.num=50'
FirstMol 'mol.num=2'
SelectMol ''
SelectBond 'bb and res.num=26..153'
DefPropAtom 'displayed' 'selected'
DefPropBond 'displayed' 'selected'
DefPropDist 'displayed' 'selected'
DefPropPrim 'displayed' 'selected'
DefPropMol 'displayed' 'selected'
SelectAtom 'bb and res.num=60..153'
Fit to_first 'selected'
```

9.8 Statistical analysis of Rv1827 pThr 22

PROCHECK-NMR was used to evaluate torsion angles of the final ensemble of structure of Rv1827 pThr 22, and the Ramachandran plot of Φ and Ψ torsion angles is shown in Figure 9.1. A summary of the statistical analysis for the final ensemble of lowest energy structures of Rv1827 pThr 22 are shown in Table 9.1.

	Rv1827 pThr 22
NMR distance and dihedral constraints	
Distance constraints	
Total NOE	1998
Intra-residue	659
Inter-residue	
Sequential (|*i*-*j*|=1)	385
Medium-range (|*i*-*j*| ≥ 2 ≤ 4)	179
Long-range (|*i*-*j*| ≥ 5)	775
Hydrogen bonds	82
Total dihedral angle restraints	100
φ	50
ψ	50
Structure statistics	
Violations (mean and s.d.)	
Distance constraints (>0.5) (Å)	1.778 ± 1.114
Dihedral angle constraints (>5) (deg)	0.389 ± 0.502
Deviations from idealized geometry	
Bond lengths (Å)	0.0058 ± 0.0001
Bond angles (deg)	0.7309 ± 0.0198
Impropers (deg)	2.0702 ± 0.0911
Average pairwise r.m.s.d. (Å)[a]	
Backbone	0.640 ± 0.090
Heavy	1.370 ± 0.130
r.m.s.d. from the mean structure (Å)[a]	0.444 ± 0.066
Residues in most favored regions [A,B,L][b]	70.1%
Residues in additional allowed regions[a,b,l,p][b]	24.5%
Residues in generously allowed regions [~a,~b,~l,~p][b]	3.0%
Residues in disallowed regions[b]	2.4%

Rv1827 pThr 22, PknB-phosphorylated Rv1827 full length; NMR, nuclear magnetic resonance; NOE, nuclear overhauser effect.
[a]Pairwise r.m.s. deviation and r.m.s. deviation from the mean structure were calculated among 20 lowest energy structures out of 50 refined structures (residues 21-25 and 55-149).
[b]Structure quality was analysed using PROCHECKnmr (Laskowski et al, 1996) over structured regions (residues 21-25 and 55-149).

Table 9.1 Rv1827 pThr 22 solution structure statistics for 20 lowest energy structures shown in Figure 6.5 (D).

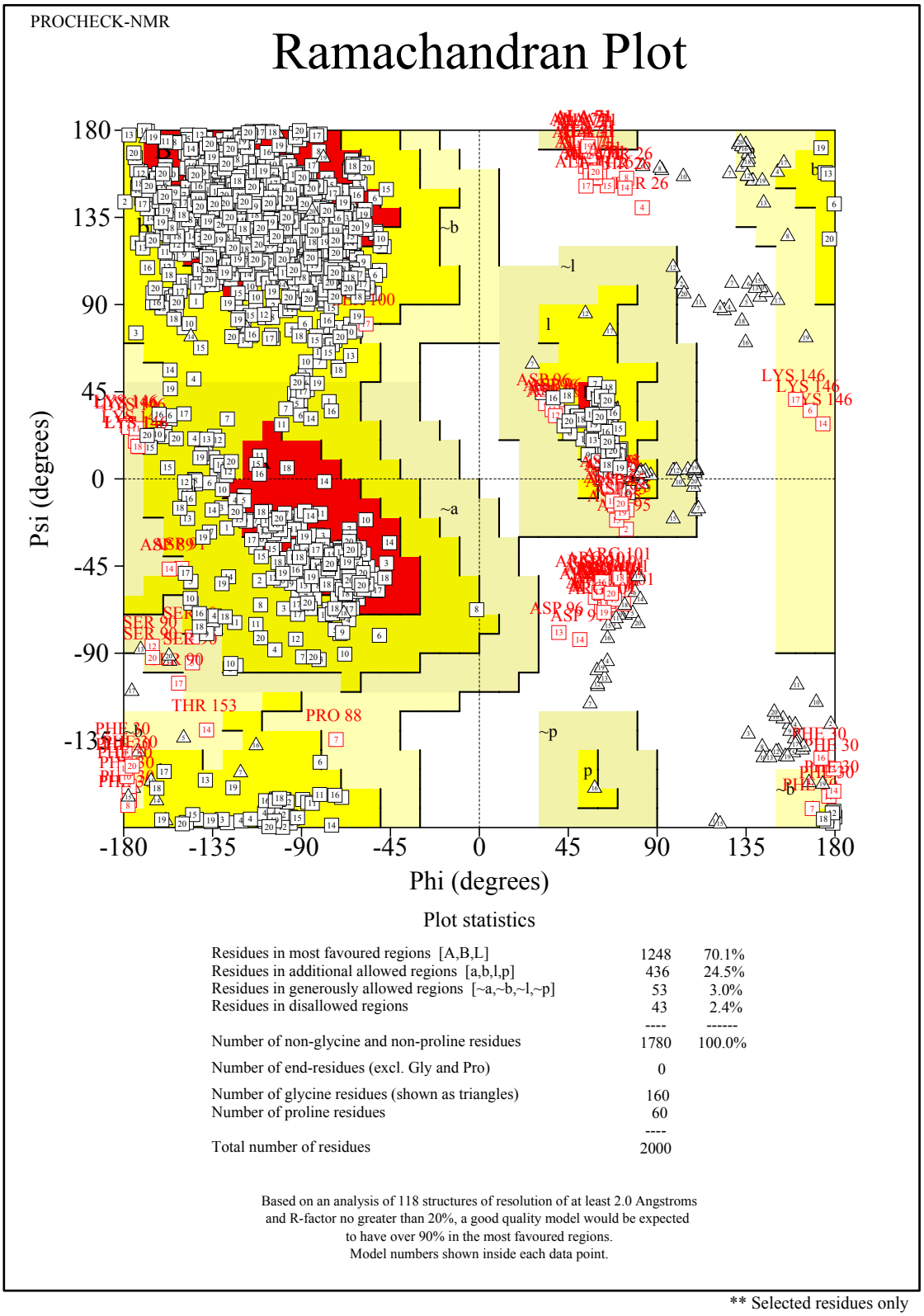

Figure 9.1. Ramachandran plot of the final ensemble of 20 lowest energy structures shown in Figure 6.5 (D).

9.9 Tau values used in T_1 and T_2 relaxation experiments

Rv1827		Rv1827 pThr 22	
T_1 (ms)	T_2 (ms)	T_1 (ms)	T_2 (ms)
8.6	8.6	8.6	8.6
96.9	17.2	96.9	17.2
193.2	34.4	193.2	25.8
393.8	68.9	297.6	34.4
696.6	94.7	393.8	51.6
997.6	120.5	594.5	77.5
1247	146.3	795.1	94.7
1496.4	17.2	995.7	120.5
193.2		1244.6	146.3
		1493.3	34.4
		393.8	

Table 9.2 Tau values used in the T_1 and T_2 relaxation experiments performed on Rv1827 and the PknB-phosphorylated form (Rv1827 pThr 22).

10 Bibliography

Andreotti, A. H., et al., 1997. Regulatory intramolecular association in a tyrosine kinase of the Tec family. Nature. 385**,** 93-7.

Asakura, Y., et al., 2007. Altered metabolic flux due to deletion of odhA causes L-glutamate overproduction in Corynebacterium glutamicum. Appl Environ Microbiol. 73**,** 1308-19.

Atsumi, S., Little, J. W., 2006a. Role of the lytic repressor in prophage induction of phage lambda as analyzed by a module-replacement approach. Proc Natl Acad Sci U S A. 103**,** 4558-63.

Atsumi, S., Little, J. W., 2006b. A synthetic phage lambda regulatory circuit. Proc Natl Acad Sci U S A. 103**,** 19045-50.

Av-Gay, Y., Everett, M., 2000. The eukaryotic-like Ser/Thr protein kinases of Mycobacterium tuberculosis. Trends Microbiol. 8**,** 238-44.

Bax, A., Grzesiek, S., 1993 Methodological advances in protein NMR. Accounts of Chemical Research. 26**,** 131-138.

Bernstein, N. K., et al., 2005. The molecular architecture of the mammalian DNA repair enzyme, polynucleotide kinase. Mol Cell. 17**,** 657-70.

Boggon, T. J., Eck, M. J., 2004. Structure and regulation of Src family kinases. Oncogene. 23**,** 7918-27.

Boitel, B., et al., 2003. PknB kinase activity is regulated by phosphorylation in two Thr residues and dephosphorylation by PstP, the cognate phospho-Ser/Thr phosphatase, in Mycobacterium tuberculosis. Mol Microbiol. 49**,** 1493-508.

Boshoff, H. I., Barry, C. E., 3rd, 2005. Tuberculosis - metabolism and respiration in the absence of growth. Nat Rev Microbiol. 3**,** 70-80.

Bott, M., 2007. Offering surprises: TCA cycle regulation in Corynebacterium glutamicum. Trends Microbiol. 15**,** 417-25.

Brunger, A. T., et al., 1998. Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr. 54**,** 905-21.

Byeon, I. J., et al., 2005. Sequential phosphorylation and multisite interactions characterize specific target recognition by the FHA domain of Ki67. Nat Struct Mol Biol. 12**,** 987-93.

Byeon, I. J., et al., 2001. Solution structure of the yeast Rad53 FHA2 complexed with a phosphothreonine peptide pTXXL: comparison with the structures of FHA2-pYXL and FHA1-pTXXD complexes. J Mol Biol. 314**,** 577-88.

Canagarajah, B. J., et al., 1997. Activation mechanism of the MAP kinase ERK2 by dual phosphorylation. Cell. 90**,** 859-69.

Cohen, P., 2000. The regulation of protein function by multisite phosphorylation--a 25 year update. Trends Biochem Sci. 25**,** 596-601.

Cordwell, S. J., 1999. Microbial genomes and "missing" enzymes: redefining biochemical pathways. Arch Microbiol. 172**,** 269-79.

Cornilescu, G., et al., 1999. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J Biomol NMR. 13**,** 289-302.

Cowley, S., et al., 2004. The Mycobacterium tuberculosis protein serine/threonine kinase PknG is linked to cellular glutamate/glutamine levels and is important for growth in vivo. Mol Microbiol. 52**,** 1691-702.

Dar, A. C., et al., 2005. Higher-order substrate recognition of eIF2alpha by the RNA-dependent protein kinase PKR. Cell. 122**,** 887-900.

De Bondt, H. L., et al., 1993. Crystal structure of cyclin-dependent kinase 2. Nature. 363**,** 595-602.

DeClue, J. E., et al., 1987. A conserved domain regulates interactions of the v-fps protein-tyrosine kinase with the host cell. Proc Natl Acad Sci U S A. 84**,** 9064-8.

Delaglio, F., et al., 1995. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR. 6**,** 277-93.

Dey, M., et al., 2005. Mechanistic link between PKR dimerization, autophosphorylation, and eIF2alpha substrate recognition. Cell. 122**,** 901-13.

Drews, S. J., et al., 2001. A protein kinase inhibitor as an antimycobacterial agent. FEMS Microbiol Lett. 205**,** 369-74.
Dunlap, J. C., 1999. Molecular bases for circadian clocks. Cell. 96**,** 271-90.
Durocher, D., Jackson, S. P., 2002. The FHA domain. FEBS Lett. 513**,** 58-66.
Eckhart, W., et al., 1979. An activity phosphorylating tyrosine in polyoma T antigen immunoprecipitates. Cell. 18**,** 925-33.
Elowitz, M. B., Leibler, S., 2000. A synthetic oscillatory network of transcriptional regulators. Nature. 403**,** 335-8.
Engh, R. A., Bossemeyer, D., 2001. The protein kinase activity modulation sites: mechanisms for cellular regulation - targets for therapeutic intervention. Adv Enzyme Regul. 41**,** 121-49.
England, P., et al., 2008. The FHA-containing protein GarA acts as a phosphorylation-dependent molecular switch in mycobacterial signaling. FEBS Lett.
Fernandez, P., et al., 2006. The Ser/Thr protein kinase PknB is essential for sustaining mycobacterial growth. J Bacteriol. 188**,** 7778-84.
Ferrell, J. E., Jr., 2002. Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. Curr Opin Cell Biol. 14**,** 140-8.
Fiuza, M., et al., 2008. From the characterization of the four serine/threonine protein kinases (PknA/B/G/L) of Corynebacterium glutamicum toward the role of PknA and PknB in cell division. J Biol Chem. 283**,** 18099-112.
Gay, L. M., et al., 2006. A conserved dimer and global conformational changes in the structure of apo-PknE Ser/Thr protein kinase from Mycobacterium tuberculosis. J Mol Biol. 360**,** 409-20.
Gerhart, J., et al., 1984. Cell cycle dynamics of an M-phase-specific cytoplasmic factor in Xenopus laevis oocytes and eggs. J Cell Biol. 98**,** 1247-55.
Gerisch, G., et al., 1975. Control of cell-contact sites by cyclic AMP pulses in differentiating Dictyostelium cells. Nature. 255**,** 547-9.
Gomez, J. E., McKinney, J. D., 2004. M. tuberculosis persistence, latency, and drug tolerance. Tuberculosis (Edinb). 84**,** 29-44.
Gonfloni, S., et al., 2000. Crosstalk between the catalytic and regulatory domains allows bidirectional regulation of Src. Nat Struct Biol. 7**,** 281-6.
Good, M. C., et al., 2004. Sensor domain of the Mycobacterium tuberculosis receptor Ser/Thr protein kinase, PknD, forms a highly symmetric beta propeller. J Mol Biol. 339**,** 459-69.
Gordon, E., et al., 2000. The crystal structure of the penicillin-binding protein 2x from Streptococcus pneumoniae and its acyl-enzyme form: implication in drug resistance. J Mol Biol. 299**,** 477-85.
Greenstein, A. E., et al., 2007. Allosteric Activation by Dimerization of the PknD Receptor Ser/Thr Protein Kinase from Mycobacterium tuberculosis. J Biol Chem. 282**,** 11427-35.
Halouska, S., et al., 2007. Use of NMR metabolomics to analyze the targets of D-cycloserine in mycobacteria: role of D-alanine racemase. J Proteome Res. 6**,** 4608-14.
Hess, B., Boiteux, A., 1971. Oscillatory phenomena in biochemistry. Annu Rev Biochem. 40**,** 237-58.
Higgins, J., 1964. A Chemical Mechanism for Oscillation of Glycolytic Intermediates in Yeast Cells. Proc Natl Acad Sci U S A. 51**,** 989-94.
Hofmann, K., Bucher, P., 1995. The FHA domain: a putative nuclear signalling domain found in protein kinases and transcription factors. Trends Biochem Sci. 20**,** 347-9.
Hu, Y., et al., 2006. Deletion of the Mycobacterium tuberculosis alpha-crystallin-like hspX gene causes increased bacterial growth in vivo. Infect Immun. 74**,** 861-8.
Huang, W., Erikson, R. L., 1994. Constitutive activation of Mek1 by mutation of serine phosphorylation sites. Proc Natl Acad Sci U S A. 91**,** 8960-3.
Hubbard, S. R., 1997. Crystal structure of the activated insulin receptor tyrosine kinase in complex with peptide substrate and ATP analog. Embo J. 16**,** 5572-81.
Hubbard, S. R., et al., 1994. Crystal structure of the tyrosine kinase domain of the human insulin receptor. Nature. 372**,** 746-54.
Huse, M., Kuriyan, J., 2002. The conformational plasticity of protein kinases. Cell. 109**,** 275-82.
Jeffrey, P. D., et al., 1995. Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex. Nature. 376**,** 313-20.

Jones, G., Dyson, P., 2006. Evolution of transmembrane protein kinases implicated in coordinating remodeling of gram-positive peptidoglycan: inside versus outside. J Bacteriol. 188**,** 7470-6.
Kang, C. M., et al., 2005. The Mycobacterium tuberculosis serine/threonine kinases PknA and PknB: substrate identification and regulation of cell shape. Genes Dev. 19**,** 1692-704.
Kay, L. E., et al., 1989. Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. Biochemistry. 28**,** 8972-9.
Kobashigawa, Y., et al., 2007. Structural basis for the transforming activity of human cancer-related signaling adaptor protein CRK. Nat Struct Mol Biol. 14**,** 503-10.
Koch, C. A., et al., 1989. The common src homology region 2 domain of cytoplasmic signaling proteins is a positive effector of v-fps tyrosine kinase function. Mol Cell Biol. 9**,** 4131-40.
Lee, G. I., et al., 2003a. NMR structure of the forkhead-associated domain from the Arabidopsis receptor kinase-associated protein phosphatase. Proc Natl Acad Sci U S A. 100**,** 11261-6.
Lee, S. J., et al., 2003b. Rad53 phosphorylation site clusters are important for Rad53 regulation and signaling. Mol Cell Biol. 23**,** 6300-14.
Levinson, A. D., et al., 1978. Evidence that the transforming gene of avian sarcoma virus encodes a protein kinase associated with a phosphoprotein. Cell. 15**,** 561-72.
Li, H., et al., 2004. 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. J Mol Biol. 335**,** 371-81.
Li, J., et al., 2002. Structural and functional versatility of the FHA domain in DNA-damage signaling by the tumor suppressor kinase Chk2. Mol Cell. 9**,** 1045-54.
Liang, X., Van Doren, S. R., 2008. Mechanistic insights into phosphoprotein-binding FHA domains. Acc Chem Res. 41**,** 991-9.
Lichtarge, O., et al., 1996. An evolutionary trace method defines binding surfaces common to protein families. J Mol Biol. 257**,** 342-58.
Lichtarge, O., Sowa, M. E., 2002. Evolutionary predictions of binding surfaces and interactions. Curr Opin Struct Biol. 12**,** 21-7.
Lin, C. C., et al., 2008. Pellino Proteins Contain a Cryptic FHA Domain that Mediates Interaction with Phosphorylated IRAK1. Structure. 16**,** 1806-16.
Linge, J. P., et al., 2003. ARIA: automated NOE assignment and NMR structure calculation. Bioinformatics. 19**,** 315-6.
Mahajan, A., et al., 2008. Structure and function of the phosphothreonine-specific FHA domain. Sci Signal. 1**,** re12.
Manning, G., et al., 2002a. Evolution of protein kinase signaling from yeast to man. Trends Biochem Sci. 27**,** 514-20.
Manning, G., et al., 2002b. The protein kinase complement of the human genome. Science. 298**,** 1912-34.
Massague, J., Gomis, R. R., 2006. The logic of TGFbeta signaling. FEBS Lett. 580**,** 2811-20.
Miyazono, K., et al., 2004. Coordinate regulation of cell growth and differentiation by TGF-beta superfamily and Runx proteins. Oncogene. 23**,** 4232-7.
Moarefi, I., et al., 1997. Activation of the Src-family tyrosine kinase Hck by SH3 domain displacement. Nature. 385**,** 650-3.
Momiji, H., Monk, N. A., 2008. Dissecting the dynamics of the Hes1 genetic oscillator. J Theor Biol. 254**,** 784-98.
Monk, N. A., 2003. Oscillatory expression of Hes1, p53, and NF-kappaB driven by transcriptional time delays. Curr Biol. 13**,** 1409-13.
Niebisch, A., et al., 2006. Corynebacterial protein kinase G controls 2-oxoglutarate dehydrogenase activity via the phosphorylation status of the OdhI protein. J Biol Chem. 281**,** 12300-7.
Nott, T. J., et al., 2009. An intramolecular switch regulates phosphoindependent FHA domain interactions in Mycobacterium tuberculosis. Sci Signal. 2**,** ra12.
Novak, B., Tyson, J. J., 2008. Design principles of biochemical oscillators. Nat Rev Mol Cell Biol. 9**,** 981-91.

O'Hare, H. M., et al., 2008. Regulation of glutamate metabolism by protein kinases in mycobacteria. Mol Microbiol.
Oliver, A. W., et al., 2006. Trans-activation of the DNA-damage signalling protein kinase Chk2 by T-loop exchange. Embo J. 25**,** 3179-90.
Olsen, L. F., Degn, H., 1977. Chaos in an enzyme reaction. Nature. 267**,** 177-8.
Olsen, L. F., Degn, H., 1978. Oscillatory kinetics of the peroxidase-oxidase reaction in an open system. Experimental and theoretical studies. Biochim Biophys Acta. 523**,** 321-34.
Oppenheim, A. B., et al., 2005. Switches in bacteriophage lambda development. Annu Rev Genet. 39**,** 409-29.
Orr, J. W., Newton, A. C., 1994. Requirement for negative charge on "activation loop" of protein kinase C. J Biol Chem. 269**,** 27715-8.
Ortiz-Lombardia, M., et al., 2003. Crystal structure of the catalytic domain of the PknB serine/threonine kinase from Mycobacterium tuberculosis. J Biol Chem. 278**,** 13094-100.
Pallen, M., et al., 2002. Bacterial FHA domains: neglected players in the phospho-threonine signalling game? Trends Microbiol. 10**,** 556-63.
Pawson, T., 1988. Non-catalytic domains of cytoplasmic protein-tyrosine kinases: regulatory elements in signal transduction. Oncogene. 3**,** 491-5.
Pawson, T., 2004. Specificity in signal transduction: from phosphotyrosine-SH2 domain interactions to complex cellular systems. Cell. 116**,** 191-203.
Pawson, T., 2007. Dynamic control of signaling by modular adaptor proteins. Curr Opin Cell Biol. 19**,** 112-6.
Pawson, T., et al., 1988. Structure-function relationships in cellular and viral fps/fes cytoplasmic protein-tyrosine kinases. Adv Exp Med Biol. 234**,** 55-64.
Pawson, T., Nash, P., 2000. Protein-protein interactions define specificity in signal transduction. Genes Dev. 14**,** 1027-47.
Pawson, T., Nash, P., 2003. Assembly of cell regulatory systems through protein interaction domains. Science. 300**,** 445-52.
Pennell, S., Smerdon, S. J., 2008. Pellino Proteins Splitting Up the FHAmily! Structure. 16**,** 1752-4.
Piotto, M., et al., 1992. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. J Biomol NMR. 2**,** 661-5.
Prabhakaran, K., et al., 2000. Regulation by protein kinase of phagocytosis of Mycobacterium leprae by macrophages. J Med Microbiol. 49**,** 339-42.
Pye, K., Chance, B., 1966. Sustained sinusoidal oscillations of reduced pyridine nucleotide in a cell-free extract of Saccharomyces carlsbergensis. Proc Natl Acad Sci U S A. 55**,** 888-94.
Rao, P. K., et al., 2008. Protein dynamics in iron-starved Mycobacterium tuberculosis revealed by turnover and abundance measurement using hybrid-linear ion trap-Fourier transform mass spectrometry. Anal Chem. 80**,** 6860-9.
Rensing, L., 1972. Periodic geophysical and biological signals as Zeitgeber and exogenous inducers in animal organisms. Int J Biometeorol. 16 Suppl**,** 113-25.
Ricagno, S., et al., 2007. The crystal structure of FdxA, a 7Fe ferredoxin from Mycobacterium smegmatis. Biochem Biophys Res Commun. 360**,** 97-102.
Rodbell, M., 1980. The role of hormone receptors and GTP-regulatory proteins in membrane transduction. Nature. 284**,** 17-22.
Rosen, M. K., et al., 1995. Direct demonstration of an intramolecular SH2-phosphotyrosine interaction in the Crk protein. Nature. 374**,** 477-9.
Sadowski, I., et al., 1988. v-fps protein-tyrosine kinase coordinately enhances the malignancy and growth factor responsiveness of pre-neoplastic lung fibroblasts. Oncogene. 2**,** 241-7.
Sadowski, I., et al., 1986. A noncatalytic domain conserved among cytoplasmic protein-tyrosine kinases modifies the kinase function and transforming activity of Fujinami sarcoma virus P130gag-fps. Mol Cell Biol. 6**,** 4396-408.
Sambrook, J. F. E. F. M., 1989. Molecular Cloning. A Laboratory Manual. . Cold Spring Harbor Laboratory Press.
Sardiwal, S., et al., 2005. A GAF domain in the hypoxia/NO-inducible Mycobacterium tuberculosis DosS protein binds haem. J Mol Biol. 353**,** 929-36.

Scherr, N., et al., 2007. Structural basis for the specific inhibition of protein kinase G, a virulence factor of Mycobacterium tuberculosis. Proc Natl Acad Sci U S A. 104**,** 12151-6.
Schultz, C., et al., 2007. Glutamate production by Corynebacterium glutamicum: dependence on the oxoglutarate dehydrogenase inhibitor protein OdhI and protein kinase PknG. Appl Microbiol Biotechnol.
Sel'kov, E. E., 1968. Self-oscillations in glycolysis. 1. A simple kinetic model. Eur J Biochem. 4**,** 79-86.
Shiloh, M. U., et al., 2008. Mycobacterium tuberculosis senses host-derived carbon monoxide during macrophage infection. Cell Host Microbe. 3**,** 323-30.
Shui, W., et al., 2009. Quantitative proteomic profiling of host-pathogen interactions: the macrophage response to Mycobacterium tuberculosis lipids. J Proteome Res. 8**,** 282-9.
Sicheri, F., et al., 1997. Crystal structure of the Src family tyrosine kinase Hck. Nature. 385**,** 602-9.
Sohaskey, C. D., Wayne, L. G., 2003. Role of narK2X and narGHJI in hypoxic upregulation of nitrate reduction by Mycobacterium tuberculosis. J Bacteriol. 185**,** 7247-56.
Songyang, Z., Liu, D., 2001. Peptide library screening for determination of SH2 or phosphotyrosine-binding domain sequences. Methods Enzymol. 332**,** 183-95.
Sousa, E. H., et al., 2007. DosT and DevS are oxygen-switched kinases in Mycobacterium tuberculosis. Protein Sci. 16**,** 1708-19.
Stavridi, E. S., et al., 2002. Crystal structure of the FHA domain of the Chfr mitotic checkpoint protein and its complex with tungstate. Structure. 10**,** 891-9.
Tian, J., et al., 2005a. Variant tricarboxylic acid cycle in Mycobacterium tuberculosis: identification of alpha-ketoglutarate decarboxylase. Proc Natl Acad Sci U S A. 102**,** 10670-5.
Tian, J., et al., 2005b. Mycobacterium tuberculosis appears to lack alpha-ketoglutarate dehydrogenase and encodes pyruvate dehydrogenase in widely separated genes. Mol Microbiol. 57**,** 859-68.
Ubersax, J. A., Ferrell, J. E., Jr., 2007. Mechanisms of specificity in protein phosphorylation. Nat Rev Mol Cell Biol. 8**,** 530-41.
Villarino, A., et al., 2005. Proteomic identification of M. tuberculosis protein kinase substrates: PknB recruits GarA, a FHA domain-containing protein, through activation loop-mediated interactions. J Mol Biol. 350**,** 953-63.
Walburger, A., et al., 2004. Protein kinase G from pathogenic mycobacteria promotes survival within macrophages. Science. 304**,** 1800-4.
Wang, P., et al., 2000. II. Structure and specificity of the interaction between the FHA2 domain of Rad53 and phosphotyrosyl peptides. J Mol Biol. 302**,** 927-40.
Wayne, L. G., Sohaskey, C. D., 2001. Nonreplicating persistence of mycobacterium tuberculosis. Annu Rev Microbiol. 55**,** 139-63.
Wehenkel, A., et al., 2008. Mycobacterial Ser/Thr protein kinases and phosphatases: physiological roles and therapeutic potential. Biochim Biophys Acta. 1784**,** 193-202.
Wehenkel, A., et al., 2006. The structure of PknB in complex with mitoxantrone, an ATP-competitive inhibitor, suggests a mode of protein kinase regulation in mycobacteria. FEBS Lett. 580**,** 3018-22.
Weldingh, K., Andersen, P., 1999. Immunological evaluation of novel Mycobacterium tuberculosis culture filtrate proteins. FEMS Immunol Med Microbiol. 23**,** 159-64.
Weldingh, K., et al., 1998. Two-dimensional electrophoresis for analysis of Mycobacterium tuberculosis culture filtrate and purification and characterization of six novel proteins. Infect Immun. 66**,** 3492-500.
Wrighton, K. H., et al., 2009. Phospho-control of TGF-beta superfamily signaling. Cell Res. 19**,** 8-20.
Wu, J. W., et al., 2001. Crystal structure of a phosphorylated Smad2. Recognition of phosphoserine by the MH2 domain and insights on Smad function in TGF-beta signaling. Mol Cell. 8**,** 1277-89.
Xu, W., et al., 1997. Three-dimensional structure of the tyrosine kinase c-Src. Nature. 385**,** 595-602.

Yaffe, M. B., Smerdon, S. J., 2004. The use of in vitro peptide-library screens in the analysis of phosphoserine/threonine-binding domain structure and function. Annu Rev Biophys Biomol Struct. 33**,** 225-44.

Yamaguchi, H., Hendrickson, W. A., 1996. Structural basis for activation of human lymphocyte kinase Lck upon tyrosine phosphorylation. Nature. 384**,** 484-9.

Yeats, C., et al., 2002. The PASTA domain: a beta-lactam-binding domain. Trends Biochem Sci. 27**,** 438.

Young, T. A., et al., 2003. Structure of Mycobacterium tuberculosis PknB supports a universal activation mechanism for Ser/Thr protein kinases. Nat Struct Biol. 10**,** 168-74.

Zhang, W., Durocher, D., 2008. Dun1 counts on rad53 to be turned on. Mol Cell. 31**,** 1-2.