Role of the PAS2 domain of the NifL regulatory protein in redox signal transduction

A thesis submitted to the University of East Anglia for the degree of Doctor of Philosophy.

By

Peter Andrew Slavny

Department of Molecular Microbiology

John Innes Centre

Norwich Research Park, Colney Lane, Norwich, NR4 7UH

September 2010

Acknowledgements

I would like to thank Ray (Professor Ray Dixon FRS) for his expert guidance, support and infectious enthusiasm for science, which has been a source of motivation and inspiration throughout my PhD. I would also like to acknowledge the huge contributions made by colleagues (past and present) in the Dixon lab: Richard Little, Paloma Salinas, Matt Bush, Nick Tucker, Marco Schüler de Oliveira and Choni Contreras. Especially Richard for the time and effort he expended training me to work in a laboratory, it was an extensive task. Thanks also to my advisor, Prof. Mark Buttner, for his help and support. The AUC analysis presented in this work was performed in collaboration with Dr. Tom Clarke at the UEA. All of the above have been extremely generous with their time and without their help it would not have been possible to complete this thesis. I am also grateful to everyone in the Department of Molecular Microbiology at the John Innes Centre for creating a friendly atmosphere in which to work and to the BBSRC for funding this research.

Role of the PAS2 domain of the NifL regulatory protein in redox signal transduction

Abstract

Per-Arnt-Sim (PAS) domains play a critical role in signal transduction in multi-domain proteins by sensing diverse environmental signals and regulating the activity of output domains. Multiple PAS domains are often found within a single protein. However, the role of duplicate PAS domains in signalling is poorly understood. The NifL regulatory protein from *Azotobacter vinelandii* provides a typical example as it contains tandem PAS domains, the most N-terminal of which, PAS1, binds a FAD co-factor and is responsible for redox sensing, whereas the second PAS domain, PAS2, has no apparent co-factor and its function is unknown. NifL regulates the activity of the transcriptional activator, NifA, in response to changes in redox potential and fixed nitrogen status. Here, genetic and biochemical approaches were used to investigate the role of the PAS2 domain in the function of the NifL protein. Amino acid substitutions in the PAS2 domain were identified that either lock NifL in a form that constitutively inhibits NifA or that fail to respond to the redox status, suggesting that PAS2 plays a pivotal role in transducing the redox signal from PAS1 to the C-terminal output domains of NifL. A combination of biochemical experiments indicates that the isolated PAS2 domain is dimeric in solution. It was observed that PAS2 dimerisation is maintained in the redox signal transduction mutants, but is inhibited by substitutions in PAS2 that lock NifL in the inhibitory conformer. Limited proteolysis experiments suggest that the PAS2 substitutions influence conformational changes induced in response to the redox state of the FAD co-factor in the PAS1 domain. Further, mutagenic analysis of an inter-domain linker helix that connects the PAS domains of NifL suggests that this region of the protein is important in redox signalling. Overall, these results support a model for signal transduction in NifL, whereby redox-dependent conformational changes in PAS1 are relayed to the C-terminal output domains via changes in the quaternary structure of the PAS2 domain.

General Abbreviations

AAA+	ATPases associated with various cellular activities
AC	Adenylate cyclase
ADP	Adenosine diphosphate
AhR	Aryl hydrocarbon receptor
AMP-PNP	5'-adenylyl-beta, gamma-imidodiphosphate
ARNT	Aryl hydrocarbon receptor nuclear translocator
ATP	Adenosine triphosphate
AUC	Analytical ultracentrifugation
BACTH	Bacterial adenylate cylase two-hybrid
PBPb	Bacterial periplasmic substrate-binding domain
Bph	Bacteriophytochrome
BSA	Bovine serum albumin
CBS	Cystathionine-beta-synthase domain
Cu-Phe	Copper *o*-phenanthroline
DGC	Diguanylate cyclase
DHp	Dimerisation and histidine phosopho-transfer domain
c-di--GMP	Bis-(3'-5')-cyclic dimeric guanosine monphosphate
DLS	Dynamic light scattering
DNA	Deoxyribonucleic acid
dNTP	Deoxy nucleotide triphosphate
DOS	Direct oxygen sensor
DTT	Dithiothreitol
EBP	Enhancer binding protein
EDTA	Ethylenediaminetetraacetic acid
FAD	Flavin adenine dinuceotide oxidised form
$FADH_2$	Flavin adenine dinuceotide reduced form
Fe (II)	Ferrous iron
Fe (III)	Ferric iron
FHA	Folkhead-associated domain
FIST	F-box and intracellular signal transduction domain
FMN	Flavin mononucleotide
GAF	cGMP-specific and regulated cyclic nucleotide phosphodiesterase, *Anabaena* Adenylate cyclase and *E. coli* transcription factor FhlA
GTP	Guanosine triphosphate
GTP-TR	Guanosine 5′-triphosphate- Texas red (sulphorhodamine 101 acid chloride)
HATPase	Histidine kinase-like ATPase domain
HIF	Hypoxia inducible factor
HisKA	Histidine kinase A phosphoacceptor domain
HPK	Histidine protein kinase
HPt	Histidine phospho-transfer domain
HTH	Helix-turn-helix domain
IPTG	Isopropyl-β-*D*-thiogalactopyranoside
K_d	Dissociation constant
LB	Luria-Bertani broth
NEM	*N*-ethylmaleimide
NFDM	Nitrogen-free Davis and Mingioli medium
NMR	Nuclear magnetic resonance

ONPG	ortho-nitrophenyl-β-galactoside
PAGE	Polyaccrylamide gel electrophoresis
PAS	Per-ARNT-Sim domain
PCR	Polymerase chain reaction
PDB	Protein data bank
PDE	Phosphodiesterase
Pfam	Protein families database
pGpG	5'-Phosphoguanylyl-(3'-5')-guanosine
PYP	Photoactive yellow protein
REC	Response regulator receiver domain
RR	Response regulator
SAP	Shrimp alkaline phosphatase
SDS	Sodium dodecyl sulphate
SEC	Size exclusion chromatography
SMART nrdb	Simple modular architecture research tool non-redundant database
sMMO	soluble methane monooxygenase
STAS	sulphate transporter and antisigma factor antagonist domain
TBE	Tris borate EDTA
TCS	Two-component system
TEMED	N,N,N',N'-Tetramethylethylenediamine
TLC	Thin layer chromatography
TM	Trans-membrane
Tris	Tris (hydroxymethyl) aminomethane
UTase/UR	Uridylyltransferase/uridylyl-removing enzyme
WT	Wild type
X-gal	5-bromo-4chloro-3-indolyl-B-D-galactopyranoside

Contents

Index to Tables and Figures

Chapter 1 – Introduction

Chapter 1 - Introduction

1.1 Signal transduction in bacteria

Signal transduction can be defined as the process by which organisms link environmental stimuli to adaptive responses. The evolution of prokaryotes has selected for microbes which are most competent to sense and respond to their environment. Thus, efficient signal transduction is crucial to bacterial survival. There are many molecular mechanisms by which this is achieved. Arguably the simplest of these is the "one-component system". One-component systems consist of a single modular protein that contains (at least) two domains: a sensory (input) domain responsible for detection of environmental stimuli and an output (effector) domain that, when activated, elicits a cellular response (Figure 1.1). This modular domain architecture allows for one-component systems to respond to an extensive repertoire of signal inputs. Indeed, a recent analysis of 145 prokaryotic genomes revealed the presence of approximately 17,000 putative one-component systems (Ulrich et al., 2005). In the archetypal one-component system, a membrane-permeable signalling molecule enters the cell via passive or facilitated diffusion and binds to the sensory domain of the cytosolic signalling protein. Subsequent intra-molecular signal relay results in activation of the output domain, thus triggering a cellular response. By far the most common output from these systems is a change in gene expression, although other outputs include regulation of cyclic nucleotide levels and protein phosphorylation (Ulrich et al., 2005).

There is one obvious disadvantage inherent to signal transduction by one-component systems; stimulus detection is limited to the cytosol as the presence of transmembrane regions in the signalling protein would prevent the output domain from accessing its target. In the late 1980s researchers began using the term "two-component" to describe an emerging class of bacterial regulatory system in which the modules responsible

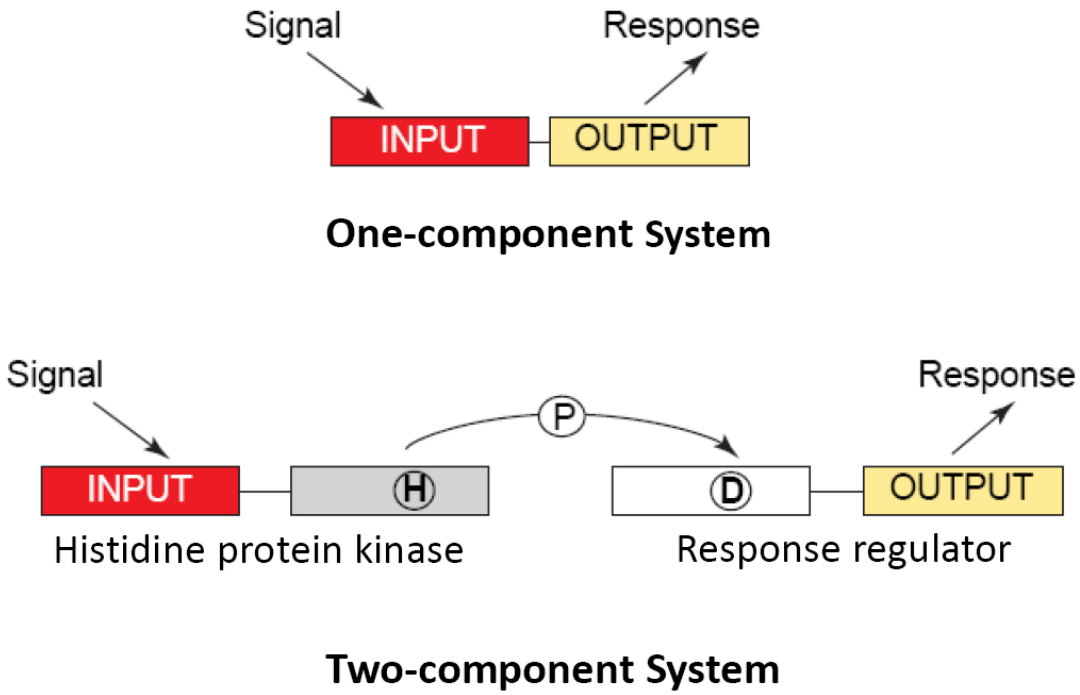


Figure 1.1 Modular organisation of bacterial signal transduction systems. Figure adapted from Ulrich et al., 2005.

for sensing and output are distributed between two proteins (Ninfa and Magasanik, 1986; Nixon et al., 1986). Two-Component systems (TCSs) are now known to be prevalent in eubacteria and archaea, and can also be found in some eukaryotes. The archetypal TCS consists of a histidine protein kinase (HPK) and cognate response regulator (RR). The HPK senses environmental cues and subsequently relays the signal to the RR via a phospho-relay system (discussed in Chapter 1.3). The RR then induces a cellular response, usually through a change in gene expression (Figure 1.1). This uncoupling of the input and output functions allows the sensor protein (the HPK) to be embedded in the membrane and directly sense extracellular stimuli. In fact, the aforementioned analysis of 145 bacterial genomes indicated that over 73% of the putative HPKs identified were likely to be membrane-integral (Ulrich et al., 2005). The prevalence of proteins involved in two-component signalling varies widely between different bacteria. *Bradyrhyzobium japonicum* has 80 HPKs and 91 RRs (Hagiwara et al., 2004), whilst two-component proteins appear to be absent in *Mycoplasma genitalium* (Mizuno, 1998). However, *M. genitalium* is atypical in this respect and the average bacterial genome contains 10-50 predicted TCSs. This is exemplified by *Escherichia coli*, which expresses 29 HPKs and 32 RRs (Mizuno, 1997; Szurmant et al., 2007). In total, several thousand genes are understood to encode proteins involved in two-component signalling systems in all denominations of life.

1.2 - PAS domains

As mentioned above, sensory modules are widely utilised by bacterial signal transduction proteins. Sensory modules are extremely diverse, reflecting the vast array of signals they detect. An example of a prevalent sensory module is the Per-ARNT-Sim (PAS) domain. PAS domains are ubiquitous signalling modules found in all kingdoms of life. The acronym PAS is derived from the names of three proteins in which PAS domains were first identified: the *Drosophila* period clock protein (Per), mammalian aryl hydrocarbon receptor nuclear translocator (ARNT) and *Drosophila* single-minded protein (Sim) (Nambu et al., 1991). PAS domains can detect a plethora of stimuli including light, oxygen, redox potential, proton motive force, metal ions and various small molecules as well as modulating protein-protein interactions (Huang et al., 1998; Zhulin and Taylor, 1998; Zhulin et al., 1997). In order to detect these stimuli, they can bind a variety of co-factors including FAD, FMN, $[4Fe\text{-}4S]^{2+}$ clusters, heme and 4-hydroxycinnamic acid or bind directly to signalling molecules (examples of all are discussed below). As a result of the extraordinary range of stimuli that PAS domains are able to detect, PAS-containing proteins have crucial roles in many cellular processes. For example, the Per and ARNT proteins mentioned above (from which the PAS acronym is derived) are involved in maintaining circadian rhythm and transcription regulation respectively. At the time of writing, PAS domains are recognised in around 14,000 proteins by the SMART (nrdb) and Pfam databases.

In addition to signal perception, PAS domains modulate the activity of output (effector) domains. Figure 1.2 shows the domain architectures of selected proteins from the SMART nrdb in which PAS domains are coupled to various effector domains including: DNA binding domains, guanylate cyclases, exonucleases, methyl acceptors and transferases, phosphatases and kinases (of histidine and serine/threonine residues), c-di-

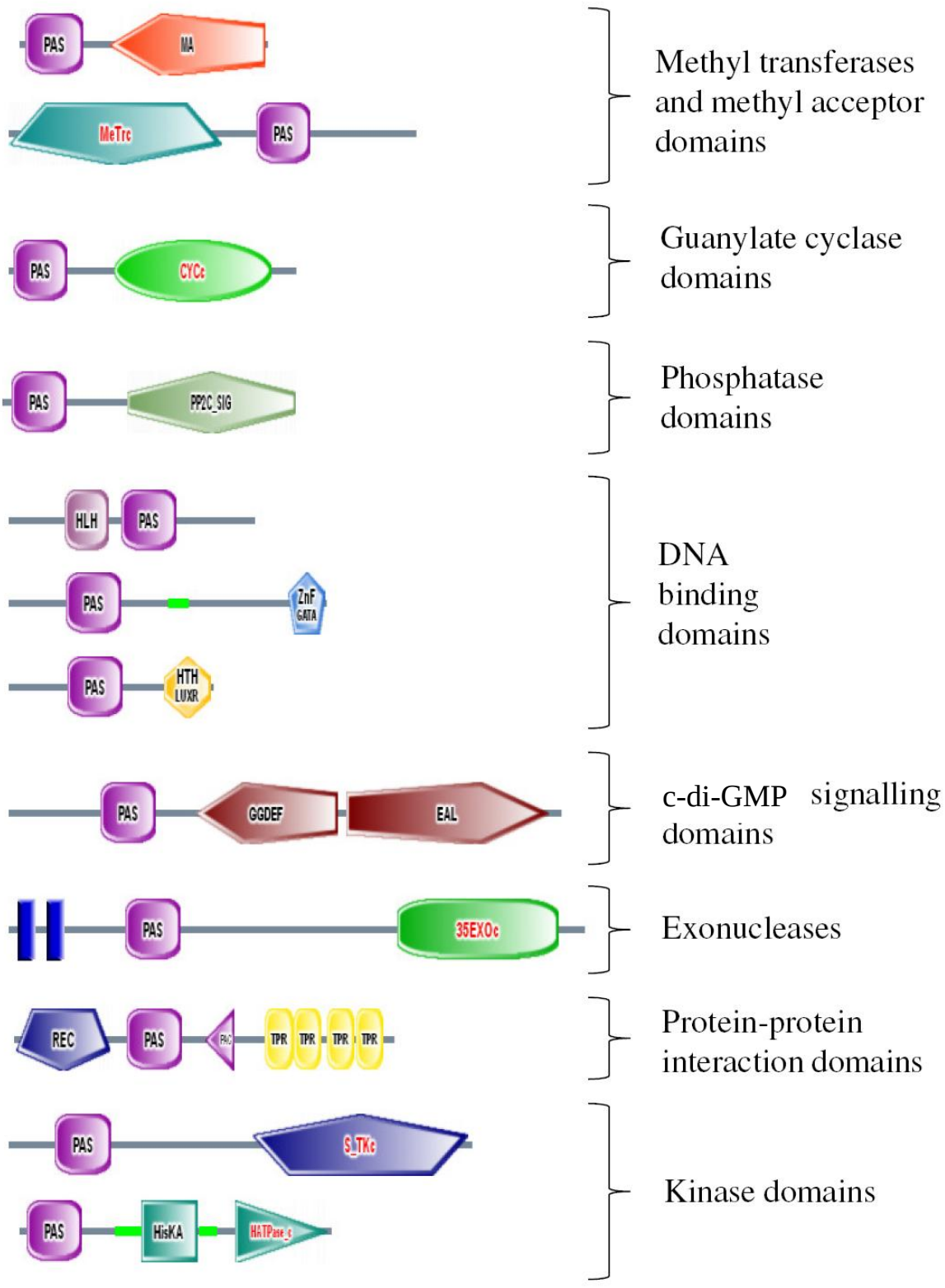


Figure 1.2. PAS-associated output domains. Domain architectures of selected PAS-containing proteins (from the SMART nrdb) to show the variety of output domains with which PAS domains are associated. PAC domains are sequence motifs that form part of the 3-dimentional PAS fold.

GMP signalling domains and protein-protein interaction modules. Moreover, searches using the Pfam database also indicate that PAS domains are often located adjacent to σ^{54} activating domains, ion transport domains, Kelch repeats and STAS (sulphate transporter and antisigma factor antagonist) domains[1]. In most cases the PAS domain is directly adjacent to the output domain. Although the proteins shown in Figure 1.2 are predominantly cytoplasmic, examples of proteins with similar architectures that contain additional trans-membrane (TM) domains are abundant (with the obvious exception of the transcription factors). Of the relatively few studied PAS-containing proteins with TM regions, the PAS domain is usually located on the cytoplasmic side of the cell membrane. However, PAS domains are not exclusively cytoplasmic and several are known to be located in the periplasm (Cho et al., 2006; Kaspar and Bott, 2002; Kaspar et al., 1999; Pappalardo et al., 2003). In such cases, signals must be relayed from the PAS domain through the trans-membrane regions of the protein and ultimately to cytoplasmic effector domains. In addition to the regulation of covalently attached output domains, some PAS domains occur as small, single domain proteins. Presumably, these PAS domains are involved in signal transduction via protein-protein interactions. In other words, they may exert *in trans* affects on the activity of output domains in other proteins.

Multiple PAS domains are often present in tandem within a single protein. Examples of proteins containing 2, 3 or 4 tandem PAS domains are shown in Figure 1.3A. Proteins with 6 or more PAS domains are not uncommon (the SMART nrdb contains 95 such proteins) and several proteins containing 10 or even 15 (UniProt indentifiers A0YNE5_9CYAN and A3IRJ7_9CHRO respectively) adjacent PAS domains can be found. The SMART and Pfam databases both indicate that a total of over 21,000 PAS domains are present in around 14,000 proteins, suggesting that a high proportion of PAS-

[1] Discrepancies between the two databases stem primarily from differences in recognition of the output domains. For example, the SMART database shows proteins containing PAS and AAA^+ domains together but fails to recognise the adjacent HTH domain and thus σ^{54} activators are not detected.

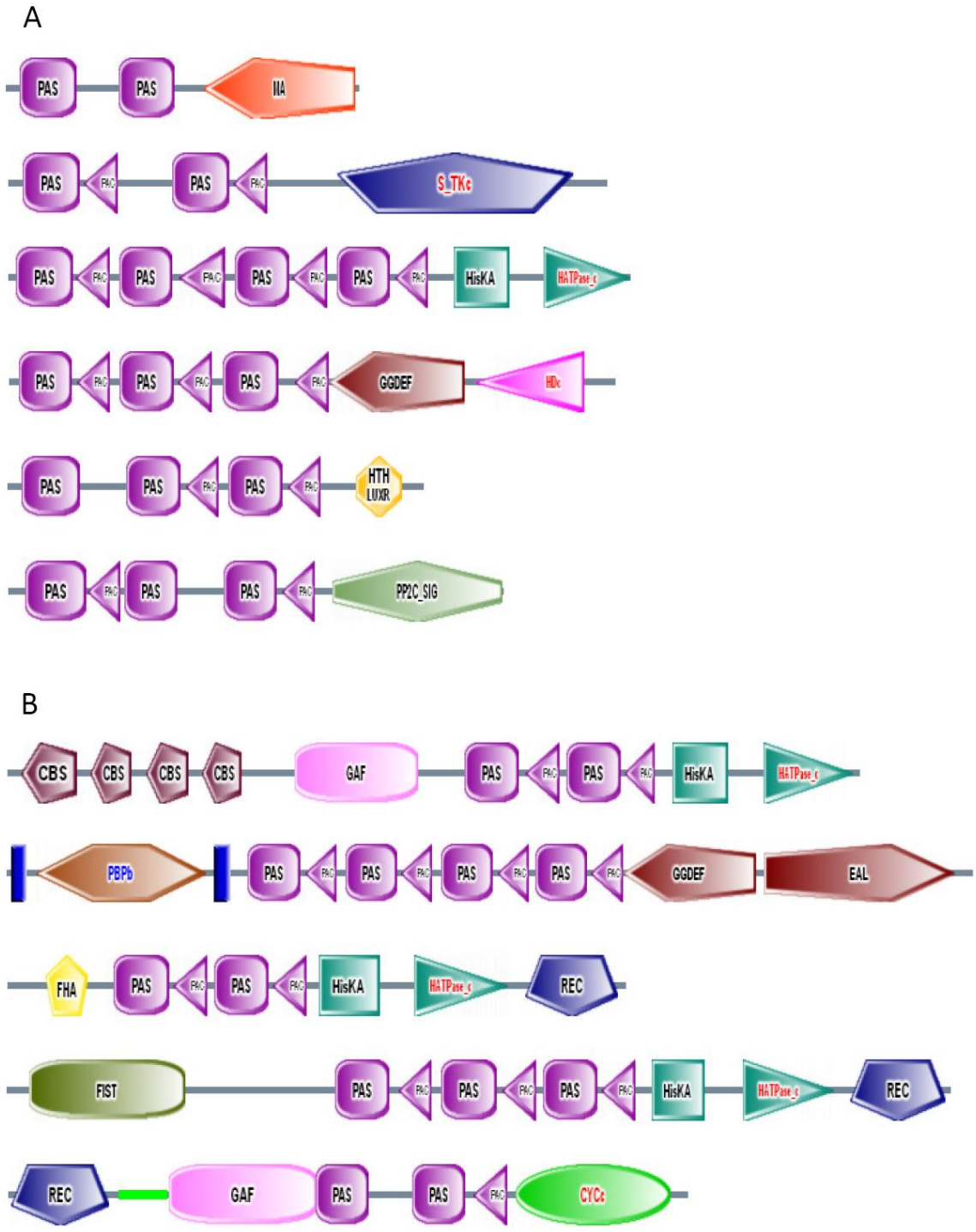


Figure 1.3 Example domain architectures from the SMART nrdb of (A) proteins containing tandem PAS domains and (B) complex modular proteins in which PAS domains are combined with multiple sensory/signalling domains. PAC domains are sequence motifs that form part of the 3-dimentional PAS fold.

containing proteins have more than one PAS domain. In addition to their location alongside output domains, PAS domains are often found in complex multi-domain proteins with many sensory domains (Figure 1.3B). For example, the first domain architecture shown in Figure 1.3B is that of a sensor protein from *Nodularia spumigena* that contains 2 PAS domains in combination with 4 cystathionine-beta-synthase (CBS) domains (which sense adenosine derivatives) and a GAF domain (a sensory module similar to a PAS domain). Other sensory modules found in combination with PAS domains include FHA domains (which recognise phosphopeptides) and FIST domains (thought to bind small ligands/amino acids). Combinations of PAS domains with various sensory modules and with other signalling domains, such as phosphate receiver domains (REC) and periplasmic substrate-binding (PBPb) domains, give rise to extremely complex modular proteins in which the activity of effector domains is regulated by many signals.

PAS domains were originally recognised as imperfect sequence repeats termed "PAS repeats", "S boxes" or the "PAS motif". Further studies revealed that this motif was the most highly conserved region of a larger domain containing a second, less well conserved motif, called a "PAC motif" or "S_2 box" (Ponting and Aravind, 1997; Zhulin et al., 1997). Thus PAS domains were originally defined on the basis of primary sequence motifs. As the availability of structural information increased, it became clear that these motifs represent a conserved three-dimensional fold and that PAS domains exhibit relatively little sequence homology. On average, the sequence identity between any two PAS domains is below 20% (Möglich et al., 2009b). Consequently, the definition was revised and classification of PAS domains now depends on conserved structural elements, specifically, an α/β fold of approximately 110 amino acids in which an anti-parallel β-sheet is flanked by several α helices. (Hefti et al., 2004; Pellequer et al., 1998).

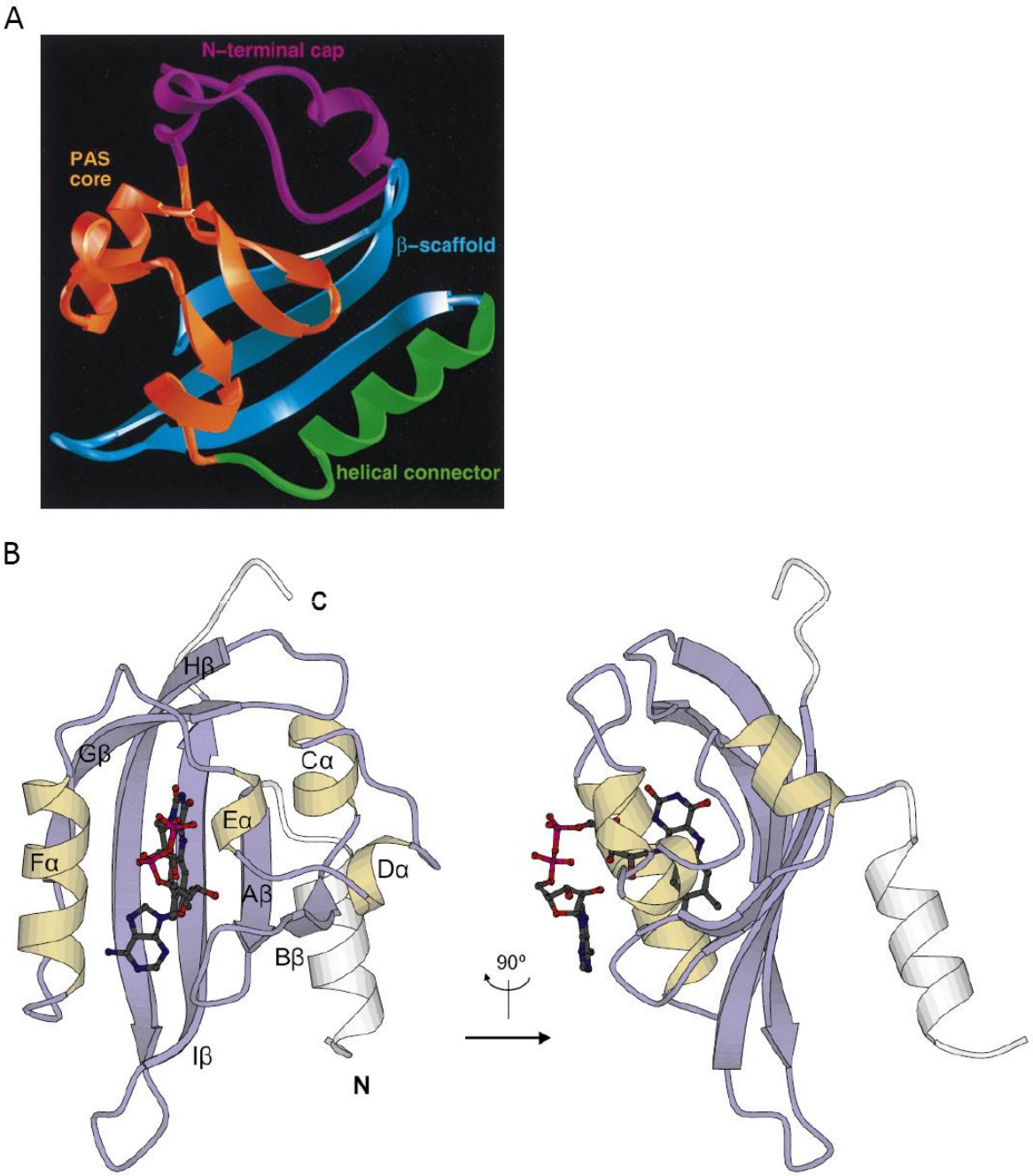


Figure 1.4 (A) Generalised PAS domain structure divided into four conserved regions: the PAS core (orange), the β-scaffold (blue), the helical connector (green) and the N-terminal cap (purple) (Pellequer et al., 1998). (B) Structure of *Azotobacter vinelandii* NifL PAS1 illustrating the annotation of conserved secondary structure elements in PAS domains (Möglich et al., 2009). This domain contains a FAD co-factor (shown here as a ball and stick diagram).

The first PAS domain structure to be solved was that of Photoactive yellow protein (PYP) from *Halorhodospira halophila* (Borgstahl et al., 1995). Initially, PYP was considered the archetypal PAS domain (Pellequer et al., 1998). Based on PYP and several other early structures, PAS domains were divided into four main parts: the β-scaffold, the helical connector, the PAS core and the N-terminal cap (Figure 1.4A). Although this nomenclature has since been superseded by the labelling of secondary structural elements (Figure 1.4B), some of these terms persist in the literature and are therefore worthy of a brief explanation. The β-strands comprising the central β-sheet (Aβ, Bβ, Gβ, Hβ and Iβ in Figure 1.4B) are shared between the β-scaffold and the PAS core. The PAS core and β-scaffold are connected by an extended α-helix called the helical connector. The N-terminal cap is a highly variable region located on the N-terminal side of the PAS core. However, the N-terminal cap is absent from several PAS domains (Vreede et al., 2003). At the time of writing, structures of 47 PAS domains had been deposited in the protein data bank (PDB). Figure 1.4B shows the structure of a PAS domain from the *Azotobacter vinelandii* NifL protein as an example of the more recent nomenclature used to describe PAS domain structures (Key et al., 2007a; Möglich et al., 2009b). Laboratories studying various PAS-containing systems employ both sets of nomenclature and so both will referred to in this chapter. The conserved secondary structural features shown in Figure 1.4B are annotated Aβ to Iβ (β strands are shown in blue and α-helices shown in gold). The most N-terminal α-helix in the structure (shown in white), which may be described as belonging to the N-terminal cap, is not conserved in all PAS domains and is therefore designated A'α. It should also be remembered that significant variation exists between PAS domains, particularly in the helical regions, and that small regions of secondary structure not shown in Figure 1.4B may be present in other PAS domains. Many co-factor binding PAS domains have a cleft between the inner face of the β-sheet and the Eα and Fα helices in

which the co-factor is located. These helices are among the most variable regions of the PAS fold to accommodate the diverse chemistry of the various prosthetic groups to which PAS domains bind (Möglich et al., 2009b).

1.2.1 Gas sensing PAS domains

(i) FixL

FixL is an oxygen sensing histidine protein kinase (HPK) that regulates genes for nitrogen fixation (*nif* and *fix* genes) via its cognate response regulator (RR), FixJ. This system has been best studied in *Sinorhizobium meliloti* and *Bradyrhizobium japonicum*. FixL proteins derived from both species have an N-terminal cytoplasmic PAS domain that contains a covalently bound Fe(II) heme group, and is required for oxygen sensing (Gilles-Gonzalez et al., 1991; Tuckerman et al., 2002; Tuckerman et al., 2001). However, the overall domain architectures of the FixL proteins from *S. meliloti* and *B. japonicum* are not identical. *S. meliloti* FixL (*Sm*FixL) is a membrane-bound protein containing 4 N-terminal TM domains, a PAS domain and C-terminal histidine kinase effector domains. In contrast, *B. japonicum* FixL (*Bj*FixL) is a cytoplasmic protein that lacks the TM regions found in *Sm*FixL but contains an additional N-terminal PAS domain. This extra PAS domain in *Bj*FixL has no known co-factor and its function is unclear.

The crystal structure of the heme-binding PAS domain from *Bj*FixL is shown in Figure 1.5A. Under low oxygen conditions (i.e. when the heme iron is unliganded) the HPK autophosphorylates, leading to transcription of nitrogen fixation genes. Therefore, the PAS domain negatively regulates the activity of the histidine kinase output domains in response to oxygen. In addition to the physiological ligand, oxygen, the heme group is able

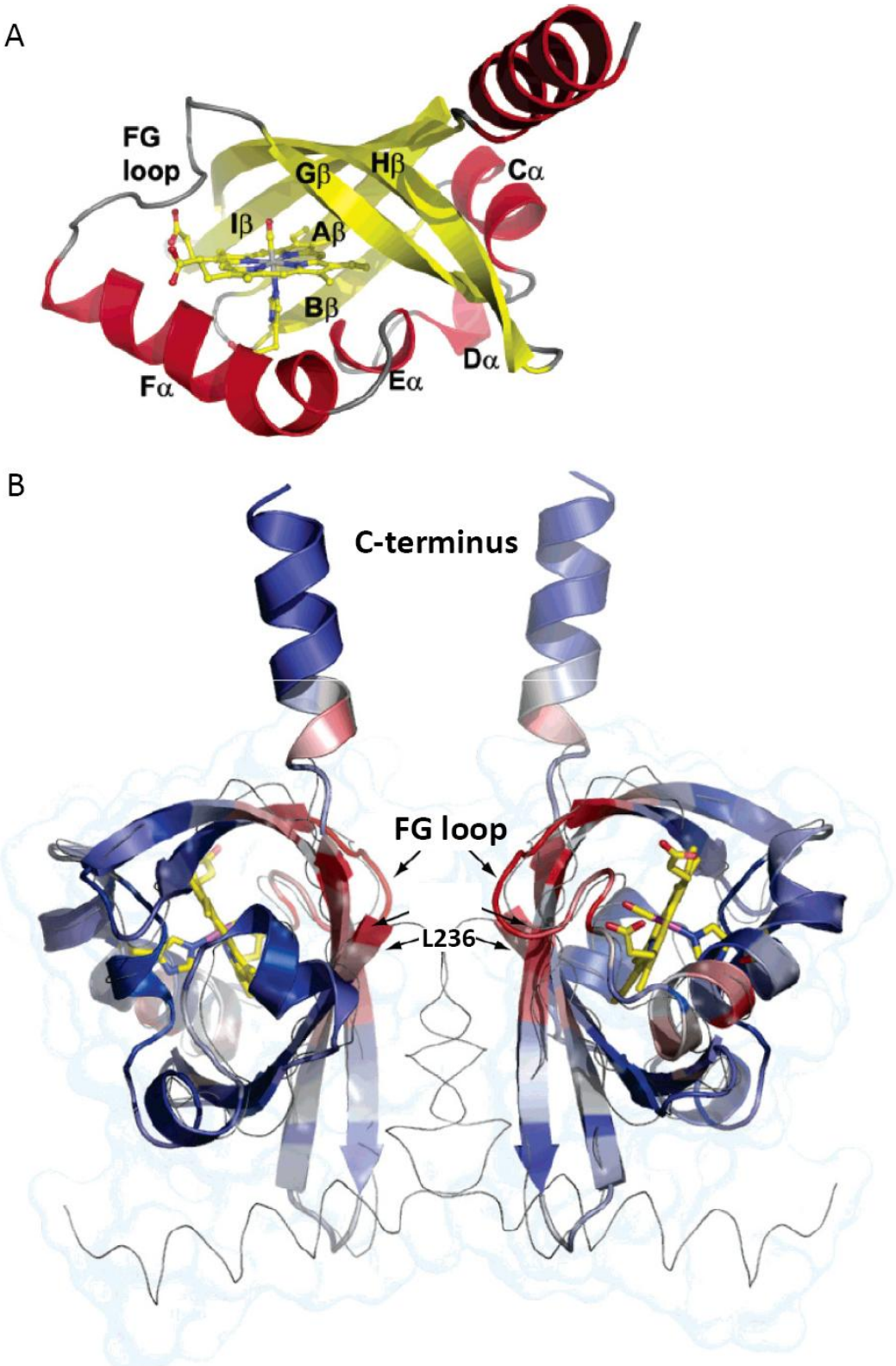


Figure 1.5 (A) Ribbon diagram of the structure of the heme-binding PAS domain from *B. japonicum* FixL (Key et al., 2007b). (B) Model of the dimeric structure of this domain (Key and Moffat, 2005). The regions shown in red are those which exhibit the greatest displacement of main chain carbon atoms in response to ligand binding. Unlabeled arrows indicate the N-terminus.

to bind carbon monoxide (CO) and nitric oxide (NO) to influence kinase activity. It has been proposed that ligand binding induces a structural change in the loop joining the Fα helix (the helical connector) to Gβ strand (in the PAS core), known as the FG loop (Figure 1.5A) and that this change is driven by a flattening of the porphyrin ring (Gong et al., 1998). However, recent studies on CO-bound signalling intermediates in *Bj*FixL have demonstrated movement of residues in the Hβ and Iβ strands (from the PAS core and β-scaffold respectively) in response to ligand binding (Key and Moffat, 2005; Key et al., 2007b). The signal is thought to be propagated via re-orientation of residues that are clustered around a conserved interaction surface found in several heme-binding PAS domains (discussed below) (Kurokawa et al., 2004; Park et al., 2004). This surface may serve as a dimerisation interface in FixL (Figure 1.5B) (Erbel et al., 2003; Key and Moffat, 2005). These findings have prompted the suggestion that adjustments to PAS structure initiated by ligand binding may result in re-orientation of the kinase core domains relative to each other, thus inhibiting autophosphorylation (Key and Moffat, 2005).

Ligand-dependent conformational changes in the *Bj*FixL PAS domain are generated, at least in part, by an alteration in the position of a leucine side chain (Leu236) that sterically occludes the ligand binding pocket. The next step in the signal transduction pathway has been the subject of intense study. Specifically, the role of two arginine residues in the initial steps of signal propagation have been addressed. Substitution of a proximal arginine located in the Fα helix with alanine (R206A in *Bj*FixL) impairs the transmission of signals between the PAS domain and the output domains. The wild-type protein exhibits a >2000-fold reduction in catalytic activity in response to ligand binding. This is reduced to a 140-fold reduction in the R206A variant (Gilles-Gonzalez et al., 2006). This arginine residue is well conserved in heme-binding PAS domains and the equivalent amino acid in *S. meliloti* FixL (R200) contributes to the kinetic stability of the inhibitory

conformer (Reynolds et al., 2009). The role of a second conserved arginine (R220 in *Bj*FixL) in the early stages of signalling has also been studied. Several amino acid substitutions were made at position 220 in *Bj*FixL and their influence on ligand (NO, CO and O_2) release after photolysis was examined (Jasaitis et al., 2006). All substitutions diminished the strain placed on the heme molecule on dissociation of the NO and CO ligands and all mutant proteins differed from the wild-type in the absorption spectra obtained after decay of the O_2 liganded complex. This implies that the distal R220 residue in *Bj*FixL contributes to formation of the primary signalling intermediate. It was also observed that all substitutions increased the yield of dissociated O_2 following decay. Wild-type *Bj*FixL allowed approximately 10% of dissociated O_2 to escape whilst the yield from the R220A variant was almost 100%, suggesting that R220 cages the O_2 molecule near the heme in the wild-type protein. Jasaitis and colleagues combined these results with molecular dynamics simulations to show that, in the first 50 ps following ligand dissociation, movement of the R220 side chain and the O_2 molecule away from the heme binding pocket may constitute the second step in signal transduction (Jasaitis et al., 2006).

It should be remembered that sensing and catalysis by FixL is likely to occur within the FixL-FixJ complex (Gilles-Gonzalez and Gonzalez, 2004; Saito et al., 2003; Tuckerman et al., 2002). Further, the presence of *Sm*FixJ influences the regulation of *Sm*FixL activity in response to O_2 by the *Sm*FixL PAS domain (Tuckerman et al., 2002). Several substitutions in the kinase domain of *Sm*FixL have been shown to impair the inhibitory affect of O_2 on FixL autophosphorylation in the presence of FixJ, whilst retaining O_2 sensitivity when FixJ is absent (Saito et al., 2003). The activities of the output (kinase core) domains and the sensory PAS domain of *Sm*FixL are also linked by the ability of ADP to allosterically reduce oxygen affinity (Nakamura et al., 2004). When ATP is hydrolysed in the kinase core region (discussed below) of one FixL subunit, the

remaining ADP molecule lowers the oxygen affinity of the PAS domain in the opposing subunit. This is assumed to stimulate ATP hydrolysis by the second subunit, thus connecting the sensory and catalytic functions of FixL. Taken together, these findings imply that PAS domains are capable of a form of bilateral inter-domain communication in which changes in the conformation of effector domains can be sensed as well as induced.

(ii) EcDOS

E. coli direct oxygen sensor (*Ec*DOS) is a heme-regulated enzyme involved in bis-(3'-5')-cyclic dimeric GMP (c-di-GMP) signalling. Cyclic di-GMP was discovered as a ubiquitous second messenger and c-di-GMP signalling is now a rapidly progressing field. At the time of writing, c-di-GMP signalling has been implicated in biofilm formation, cell motility, long-term stress responses, secondary metabolite synthesis, cell cycle control and virulence (Schirmer and Jenal, 2009). *Ec*DOS contains two C-terminal output domains, namely EAL and GGDEF domains. These function as a phosphodiesterase (PDE) and diguanylate cyclase (DGC) respectively. DGCs catalyse conversion of GTP to c-di-GMP and PDEs catalyse the degradation of c-di-GMP to 5'-Phosphoguanylyl-(3'-5')-guanosine (pGpG). Despite the presence of these two seemingly antagonistic effector domains, only PDE activity has been reported in *Ec*DOS. This may be due to catalytic inactivity of the GGDEF domain, as there are many examples of proteins containing tandem GGDEF and EAL domains in which one domain is inactive (Schmidt et al., 2005). However, the possibility that the GGDEF domain is active under conditions not yet studied cannot be eliminated. Alternatively, an inactive GGDEF domain may retain the ability to bind its substrate (GTP) and modulate the PDE activity of the adjacent EAL domain in response to substrate binding. That is, the GGDEF domain could potentially regulate PDE activity in response to the changing GTP levels. Precedents can be found for tandem GGDEF and

EAL domain pairs that perform all the functions mentioned above (Christen et al., 2005; Kumar and Chatterji, 2008; Schirmer and Jenal, 2009). The PDE activity of *Ec*DOS is regulated by a sensor region containing two PAS domains in tandem. The most N-terminal PAS domain, PASA, contains a covalently bound heme co-factor and is involved in signal perception, while the second PAS domain, PASB, has no apparent co-factor and its function is unknown.

The activity of the output domains is regulated by binding of gases to the Fe(II) heme moiety of PASA (Sasakura et al., 2006). Specifically, the Fe(II) heme can bind O_2, CO or NO (at similar affinities) to stimulate the PDE activity of *Ec*DOS by up to 8-fold (Tanaka et al., 2007). However, the relative cellular concentrations of these molecules imply that O_2 is the physiologically relevant ligand (Liebl et al., 2003; Sasakura et al., 2006; Taguchi et al., 2004). Several structural studies have enhanced our understanding of the molecular mechanisms underpinning signal perception and transduction. The crystal structure of the PASA domain has been solved in the oxy and deoxy form (Park et al., 2004). The *Ec*DOS PASA domain forms a dimer, mediated by a dimerisation interface consisting mostly of residues in the A'α helix. One protomer of the PASA domain is shown in Figure 1.6A (the A'α helix is shown in white). In each protomer, the heme co-factor is located between the Fα helix (on the proximal side) and the Gβ and Hβ strands (on the distal side). In the oxy and deoxy state, the heme is six-coordinate (in contrast to the FixL heme which is five-coordinate in the deoxy state) with H77 occupying the fifth, proximal, heme coordination site. Differences between the on (oxy) and off (deoxy) state arise from switching of the sixth distal ligand from O_2 to M95 (Figure 1.6B). In the on state, O_2 is the distal ligand and an arginine residue (R97) from the Gβ sheet forms hydrogen bonds with the O_2 molecule (Park et al., 2004). Mutational studies have confirmed the importance of this arginine in the signalling mechanism and ligand

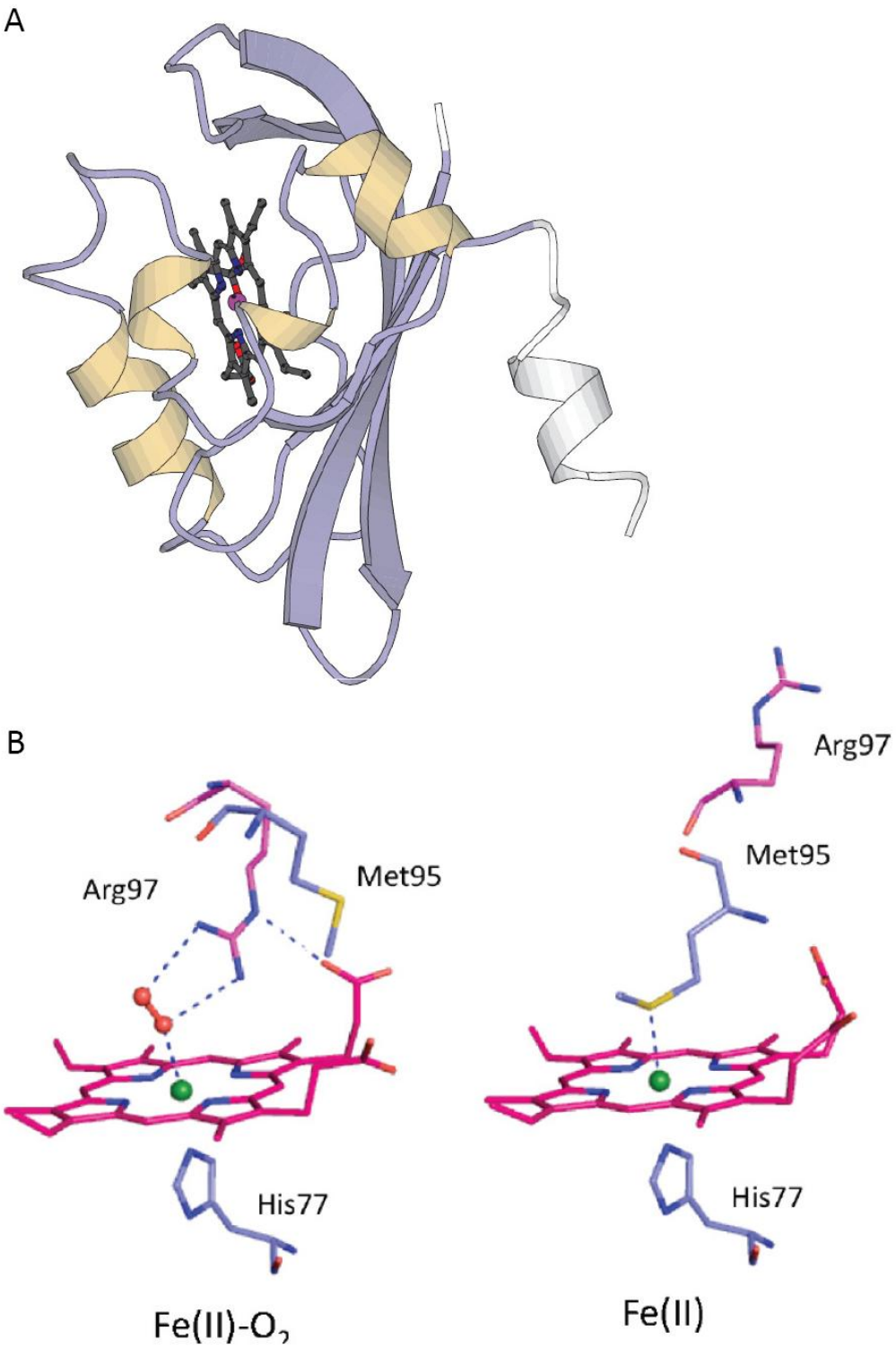


Figure 1.6 (A) Crystal structure of the *Ec*DOS PASA domain (PDB entry 1V9Z) (Möglich et al., 2009b). (B) Comparison of crystal structures of the O_2 liganded and unliganded heme from the *Ec*DOS PASA domain (Ishitsuka et al., 2008).

recognition (El-Mashtoly et al., 2008; Ishitsuka et al., 2008; Tanaka and Shimizu, 2008). Dissociation of the O_2 ligand results in numerous conformational changes. The side chain of R97 rotates by almost 180^o towards the protein surface to form a salt bridge with R112 and E98 (from the Hβ and Gβ strands respectively). A second residue, M95, also undergoes a near 180^o rotation so that its sulphur atom may replace the O_2 molecule as the distal ligand of the heme (Fig. 1.6B). The switching of heme ligands during the transition between the oxy and deoxy state is accompanied by movements in the FG loop (as in FixL), the Gβ strand and the loop connecting the Hβ and Iβ strands (the HI loop) (Park et al., 2004).

Unlike the FixL Fe(II) heme discussed above, the *Ec*DOS Fe(II) heme is relatively prone to autoxidation to form an Fe(III) heme complex (Taguchi et al., 2004; Tanaka and Shimizu, 2008; Tanaka et al., 2007). Oxidation of the heme group inhibits PDE activity and could potentially have a role in deactivating *Ec*DOS under conditions of oxidative or nitrosative stress. Oxidation of the heme results in a replacement of M95 as the distal ligand (see above) with a water molecule (Kurokawa et al., 2004). This triggers a reorganisation of the hydrogen bonding network surrounding the heme moiety and a change in the rigidity of the FG loop. Further, it has been demonstrated that external ligands (cyanide and imidazole) can bind the heme Fe(III) complex in PASA to stimulate PDE activity *in vitro* (Tanaka and Shimizu, 2008), although the physiological relevance of this, if any, is unclear. Tanaka and Shimizu propose that *Ec*DOS may exist in three states: (i) the inactive Fe(III) form, (ii) the resting Fe(II) form and (iii) the active Fe(II)-O_2 form (Tanaka and Shimizu, 2008). However, this is not a universally accepted hypothesis and the relevance of heme autoxidation to signalling *in vivo* is still a subject of debate.

Detailed studies on the *Ec*DOS apo-protein, heme-free mutant proteins and N-terminal truncations lacking the PASA domain have revealed that *Ec*DOS is active in the

absence of heme (or the entire PASA domain) and that heme-bound PASA represses the activity of the EAL domain (Yoshimura et al., 2003). This inter-domain repression is then released by ligand binding to the heme Fe(II) complex. In other words, the PASA domain negatively regulates catalysis and binding of O_2 relieves inhibition.

(iii) NreB

NreB is a cytoplasmic histidine protein kinase that regulates genes involved nitrate/nitrite respiration (*narGHJI*, *narT* and *nirRBD*) via its cognate response regulator, NreC (Fedtke et al., 2002). The NreB protein from *Staphylococcus carnosus* is the only member of its family to be studied, although homologs are present in all *Staphylococcus* species and many other Gram-positive bacteria. NreB consists of an N-terminal PAS domain and C-terminal histidine kinase domains. The PAS domain binds an oxygen liable $[4Fe\text{-}4S]^{2+}$ cluster. Exposure of anaerobically purified NreB to air results in degradation of the $[4Fe\text{-}4S]^{2+}$ cluster that correlates precisely with a decrease in kinase activity (Müllner et al., 2008). Additionally, *in vitro* insertion of the cluster restores kinase activity (Kamps et al., 2004). It is not clear whether the PAS domain inhibits kinase activity and incorporation of the $[4Fe\text{-}4S]^{2+}$ cluster is required to relieve that inhibition, or whether the presence of the cluster is necessary for kinase activity. NreB proteins contain four conserved cysteine residues (C59, C62, C74 and C77 in *S. carnosus* NreB) which ligate the $[4Fe\text{-}4S]^{2+}$ cluster. Substitution of any of these cysteines for alanine or serine results in a loss of activity *in vivo* (Müllner et al., 2008). It has been demonstrated that the $[4Fe\text{-}4S]^{2+}$ -containing form of NreB predominates in cells growing anaerobically, whilst the NreB apo-protein is prevalent in aerobically grown cells (Reinhart et al., 2009). The conversion between the apo-protein and cluster associated protein is therefore a physiologically relevant switch.

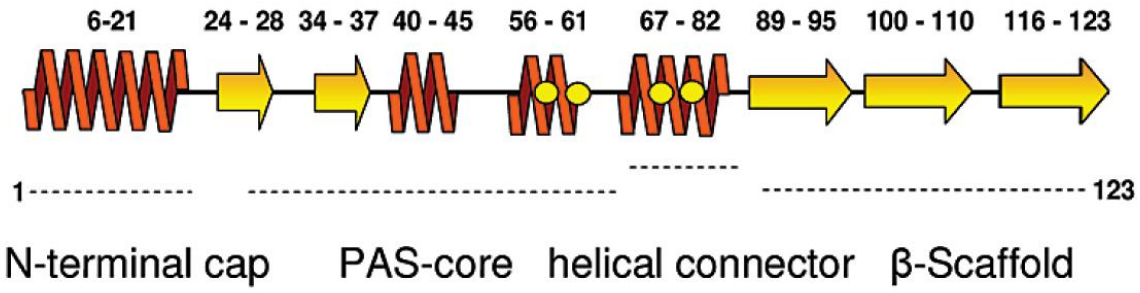


Figure 1.7 Prediction of the secondary structure features of the NreB PAS domain. The positions of the cysteine residues that coordinate the $[4Fe\text{-}4S]^{2+}$ cluster are indicated by yellow circles (Müllner et al., 2008).

Figure 1.7 shows the predicted secondary structure of the NreB PAS domain and the positions of the cysteines that coordinate the $[4Fe\text{-}4S]^{2+}$ cluster (yellow circles). The cysteines are shared evenly between the C-terminal end of the Eα helix (in the PAS core) and the Fα helix (helical connector). The most N-terminal pair of cysteines is located in a region similar to that of the conserved proximal histidine residue that coordinates the heme group of *Ec*DOS and FixL. The cysteine residues in the Fα helix may be located in a position analogous to that of the methionine residue from *Ec*DOS PASA that ligates the (deoxy Fe(II)) heme (Müllner et al., 2008). These similarities imply that the NreB PAS domain may accommodate its co-factor in a similar manner to the PAS domains of FixL and *Ec*DOS, despite the different chemical properties of these moieties. To date, NreB contains the only known $[4Fe\text{-}4S]^{2+}$ cluster-binding PAS domain.

1.2.2 Ligand binding PAS domains

(i) CitA and DcuS

CitA is an integral membrane histidine protein kinase that senses extracellular citrate to regulate the transcription of genes involved in (anaerobic) citrate metabolism and transport, via its cognate response regulator, CitB. CitA activity is regulated by direct binding of citrate to a periplasmic PAS domain (PASp) at high affinity (Kaspar and Bott, 2002). The structure of this domain from the *Klebsiella pneumoniae* CitA protein (shown in Figure 1.8A) was solved in 2003 and was the first example of a PAS domain located outside the cytoplasm (Reinelt et al., 2003). It was observed that the structure of PASp differed from the available structures of cytoplasmic PAS domains. PASp is dimeric in the crystal structure and two of the three N-terminal helices (that constitute the N-terminal cap) form the dimerisation interface. These α-helices of CitA are longer than their

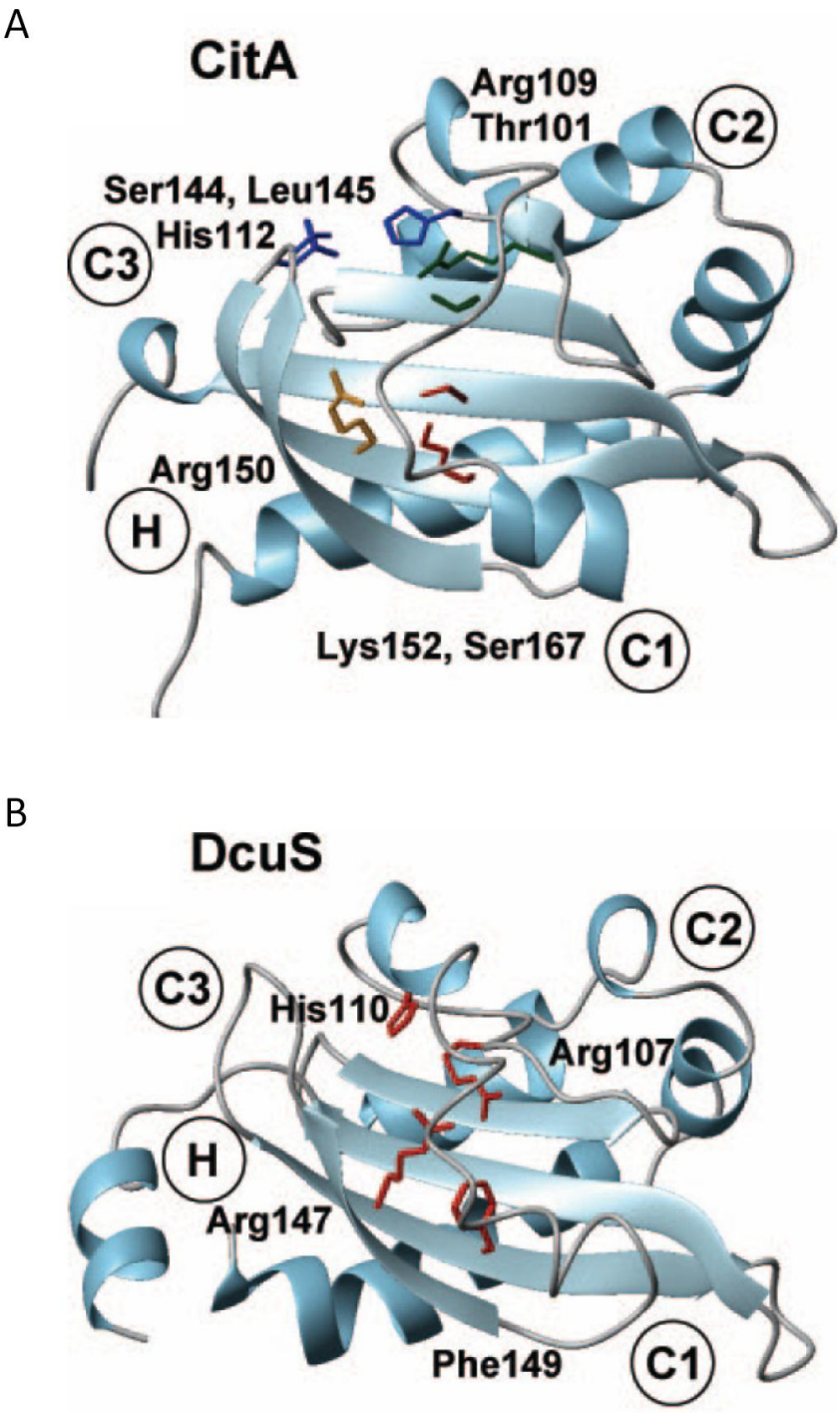


Figure 1.8. Ribbon diagrams illustrating structures of the ligand binding PAS domains from (A) CitA and (B) DcuS. Ligand binding sites are labelled C1 – C3 and the key residues involved at each site are shown (Masher et al., 2006).

counterparts in PYP whereas PAS domains such as *Ec*DOS PASA and NifL PAS1 (discussed below) contain only one N-terminal helix (Key et al., 2007a; Kurokawa et al., 2004; Park et al., 2004). The N-terminal cap α-helices are not necessary for PYP function (see below) and are completely lacking in FixL, whilst the N-terminal helices of NifL PAS1 and *Ec*DOS PASA each contribute to an important dimerisation interface (Key et al., 2007a; Kurokawa et al., 2004; Park et al., 2004; Vreede et al., 2003). The contrast between these PAS domains illustrates the structural and functional variability of the N-terminal cap. The PAS core region of CitA also differs notably from that of many cytoplasmic PAS domains. CitA lacks the Cα and Eα helices. Further structural differences accommodate citrate binding at three sites (C1, C2 and C3 in Figure 1.8). Smaller inter-strand loops in the characteristic PAS β-sheet, which forms the bottom of the citrate binding pocket, facilitate closer proximity of the ligand (to the β-sheet) and a more tightly closed pocket. The carboxylate groups of the citrate are each ligated by one basic residue (K152 at C1, R109 at C2 and H112 at C3) in addition to a minimum of one hydroxyl side chain (Figure 1.8) (Reinelt et al., 2003). The binding site also includes an α-helix in the PAS core and two loops. These loops, termed the major and minor loops, form a tight lid over the bound citrate (Reinelt et al., 2003; Sevvana et al., 2008).

A more detailed understanding of the signalling mechanism has been gleaned from a recent study, in which the structures of the citrate-bound and citrate-free *K. pneumoniae* PASp were examined by crystallography and nuclear magnetic resonance (NMR) spectroscopy (Sevvana et al., 2008). Several structural changes were found to accompany ligand binding (Figure 1.9). Residues 100-103 in the minor loop adopt a type I β-turn when citrate is bound and ligand dissociation appears to trigger a reorganisation of the loop to form a type II β-turn. This may be important in the signalling mechanism as the backbone amide of residue 102 and the side chain of residue 101 are involved in citrate binding

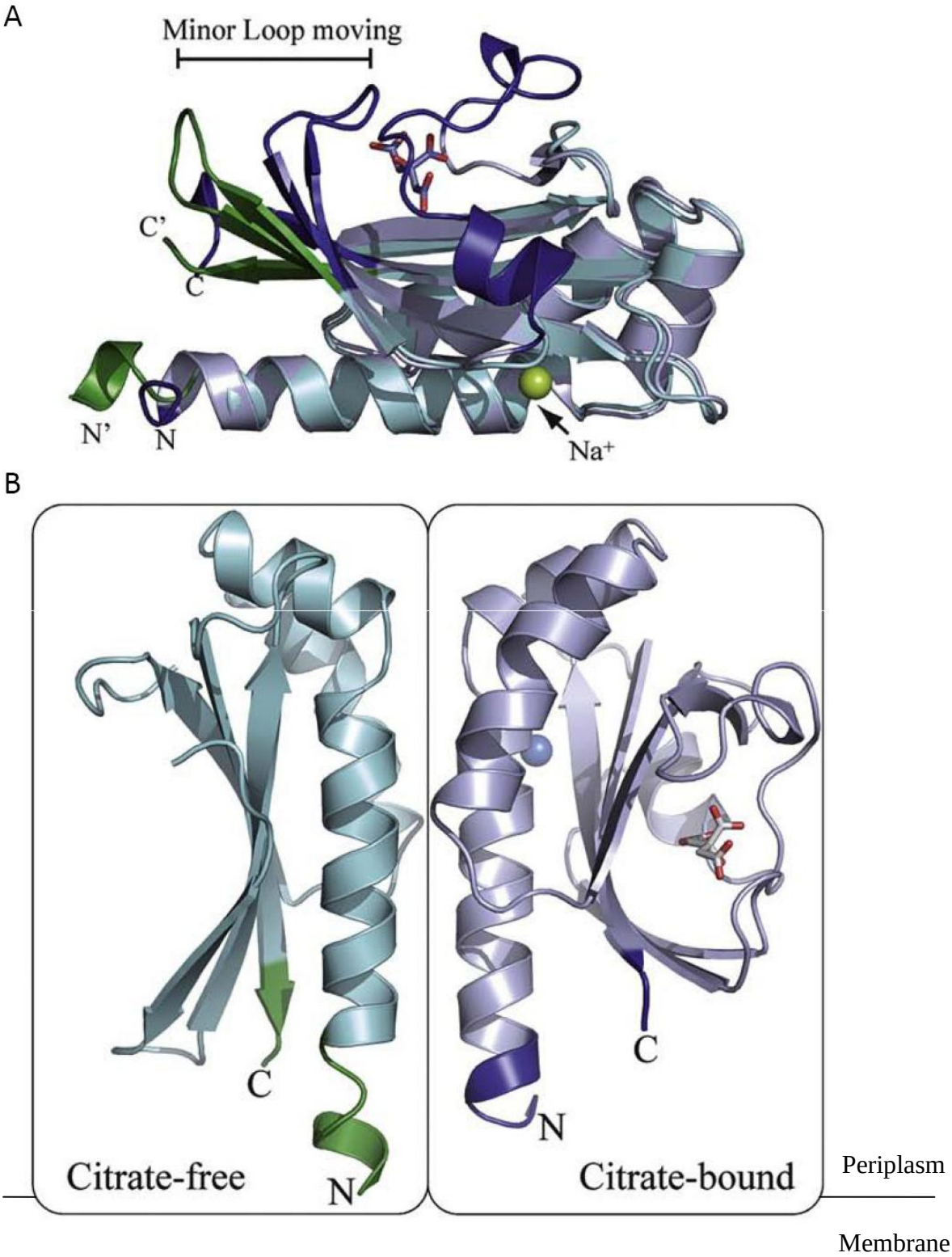


Figure 1.9. (A) Structure of the ligand bound CitA PASp protomer (light blue with differences highlighted in dark blue) superimposed onto the ligand-free CitA PASp protomer (cyan with difference highlighted in green). (B) Comparison of citrate-bound and citrate-free CitA PASp showing contraction of the domain in response to ligand binding (Sevvana et al., 2008).

(Sevvana et al., 2008). Movement in the minor loop is concomitant with a flattening of the central β-sheet (Figure 1.9A). Overall, the evidence implies that the β-sheet and minor loop adopt an open (or fluctuating) conformation in the absence of citrate and that ligand binding prompts a conformational change in these regions to form a lid over the binding pocket (Figure 1.9A). There are also ligand-dependent changes in the major loop. In the absence of citrate, the major loop is disordered in the crystal structure and appears to fluctuate between multiple conformations in solution. Ligand binding stabilises the structure of the major loop, which contributes significantly to ligation of the citrate moiety (Sevvana et al., 2008). Another potentially important observation is that PASp binds a Na^+ ion in the citrate-bound state but this ion is absent in the citrate-free structure. Although the failure to resolve the Na^+ ion in citrate-free PASp does not strictly eliminate the possibility that one is present and the identity of the modelled ion cannot be completely certain, metal ion binding by PAS domains is not without precedent (Cheung et al., 2008; Cho et al., 2006) and Na^+ sensing by PASp would have a clear physiological purpose. Na^+ is important in citrate transport and metabolism and CitB dependent gene expression requires both Na^+ and citrate (Bott et al., 1995; Meyer et al., 2001; Pos and Dimroth, 1996; Sevvana et al., 2008). If Na^+ binding is part of the signalling mechanism, CitA PASp would be the first example of a single PAS domain to integrate signals from multiple ligand binding events. Overall, citrate binding appears to induce a contraction of the central β-sheet that pulls the C-terminal region of PASp away from the membrane (Figure 1.9B). It has been postulated that this results in a "piston-type" movement of the TM regions that regulates activity of cytoplasmic output domains (Sevvana et al., 2008).

In addition to the N-terminal periplasmic PAS domain, CitA has two TM regions, a cytoplasmic PAS domain (PASc) and C-terminal histidine kinase domains (Mascher et al., 2006). The function of the PASc domain remains unclear. PASp is not required for kinase

activity as N-terminal truncations of *K. pneumoniae* CitA lacking PASc, the TM domains and the periplasmic region are active *in vitro* (Kaspar et al., 1999). This is particularly interesting given that citrate is required for CitA activity *in vivo*. A recent study has presented evidence that the activity of the cytoplasmic region of *E. coli* CitA (containing the PASc and histidine kinase domains) is dependent on redox conditions *in vitro* (Yamamoto et al., 2009). The authors showed that kinase activity increases concomitantly with increasing dithiothreitol (DTT) concentration whereas the protein is inactive in the absence of DTT. Substitution of a cysteine residue (C529) in the kinase domain for alanine results in constitutive kinase activity (i.e. DTT is no longer required for activity), implying a role for the kinase domain in redox sensing. However, these findings are contradictory to the work on the *K. pneumoniae* CitA mentioned above, in which activity was observed for the kinase domains in the absence of any reductant (Kaspar et al., 1999). Possible causes of this discrepancy include differences in the constructs tested (one of which lacks PASc and is fused to MBP) and mechanistic differences between the CitA proteins from *E. coli* and *K. pneumoniae*. Despite these uncertainties regarding redox sensing by the CitA protein, CitB dependent transcription is known to require a low oxygen tension *in vivo* (Bott et al., 1995).

DcuS, like CitA, is a membrane-embedded histidine protein kinase that contains a periplasmic ligand-binding PAS domain (DcuS PASp), a cytoplasmic PAS domain (DcuS PASc) and C-terminal histidine kinase domains. DcuS, together with its cognate RR, DcuR, regulates the transcription of fermentation genes (Golby et al., 1999; Zientz et al., 1998). DcuS PASp binds a broader range of ligands than CitA PASp, including citrate and C_4-dicarboxylates. The structure of the DcuS PASp domain from *E. coli* has been determined by NMR spectroscopy (Figure 1.8B) (Pappalardo et al., 2003). The topology of the three major β-strands in CitA PASp and DcuS PASp are similar and, in both proteins,

the central β-sheet is flanked by the N-terminal α-helices on one side and the PAS core on the other (Pappalardo et al., 2003). The residues shown in Figure 1.8B (R107, H110, R147 and F120) are required for ligand-mediated activation of DcuS (Kneuper et al., 2005) and are homologous to the basic residues involved in ligand binding in CitA. Overall, the binding pocket of DcuS PASp bears a strong resemblance to that of CitA PASp (Mascher et al., 2006; Pappalardo et al., 2003; Reinelt et al., 2003). Ligand binding to DcuS PASp generates a signal that is relayed to the cytoplasmic PAS domain via two transmembrane helices. The PASc domain has no known role in signal perception but the plasticity of this domain is believed to be important for signal transduction to the histidine kinase domains. Several amino acid substitutions that result in ligand-independent (constitutive) activation of DcuS have been indentified in the PASc domain, implying that this domain is important in signalling. When DcuS PASc was modelled on the dimeric crystal structure of the NifL PAS1 domain (discussed below), it was observed that the substituted residues were located close to the α-helical N-terminal cap that forms part of the dimerisation interface (Etzkorn et al., 2008). This suggests a model in which signal perception by DcuS PASp impacts upon the stability of the dimer interface in DcuS PASc. Presumably, these changes in PASc modulate the activity of the C-terminal histidine kinase domains. Therefore, DcuS PASc is the first example of a PAS domain that appears to be involved in signal relay rather than stimulus perception. It is possible that the CitA PASc domain may have a similar function.

(ii) Other ligand-binding PAS domains

Despite the relatively small number of PAS domains studied, a diverse range of ligands have been identified. These include carboxylic acids, amino acids, divalent metal ions and aromatic hydrocarbons (Cho et al., 2006; Denison et al., 2002; Glekas et al.,

2009). It is likely that continued study of PAS-containing proteins will reveal new small molecule ligands. This section will summarise some examples of ligand-binding PAS domains from prokaryotic systems.

In addition to the PAS domains mentioned above, which bind specifically to citrate (CitA PASp) or non-specifically to a broader range of carboxylic acids (DcuS PASp), there are ligand-binding PAS domains that respond specifically to C_4-dicarboxylates. An example of this is the DctB protein. DctB is a histidine protein kinase that regulates the transcription of rhizobial C_4-dicarboxylate transport (*dct*) genes. In contrast to CitA and DcuS, DctB has two periplasmic PAS domains, known as the membrane-distal and membrane-proximal PAS domains. The membrane-distal PAS domain binds C_4-dicarboxylates, while the membrane-proximal PAS domain is not associated with any ligand or co-factor (Zhou et al., 2008). Ligand binding to the membrane-distal PAS domain induces a tightening of the binding pocket and movements in several loop regions, mirroring the ligand-dependent conformational changes observed in CitA PASp (discussed above).

The N-terminal PAS domains of CitA, DcuS and DctB are structurally similar to a periplasmic PAS domain found in the histidine protein kinase, PhoQ. This protein senses the extracellular concentration of divalent cations to regulate virulence and stress response genes in several Gram-negative pathogenic bacteria and the Mg^{2+} starvation response in *E. coli* (Monsieurs et al., 2005; Zwir et al., 2005). PhoQ activity is sensitive to changes in the concentration of Mg^{2+} and Ca^{2+} ions both *in vitro* and *in vivo*. The protein is active when extracellular concentrations of these ions are low, whilst increases in Mg^{2+} and Ca^{2+} levels result in diminished kinase activity (Sanowar and Le Moual, 2005; Vescovi et al., 1997). Crystal structures are available for the Ca^{2+}-bound PAS domain from *Salmonella typhimurium* PhoQ and the *E. coli* PhoQ PAS domain bound to Ni^{2+} ions (Cheung et al.,

2008; Cho et al., 2006). Both proteins contain a conserved cluster of acidic residues in close proximity to the inner membrane. These residues appear to directly bind divalent cations. It has been postulated that dissociation of the cations results in an electrostatic repulsion between the acidic surface of the PAS domain and the plasma membrane that might drive conformational changes leading to PhoQ activation (Cho et al., 2006).

Recent work on the *Bacillus subtilis* chemoreceptor McpB has revealed a periplasmic sensor region that is likely to contain tandem PAS domains, similar to those found in DctB (Glekas et al., 2009). This sensor region was shown to bind asparagine with a K_d of 14 μM. Mutations that decrease the affinity of the sensor region for asparagine were identified in the membrane-distal PAS domain. Moreover, the decreased affinity of the isolated sensor domain for asparagine *in vitro* correlated to reduced chemotactic responses in swarm plates and capillary assays (Glekas et al., 2009). These results suggest that the membrane-distal PAS domain in the periplasmic sensor region of McpB binds asparagine to regulate chemotaxis in *B. subtilis*.

Overall, subtle adaptations to the PAS fold can facilitate binding of chemically diverse ligands. These adaptations range from changes in the chemical properties of amino acid side chains located in the central cleft to the incorporation of clusters of charged residues on the outer surface of the domain. The mechanism by which ligand binding to PAS domains is coupled to conformational changes in output domains apparently varies between proteins. Diversity in the mechanisms by which PAS domains can sense ligands and relay ligand-binding events to output domains highlights their adaptability with regard to signalling.

1.2.3 Redox Sensing PAS domains

(i) NifL

The NifL regulatory protein controls the transcription of *nif* genes (required for biosynthesis of the molybdenum dependant nitrogenase) in γ-proteobacteria via interactions with the transcriptional activator, NifA. The NifL-NifA system is further discussed in section 1.4 whereas this section focuses on the N-terminal sensory region that contains two PAS domains in tandem. The most N-terminal PAS domain, PAS1, senses cellular redox status and binds a FAD co-factor. NifL is inactive when the FAD moiety is fully reduced (to $FADH_2$). Oxidation of the PAS1 co-factor activates the NifL protein to inhibit transcriptional activation by NifA in the presence of excess oxygen (Hill et al., 1996). The second PAS domain, PAS2, has no apparent co-factor and, prior to the work performed in this thesis, its function was unknown.

The crystal structure of the PAS1 domain (residues 21 – 140 of *A. vinelandii* NifL) has recently been solved at 1.04 Å resolution (Figure 1.10A). The structure reveals a typical α/β PAS fold that accommodates a non-covalently bound FAD co-factor and forms a dimer in the asymmetric unit. Dimerisation of the NifL PAS1 domain is mediated by an amphipathic A'α helix. The A'α helices of each protomer interact with each other as well as the hydrophobic surface of the β-sheet from the opposing subunit. This extended dimerisation interface buries 2066 $Å^2$ of hydrophobic surface area and is highly similar to that observed in the crystal structures of *Ec*DOS PASA and the *Sm*FixL PAS domain (Key et al., 2007a). Association of the FAD co-factor is stabilised by an extensive hydrogen bonding network. Hydrogen bonds connect the isoalloxazine ring to an asparagine residue in the Gβ strand (N102) and the ribose and adenine portions of the FAD molecule to residues (W87 and R80) in the Fα helix (Key et al., 2007a). The co-factor is connected to the external environment via a cavity running through the protein, providing a possible

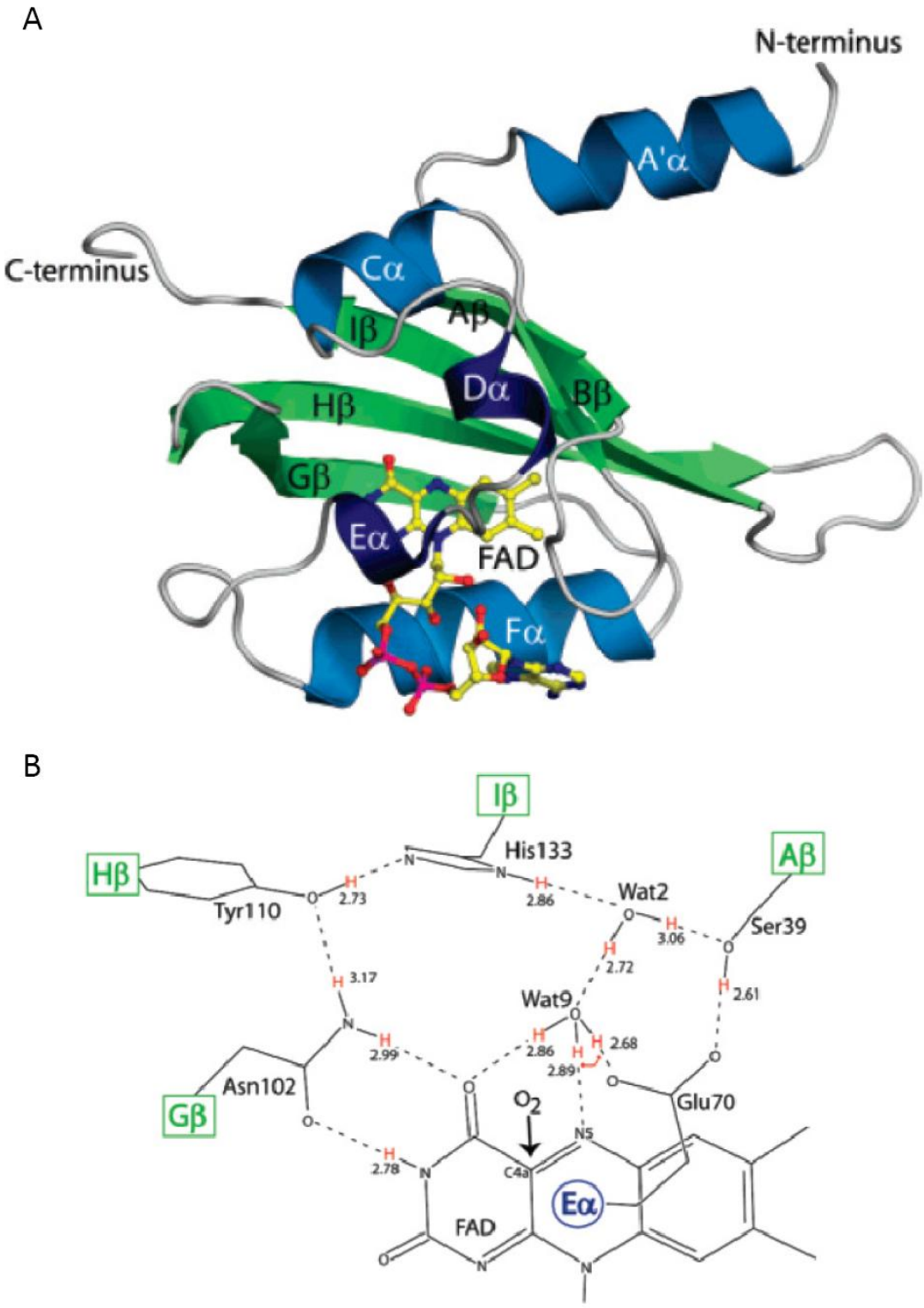


Figure 1.10. (A) Ribbon diagram of the NifL PAS1 domain and (B) the hydrogen bonding network within the oxidised flavin binding pocket (Key et al., 2007a).

route of entry for molecular oxygen. This cavity contains two water molecules that form hydrogen bonds with the FAD co-factor and several amino acid side chains (Figure 1.10B). Analysis of the PAS1 structure has led to a model of signal perception whereby diatomic oxygen attacks the C4α carbon atom of the isoalloxazine ring leading to deprotonation of the N5 atom and thereby triggering a re-organisation of the hydrogen bonding network (Figure 1.10B). It is this shift in the pattern of hydrogen bonding that constitutes the initial structural change associated with signal perception. Switching of FAD between its reduced and oxidised form necessitates generation of an intermediate hydroperoxy species. The structure suggests two possible catalysts for the production of this intermediate, the glutamic acid at position 70 or a nearby water molecule (Figure 1.10B). Substitution of this glutamic acid for alanine blocks redox sensing by NifL (Salinas P., Little R. and Dixon R., unpublished data). Changes in the position of side chains from several residues in the central β-sheet (E70, H133 and S39) have been observed after an extended period of X-ray illumination, indicating a structural change in the β-sheet upon photoreduction. This provides a potential mechanism by which signals may be propagated to influence the conformation of other regions of the protein (Key et al., 2007a).

(ii) MmoS

MmoS is a sensor protein that regulates expression of the soluble methane monooxygenase (sMMO) in *Methylococcus capsulatus* (Bath). Under conditions of copper starvation, MmoS activates transcription of genes involved in sMMO biosynthesis (Csaki et al., 2003). MmoS is a complex modular protein that is predicted to contain nine discrete domains. The domain architecture of MmoS is shown in Figure 1.11A. The N-terminus of MmoS is anchored to the cell membrane via a transmembrane domain. The cytoplasmic

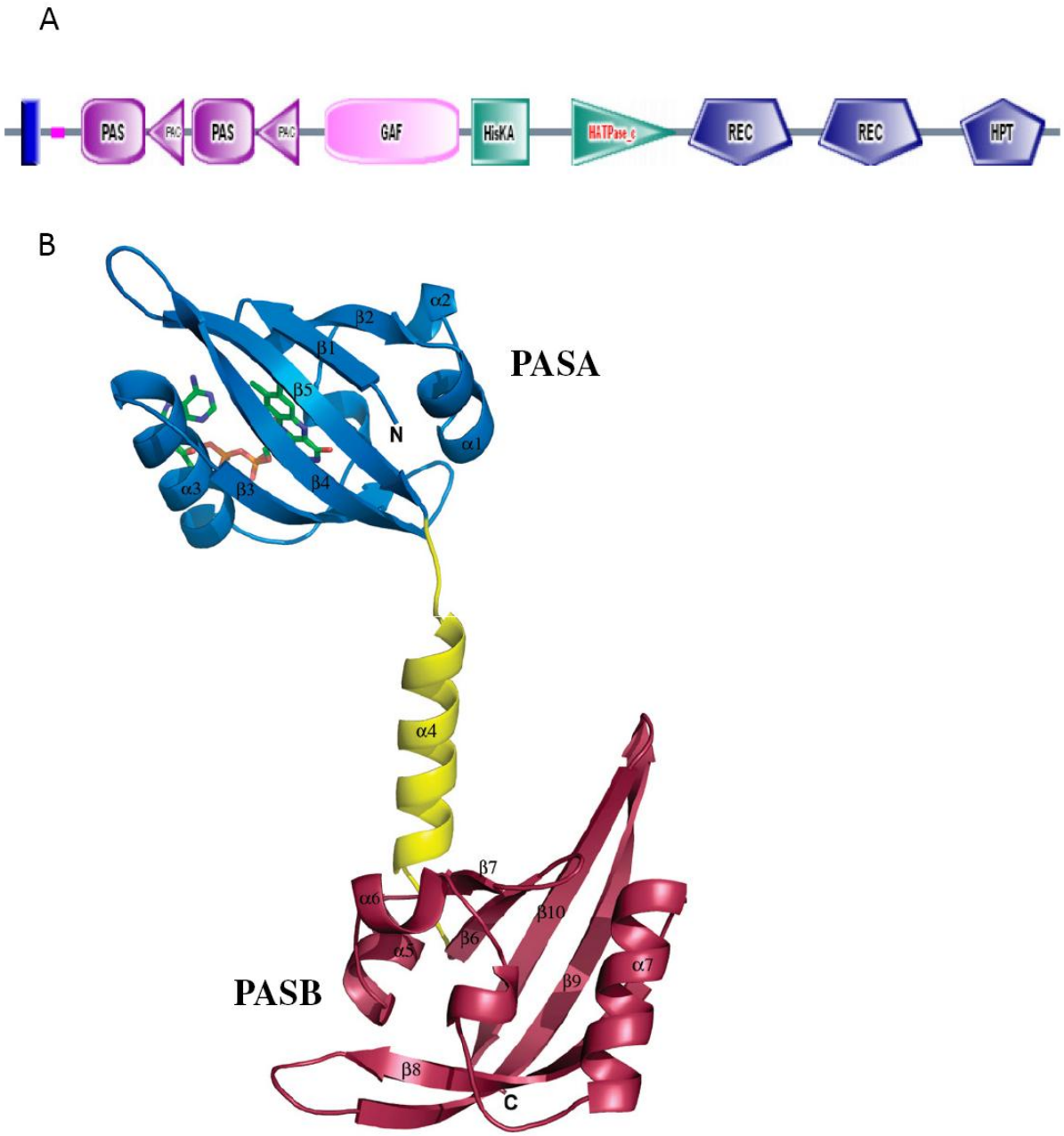


Figure 1.11. (A) Domain architecture of *Methylococcus capsulatus* (Bath) MmoS from the SMART nrdb. (B) Crystal structure of the MmoS PAS domains (Ukaegbu & Rosenzweig, 2009).

portion of the protein contains tandem PAS domains, a GAF domain, histidine kinase domains (HisKA and HATPase modules), two phosphate receiver (REC) domains and a C-terminal histidine phosopho-transfer (HPt) domain (Figure 1.11A). It has been proposed that depletion of copper activates the kinase domains, resulting in autophosphorylation and subsequent phospho-transfer to the HPt domain. These events are likely to stimulate transcription of sMMO biosynthesis genes via the sequential activation of two further proteins (MmoQ and MmoR) (Ukaegbu et al., 2006).

The crystal structure of a fragment of the MmoS protein containing the two PAS domains was solved in 2009. This structure is of particular interest as it is the only one available to date showing two cytoplasmic prokaryotic PAS domains in tandem. The protein crystallised as a monomer in the asymmetric unit and the two PAS domains are connected by an α-helical linker (Figure 1.11B). The structure revealed that the N-terminal PAS domain (PASA) binds a FAD co-factor, while the more C-terminal PAS domain (PASB) has no co-factor or obvious ligand-binding pocket (Ukaegbu and Rosenzweig, 2009). The redox potential of the MmoS FAD group is similar to that of NifL and it has been postulated that oxidation/reduction of the PASA co-factor in MmoS regulates the activity of the C-terminal output domains (Ukaegbu et al., 2006). The function of the second PAS domain remains unclear. However, given the lack of any co-factor or ligand binding pocket, it would appear that the PASB domain has a role other than signal perception.

1.2.4 Light sensing PAS domains

(i) PYP

Photoactive yellow protein (PYP) was first discovered in *Halorhodospira halophila* and is thought to have a role in the phototaxis of purple bacteria; *H. halophila* are

negatively phototactic towards blue light (Meyer, 1985; Sprenger et al., 1993). However, PYP has also been identified in *Rhodobacter capsulatus* where *pyp* and its two biosynthetic genes are clustered with the *gvp* genes (encoding proteins required for gas containing vesicle formation) which determine cell buoyancy. Expression of *gvp* genes is responsive to changes in light availability in many bacterial species and it has been proposed that *R. capsulatus* PYP (*Rc*PYP) may be involved in the regulation of cell buoyancy (Kyndt et al., 2004a). Differences in the proposed function of *H. halophila* PYP (*Hh*PYP) and *Rc*PYP *in vivo* correlate to biochemical differences between the proteins. *Hh*PYP and *Rc*PYP have differing absorption spectra and photocycle kinetics (Kyndt et al., 2004a). PYP exemplifies a class of proteins that consist of a single PAS domain only and lack distinct output domains. It seems likely that proteins of this type transduce signals via stimulus-dependant interactions with cellular targets. However, at the time of writing, the details of signal transduction by PYP remain unclear.

PYP is a small photoreceptor protein consisting of 125 amino acids. Exposure of PYP to visible light induces bleaching of the protein's characteristic yellow colour and incubation of the bleached protein in the dark restores its colour within 1 second (Meyer et al., 1987). PYP contains a 4-hydroxycinnamic acid co-factor that is covalently attached to a conserved cysteine residue in the protein moiety via a thioester bond (Baca et al., 1994). This chromophore contains an isomerisable double bond and an ionisable oxygen atom. Light absorption results in protonation of the oxygen atom and concomitant isomerisation of the 4-hydroxycinnamic acid group from the *trans* form to the *cis* form (Figure 1.12A) (Genick et al., 1997; Genick et al., 1998; Kort et al., 1996). Co-factor isomerisation triggers an alteration in the hydrogen bonding pattern in the PYP active site (Figure 1.12B) and initiates a cycle of rapid chemical and conformational changes known as the photocycle. During the cycle, several transient signalling intermediates are formed before

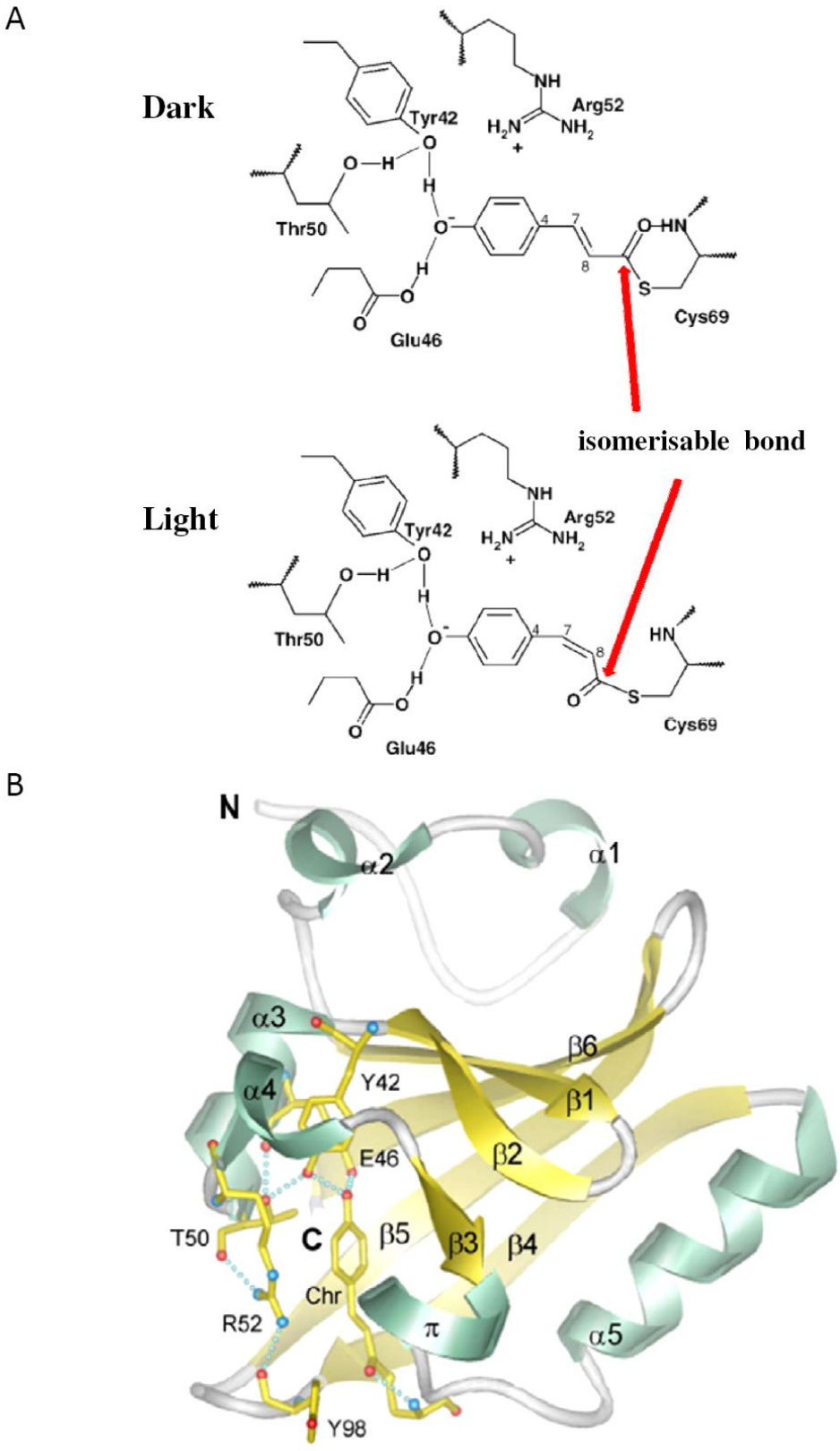


Figure 1.12. (A) Chemical changes in the photoactive yellow protein (PYP) active site in the dark state and light state (Groot et al., 2003) and (B) PYP crystal structure showing the active site hydrogen bonding network (Brudler et al., 2006).

PYP reverts to the dark state (PYP_{dark}). As no activity or downstream interactions have been identified it is difficult to discern which of these conformations represents the "on state". The current hypothesis, based on spectroscopic analogies with rhodopsins, is that a signalling state known as PYP_M (or I_2) represents the active form of PYP. In this state, there are significant alterations in the global conformation and surface properties of PYP. Hydrophobic sites that are buried when PYP is in the dark state become surface exposed in PYP_M and it has been postulated that this change in surface mediates interaction with a transducer protein (Hendriks et al., 2002; Hoff et al., 1999; Imamoto and Kataoka, 2007).

Interestingly, PYP-like domains have also been found in larger proteins that contain additional output domains. Two of these "chimeric" proteins have been studied to date, namely PYP phytochrome-related (Ppr) protein from *Rhodospirillum centenum* and PYP/bacteriophytochrome/diguanylate cyclase (Ppd) from *Thermochromatium tepidum* (Jiang et al., 1999; Kyndt et al., 2004b). Both proteins contain an N-terminal sensor region consisting of a PYP-like domain adjacent to a bacteriophytochrome (Bph) domain. These two sensory domains regulate the activity of C-terminal output domains. Ppr contains histidine kinase effector domains and Ppd has C-terminal GGDEF and EAL domains. There is limited information on intra-molecular signal relay in Ppd, whereas inter-domain communication in Ppr has been the focus of several studies (Jiang et al., 1999; Kamikubo et al., 2008; Kyndt et al., 2007). The activities of the PYP-like and Bph domains of Ppr (which sense blue and red light respectively) appear to be antagonistic; activation of the PYP domain with blue light accelerates recovery (i.e. decay of the activated photocycle intermediate) of the Bph domain after illumination with red light. Conversely, the presence of a functional Bph domain accelerates recovery of the PYP domain after blue light illumination (Kyndt et al., 2007). These results strongly imply that a form of inter-domain communication occurs between the PYP-like and Bph domains of Ppr. Moreover, a recent

study of the full-length Ppr protein demonstrated that the presence of the C-terminal histidine kinase domains can accelerate recovery of the PYP-like domain in the absence of a functional Bph domain, suggesting communication between the PYP-like domain and the output domains (Kamikubo et al., 2008).

PYP is an unusual PAS domain in that it is known to exist as both a distinct protein and a protein domain. The SMART and Pfam databases contain other examples of hypothetical proteins that appear to consist of a single PAS domain (or pair of PAS domains) but none have yet been characterised. PYP not only provides a valuable model for studying the mechanisms by which stimuli induce changes in the signalling state of PAS domains but also provides insight into how PAS domains can be incorporated into complex modular proteins to facilitate integration of multiple signals.

(ii) YtvA

YtvA mediates induction of the general stress response in *Bacillus subtilis* in response to blue light (Akbar et al., 2001; Avila-Perez et al., 2006; Gaidenko et al., 2006). The *B. subtilis* general stress response is controlled by the alternative sigma factor σ^B. In unstressed cells, σ^B activity is inhibited by the anti-σ factor, RsbW. Exposure to a variety of stresses results in inhibition of RsbW by the anti-anti-σ factor, RsbV. Under these conditions, σ^B is released to promote transcription of stress resistance genes. RsbV activity is regulated by two discrete pathways responding to energy stress and environmental stress. YtvA activates the environmental stress pathway. Signal input to this pathway is multi-faceted and will not be discussed in depth in this chapter. One mode of environmental stress detection involves a large protein complex known as the "stressosome", in which the phosphorylation states of several STAS (sulphate transporter and antisigma factor antagonist) domain containing proteins are thought to be responsive to

multiple forms of stress (Marles-Wright et al., 2008). The next step in signal transduction is the release of a "stressosome" component (called RsbU) that indirectly activates RsbV, resulting in activation of σ^B. YtvA has been shown to co-purify with several component proteins of the "stressosome", although the relevance of this to signalling remains unclear (Gaidenko et al., 2006).

The YtvA protein consists of two domains, a sensory N-terminal PAS domain (Y-PAS) and a C-terminal STAS output domain. The STAS domain has been shown to bind GTP and could potentially mediate interactions between YtvA and "stressosome" proteins (Buttani et al., 2006). However, the function and molecular mechanism of signal transduction by the STAS domain are poorly understood. The PAS domain senses blue light via a flavin mononucleotide (FMN) chromophore (Figure 1.13A). The crystal structure of the YtvA PAS domain has been solved in both the illuminated state and the ground state (Möglich and Moffat, 2007). The domain is dimeric in the asymmetric unit and adopts the canonical PAS fold with an additional C-terminal α-helix, called the Jα helix, which extends outward from the globular dimer. Illumination results in the formation of a thioester bond between the C4a atom of the FMN co-factor and a cysteine residue (C62) in the Eα helix (Figure 1.13B). This initial structural change in the active site is propagated by movements in the Eα and Jα helices as well as several loop regions. Overall, signal perception triggers a quaternary structural change whereby Y-PAS subunits undergo a 5^o rotation relative to one another in a "scissor-like" movement (Möglich and Moffat, 2007).

Recent evidence suggests that the Y-PAS Jα helices form a coiled-coil α-helical linker between the PAS domain and the STAS domain in the dimeric YtvA protein and it has been postulated that signals are transmitted between these domains via changes in the

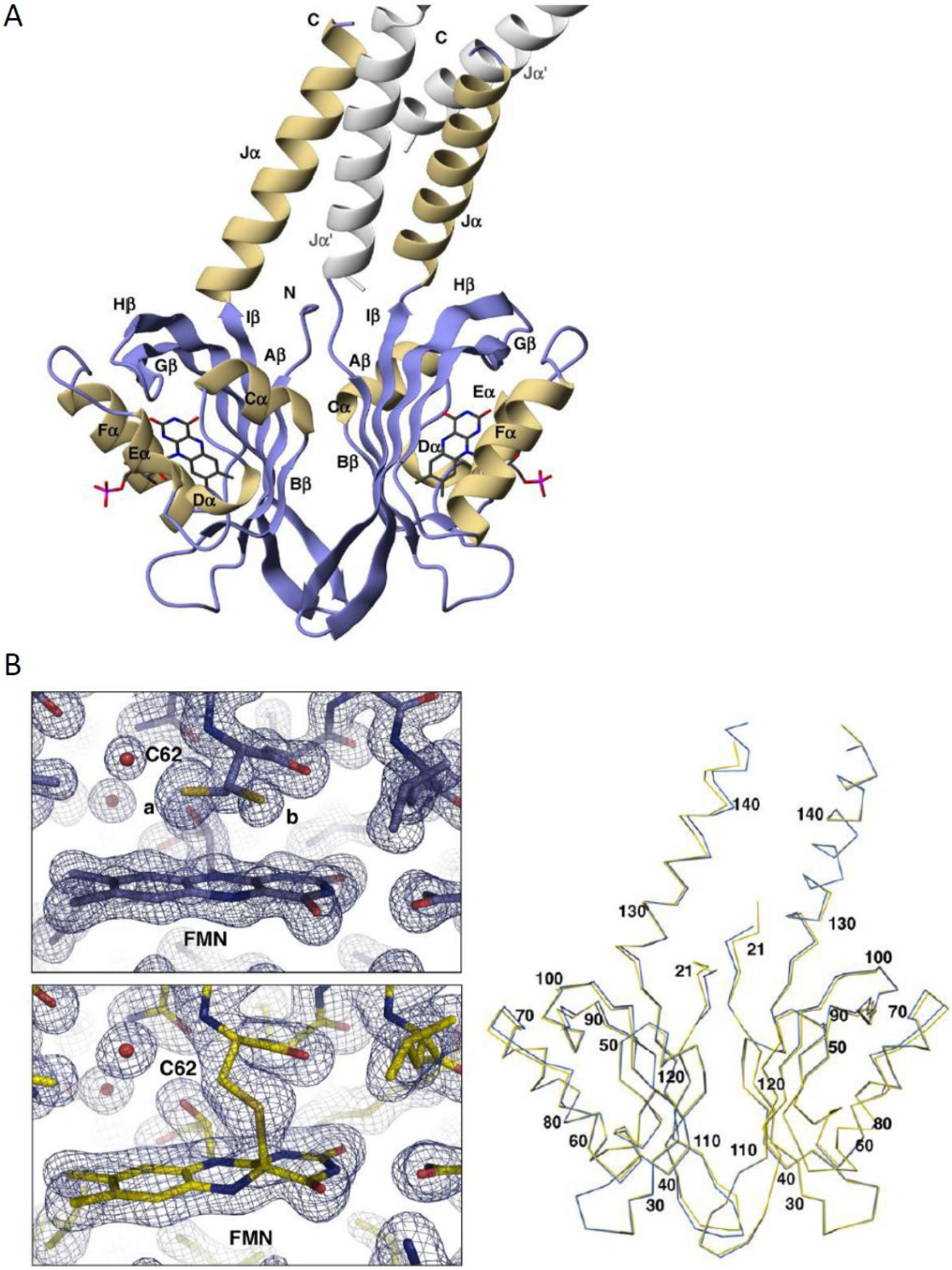


Figure 1.13. (A) Crystal structure of the YtvA PAS domain (Y-PAS). (B) Light dependent structural changes in Y-PAS. The panels on the left show electron density maps of the Y-PAS FMN-binding cavity in the ground state (upper panel) and after blue light illumination (lower panel). C^{α} traces of Y-PAS in the illuminated state (yellow) and ground state (blue) are shown on the right.

quaternary structure or stability of the Jα helices (Möglich and Moffat, 2007). Moreover, Moffat and colleagues have demonstrated that a chimeric protein containing Y-PAS fused to the FixL output (histidine kinase) domains exhibits light dependant histidine kinase activity *in vivo* and *in vitro* (Möglich et al., 2009a). This supports the hypothesis that the Jα helix, which is present in both the native oxygen-sensing PAS domain of FixL and Y-PAS, has a role in signal relay that is conserved in these proteins. Further evidence for inter-domain communication between the PAS and STAS domains of YtvA has been provided by spectroscopic studies using fluorescent GTP analogues (Buttani et al., 2007; Buttani et al., 2006). Light dependent changes have been observed in the spectroscopic properties of the GTP-TR bound protein, indicating that the sensory state of the PAS domain influences the conformation of the GTP binding site in the STAS domain. Moreover, mutational analysis indicates that light-dependant GTP binding is important for YtvA function *in vivo* (Avila-Perez et al., 2009).

1.2.5 PAS domains and protein-protein interactions

In addition to their role in stimulus detection and signal relay, PAS domains can also mediate protein-protein interactions. Many PAS-containing proteins transduce signals via switching of binding partners and, in several systems, PAS domains are thought to modulate binding partner specificity (Huang et al., 1993; Lindebro et al., 1995; Pongratz et al., 1998; Rowlands and Gustafsson, 1997). A well studied example of this is the aryl hydrocarbon receptor (AhR). AhR is a eukaryotic transcription factor found in numerous species and diverse tissue types. AhR activity influences various signalling pathways involved in many cellular processes, including cell cycle regulation, development and apoptosis (Puga et al., 2009). However, AhR is best characterised for its role in the xenobiotic enzyme induction pathway, which has been studied since the 1970's (Schmidt

and Bradfield, 1996). In its inactive state, AhR is found in the cytoplasm in complex with several other proteins (Petrulis and Perdew, 2002). Ligand binding to one of two PAS domains triggers a conformational change in the protein that exposes a nuclear localisation sequence, resulting in movement of AhR from the cytoplasm into the nucleus (Henry and Gasiewicz, 2003; Hord and Perdew, 1994). AhR then dissociates from the complex and dimerises with a second PAS-containing protein called ARNT (Aryl hydrocarbon receptor nuclear translocator) to form a transcriptionally active hetero-dimer (Hankinson, 1995). The N-terminal PAS domains of AhR are important in each of these signal transduction steps. They mediate protein-protein interactions with ARNT as well as at least one component of the cytoplasmic signalling complex (Perdew, 1988; Reisz-Porszasz et al., 1994). ARNT, like AhR, contains two PAS domains that modulate switching of interaction partners. ARNT is capable of forming a homo-dimer or interacting with one of with three other PAS-containing proteins (including AhR) to form hetero-dimers (Lees and Whitelaw, 1999; Moffett et al., 1997; Pollenz et al., 1994). Interaction with one of these partners, namely hypoxia-inducible factor (HIF-α), is achieved by hetero-dimerisation of the PAS domains from each protein (Erbel et al., 2003; Yang et al., 2005). Isolated PAS domains from these proteins also interact *in vitro* (Erbel et al., 2003). Therefore, the ARNT PAS domains play a role in the interactions with at least two of its three binding partners. Thus, PAS domains can function as protein-protein interaction modules and, in this capacity, they are important in many signalling pathways that control diverse biological processes including the hypoxic response and cell cycle regulation in eukaryotes.

1.2.6 Common aspects of PAS domain signalling

Despite the versatility of PAS domains with respect to their biological function, their highly conserved structure implies that some aspects of the signal transduction

mechanism are likely to be, at least partially, conserved. Signal-dependant structural changes in the central β-sheet have been reported in PAS domains from a diverse range of proteins. These include light sensing PAS domains from plants (phototropins), fungi (*N. crassa* Vivid) and bacteria (YtvA and PYP) as well as bacterial ligand-binding PAS domains (CitA) and eukaryotic protein-protein interaction PAS domains (ARNT) (Evans et al., 2009; Harper et al., 2003; Möglich and Moffat, 2007; Rajagopal et al., 2005; Sevvana et al., 2008; Zoltowski et al., 2007). This is consistent with the importance of the β-sheet in co-factor binding and dimerisation of many PAS domains and implies that the central β-sheet has a conserved role in signal propagation in PAS domains of varying function. This is particularly interesting given that the output domains of these proteins, to which the signalling state of the PAS domain(s) must be relayed, are dissimilar in structure and function. In a recent review, Möglich and colleagues note that the tertiary structural uniformity of the PAS core is in stark contrast to the structurally diverse output domains, suggesting that signal transmission is not dependent on tertiary structural recognition between domains (Möglich et al., 2009b). The authors also point out that the majority of PAS-associated output domains function as oligomers and that alterations in quaternary structure could therefore provide a common mechanism of signal transduction. There is a considerable body of evidence to support this hypothesis. Quaternary structural changes in PAS domains from FixL, CitA, DctB and KinA regulate the activity of histidine kinase output domains, whilst signal-dependent alterations in the quaternary structure of PAS domains from *Ec*DOS and YtvA module the activity of EAL and STAS output domains respectively (Ayers and Moffat, 2008; Kurokawa et al., 2004; Lee et al., 2008; Möglich and Moffat, 2007; Zhou et al., 2008). The importance of quaternary re-arrangements to the signalling mechanism has also been demonstrated in PAS-containing proteins from plants and mammals (Erbel et al., 2003; Evans et al., 2009; Nakasako et al., 2008). Moreover, the

oxygen-sensing PAS domain from *Bj*FixL can adopt two distinct quaternary structures with similar free energy (Ayers and Moffat, 2008). The authors also identified an extended dimerisation interface that is conserved in PAS domains from *Ec*DOS, YtvA, *Av*NifL and *Cr*Phot (*Chlamydomonas reinhardtii* phototropin) and may facilitate switching between alternative quaternary arrangements. Overall, the available information suggests a model for PAS domain signalling whereby signal perception induces changes in the association state or orientation of PAS subunits to influence the activity of output domains.

1.3 Histidine Protein Kinases

As discussed in section 1.1, bacteria utilise two-component systems (TCSs) to adapt their physiology according to changes in their environment. A typical two-component system consists of a histidine protein kinase (HPK) and cognate response regulator (RR). The HPK phosphorylates the RR in response to environmental stimuli and the phosphorylated RR then elicits a cellular response, often via a change in gene expression. This section will focus on HPKs due to their relevance to the system studied in this thesis. RRs will not be considered in detail.

Since their discovery in the 1980s, HPKs have been shown to control a plethora of cellular processes in bacteria, including motility, various metabolic switches, virulence, nutrient uptake and many aspects of development (Mascher et al., 2006). Several genome analysis studies have highlighted the importance of two-component signalling systems in model organisms (Fabret et al., 1999; Hutchings et al., 2004; Rodrigue et al., 2000). For example, analysis of the *Streptomyces coelicolor* A3(2) genome indicated the presence of at least 67 TCSs as well as 17 unpaired HPKs and 13 orphan RRs. Of the 84 HPKs, 74 have unknown function. It is thought that the remaining 10 HPKs play roles in the regulation of development, aspects of secondary metabolism, responses to cell wall damage, osmoadaption and the osmotic shock response, the phosphate starvation response, chitinase production and vancomycin resistance (Hutchings et al., 2004). A similar genomic analysis revealed the presence of 36 HPKs and 35 RRs in *Bacillus subtilis* (Fabret et al., 1999). Microarray analysis has since been used to determine the regulons of 27 *B. subtilis* TCSs (Kobayashi et al., 2001; Ogura et al., 2001). The size of these regulons varies considerably between TCSs, ranging from 4 to 98 genes (Kobayashi et al., 2001). Overall, the genes that comprise these TCS controlled regulons are extremely diverse in function and impact most aspects of cellular physiology.

1.3.1 Phosphochemistry

The reactions of HPKs can be split into two stages (Figure 1.14). The first stage is autophosphorylation of the HPK via conversion of ATP to ADP (Figure 1.14, step 1). During this step the γ-phosphoryl group is moved to a conserved histidine residue within the HPK. The second reaction is called the phospho-transfer reaction (Figure 1.14, step 2), in which the phosphate group is switched to a conserved aspartate residue in the RR. Divalent metal ions are a necessity for both reactions. There is a significant difference in the chemistry of phosphorylated histidines compared to their Ser/Thr/Tyr counterparts, namely they are phosphoramidates rather than phosphoesters. Hydrolysis of the phosphoester bond has a free energy (ΔG^o) of between -6.5 and -9.5 kcal mol^{-1} in contrast to a ΔG^o of -12 to -14 kcal mol^{-1} for the P-N bond of phospho-histidine (Stock et al., 1990). As a result, the physiological functions of HPKs tend to fill niches not suited to Ser/Thr/Tyr kinases. For example, the high energy N~P bond of phosphohistidines is apposite for phospho-transfer.

1.3.2 Domain Architecture

HPKs are modular proteins with highly variable domain architectures, reflecting the array of different signals they perceive and transduce. HPKs can also differ dramatically in size; the smallest are less than 40 kDa whilst larger HPKs can exceed 200 kDa. Despite this variability, all HPKs consist of two main regions known as the sensor region and the core transmitter region. Both of these are discussed below in detail. HPKs are grouped into two distinct categories, namely orthodox HPKs and hybrid HPKs (Parkinson and Kofoid, 1992). In orthodox HPKs, the conserved histidine residue in the core transmitter region is the sole site of phosphorylation (Figure 1.15 and Figure 1.17). Of the 29 HPKs in *E. coli*, 24 are orthodox HPKs (Mizuno, 1997). In the second category,

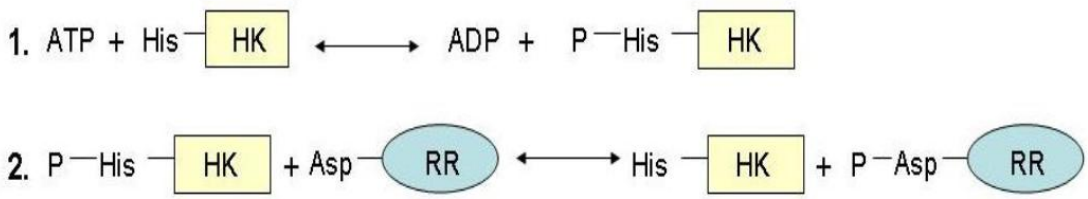


Figure 1.14. The two reactions of histidine protein kinases: (1) autophosphorylation and (2) phospho-transfer.

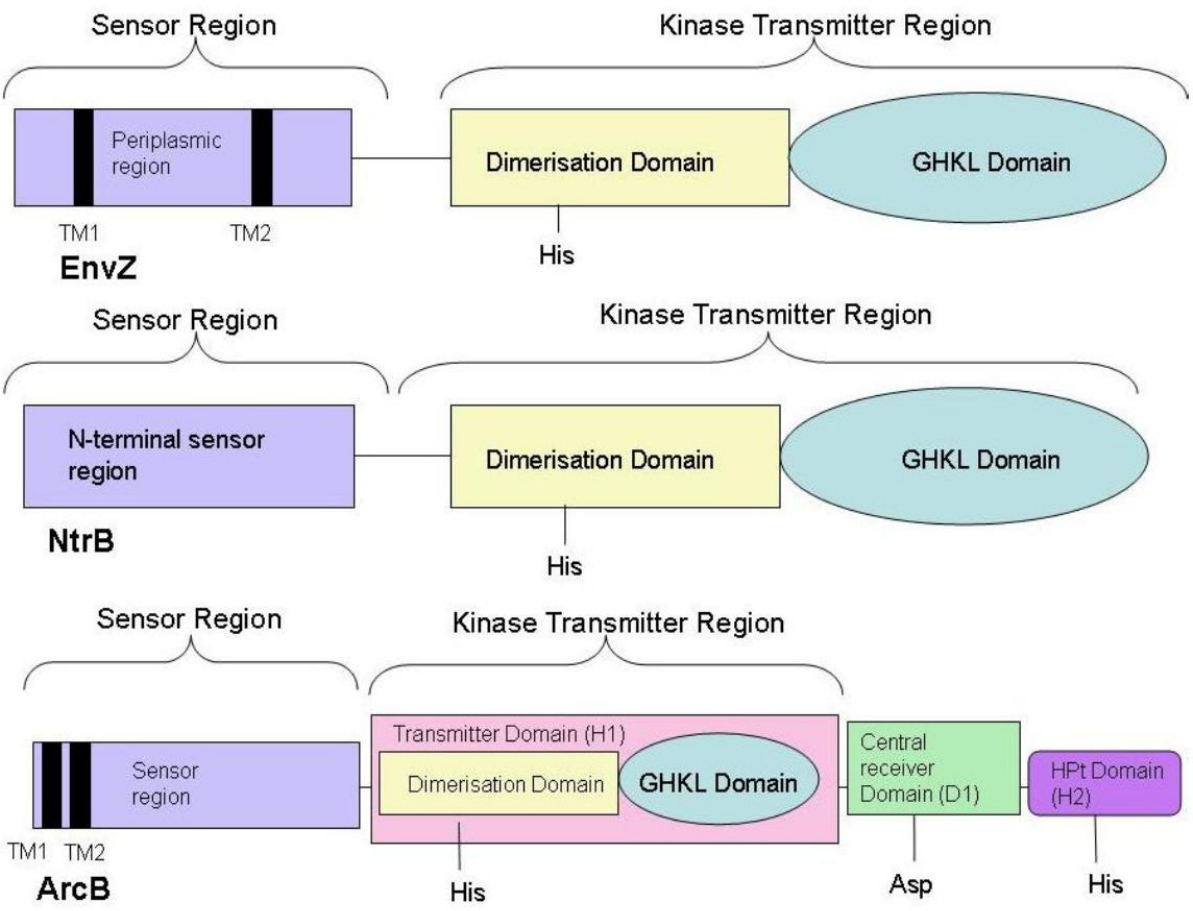


Figure 1.15. Domain architectures of three well studied HPKs. All contain sensor regions, core kinase domains and a conserved histidine residue. EnvZ, like the majority of HPKs, has a periplasmic sensor region between two transmembrane helices (TM1 and TM2). NtrB is an entirely cytoplasmic HPK. ArcB is a hybrid HPK with two additional modules: the central receiver domain (or D1 domain) and the histidine phospho-transfer (HPt) domain. After autophosphorylation of the conserved histidine residue in the kinase transmitter region, the phosphate group is transferred first to an aspartate residue in the D1 domain and then to the histidine of the HPt domain, before finally being switched to the RR. The two TM regions of ArcB are separated by only 16 amino acids and function as anchorage to the membrane rather than enclosing a periplasmic sensor region, as in EnvZ.

hybrid kinases (exemplified by ArcB in Figure 1.15), the site of autophosphorylation is not in the core transmitter region and/or there is a more complicated phospho-relay system involving other histidine-containing domains or proteins (Stock et al., 2000). Indeed, it is not uncommon for hybrid kinase systems to consist of more than two components. This chapter will focus on orthodox HPKs.

In vitro, HPKs form homodimers and in most cases autophosphorylation (Figure 1.14, step 1) is thought to occur in *trans* between subunits (Ninfa et al., 1993; Surette et al., 1996; Swanson et al., 1993). It is also possible for HPKs to catalyse the dephosphorylation of their partnered RR (though most RRs exert their own autophosphatase activity) (Kanamaru et al., 1989; Keener and Kustu, 1988). The relative rates of these reactions determine the kinetics and efficacy of the response. These, in turn, are controlled via signals initiated by the sensor region of the HPK, in response to environmental cues.

1.3.3 The Sensor Region

Sensor regions are poorly conserved between different HPKs and there are no ubiquitous motifs. The mechanisms by which environmental signals are perceived are extremely variable, reflecting the diversity of the stimuli to which HPKs respond. The sensor regions of many HPKs incorporate more than one sensory module and integrate multiple signals. Modules commonly recruited to HPK sensor regions include PAS and GAF domains. In fact, 33% of HPKs contain a cytoplasmic PAS domain whilst 9% contain a GAF domain (Szurmant et al., 2007). It should be remembered that signal perception is not strictly limited to the sensor region. An example of this is the NtrB protein, a HPK that, together with its cognate RR (NtrC), regulates the transcription of genes involved in nitrogen metabolism and uptake in *E. coli*. The signal for nitrogen status that controls NtrB

activity is sensed by a P_{II} signal transduction protein, which interacts directly with the kinase core region (Pioszak et al., 2000).

1.3.4 The Kinase Transmitter Region

The kinase transmitter region contains two domains: the GHKL domain and the dimerisation and histidine phospho-transfer (DHp) domain. In an orthodox HPK, the GHKL domain consists of approximately 250 amino acids and is responsible for nucleotide binding and kinase activity (Parkinson and Kofoid, 1992). The dimerisation domain contains a conserved His residue that is the site of autophosphorylation by the GHKL domain (Stock et al., 2000). The kinase transmitter region is more highly conserved than the sensor region and structural information is available on the GHKL and DHp domains from several HPKs.

(i) Structure and function of dimerisation domains

As mentioned above, HPKs are homodimeric and proper dimerisation is required for activity. Dimerisation is mediated by the DHp domain (also known as the dimerisation domain). This domain also contains a conserved histidine residue which becomes phosphorylated when the HPK is active. To date, high resolution structures of only three DHp domains from orthodox HPKs have been characterised: those of *E. coli* EnvZ (Tomomori et al., 1999), *Thermotoga maritima* HK853 (Marina et al., 2005) and *B. subtilus* DesK (Albanesi et al., 2009). Dimerisation domains exhibit some sequence homology and contain a consensus sequence hxxxhxHahhpPhxxh (Figure 1.16). The histidine and proline from this sequence are conserved in all HPKs and there is a high

Figure 1.16. Multiple sequence alignment to illustrate the regions of homology in the dimerisation domains of various HPKs. Secondary structural elements are indicated above. Conserved residues are shown in red and partially conserved hydrophobic residues are indicated by an asterisk (Tomomori et al., 1999).

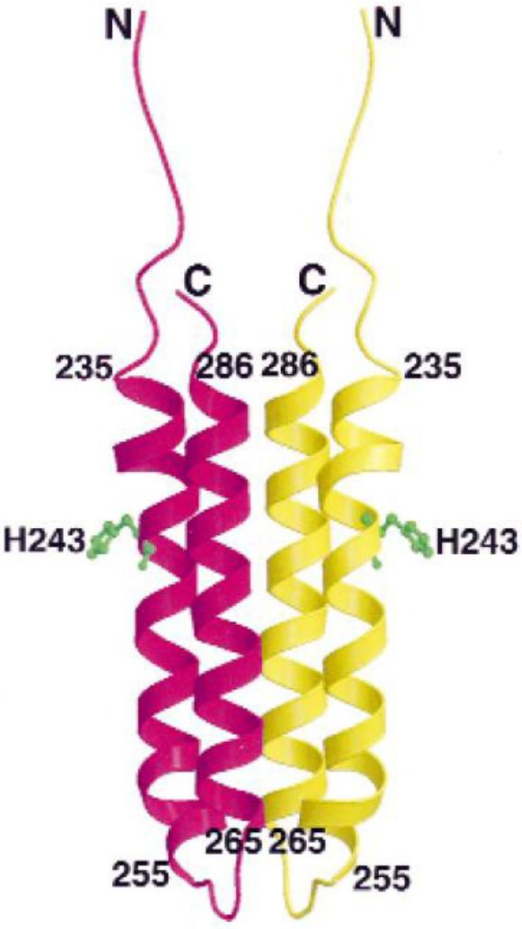


Figure 1.17. Ribbon diagram of the four helix bundle formed by two DHp domain subunits in EnvZ. The histidine residues that are autophosphorylated upon HPK activation are shown in green (Tomomori et al., 1999).

degree of sequence homology between HPKs in the surrounding hydrophobic residues. The secondary structure consists of two α-helices (helix I and helix II). Helix I contains the aforementioned consensus sequence, while helix II contains an additional patch of conserved hydrophobic amino acids (Tomomori et al., 1999).

The first published structure of a dimerisation domain was that of *E. coli* EnvZ (Tomomori et al., 1999). EnvZ and its cognate RR, OmpR, regulate the transcription of genes involved in osmotic homeostasis (such as *ompF* and *ompC* which encode outer-membrane porins) in response to changes in extracellular osmolarity. The DHp domains from each EnvZ subunit combine to form a symmetrical four helix bundle in the EnvZ homodimer (Figure 1.17). There are two inter-subunit surfaces on opposing sides of the four helix bundle (i.e. one from either subunit), each containing an acidic cluster, and two intra-subunit surfaces. The conserved histidines (H243 in *E. coli* EnvZ) are situated on the edge of the molecule, between these surfaces. The inter-subunit surface also incorporates a hydrophobic cluster which, together with the acidic cluster, has been postulated to mediate interactions with the GHKL domain (see below) and OmpR. Moreover, substitutions in the dimerisation domain that impede EnvZ function are predominantly located near the inter-subunit surface, emphasising the importance of dimerisation to the kinase function of the GHKL domain (Portnoy et al., 1999; Tomomori et al., 1999).

(ii) Structure and function of GHKL domains

The catalytic domain of HPKs is called the GHKL domain. This domain binds ATP and catalyses hydrolysis of the γ-phosphate and phosphorylation of the DHp domain (Figure 1.14). The GHKL domain is defined by four conserved sequence motifs, namely the N, G_1 (or D), F and G_2 boxes (Figure 1.18). These motifs are not confined to kinase core domains in HPKs and form the ATP binding sites of structurally homologous domains

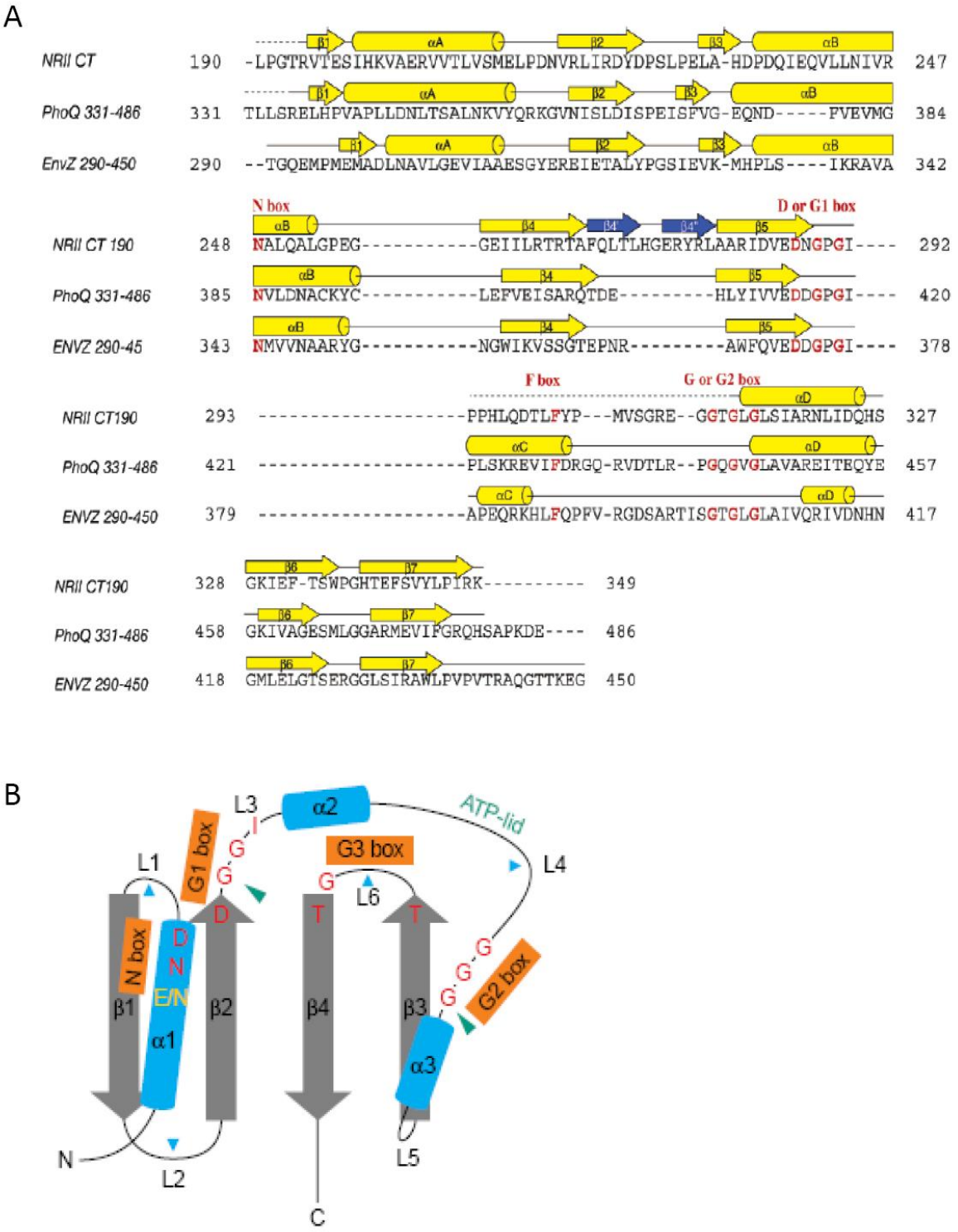
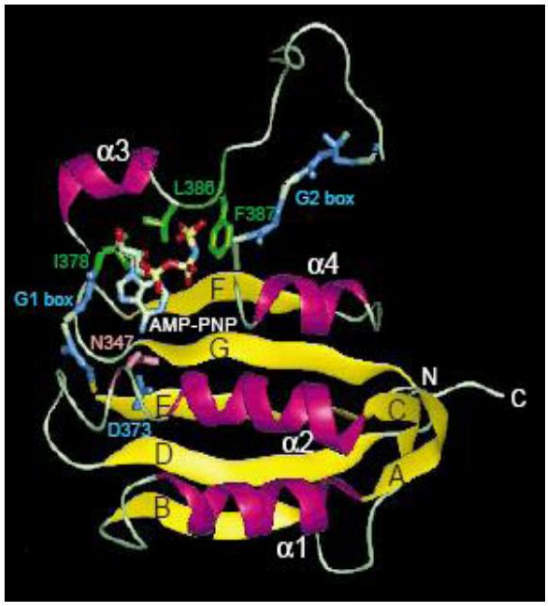


Figure 1.18. (A) Sequence and secondary structure alignment of the GHKL domains from NtrB (or NRII), EnvZ and PhoQ with conserved boxes shown in red (Song et al., 2004). (B) Generalised topology of GHKL domains (Dutta and Inouye, 2000). β-strands are coloured gray, α-helices blue and the N, G_1, G_2 and G_3 boxes are marked in orange. Fully conserved residues are red whilst partially conserved amino acids are shown in yellow.

in DNA gyrase B, Hsp90 and MutL (Bilwes et al., 1999; Marina et al., 2001; Tanaka et al., 1998). A fifth region of homology, termed the G3 box (Figure 1.18B), was defined more recently (Dutta and Inouye, 2000). If the GHKL domains of EnvZ, PhoQ (an orthodox HPK involved in the phosphate starvation response) and CheA (a hybrid HPK involved in chemotaxis) are superimposed, approximately 70% of the residues are identically positioned in all three HPKs and the majority of these are clustered around the five conserved boxes. The identical residues outside of these regions are predominantly hydrophobic, buried amino acids that contribute to formation of the core of the molecule (Marina et al., 2001).

Crystal structures of the GHKL domains of DesK, CheA, PhoQ and (a mutant form of) NtrB are available (Albanesi et al., 2009; Bilwes et al., 1999; Marina et al., 2001; Song et al., 2004; Tanaka et al., 1998) in addition to the NMR structure of EnvZ (Tanaka et al., 1998). All revealed an autonomously folding two-layer α/β sandwich. In EnvZ and PhoQ, this sandwich consists of a five stranded β-sheet (Figure 1.19A , EnvZ strands B, D, E, F and G and Figure 1.19B, PhoQ strands βB, βD, βE, βG and βF) and 3 α-helices (α1, α2 and α3 in Figure 1.19B) that enclose a central hydrophobic core. The non-hydrolysable ATP analogue (AMP-PNP) utilised in crystallisation is located in a deep cavity at one end of the molecule, while the opposing end is sealed by a small anti-parallel β-sheet comprised of strands A and C from EnvZ (Figure 1.19A) or βA and βC from PhoQ (Figure 1.19B). The structure of EnvZ indicated a high degree of flexibility in this ATP binding region, and the adjacent loops are known to undergo structural changes in MutL upon ATP binding and hydrolysis (Ban et al., 1999). This has prompted the suggestion that nucleotide binding in HPKs may induce analogous changes in conformation (Stock, 1999).

A



B

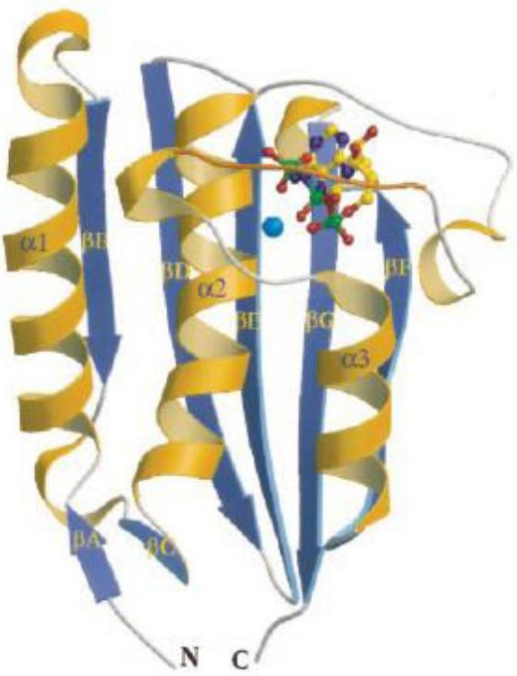


Figure 1.19. (A) Structure of the EnvZ GHKL domain bound to AMP-PNP as determined by NMR (Tanaka et al., 1998). (B) Structure of the GHKL domain of PhoQ complexed with AMP-PNP and a magnesium ion co-factor as determined by X-ray crystallography (Marina et al., 2001).

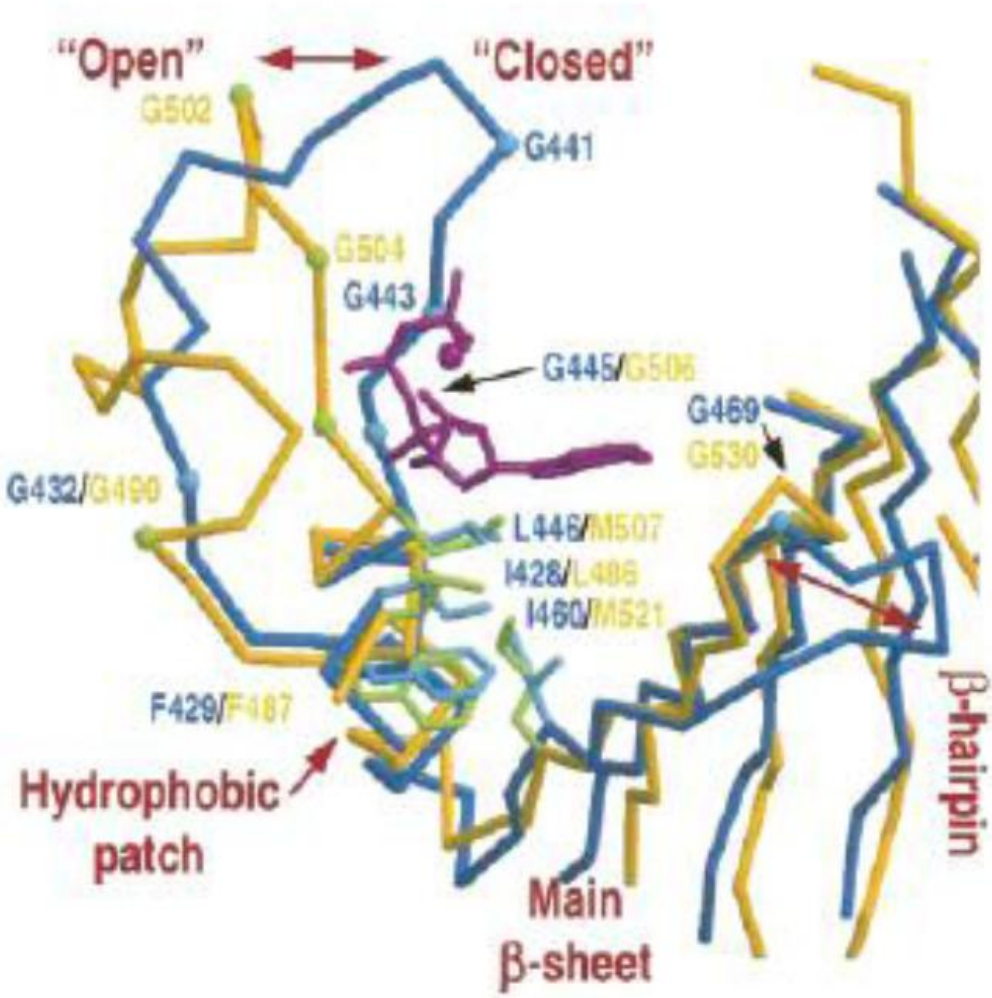


Figure 1.20. The carbon backbones of PhoQ (blue) and CheA (yellow) superimposed to illustrate the "open" and "closed" conformations of the ATP lid (Marina et al., 2001).

The extended loop that covers the AMP-PNP molecule in Figure 1.19b has been termed the ATP lid (Dutta and Inouye, 2000). It contains the G_2 and F boxes. The ATP lid is highly flexible and its motility is aided by the three glycine residues that constitute the G_2 box. The F box functions as an N-terminal anchor for the ATP lid in PhoQ and EnvZ. The C-terminal end is tethered by a conserved hydrophobic patch (Figure 1.20). In PhoQ, the γ-phosphate group forms hydrogen bonds with amino acid side chains from the ATP lid and the N box. Moreover, three residues make extensive contacts with both the ATP analogue and the chelated magnesium ion co-factor, indicating that ATP binding may encourage a more "closed" conformation (Marina et al., 2001). The difference between the "open" and "closed" positions of the ATP lid is clearest when comparing the AMP-PMP bound PhoQ structure to that of the ligand-free CheA protein (Figure 1.20). Nucleotide binding is believed to be the main effecter of this change in conformation as the anchoring residues at either terminus appear to be largely super-imposable (Marina et al., 2001). The ATP loop is thought to be vital for proper HPK function and mutagenesis of its proposed hinge region eliminates kinase activity in EnvZ (Yang and Inouye, 1993).

Despite the characteristic sequence motifs of GHKL domains in HPKs, there remains some significant variability between them. 11 separate categories have been described (Grebe and Stock, 1999). NtrB contains a short β-hairpin between strands β4 and β5 which is comprised of two β-strands (β4' and β4") and is completely absent from all other HPKs of known structure (Figure 1.18A). This novel structural feature has been suggested as the site of NtrB interaction with the P_{II} signalling protein, and several substitutions in this vicinity significantly impair P_{II} binding (Pioszak and Ninfa, 2003; Song et al., 2004).

(iii) Domain interactions in the transmitter region

In 2005 the crystal structure of the entire kinase transmitter region of a HPK from *Thermotoga maritima* (HK853) was published. As predicted, the dimerisation interface was confined to the DHp domain. Interactions between the GHKL and DHp domains occur via conserved, buried hydrophobic residues and a coiled coil linker region. Helix II of the dimerisation domain interacts with two helices in the GHKL domain analogous to the α1 and α2 helices from the GHKL domains of PhoQ and EnvZ (Figure 1.19). Additionally, the phenylalanine residue that constitutes the F box of the GHKL domain is orientated towards a hydrophobic pocket in helix I. This extensive structural connection may allow for the large movement of the GHKL domain, relative to the H domain, that is necessary to bring the catalytic region into close enough proximity of the conserved histidine to achieve autophosphorylation. This led to a model of HPK autophosphorylation in which stimulus perception (or the absence thereof) causes the GHKL domain of one subunit to shift towards the DHp domain of the other (Figure 1.21). After autophosphorylation, a second shift in the position of the GHKL domain, away from the phospho-histidine, is required to allow docking of the RR for phospho-transfer. HPKs, therefore, must undergo large domain movements in order to adopt multiple conformational states (Marina et al., 2005).

Our understanding of domain interactions in the transmitter region has recently been enhanced by three further structural studies (Albanesi et al., 2009; Bick et al., 2009; Casino et al., 2009). The crystal structure of the cytoplasmic region of *T. maritima* HK853 in complex with its cognate response regulator is now available (Casino et al., 2009). Thus, structures of the DHp and GHKL domain are known in two discrete signalling states. Comparison of these structures implies that, contrary to the accepted paradigm, the autophosphorylation reaction in the HK853 dimer occurs in *cis*. This observation was

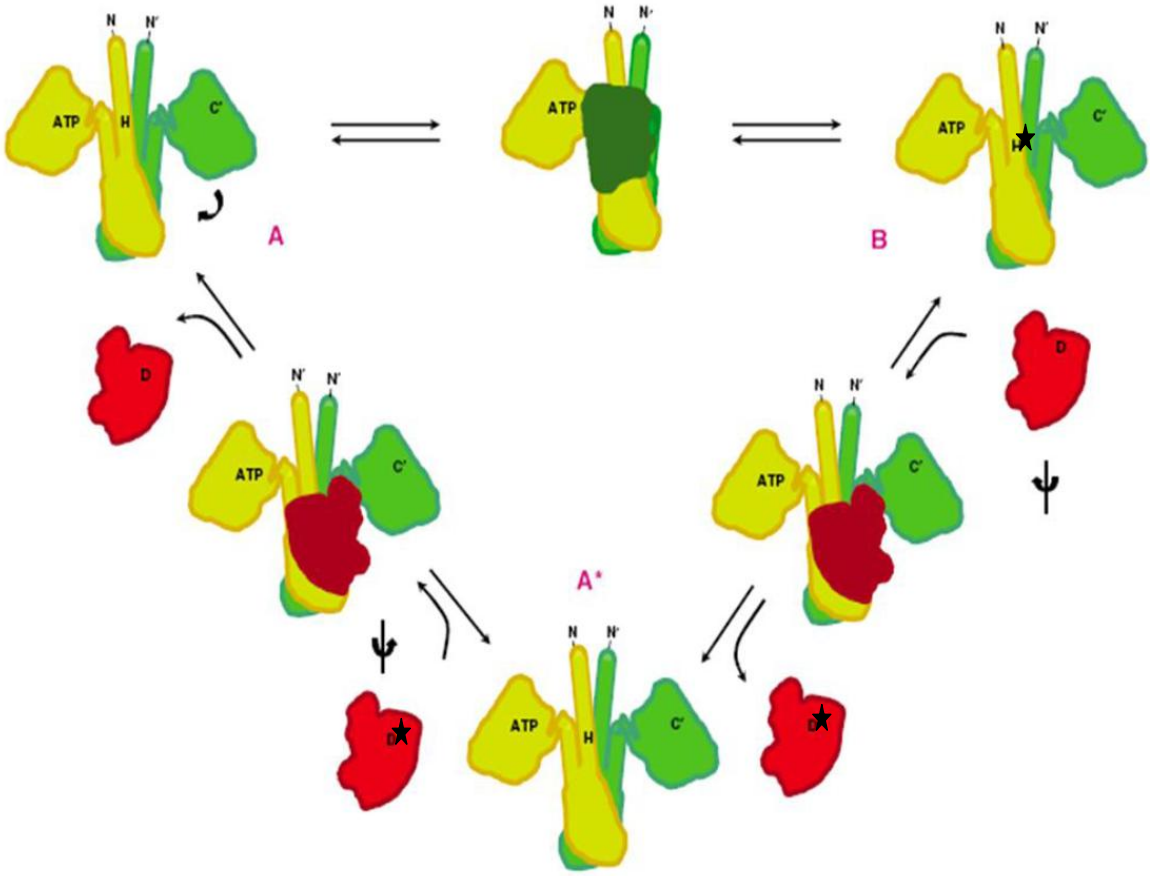


Figure 1.21. Model of the HPK domain re-arrangements that occur during the autophosphorylation (A→B), phospho-transfer (B→A*) and phosphatase (A*→A) reactions (Marina et al., 2005). HPK protomers are shown in yellow and green with the N and C termini indicated. The RR is red and the black star denotes a phosphoryl group. The approximate positions of the phospho-accepting histidine on the HPK (H) and phospho-accepting aspartate on the RR (D) are also indicated.

confirmed biochemically in both HK853 and a second HPK that shares a high degree of sequence homology with HK853, *Staphylococcus aureus* PhoR (Casino et al., 2009). The relative frequency of this *cis* mechanism compared to the *trans* mechanism established for EnvZ, NtrB and CheA is not known (Ninfa et al., 1993; Surette et al., 1996; Swanson et al., 1993). These results necessitate a minor refinement of the above model of HPK domain interactions (Figure 1.21). There is still thought to be a signal-dependent movement of the GHKL domain relative to the dimerisation domain. However, in some cases this may entail a shift in the position of the GHKL domain towards the conserved histidine residue in the dimerisation domain of the same protomer, rather than that of the opposing protomer as previously suggested. The recent structural data support a model for signal transduction in HK853 in which a series of rotational movements between pairs of helices, beginning in the transmembrane region and extending to the upper part of the dimerisation domain, result in altered interactions between the GHKL and DHp domains (Casino et al., 2009). This model is consistent with newly available structural information on two other HPKs, *B. subtilis* KinB and DesK (Albanesi et al., 2009; Bick et al., 2009). The structure of DesK is available in three discrete conformations, thought to correspond to the unphosphorylated, phosphorylated and phosphatase competent states (Albanesi et al., 2009). In this system, rotational movements in the DHp domain appear to mediate the transition between these states by modulating interactions between the DHp domain and the RR or GHKL domain. Overall, structural plasticity in the dimerisation domain is important to signal transduction in several systems and may allow HPKs to undergo the sizable conformational changes needed to accommodate multiple signalling states.

1.4 The NifL-NifA system

Biological nitrogen fixation requires the reduction of atmospheric dinitrogen to ammonia under physiological conditions. Nitrogen fixation is an essential component of the nitrogen cycle and maintains nitrogen levels in the biosphere. In addition to its environmental importance, biological nitrogen fixation is of great consequence to agriculture as the availability of fixed nitrogen is often the limiting factor for crop yield. Biological nitrogen fixation relies on the enzymatic activity of nitrogenase and is an energetically costly and kinetically slow process (Thorneley and Lowe, 1983). Nitrogen fixation is thought to consume 40 mol of ATP per mol of ammonia generated *in vivo* and microbial growth in the absence of a fixed nitrogen source requires a high concentration of nitrogenase (Hill, 1992). Thus, biological nitrogen fixation is only advantageous in specific environments and may damage the competitiveness of the cell if attempted under sub-optimal conditions. This, in conjunction with the irreversible inactivation of nitrogenase upon exposure to oxygen, necessitates stringent transcriptional control of nitrogenase biosynthesis genes in response to the cellular levels of oxygen, fixed nitrogen and carbon. Members of the γ-subgroup of proteobacteria achieve this using the NifL-NifA system.

The NifL-NifA system is best studied in *Azotobacter vinelandii*. When first sequenced, the *nifL* was thought to encode a HPK on the basis of sequence homology (Blanco et al., 1993; Drummond and Wootton, 1987). However, mutational analysis of the conserved histidine residue in NifL demonstrated its redundancy in the signalling mechanism (Woodley and Drummond, 1994). NifA is a transcriptional activator that, under conditions conducive to nitrogen fixation, stimulates the transcription of *nif* genes (required for biosynthesis of the molybdenum-dependent nitrogenase). When environmental circumstances do not favour nitrogen fixation, NifL inhibits NifA activity via formation of an inhibitory protein-protein complex. The stability of this complex, and

thus the activity of NifA, is controlled in response to the redox, carbon and fixed nitrogen status of the cell (Dixon and Kahn, 2004; Martinez-Argudo et al., 2004c; Schmitz et al., 2002). All these stimuli, together with the binding of small effector molecules to both proteins, are integrated on a molecular level by the NifL-NifA system via a complicated set of domain interactions that control *nif* gene transcription.

1.4.1 Domain Architecture of NifL

NifL is a modular protein that consists of four discrete domains. The C-terminal region of NifL contains a GHKL (nucleotide binding) domain (Figure 1.22B) with its characteristic N, G1, F and G2 boxes (Blanco et al., 1993; Drummond and Wootton, 1987). However, this domain does not hydrolyse ATP and no autophosphorylation reaction occurs in NifL (Söderbäck et al., 1998). A conserved histidine residue in a region similar to the DHp domains of HPKs is also apparent. However, the mechanism of signal transduction in NifL deviates from that of HPKs, as substitution of this histidine for alanine, phenylalanine, serine, lysine or valine (among others) has no effect on NifL-NifA interactions (Woodley and Drummond, 1994). This DHp-like domain in NifL is known as the H domain (Figure 1.22B). Secondary structure predictions indicate that the NifL H domain may form an anti-parallel four-helix bundle within the NifL dimer, similar to those found in HPKs. However, it should be remembered that evidence regarding the oligomerisation state of NifL is not conclusive. The sensory N-terminal region of NifL contains tandem PAS domains. As discussed in section 1.2.3, the most N-terminal of these PAS domains, PAS1, is responsible for redox sensing (Hill et al., 1996, Söderbäck et al., 1998; Key et al., 2007a), whilst the function of the second PAS domain, PAS2, remains unclear. However, preliminary evidence suggests a role for PAS2 in signal relay (see section 1.5).

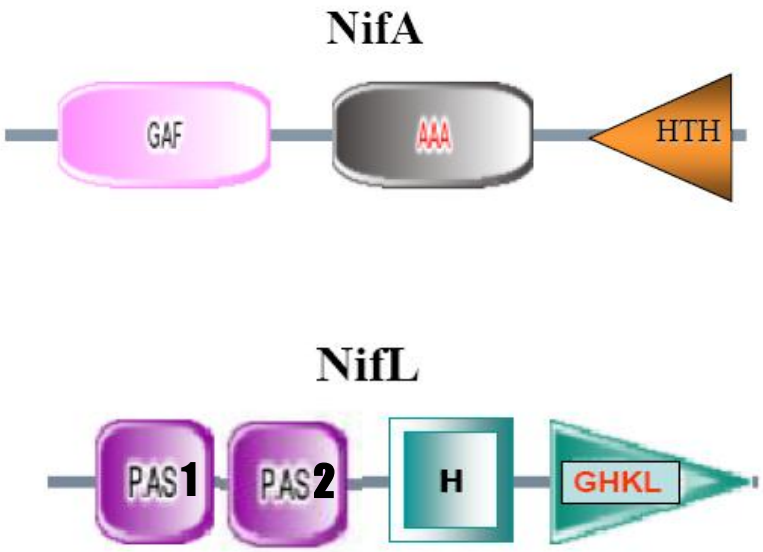


Figure 1.22 Domain architectures of the (A) NifA and (B) NifL proteins from *Azotobacter vinelandii*.

1.4.2 Domain Architecture of NifA

The *nif*-specific transcriptional activator, NifA, is a member of the bacterial enhancer-binding protein (EBP) family. EBPs are transcriptional activators that recognise enhancer binding sites approximately 100 base pairs up or downstream of sigma54 (σ^{54}) dependent promoters (Morett and Segovia, 1993). Transcription initiation at σ^{54} dependent promoters is atypical in that the RNA polymerase holoenzyme alone is not competent to initiate transcription. An additional transcriptional activator, specifically an EBP, is absolutely required in order to generate an open promoter complex. The EBP utilizes the energy from ATP hydrolysis to catalyse isomerisation of the σ^{54}-RNA polymerase complex, into a transcriptionally competent state (Buck et al., 2000). Nucleotide hydrolysis and interaction of the EBP with σ^{54} are mediated by a protein domain that is conserved in all EBPs, and belongs to the AAA+ (ATPases associated with various cellular activities) superfamily of ATPases (Buck et al., 2000; Zhang et al., 2002). These domains are found in all kingdoms of life and are involved in numerous cellular processes. Their primary function is to convert the chemical energy stored in ATP into mechanical work. In EBPs, ATP hydrolysis drives a series of conformation changes that promote interaction between a conserved (GAFTGA) motif in the AAA+ domain and σ^{54} (Morett and Segovia, 1993; Neuwald et al., 1999; Rappas et al., 2005). NifA is a typical EBP, consisting of three domains: a C-terminal helix-turn-helix (HTH) domain, a AAA+ domain and an N-terminal GAF domain (Figure 1.22A). The HTH domain is a DNA-binding module that recognises specific enhancer elements 100 base pairs upstream of *nif* promoters (Morett and Segovia, 1993; Ray et al., 2002). The GAF domain of NifA binds 2-oxoglutarate and modulates the response of NifA to NifL (Barrett et al., 2001; Little and Dixon, 2003; Martinez-Argudo et al., 2004a). GAF domains have a similar tertiary structure to that of PAS domains (see above) and are thought to be of shared ancestry (Ho et al., 2000).

1.4.3 Factors influencing NifL-NifA interactions

(i) Nucleotide Binding

As mentioned above, the GHKL domain of NifL binds adenosine nucleotides but does not hydrolyse ATP (Söderbäck et al., 1998). NifL is incompetent to bind NifA in the absence of nucleotide *in vitro* (Eydmann et al., 1995) and ADP binding has been shown to stabilise the NifL-NifA binary complex (Eydmann et al., 1995; Money et al., 1999). NifL has a higher affinity for ADP (K_d = 16 μM) than for ATP (K_d = 130 μM). However, the relevance of this *in vivo* is unclear as the cytoplasmic concentrations of both nucleotides are significantly above the dissociation constants for their interactions with NifL (Söderbäck et al., 1998). Moreover, the proportion of cellular nucleotide that is unliganded (i.e. available for interaction with NifL) is not accurately known. Partial protease digestion experiments indicate that nucleotide binding to the GHKL domain of NifL induces a conformational change in the C-terminal region of the protein (Söderbäck et al., 1998). Mutant NifL proteins that are deficient in nucleotide binding are incompetent to inhibit NifA activity *in vivo* and show diminished affinity for NifA *in vitro*. Furthermore, limited proteolysis experiments indicate that the conformational changes in NifL associated with ADP binding are absent in these mutant proteins (Perry et al., 2005). Taken as a whole, the available data indicate that adenosine nucleotides bind to the GHKL domain of NifL causing a conformational change that significantly increases the affinity (and stability) of the NifL-NifA interaction. NifL may or may not sense changes in the ATP/ADP ratio as an indication of energy status *in vivo*.

(ii) The redox signal

Owing to the extreme sensitivity of nitrogenase to oxygen, transcription of *nif* genes may be disadvantageous to the cell under oxidising conditions, even when other

environmental factors dictate that nitrogen fixation is favourable. Therefore, it is appropriate that NifL is competent to inhibit NifA under oxidising conditions. Spectroscopic studies of the N-terminal PAS domain (PAS1) of NifL show absorption peaks at 360 and 445 nm, and shoulders at 420 and 470 nm. These spectral features are characteristic of flavoproteins. Denaturation and further TLC analysis indicated that the prosthetic group is FAD (Hill et al., 1996). Oxidation of this prosthetic group induces a conformational change in NifL that promotes formation of the inhibitory NifL-NifA complex. Upon full reduction of FAD (to $FADH_2$), this inhibition is removed (Hill et al., 1996). Reduction of this redox-sensing group can be achieved *in vitro* using several redox donors and enzymes. The redox potential of these reactions is around ~225mV at pH 8 (Macheroux et al., 1998). However, the relevant redox donor and oxidant *in vivo* are not known. Molecular oxygen is a plausible candidate for the role of electron acceptor as NifL is quickly oxidised upon contact with air to yield hydrogen peroxide (Little et al., 1999). Regardless of the oxidant, oxidation of the PAS1 co-factor is thought to trigger a conformational change in the PAS1 domain via re-organisation of a hydrogen bonding network that surrounds the FAD moiety (Key et al., 2007a). The molecular events that underpin signal perception by the PAS1 domain are discussed in detail in section 1.2.3. The physiological redox donor for the *Klebsiella pneumoniae* NifL (*Kp*NifL) protein is likely to be the menaquinone pool (Thummer et al., 2007). In *K. pneumoniae*, the reduced form of NifL associates with the plasma membrane. This redox-dependent membrane sequestration is important for the release of *Kp*NifA from inhibition by *Kp*NifL under nitrogen fixing conditions (Klopprogge et al., 2002). By contrast, the *Azotobacter vinelandii* NifL protein remains in the cytoplasm irrespective of environmental signals and regulation of NifA activity is mediated solely through signal-dependent conformational changes (Klopprogge et al., 2002).

(iii) GlnK Interactions

In 1998, it was demonstrated that *Azotobacter vinelandii* NifL can sense the cellular nitrogen status independently of the redox signal. Truncated constructs of NifL, lacking the flavin-containing PAS1 domain, are competent to inhibit NifA in response to excess fixed nitrogen (Little et al., 2000; Söderbäck et al., 1998). This is consistent with the large and unnecessary energetic cost that would be incurred by a cell producing nitrogenase in nitrogen-replete conditions, whatever the cellular redox and carbon status. In the *Azotobacter vinelandii* NifL-NifA system, nitrogen sensing occurs via GlnK, a member of the P_{II} signal transduction protein family (van Heeswijk et al., 1995). Under conditions of fixed nitrogen excess, GlnK binds NifL to promote formation of an inhibitory ternary complex with NifA (Little et al., 2002; Little et al., 2000; Rudnick et al., 2002). GlnK is covalently modified by the uridylyltransferase/uridylyl-removing (UTase/UR) enzyme (encoded by the *glnD* gene) depending on cytoplasmic concentrations of glutamine (Arcondeguy et al., 2001). Glutamine is a common signal of cellular nitrogen status and its concentration within the cell increases in proportion to the availability of fixed nitrogen in enteric bacteria (Hu et al., 1999; Ikeda et al., 1996). Under nitrogen-limiting conditions, when cytoplasmic glutamine levels are comparatively low, the UTase activity of GlnD is favoured, resulting in uridylylation of GlnK. This prevents interaction of GlnK with NifL *in vitro*, allowing NifA to dissociate from NifL (and thus activate transcription) if the oxygen and carbon signals are appropriate (Little et al., 2000). By contrast, when fixed nitrogen is readily available, glutamine interacts with GlnD to increase UR activity. Hence, GlnK is deuridylylated and is able to promote formation of the inhibitory GlnK-NifL-NifA ternary complex (Figure 1.23). Uridylylation of GlnK is vital for the release of NifA from inhibition by NifL *in vivo* as strains with impaired UTase activity are not capable of fixing nitrogen. Nitrogen fixation can be restored by insertion

mutations that inactivate NifL (Contreras et al., 1991). GlnK is a trimeric protein with three uridylylation sites (one on each subunit), located on tyrosine residues in a surface-exposed loop (known as the T-loop). Substitutions in the T-loop result in forms of GlnK that are deficient in their interactions with NifL (Little et al., 2002). When deuridylylated, native GlnK interacts specifically with the GHKL domain of *A. vinelandii* NifL and neither of the N-terminal PAS domains, nor the central H region, are required for this interaction (Little et al., 2002).

The GlnK trimer, like other members of the P_{II} protein family, contains three 2-oxoglutarate binding sites and three ATP binding sites. These low-molecular-mass effector molecules modulate the activity of the P_{II} protein (Ninfa and Atkinson, 2000; Radchenko et al., 2010). Increasing levels of 2-oxoglutarate have been demonstrated to promote interaction between *A. vinelandii* GlnK and NifL *in vitro* (Little et al., 2002). This affect occurs within the physiological range and, contrary to the P_{II} proteins from *E. coli*, there is no negative co-operativity in 2-oxoglutarate binding to *A. vinelandii* GlnK. It is possible that more than one molecule of 2-oxoglutarate is required to significantly increase the affinity of GlnK for NifL. Although ATP and Mg^{2+} are required for the GlnK-NifL interaction *in vitro*, the presence of nucleotide binding sites on both NifL and GlnK makes it difficult to dissect the role of ATP in their association (Little et al., 2002).

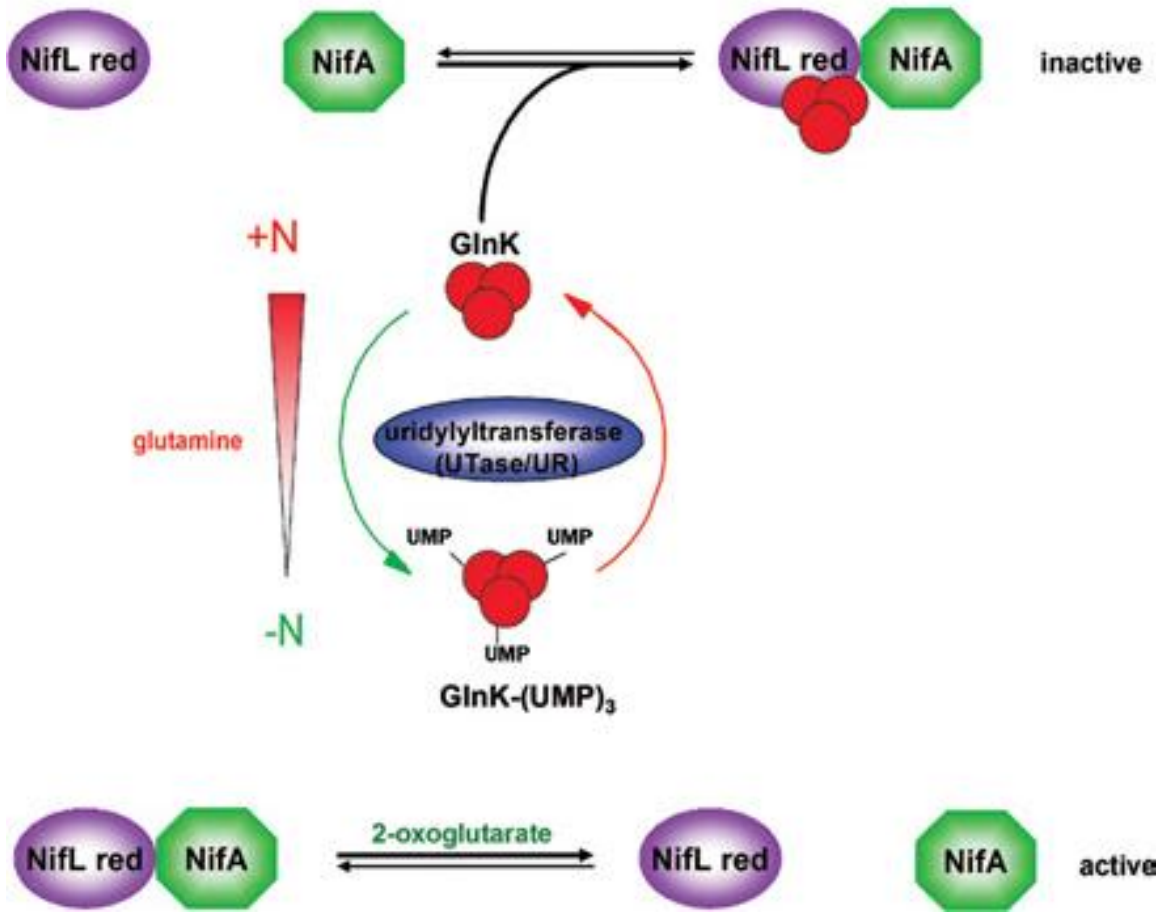


Figure 1.23. Influence of nitrogen availability on GlnK interactions with NifL/NifA. Glutamine levels dictate the uridylylation state of GlnK via affects on UTase/UR activity of GlnD. In its deuridylylated form GlnK promotes assembly of the inhibitory NifA-NifL-GlnK ternary complex. Covalent modification of GlnK prevents the NifL-GlnK interaction, allowing NifA to escape inhibition if 2-oxoglutarate levels are sufficiently high (Martinez-Argudo et al., 2005).

(iv) 2-Oxoglutarate

2-oxoglutarate is an intermediate in the tricarboxylic acid cycle making it an appropriate signal of the carbon status. However, its occurrence in cellular metabolism extends to nitrogen assimilation and thus 2-oxoglutarate can also be thought of as an indirect signal of the nitrogen status. It provides the carbon skeletons for nitrogen assimilation, and so forms a nexus between carbon and nitrogen metabolism. The 2-oxoglutarate concentration within the cell increases with carbon availability and diminishes when fixed nitrogen is replete.

In addition to its role in activating the GlnK signal transduction protein, 2-oxoglutarate binds to the GAF domain of NifA, inducing a conformational change in NifA that antagonises the effect of nucleotide binding to NifL. At high concentrations of 2-oxoglutarate, this allows dissociation of NifA from the reduced, nucleotide-bound form of NifL (Little and Dixon, 2003). However, when 2-oxoglutarate concentrations are relatively low, reduced NifL is competent to inhibit NifA when nucleotide is present (Little et al., 2000). This influence of 2-oxoglutarate on NifA activity is only evident in the presence of NifL. Hence, 2-oxoglutarate must influence the stability of the NifL-NifA binary complex rather than directly altering NifA activity (Martinez-Argudo et al., 2004c). 2-oxoglutarate binds the NifA GAF domain with a K_d of 60 μM *in vitro* (Little and Dixon, 2003). It has been suggested that the physiological concentration of this effector molecule in *E. coli* ranges from 100 μM to 1 mM, depending on carbon and nitrogen status of the cell (Senior, 1975). However, more recent work indicates that the minimum cellular concentration under conditions of nitrogen excess may be <50μM (Reyes-Ramirez et al., 2001) and a decrease in 2-oxoglutarate concentration from 1.4 mM to 0.3 mM occurs within 2 minutes of administering an ammonium shock to N-limited *E. coli* cells (Radchenko et al., 2010). The responsiveness of the system to this effector is, therefore, within the physiological

range. Moreover, a variant form of the NifA protein, containing an amino acid substitution in the GAF domain that eliminates 2-oxoglutarate binding (F119S), is hypersensitive to inhibition by NifL *in vitro* and is unable to escape inhibition sufficiently to allow measurable NifA activity *in vivo*, even under conditions appropriate for nitrogen fixation (Martinez-Argudo et al., 2004b). Taken together, this data suggests that the elevated level of 2-oxoglutarate present when carbon supplies are replete and fixed nitrogen is limiting allows NifA to free itself from inhibition by reduced NifL. In other words, 2-oxoglutarate binding to NifA provides the final "push" necessary to initiate *nif* gene transcription under nitrogen fixing conditions.

1.4.4 Inter-domain interactions in NifL

The interaction between the NifL and NifA proteins may be analogous to the docking of a RR to its cognate HPK in the phosphatase conformation (Little et al., 2007; Marina et al., 2005; Martinez-Argudo et al., 2004a). Based on this analogy and experimental evidence indicating that neither the N-terminal region of NifL nor the GHKL domain alone is competent to bind NifA, it has been suggested that the H domain may provide a surface for NifL-NifA interactions. This necessitates inter-domain communication between the H domain and the sensory modules of NifL in order to transduce oxygen and fixed nitrogen signals. Additionally, the broader analogy between the signalling states of NifL and the conformational changes associated with signalling in HPKs (see section 1.21) implies that transition between the inhibitory and non-inhibitory conformers of NifL may involve movement of the GHKL domain relative to the H domain.

Recent mutagenic studies have identified two distinct classes of amino acid substitution of the H domain of NifL (Little et al., 2007; Martinez-Argudo et al., 2004a). Substitutions belonging to the first class cause NifL to inhibit NifA activity irrespective of

environmental conditions. That is, they lock the NifL protein in an inhibitory conformation (these are referred to as “locked-on” mutants). Substitutions belonging to the second class give rise to a form of NifL that does not inhibit NifA activity in the presence of excess oxygen but responds normally to fixed nitrogen. The identification of substitutions in the H domain that prohibit transduction of the redox signal, whilst leaving GlnK mediated signalling unaffected, indicates that the conformational states required for inhibition in response to oxygen and fixed nitrogen are not equivalent (Little et al., 2007). The NifA variant Y254N is capable of discriminating between these states (Reyes-Ramirez et al., 2002). NifA-Y254N is resistant to inhibition by NifL under conditions of excess fixed nitrogen but is relatively sensitive to inhibition by the oxidised conformer of NifL. Additionally, a variant form of the NifL protein that fails to bind adenosine nucleotides (containing the G480A substitution in the GHKL domain), can inhibit 2-oxoglutarate bound NifA *in vitro* in the presence of deuridylylated GlnK, but is unable to do so in response the redox signal (Perry et al., 2005). Thus, the GlnK bound form of NifL (present under conditions of fixed nitrogen excess) appears to have a diminished requirement for nucleotide binding when compared to the oxidised form of NifL. Overall, broad analogies with HPKs and the likelihood that NifL accesses different conformations when in the binary and ternary inhibitory complexes, suggests that the oxygen and nitrogen stimuli may result in movement of the GHKL domain, relative to the H domain, via different mechanisms.

1.5 Introduction to this work

The *Azotobacter vinelandii* NifL protein contains two N-terminal PAS domains and, prior to the work in this thesis, the function of the second PAS domain (PAS2) was unknown. The first evidence concerning the function of the PAS2 domain was obtained from a mutagenic study of domain interactions in NifL. As mentioned in section 1.4.4, several "locked-on" NifL variants containing substitutions in the H domain have been characterised. One such mutant protein is NifL-R306C (Martinez-Argudo et al., 2004a). By performing random mutagenesis of the *nifL* gene it has been possible to select for secondary substitutions in NifL that suppress the "locked-on" phenotype of the R306C variant. Two such second-site suppressor mutations (encoding L199P and C237R) are located in the PAS2 domain of NifL. In other words, mutant NifL proteins carrying R306C in combination with L199P or C237R in the PAS2 domain allow NifA activity under certain conditions, whereas NifL-R306C does not. Given that the PAS2 domain is not directly required for the NifL-NifA interaction, this implies that PAS2 can influence the conformation of the H domain. Colleagues in the Dixon laboratory have since substituted the leucine at position 199 by arginine or glutamic acid, both of which display "locked-on" phenotypes (i.e. they inhibit NifA activity under all conditions) similar to those exhibited by substitutions of R306. Therefore, PAS2 has a role in the conformational changes in NifL that occur during the transition between the inhibitory and non-inhibitory state. Taken together, the available evidence indicates that the PAS2 domain may have a role in inter-domain communication and signalling in NifL. As mentioned in section 1.2, it is extremely common for modular signalling proteins to contain multiple PAS domains. In many studied proteins containing tandem PAS domains, one domain has a role in signal perception whilst the function of the second domain is poorly understood. The NifL PAS2 domain is typical in this respect and further investigation into its role in signalling may

provide clues regarding the function of tandem PAS domains in modular proteins. The aim of this work was to elucidate the function of the NifL PAS2 domain using a combination of genetic and biochemical techniques. It was hoped that random mutagenesis of the DNA sequence encoding the PAS2 domain would yield mutations in *nifL* with interesting phenotypes and that biochemical analysis of the resulting NifL variants would provide insight into the function of the PAS2 domain.

Chapter 2 - Materials and methods

2.1 Suppliers

All chemicals were purchased from Sigma-Aldrich, Severn Biotech, Fisher, Melford, Bio-Rad or Merck unless otherwise stated. Restriction enzymes and other molecular biology reagents were obtained from Roche, Invitrogen or New England Biolabs. Disposable columns for DNA purification or protein buffer exchange were purchased from Qiagen and Thermo-Scientific respectively.

2.2 Strains and plasmids

All *E. coli* strains and plasmids used in this work are listed in Table 2.1.

Strain or Plasmid	**Description**	**Reference/Source**
***E. coli* Strains**		
DH5α	*sipE44* Δ(*lacU169 hsdR17 recA1 endA1 gyrA96 thi-1 relA1*)	(Hanahan, 1983)
ET8000	*rbs lacZ*::IS*1* gyr*A hutC*$^{c}_{k}$	(Reyes-Ramirez et al., 2001)
BL21 (DE3) pLysS	*F- ompT hsdS*$_B$($r_B^- m_B^-$)*gal dcm*(DE3) carrying pLysS, which encodes T7 lysosyme	(Studier et al., 1990)
BTH101	*F- cya-99 araD139 galE15 galK16 rpsL1 (Str*^{r}*) hsdR2 mcrA1 mcrB1*	(Karimova et al., 2000)
Plasmids		
pPR34	pT7-7 derivative carrying *A. vinelandii NifLA*	(Söderbäck et al., 1998)
pPR54	pPR34 derivative encoding NifL$_{(147-519)}$ and NifA	(Söderbäck et al., 1998)
pPRT22	*nifH-lacZ* reporter plasmid (in pACYC184)	(Tuli and Merrick 1988)
pPR39	pPR34 derivative encoding NifL$_{(454-519)}$ and wild-type NifA	(Söderbäck et al., 1998)
pNLG480A	NifL-G480A in pPR34	(Perry et al. 2005)
pNSK1	NifL-L199E in pPR34	This Work
pNSK2	NifL-L199R in pPR34	This Work
pRL46	NifL-L199P in pPR34	Richard Little
pUT18	BACTH system plasmid with CyaA$_{(225-399)}$, ampicillin resistance marker and MCS.	(Karimova et al., 1998)
pT25	BACTH system plasmid with CyaA$_{(1-244)}$,	(Karimova et al., 1998)

	chloramphenicol resistance marker and MCS.	
pPS1	NifL-L199E, G480A in pPR34	This Work
pPS2	NifL-L199R, G480A in pPR34	This Work
pPS3	NifL-L199E in pPR54	This Work
pPS4	NifL-L199R in pPR54	This Work
pPS5	NifL-L199G in pPR34	This Work
pPS6	NifL-L199Q in pPR34	This Work
pPS7	NifL-L199A in pPR34	This Work
pPS8	NifL-L199V in pPR34	This Work
pPS9	NifL-L199W in pPR34	This Work
pPS10	NifL-L199F in pPR34	This Work
pPS11	NifL-L196A in pPR34	This Work
pPS12	NifL-L200A in pPR34	This Work
pPS13	NifL-R201A in pPR34	This Work
pPS14	NifL-V165D in pPR34	This Work
pPS15	NifL-V165I in pPR34	This Work
pPS16	NifL-E202A in pPR34	This Work
pPS17	NifL-E202K in pPR34	This Work
pPS18	NifL-E194K in pPR34	This Work
pPS19	NifL-L200E in pPR34	This Work
pPS20	NifL-V166M in pPR34	This Work
pPS21	NifL-R240W in pPR34	This Work
pPS22	NifL-L200F in pPR34	This Work
pPS26	NifL-S192G in pPR34	This Work
pPS27	NifL-S193G in pPR34	This Work
pPS28	NifL-S195G in pPR34	This Work
pPS29	NifL-V166D in pPR34	This Work
pPS30	NifL-V166A in pPR34	This Work
pPS31	NifL-V166M, E70A in pPR34	This Work
pPS32	NifL-V166M, I26A in pPR34	This Work
pPS33	NifL-L199Q, E70A in pPR34	This Work
pPS34	NifL-L199Q, I26A in pPR34	This Work
pPS35	NifL-L200A, E70A in pPR34	This Work
pPS36	NifL-L200A, I26A in pPR34	This Work
pPS37	NifL-L200A, F27A in pPR34	This Work
pPS38	NifL-L200A, I22A in pPR34	This Work
pPS54	$NifL_{(143-519)}$ in pPR34	This Work
pPS55	$NifL_{(143-519)}$-L200A in pPR34	This Work
pPS39	NifL-L196P in pPR34	This Work
pPS40	NifL-L200P in pPR34	This Work
pPS42	NifL-L235P in pPR34	This Work
pPS43	NifL-F253L in pPR34	This Work
pPS44	NifL-A302T in pPR34	This Work
pPS45	NifL-I304T in pPR34	This Work
pPS46	NifL-Q308E in pPR34	This Work
pPS47	NifL-N177S in pPR34	This Work
pPS48	NifL-E291A in pPR34	This Work
pETM11	pET24d (Novagen) derivative with a TEV	Pinotsis et al., 2006

	protease cleavage site, polyhistidine tag and MCS	
pETNdeM11	pETM11 with a *Nde* 1 site replacing the *Nco* 1 site for more convenient cloning	(Tucker N., Unpublished work)
pPS50	$NifL_{(143-284)}$ in pETM11	This Work
pPS51	$NifL_{(143-284)}$-L200A in pETM11	This Work
pPS52	$NifL_{(143-284)}$-I153A in pETM11	This Work
pPS53	$NifL_{(143-284)}$-V157A in pETM11	This Work
pPS56	$NifL_{(143-284)}$-L199R in pETM11	This Work
pPS57	$NifL_{(143-284)}$-N177S in pETM11	This Work
pPS61	$NifL_{(147-284)}$-L199R in pT25	This Work
pPS62	$NifL_{(147-284)}$-V157A in pT25	This Work
pPS63	$NifL_{(147-284)}$-V166M in pT25	This Work
pPS66	NifL-S236P in pPR34	This Work
pPS69	$NifL_{(279-519)}$ in pT25	This Work
pPS70	NifL in pETNdeM11	This Work
pPS71	$NifL_{(143-519)}$ in pETNdeM11	This Work
pPS72	NifL-V166M in pETNdeM11	This Work
pPS73	$NifL_{(143-519)}$-V166M in pETNdeM11	This Work
pPS74	$NifL_{(143-284)}$-V166M in pETM11	This Work
pPS75	$NifL_{(143-284)}$-F253L in pETM11	This Work
pPS76	$NifL_{(143-519)}$-V157A in pPR34	This Work
pPS77	$NifL_{(143-519)}$-V166M in pPR34	This Work
pPS78	NifL-L167P in pPR34	This Work
pPS79	NifL-L271P in pPR34	This Work
pPS80	NifL-L283Q in pPR34	This Work
pPS81	NifL-286P in pPR34	This Work
pPS82	NifL-K284E in pPR34	This Work
pPS83	NifL-Q308R in pPR34	This Work
pPS84	NifL-G295Q in pPR34	This Work
pPS85	NifL-L292P in pPR34	This Work
pPS86	$NifL_{(147-284)}$-V166M in pUT18	This Work
pPS87	$NifL_{(147-284)}$-L199R in pUT18	This Work
pPS88	$NifL_{(147-284)}$- L200A in pUT18	This Work
pPS89	$NifL_{(147-284)}$-V157A in pUT18	This Work
pPS90	$NifL_{(157-278)}$ in pUT18	This Work
pPS91	$NifL_{(279-519)}$ in pUT18	This Work
pPS92	$NifL_{(147-284)}$-N177S in pUT18	This Work
pPS93	$NifL_{(1-146)}$ in pUT18	This Work
pPS94	$NifL_{(1-278)}$ in pUT18	This Work
pPS95	$NifL_{(147-284)}$-I153A in pUT18	This Work
pPS96	$NifL_{(147-284)}$-C237S in pUT18	This Work
pPS97	$NifL_{(147-284)}$-F253L in pUT18	This Work
pPS98	$NifL_{(1-146)}$ in pT25	This Work
pPS99	$NifL_{(1-278)}$ in pT25	This Work
pPS100	$NifL_{(147-284)}$-N177S in pT25	This Work
pPS101	$NifL_{(147-284)}$-L200A in pT25	This Work
pPS102	$NifL_{(147-278)}$ in pT25	This Work

pPS103	$NifL_{(147-284)}$-I153A in pT25	This Work
pPS104	$NifL_{(147-284)}$-C237S in pT25	This Work
pPS105	$NifL_{(147-284)}$-F253L in pT25	This Work
pPS106	$NifL_{(1-278)}$-L200A in pT25	This Work
pPS107	$NifL_{(1-278)}$-V157A in pT25	This Work
pPS108	$NifL_{(1-278)}$-V166M in pT25	This Work
pPS109	$NifL_{(1-278)}$-I153A in pT25	This Work
pPS110	$NifL_{(1-278)}$-F253L in pT25	This Work
pPS111	$NifL_{(1-278)}$-L200A in pUT18	This Work
pPS112	$NifL_{(1-278)}$-V157A in pUT18	This Work
pPS113	$NifL_{(1-278)}$-V166M in pUT18	This Work
pPS114	$NifL_{(1-278)}$-I153A in pUT18	This Work
pPS115	$NifL_{(1-278)}$-F253L in pUT18	This Work
pPS116	$NifL_{(1-284)}$-V166M in pETM11	This Work
pPS117	$NifL_{(1-284)}$-I153A in pETM11	This Work
pPS118	$NifL_{(143-284)}$-C181S, C237F in pETM11	This Work
pPS119	$NifL_{(143-284)}$-C181S, C237F, V157C in pETM11	This Work
pPS120	$NifL_{(143-284)}$-C181S, C237F, V166C in pETM11	This Work
pPS121	$NifL_{(143-284)}$-C181S, C237F, R240C in pETM11	This Work
pPS122	$NifL_{(143-284)}$-C181S, C237F, N177C in pETM11	This Work
pPS123	NifL-E70A, V157A in pPR34	This Work
pPS124	NifL-V166M, G480A in pPR34	This Work
pPS125	NifL-V157C in pPR34	This Work
pPS126	NifL-V166C in pPR34	This Work
pPS127	NifL-R240C in pPR34	This Work
pPS128	NifL-V157C in pRL125	This Work
pPS129	NifL-V166C in pRL125	This Work
pPS130	NifL-R240C in pRL125	This Work
pPS131	NifL-L175A in pPR34	This Work
pPS132	NifL-V251A in pPR34	This Work
pPS133	NifL-L261A in pPR34	This Work
pPS134	NifL-L262A in pPR34	This Work
pPS135	NifL-L263A in pPR34	This Work
pPS136	NifL-T264A in pPR34	This Work
pRL125	NifL-C181S, C237F, C380S, C507T in pPR34	(Little R., unpublished work)
pPS138	$NifL_{(143-284)}$-L175A in pT25	This Work
pPS139	$NifL_{(143-284)}$-L262A in pT25	This Work
pPS140	$NifL_{(143-284)}$-L175A in pUT18	This Work
pPS141	$NifL_{(143-284)}$-L262A in pUT18	This Work
pPS142	NifL- E70A, V166M in pPR34	This Work
pPS143	$NifL_{(1-284)}$-C181S, C237F, V157C in pETNdeM11	This Work
pPS144	NifL- L144P, V166M in pPR34	This Work
pPS145	NifL- H133R, V166M in pPR34	This Work

pPS146	NifL- L48P, V166M in pPR34	This Work
pPS147	NifL- Y110C, V166M in pPR34	This Work
pPS148	NifL- F54L, V166M in pPR34	This Work
pPS149	$NifL_{(1-278)}$-V119A in pUT18	This Work
pPS150	$NifL_{(1-278)}$-L130A in pUT18	This Work
pPS151	$NifL_{(1-278)}$-M132A in pUT18	This Work
pPS152	$NifL_{(1-278)}$-V119A in pT25	This Work
pPS153	$NifL_{(1-278)}$-L130A in pT25	This Work
pPS154	$NifL_{(1-278)}$-M132A in pT25	This Work
pPS155	NifL-E70A, V157C, C181S, C237F, C380S, C507T in pPR34	This Work
pPS156	$NifL_{(1-284)}$-E70A, V157C, C181S, C237F in pETNdeM11	This Work
pPS157	NifL-V157C, C181S, C237F, C380S, C507T in pETNdeM11	This Work
pPS158	$NifL_{(143-519)}$-V157C, C181S, C237F, C380S, C507T in pETNdeM11	This Work
pPS159	NifL-E70A, V157C, C181S, C237F, C380S, C507T in pETNdeM11	This Work
pPS160	NifL-E70A, V157C in pPR34	This Work
pPS161	NifL-C181S, C237F, C380S, C507T, R240C in pPR34	This Work
pPS162	$NifL_{(147-284)}$- R240W in pUT18	This Work
pPS163	$NifL_{(147-284)}$- R240W in pT25	This Work
pPS164	$NifL_{(1-284)}$- L175A in pETNdeM11	This Work
pPS165	NifL-I153A in pETNdeM11	This Work
pPS166	$NifL_{(1-145,\ 273-519)}$ (PAS2 deletion Δ146-272)	This Work
pPS167	$NifL_{(1-145,\ 276-519)}$ (PAS2 deletion Δ146-275)	This Work
pPS168	$NifL_{(1-147,\ 271-519)}$ (PAS2 deletion Δ148-270)	This Work
pPS169	$NifL_{(143-271)}$ in pETNdeM11	This Work
pPS173	NifL-L139A in pPR34	This Work
pPS174	NifL-L142A in pPR34	This Work
pPS175	NifL-V146A in pPR34	This Work
pPS180	NifL-E143A in pPR34	This Work
pPS181	NifL-R145A in pPR34	This Work
pPS182	NifL-N148A in pPR34	This Work
pPS183	NifL-Q149A in pPR34	This Work
pPS184	NifL-R150A in pPR34	This Work
pPS185	NifL-E154A in pPR34	This Work
pPS186	NifLΔL151 in pPR34	This Work
pPS187	NifLΔR150-L151 in pPR35	This Work
pPS188	NifLΔQ149-L151 in pPR34	This Work
pPS189	NifLΔN148-L151 in pPR34	This Work
pPS190	NifLΔN147-L151 in pPR34	This Work
pPS191	$NifL_{(143-519)}$-ΔN147-L151 in pPR34	This Work

Table 2.1 *E. coli* strains and plasmids used in this work.

2.3 Buffers and solutions

2.3.1 Media

Liquid media were prepared by dissolving the appropriate amount of the reagents stated below in distilled water and autoclaving at 121^{o}C and 15 psi for 15 minutes. For solid media, 1% (w/v) bactoagar was added to liquid media prior to sterilisation.

LB broth	1% (w/v) Tryptone
(pH 7.0)	0.5% (w/v) Yeast extract
	0.5% (w/v) NaCl
NFDM	2.1% (w/v) glucose
(pH 7.0)	173 μM $FeSO_4$
	875 μM $MgSO_4$
	128 μM Na_2MoO_4
Hino and Wilson buffer	1.38 M K_2HPO_4
(pH 7.0, for use with NFDM)	0.5 M KH_2PO_4
2 x YT broth	1.6% (w/v) Tryptone
(pH 7.0)	2.0% (w/v) Yeast extract
	0.5% (w/v) NaCl

2.3.2 Antibiotics

Where appropriate, antibiotics were added to the media at the following final concentrations:

Carbenicillin	100 μg ml^{-1}

Kanamycin	50 μg ml^{-1}
Chloramphenicol	35 μg ml^{-1}

2.3.3 Buffers for DNA work

TBE buffer	135 mM Tris
	45 mM Boric acid
	2.5 mM EDTA
Loading dye for electrophoresis (5x)	0.5% (w/v) Xylene cyanol ff
	0.25% (w/v) Bromophenol blue
	50% (v/v) Glycerol

2.3.4 Buffers for protein work

(i) Buffers for SDS-PAGE

Resolving buffer (4x)	1.5 M Tris-HCl (pH 8.8)
Stacking buffer (4x)	0.5 M Tris-HCl (pH 6.8)
Sample buffer (loading dye)	63 mM Tris-HCl (pH 6.8)
	2% (w/v) SDS
	10% (v/v) Glycerol
	5% (v/v) β-Mercaptoethanol
	0.001% (w/v) Bromophenol Blue
Tank buffer (running buffer)	25 mM Tris
	192 mM Glycine
	0.1% (w/v) SDS
SDS-PAGE stain	5% (v/v) Methanol

	16.5% (v/v) Acetic acid
	0.1% (w/v) Coomassie blue
SDS-PAGE destain	5% (v/v) Methanol
	16.5% (v/v) Acetic acid

(ii) Buffers for chromatography and protein storage

Nickel affinity loading buffer	25 mM KH_2PO_4 (pH 8.0)
	200 mM NaCl
	20 mM Imidazole
Nickel affinity elution buffer	25 mM KH_2PO_4 (pH 8.0)
	200 mM NaCl
	500 mM Imidazole
Analytical gel filtration buffer	50 mM Tris-HCl (pH 8.0)
	100 mM NaCl
	5% (v/v) Glycerol
Storage buffer	50 mM Tris-HCl (pH 8.0)
	50 mM NaCl
	1 mM DTT
	50% (v/v) glycerol

(iii) Buffers for western blotting

Transfer buffer	25 mM Tris (pH 8.0)
	190 mM Glycine
	20% (v/v) Methanol

TBS buffer	10 mM Tris-HCl (pH 7.5)
	100 mM NaCl
Blocking buffer	3% (w/v) BSA in TBS
TBS-Tween/Triton buffer	20 mM Tris-HCl (pH 7.5)
	500 mM NaCl
	0.05% (v/v) Tween
	0.2% (v/v) Triton X-100

(iv) Buffers for limited proteolysis experiments

TA buffer (5x)	250 mM Tris-acetate (p.H 7.9)
	500 mM Potassium acetate
	40 mM Magnesium acetate
	5 mM DTT

(v) Buffers for β-galactosidase Assays

Z-Buffer	60 mM Na_2HPO_4
	40 mM $NaH_2PO_4.2H_2O$
	10 mM KCl
	1 mM $MgSO_4.7H_2O$
Lysis Buffer	0.27% (v/v) β-Mercaptoethanol
	0.005% (w/v) SDS
	In Z-Buffer
Start buffer	13.3 mM 2-Nitrophenyl-β-galactoside
	In Z-buffer

Stop buffer 1 M Na_2CO_3

2.4 Microbiological methods

2.4.1 Preparation of chemically competent *E. coli*

A culture of the appropriate *E. coli* strain was grown overnight in a universal tube containing 5 ml LB at 37^{o}C. A 250 ml conical flask containing 50 ml LB was then inoculated with 500 µl of the overnight culture and grown at 37^{o}C until the optical density at 600 nm (OD_{600}) reached 0.3 - 0.4. Cells were then harvested by centrifugation at 3500 rpm (in a Sorvall Biofuge primo centrifuge) for 7 minutes at 4 ^{o}C. The cell pellet was gently resuspended in 12.5 ml of ice cold 0.1 M $MgCl_2$. The centrifugation step was then repeated and the resulting pellet was gently resuspended in 25 ml of ice cold 0.1 M $CaCl_2$. This cell suspension was incubated on ice for 20 minutes. Cells were then harvested by a final centrifugation at 3500 rpm for 7 minutes at 4 ^{o}C. The pellet was resuspended in 1 ml 0.1 M $CaCl_2$ and 20% (v/v) glycerol. Competent cells were stored in 150 µl aliquots at -80^{o}C until required.

2.4.2 Transformation of competent *E. coli*

Plasmid DNA (100 ng - 300 ng) was added to a 1.5 ml eppendorf tube containing 50 µl of chemically competent cells and incubated on ice for 45 minutes. The cells were then heat shocked at 42^{o}C for 90 seconds and subsequently incubated on ice for 1 minute. Next, 950 µl of 2 x YT broth was added and the cells were placed at 37^{o}C for 1 hour. 50 µl and 100 µl aliquots of the transformed cells were then spread onto agar plates containing the appropriate media and antibiotics. The agar plates were incubated at 37^{o}C overnight (except for the indicator plates used for

screening randomly generated mutants in strain ET8000, which were incubated at 30^{o}C for 72 hours).

2.4.3 Electroporation of *E. coli*

A 500 ml conical flask containing 250 ml LB was inoculated with 2.5 ml of an overnight culture (grown in a universal tube containing 5 ml LB at 37^{o}C) of *E. coli* strain DH5α and grown at 37^{o}C until the OD_{600} reached 0.5 - 0.7. The cells were then incubated on ice for 15 minutes. Next, the cells were harvested by centrifugation at 4000 rpm (in a Sorvall RC 5B plus centrifuge fitted with a SLA-3000 rotor) for 10 minutes at 4^{o}C. The supernatant was carefully discarded and the pellet was resuspended in 400 ml cold sterile water. The suspension was centrifuged again under the same conditions and this cycle of resuspension in water and centrifugation was repeated 3 times. Finally, the cells were resuspended in 500 µl cold sterile water. 3 µl (~75 ng) of butanol precipitated DNA (see below) and 200 µl of the cell suspension were mixed in a 2 mm electroporation cuvette (Geneflow Ltd.) and electroporated at 2.5 Kv, 400 Ω and 25 µF. 1 ml LB was then added and the cells were incubated at 37^{o}C for 1 hour. Transformed cells were then grown in selective media as appropriate.

2.5 DNA purification and manipulation methods

2.5.1 Purification of plasmid DNA

Preparations of plasmid DNA were carried out from 5 ml overnight cultures using the QIAprep spin miniprep kit (Qiagen) as directed by the manufacturer.

2.5.2 Butanol precipitation of DNA

DNA samples (50 – 100 μL) were thoroughly mixed with 1.2 ml butan-1-ol, and incubated at room temperature for 10 minutes. Samples were then centrifuged at 13000 rpm (in a Thermo Scientific Heraeus Pico bench-top centrifuge) for 4 minutes and the supernatant was discarded. Next, 1 ml 100% Ethanol (-20^{o}C) was added. The sample was then briefly mixed using a vortex and incubated at -20^{o}C for 20 minutes. Following this incubation, the mixture was centrifuged for 5 minutes at 13000 rpm and the supernatant was carefully discarded. 1 ml 70% Ethanol (-20^{o}C) was added to the pellet and the sample was vortexed briefly and incubated at -20^{o}C for 10 minutes. The centrifugation step was repeated once more and the pellet was dried in a Speedivac rotary evaporator for 10 minutes. Finally, the pellet was resuspended in 5 μl sterile water.

2.5.3 DNA sequencing

Dye-terminator DNA sequencing was used to ensure that mutant *nifL* genes carried no additional mutations and to identify the sequence changes that emerged after random mutagenesis (see below). Sequencing reactions were carried out using BigDye Terminator 3.1 (Applied Biosciences) as instructed by the manufacturer. Completed reactions were submitted to Genome Enterprise Ltd. for capillary electrophoresis and florescence detection.

2.5.4 Restriction endonuclease digestion

DNA samples were incubated for 3 hours with 1-10 U of the relevant restriction enzyme/s per μg of DNA. Reactions were performed in 1x buffer provided by the manufacturer at the appropriate temperature for each enzyme.

2.5.5 Agarose gel electrophoresis

Agarose gel electrophoresis was used for size determination and purification of DNA fragments. TBE buffer containing 1% (w/v) agarose was melted using a microwave and allowed to set in a gel mould from VWR International. DNA samples were mixed with 5x loading dye and gels were run in TBE buffer at 80 V - 100 V for between 45 minutes and 1.5 hours, depending on DNA size. Ethidium bromide was then added to the electrophoresis buffer to a final concentration of 5 μg ml^{-1} and the gel was stained for 15 minutes before visualisation on a short wavelength UV transilluminator. A longwave UV transilluminator was used to visualise fragments for gel excision.

2.5.5 Purification of DNA fragments

Gel slices containing the appropriate DNA fragment were excised after electrophoresis and the DNA was extracted from the gel slice using the QIAquick gel extraction kit (Qiagen) as instructed by the manufacturer. DNA fragments from PCR were purified using the QIAquick PCR purification kit (Qiagen) as directed by the manufacturer.

2.5.7 Dephosphorylation of DNA

Shrimp alkaline phosphatase (SAP) is routinely used to catalyse the removal of the 5' phosphate from DNA molecules to prevent re-circularisation of vector DNA in ligation reactions. DNA samples were incubated in 1x SAP reaction buffer provided by the enzyme manufacturer (Roche) with 2 U SAP per μg of DNA for 45 minutes at 37^{o}C. The enzyme was then inactivated by heating to 75^{o}C for 15 minutes.

2.5.8 Ligation of DNA

T4 DNA ligase is routinely used to join DNA molecules via formation of a phosphodiester bond. A 1:5 ratio of vector to insert was ligated using T4 DNA ligase (NEB) in the buffer supplied by the manufacturer. Ligation reactions were carried out overnight at 16^{o}C.

2.5.9 Site directed mutagenesis

All site directed mutations were constructed using a two-step PCR technique (Figure 2.1). This method generates point mutations in a cloned DNA fragment (carrying unique restriction sites for further cloning) and requires two pairs of primers. The first set of primers anneals either side of the region of interest. These are called external primers (Figure 2.1, red arrows). The second pair of primers are complementary to each other and carry the required mutation/s. These anneal within the region of interest and are known as mutagenic (or internal) primers (Figure 2.1, blue arrows). This method requires three separate PCR reactions in two stages (called step 1 and step 2). The first step involves two PCR reactions, each using one external

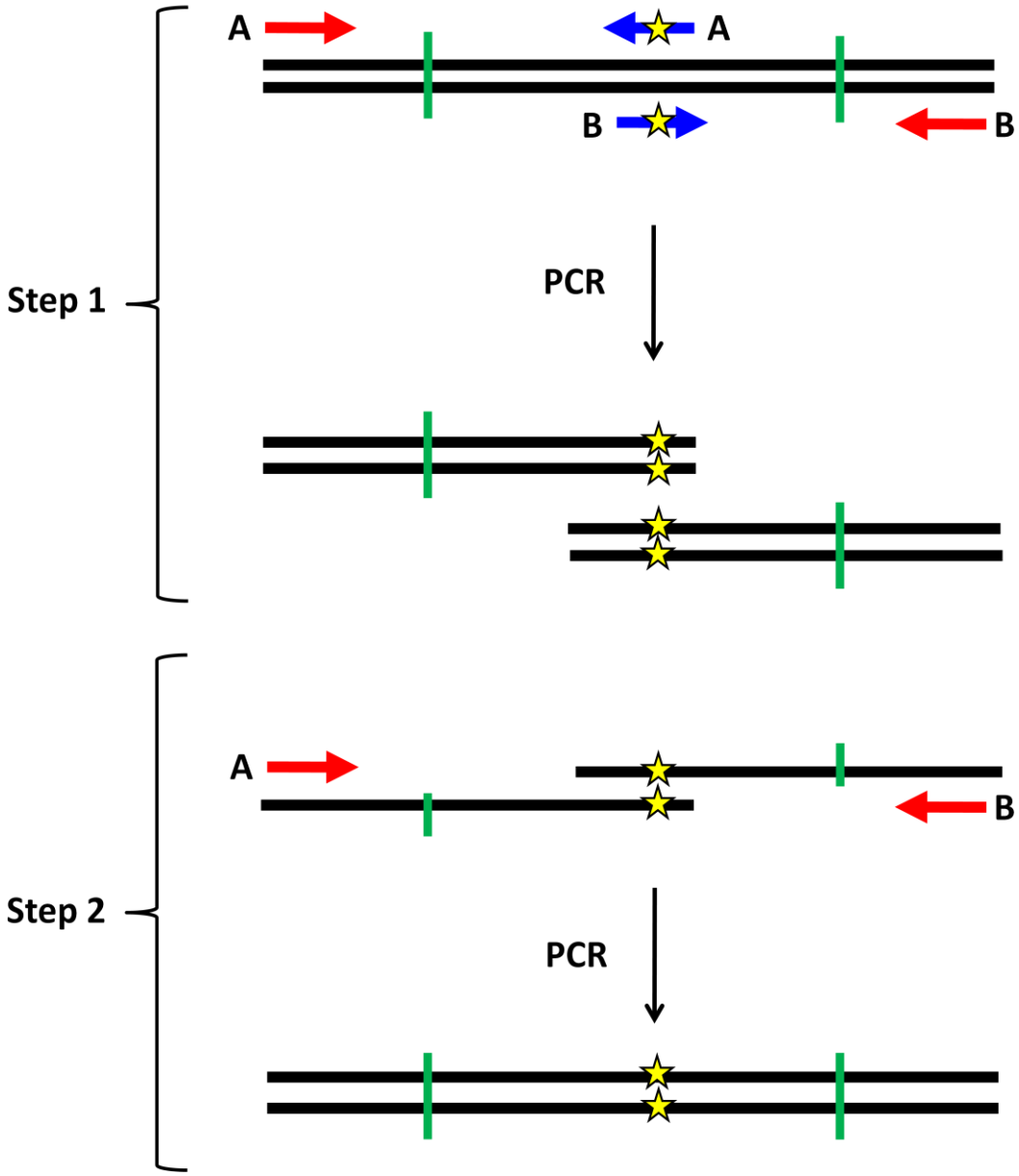


Figure 2.1. Two-step PCR method for site-directed mutagenesis. Individual DNA strands are represented as black lines and (perpendicular) green lines indicate unique restriction sites. External primers are coloured red whilst mutagenic primers are blue. The yellow stars denote point mutations. Step 1 consists of two separate PCR reactions using primer set A or B whereas step 2 consists of a single amplification using the external (red) primers and the purified PCR products from step 1.

primer and one mutagenic primer, to introduce the desired mutation into the region of interest (Figure 2.1, step 1, PCR reactions are carried out using primer pairs A and B). The second step involves combining 50 ng - 100 ng of the purified PCR products from step 1 and using this mixture as the template for another PCR amplification using the external primers, thus reassembling the mutated region of interest (Figure 2.1, step 2). To minimise the risk of unwanted mutations, high fidelity Accuzyme mix (Bioline) was used for all PCR reactions as directed by the manufacturer. All primers used for mutagenesis are listed in Table 2.2.

Primer Name	**Sequence (5' - 3')**
Nic1	GATCATGCTGCGCAGGCTGCTTTC
L1x	CCGCCGCAAGGACAAGACC
L199Ea	GGTGGCGGAGCTGCGGGAAAAC
L199Eb	GTTTTCCCGCAGCTCCGCCACC
L199Ra	GGTGGCGCGCCTGCGGGAAAAC
L199Rb	GTTTTCCCGCAGGCGCGCCACC
L199Ga	GGTGGCGGGCCTGCGGGAAAAC
L199Gb	GTTTTCCCGCAGGCCCGCCACC
L199Qa	GGTGGCGCAGCTGCGGGAAAAC
L199Qb	GTTTTCCCGCAGCTGCGCCACC
L199Aa	GGTGGCGGCGCTGCGCGAAAAC
L199Ab	GTTTTCCCGCAGCGCCGCCACC
L199Va	GGTGGCGGTGCTGCGCGAAAAC
L199Vb	GTTTTCCCGCAGCACCGCCACC
L199Wa	GGTGGCGTGGCTGCGCGAAAAC
L199Wb	GTTTTCCCGCAGCCACGCCACC
L199Fa	GGTGGCGTTCCTGCGGGAAAAC
L199Fb	GTTTTCCCGCAGGAACGCCACC
V165Da	GCGATGGACGTGCTCGAC
V165Db	GTCGAGCACGTCCATCGC
L200Fa	TGGCGCTGTTCCGGGAAAAC
L200Fb	GTTTTCCCGGAACAGCGCCA
E194Ka	CAGCAGCAAGAGCCTGGTGG
E194Kb	CCACCAGGCTCTTGCTGCTG
L200Ea	TGGCGCTGGAACGGGAAAAC
L200Eb	GTTTTCCCGTTCCAGCGCCA
E202Ka	CGCTGCTGCGGAAGAACCTC
E202Kb	GAGGTTCTTCCGCAGCAGCG
E202Aa	CGCTGCTGCGGGCGAACCTC

E202Ab	GAGGTTCGCCCGCAGCAGCG
L200Aa	TGGCGCTGGCGCGGGAAAAC
L200Ab	GTTTTCCCGCGCCAGCGCC
R201Aa	TGGCGCTGCTGGCGGAAAAC
R201Ab	GTTTTCCGCCAGCAGCGCC
L196Aa	AGCGAGAGCGCCGTGGCGC
L196Ab	AGCGCCACGGCGCTCTCGC
L151Aa	CAGCGCGCGATGATCGAG
L151Ab	CTCGATCATCGCGCGCTG
I153Aa	CTGATGGCCGAGGCGGTG
I153Ab	CACCGCCTCGGCCATCAG
V157Aa	GAGGCGGTGGCCAACGCC
V157Ab	GGCGTTGGCCACCGCCTC
S192Aa	GATGGCGGCAGCGAGAGCCTG
S192Ab	CAGGCTCTCGCTGCCGCCATC
S193Aa	GATGGCAGCGGCGAGAGCCTG
S193Ab	CAGGCTCTCGCCGCTGCCATC
S195Aa	CAGCAGCGAGGGCCTGGTG
S195Ab	CACCAGGCCCTCGCTGCTG
V166Aa	GCGATGGTGGCGCTCGAC
V166Ab	GTCGAGCGCCACCATCGC
V166Da	GCGATGGTGGACCTCGAC
V166Db	GTCGAGGTCCACCATCGC
V166Ca	GCGATGGTGTGCCTCGAC
V166Cb	GTCGAGGCACACCATCGC
V157Ca	GAGGCGGTGTGCAACGCC
V157Cb	GGCGTTGCACACCGCCTC
R240Ca	CACGGCTGCGCCATCCAC
R240Cb	GTGGATGGCGCAGCCGTG
L175Ab	GGGTTGGAGGCCATCACC
V251Aa	GGCCCACGCGTTCTTCGC
V251Ab	GCGAAGAACGCGTGGGCC
L261Aa	ACGCTACGCGCTGCTGAC
L261Ab	GTCAGCAGCGCGTAGCGT
L262Aa	GCTACCTGGCGCTGACCA
L262Ab	TGGTCAGCGCCAGGTAGC
L263Aa	CCTGCTGGCGACCATCAA
L263Ab	TTGATGGTCGCCAGCAGG
T264Aa	GCTGCTGGCCATCAACGA
T264Ab	TCGTTGATGGCCAGCAGC
L139Aa	CAGCGAAGCGCACGAACT
L139Ab	AGTTCGTGCGCTTCGCTG
L142Aa	GCACGAAGCGGAACAACG
L142Ab	CGTTGTTCCGCTTCGTGC
V146Aa	GAACAACGCGCCAACAACC
V146Ab	GGTTGTTGGCGCGTTGTTC

E143Aa	CACGAACTGGCGCAACGCGTC
E143Ab	GACGCGTTGCGCCAGTTCGTG
R145Aa	CTGGAACAAGCCGTCAACAACC
R145Ab	GGTTGTTGACGGCTTGTTCCAG
Q149Aa	GTCAACAACGCGCGCCTGATG
Q149Ab	CATCAGGCGCGCGTTGTTGAC
R150Aa	CAACAACCAGGCCCTGATGATC
R150Ab	GATCATCAGGGCCTGGTTGTTG
E154Aa	CTGATGATCGCGGCGGTGGTC
E154Ab	GACCACCGCCGCGATCATCAG
N148Aa	CGCGTCAACGCCCAGCGCCTG
N148Ab	CAGGCGCTGGGCGTTGACGCG

Table 2.2. Primers for mutagenesis.

2.5.10 Random mutagenesis of the PAS2 domain

PCR mutagenesis was carried out with *Taq* DNA polymerase (Roche) under standard reaction conditions. Reaction mixtures contained 75 ng of template pPR34, 100 ng of each primer (primers L1x and Nic1 were used, Table 2.2), 0.2 mM dNTPs, 1.5 mM $MgCl_2$ and 5 units of enzyme in a final volume of 50 μl. PCR products were purified (Qiagen kit) and cut with restriction endonucleases *Mlu* I and *Apa* I (Invitrogen) and subsequently recloned (after gel purification using a Qiagen kit) into pPR34 vector which had been cut with the same enzymes. Ligation mixtures were then butanol precipitated and electroporated into *E. coli* strain DH5α (see sections 2.5.2 and 2.4.3). The electroporation procedure described in section 2.4.3 yielded a 1.2 ml culture of transformed cells (in LB broth) containing the mutagenised pPR34 plasmid. Next, 3.8 ml of LB broth supplemented with carbenicillin (100 μg ml^{-1}) was added in order to facilitate the growth of a 5 ml overnight culture and subsequent recovery of the plasmid DNA using a QIAprep spin miniprep kit (Qiagen). The resultant plasmids (containing the PCR-generated insert) were transformed into *E. coli* strain ET8000 which contains the reporter plasmid pRT22 (carrying a *nifH-lacZ* fusion). Transformants were screened on solid NFDM medium supplemented with

Hino and Wilson buffer (5% v/v), casein hydroxylate (200 μg ml^{-1}), X-gal (5-bromo-4chloro-3-indolyl-B-D-galactopyranoside, 40 μg ml^{-1}), chloramphenicol (35 μg ml^{-1}) and carbenicillin (100 μg ml^{-1}). Mutations in *nifL* that resulted in altered NifA activity were selected (see Chapter 3.2) and their plasmid DNA recovered and sequenced to identify mutations. Plasmids of known sequence were then transformed back into the host strain for further phenotypic analysis (see section 2.9.2).

2.6 Construction of plasmids

2.6.1 Plasmids for analysis of NifL activity *in vivo*

All plasmids used to investigate NifL activity *in vivo* were derived from pPR34. The pPR34 plasmid is a pT7-7 derivative carrying transcriptionally coupled (and independently translated) copies of the *A. vinelandii nifL* and *nifA* genes under the control of a constitutive promoter (Söderbäck et al., 1998). Plasmids encoding truncated forms the NifL protein starting at residue 143 (pPS54, pPS77 and pPS191) were generated by PCR amplification using pPR34 (or a mutant derivative) as template DNA. The forward primer pPS54a was used to introduce an *Nde* I site immediately prior to the codon for NifL residue 143. The reverse primer was MS2rev, which anneals downstream of a unique *Not* I restriction site in *nifL* (primer sequences are shown in Table 2.3). PCR products were purified, digested with *Not* I and *Nde* I, and ligated into pPR34 vector which had been cut with the same enzymes and treated with SAP (section 2.5.7). Ligation mixtures were used to transform *E. coli* strain DH5α. Transformed cells were spread onto LB agar plates supplemented with carbenicillin (plasmid pPR34 carries a carbenicillin resistance cassette). This selection method was used to obtain all the pPR34 mutant derivatives created in this work. Plasmids encoding alternative truncations of the NifL protein starting at

residue 147 (pPS3 and pPS4) were created by site directed mutagenesis (described in section 2.5.9) of plasmid pPR54 (a pPR34 derivative which encodes $NifL_{(147-519)}$) (Söderbäck et al., 1998).

To create variant forms of the NifL protein containing two amino acid substitutions, plasmids carrying the mutations of interest were digested at unique restriction sites and the resulting DNA fragments were purified and appropriate combinations were ligated to yield the required double mutant. In detail, plasmid pPS124 (encoding the NifL-V166M, G480A double substitution) was constructed by digesting pPS20 (encoding the NifL-V166M variant) with the restriction endonucleases *Nde* I and *Not* I, gel purifying the resultant *nifL* fragment (Qiagen kit) and cloning this into plasmid pNLG480A (encoding the NifL-G480A variant, Perry et al., 2005) which had been digested with the same enzymes and gel purified.

Deletion mutants in *nifL* (encoding NifL variants lacking the PAS2 domain or residues in the inter-domain region between PAS1 and PAS2) were cloned using a two-step PCR technique similar to that described in section 2.5.9. As previously, two pairs of primers, including external primers that flank the region of interest, were used to generate the required DNA fragments. However, the mutagenic (or internal) primers differed from those used to generate point mutations. The forward primers consisted of two sequence elements; they contained a 3' annealing region that primed the PCR reaction and a 5' non-annealing tail (Figure 2.2). This 5' tail contained the appropriate deletion and was complementary to the reverse (internal) primer. Apart from these differences, the PCR mutagenesis was carried out as described in section 2.5.9. All constructs were confirmed by DNA sequencing.

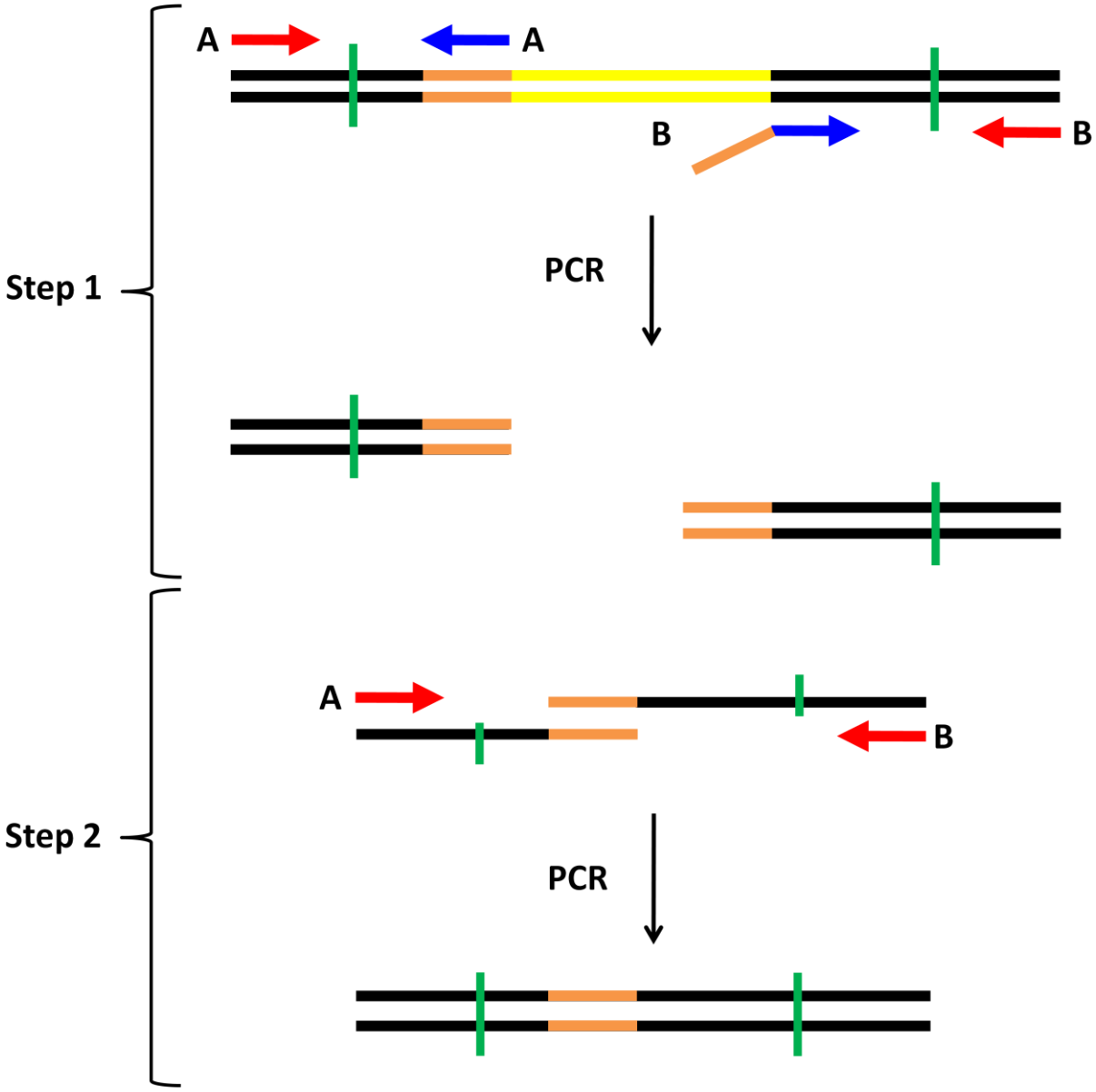


Figure 2.2. Two-step PCR method for deletion mutagenesis. Individual DNA strands are represented as black lines and (perpendicular) green lines indicate unique restriction sites. The region of the DNA sequence to be deleted is coloured yellow and the orange sequence elements are complementary to one another. External primers are coloured red whilst internal primers are blue. One of the internal primers contains a 5' "non-annealing" sequence that is complementary to a region upstream of the DNA sequence to be deleted. Step 1 consists of two separate PCR reactions using primer set A or B whereas step 2 consists of a single amplification using the external (red) primers and the purified PCR products from step 1.

Primer Name	**Sequence (5' - 3')**
P145r	GCGTTGTTCCAGTTCGTG
P147r	GTTGACGCGTTGTTCCAG
145tail-273fw	CACGAACTGGAACAACGCCAGAAGCAGCAGGATTCG CGGCTCA
145tail-276fw	CACGAACTGGAACAACGCCAGGATTCGCGGCTCAACG CGCTGA
147tail-271fw	CTGGAACAACGCGTCAACCTGCGCCAGAAGCAGCAG GATTCGC
Nic1	GATCATGCTGCGCAGGCTGCTTTC
pPS54a	GCGAATTGCACCATATGGAACAACGC
Pa1	CTAGAGAATTCGGATAGACGAGGCACC
D1f	CGCGTCAACAACCAGCGCATGATCGAGGCGGTGGTCA ACGCC G
D2f	CAACGCGTCAACAACCAGATGATCGAGGCGGTGGTCA ACGCCG
D3f	GAACAACGCGTCAACAACATGATCGAGGCGGTGGTCA ACGCCG
D4f	CTGGAACAACGCGTCAACATGATCGAGGCGGTGGTCA ACGCCG
D5f	GAACTGGAACAACGCGTCATGATCGAGGCGGTGGTCA ACGCCG
D6f	CACGAACTGGAACAACGCATGATCGAGGCGGTGGTCA ACGCCG
D1r	GCGCTGGTTGTTGACGCG
D2r	CTGGTTGTTGACGCGTTGTTC
D3r	GTTGTTGACGCGTTGTTCCAGTTC
D4r	GTTGACGCGTTGTTCCAGTTCG
D5r	GACGCGTTGTTCCAGTTCGTG
D6r	GCGTTGTTCCAGTTCGTGCAATTC

Table 2.3. Primers used for construction of pPR34 derivative plasmids

2.6.2 Plasmids for bacterial adenylate cyclase two-hybrid (BACTH) analyses

DNA fragments encoding the protein of interest flanked by a 5' *Bam*H I site and a 3' *Kpn* I site were generated by PCR. Primers NifL-BTH-F and PAS1-BTH-1F were used to amplify the *nifL* gene from sequence encoding NifL residues 1 and 147 respectively and to introduce a 5' *Bam*H I site. The reverse primers PAS2-BTH-2R and PAS1-BTH-1R were used to amplify to residues 284 or 146 of NifL respectively (the sequence of each primer is shown in Table 2.4). Plasmid pPR34 or derivative

plasmids containing the desired mutation/s were used as template DNA for PCR reactions using appropriate combinations of the primers mentioned above. PCR products were purified (Qiagen kit), digested with *Kpn* I and *Bam*H I, and cloned into the BACTH vectors pT25 and pUT18 (see section 2.9.3) which had been cut with the same enzymes and dephosphorylated. All constructs were confirmed by DNA sequencing.

Primer Name	**Sequence (5' - 3')**
NifL-BTH-F	GAGGATCCCATGACCCCGGCCAACCCGAC
PAS1-BTH-2R	CTTAGGTACCCGGTGCAATTCGCTGGT
PAS1-BTH-1R	CTTAGGTACCACGCGTTGTTCCAG
PAS2-BTH-1F	GAGGATCCCAACAACCAGCGCCTG
PAS2-BTH-2R	CTTAGGTACCTTCAGCGCGTTGAG

Table 2.4. Primers used to clone plasmids for bacterial two-hybrid work

2.6.3 Plasmids for protein overexpression

All plasmids for protein overexpression were derived from pETM11 (EMBL). DNA fragments encoding the required region of NifL were PCR amplified using primers with appropriate restriction sites engineered at their 5' ends (primer sequences are shown in Table 2.5), enabling directional cloning into the overexpression vector. For overexpression of $N_{his6}NifL_{(143-284)}$, $N_{his6}NifL_{(1-284)}$ and mutant derivatives, DNA fragments encoding the appropriate NifL residues flanked by a 5' *Nco* I site and a 3' *Bam*H I site (preceded by a stop codon) were generated by PCR using the forward primers Nco1-E143-NifL (for constructs starting at residue 143) or NifLNcoFor (for constructs starting at residue 1) and the reverse primer NifL-284-TGA-Bam. Plasmid pPR34 or derivatives from mutagenesis (see above) were used as template DNA. The *Nco* I-*Bam*H I digested fragments were then cloned into the plasmid pETM11, which had been cut with the same enzymes and dephosphorylated. For overexpression of $N_{his6}NifL_{(143-519)}$, $N_{his6}NifL$ and mutant

derivatives, fragments encoding the appropriate NifL residues flanked by a 5' *Nde* I site and a 3' *Bam*H I site were PCR amplified from pPR34 (for N_{his6}NifL), pPS54 (for $N_{his6}NifL_{(143-519)}$) or mutant derivatives. The forward primer Pa1 and the reverse primer NifL2 were used in the PCR reaction and the products were then digested with the restriction endonucleases *Nde* I and *Bam*H I. The resultant DNA fragment was cloned into a plasmid derived from pETM11 (called pETNdeM11) in which the *Nco* I site in the multiple cloning region was mutated to yield a *Nde* I site via the two-step PCR mutagenesis technique described in section 2.5.9. All constructs were confirmed by DNA sequencing.

Primer Name	**Sequence (5' - 3')**
Nco I-E143-NifL	CAAGCGCCATGGAACAACGCGTCAACAACC
NifL-284-TGA-Bam	GCAAGGATCCTCACTTCAGCGCGTTGAGCCG
Pa1	CTAGAGAATTCGGATAGACGAGGCACC
NifLNcoFor	CGCATCCATGGCCACCCCGGCCAACCCGACCCT
NifL2	CGAAGGATCCTCAGGTGGAGGCCGAGAAGGG

Table 2.5. Primers used to clone of plasmids for protein overexpression

2.7 Protein methods

2.7.1 SDS Polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE is routinely used to separate proteins on the basis on molecular weight. 12.5% polyacrylamide gels were used in this work unless otherwise stated. The resolving gel was prepared by mixing 7.5 ml of 4x resolving buffer with 9.5 ml of distilled water, 12.5 ml of 30% acrylamide (Severn Biotech), 0.3 ml of 10% (w/v) SDS solution and 150 µl of 10% (w/v) ammonium persulphate solution. 25 µl TEMED (N,N,N',N'-Tetramethylethylenediamine) was added to initiate acrylamide polymerisation and the mixture was immediately poured into the assembled gel mould (Atto corp.). For the stacking gel, 6.1 ml distilled water was mixed with 2.5

ml of 4x stacking buffer, 1.33 ml of 30% acrylamide, 0.1 ml of 10% SDS and 100 µl of 10% ammonium persulphate. 15 µl TEMED was then added and the solution was mixed and poured into the gel mould. The gel comb was inserted into the gel mould to create the wells and the gel was allowed to set. Samples for SDS-PAGE were prepared by mixing with an equal volume of SDS sample buffer and boiling for 3 minutes at 100^{o}C. Electrophoresis was carried out at approximately 170 V (constant voltage) for 45 - 60 minutes. Gels were stained in 20 ml SDS-PAGE stain for 15 minutes and subsequently destained in 40 ml SDS-PAGE destain overnight. Alternatively, protein bands were visualised by staining in 15 ml InstantBlue (Expedeon Ltd.) for 2 hours. Where pre-cast gels were used, 12% RunBlue pre-cast gels were purchased from Expedeon Ltd. and SDS-PAGE was carried out as directed by the manufacturer.

2.7.2 Overexpression of proteins for purification

All proteins were expressed from pETM11 derived plasmids in *E. coli* strain BL21 (DE3) pLysS. Chemically competent cells were transformed with the appropriate plasmid and plated onto solid LB supplemented with kanamycin and chloramphenicol (pETM11 and pLysS carry kanamycin and chloramphenicol resistance markers respectively). A single colony was used to inoculate 5 ml liquid LB medium with the same antibiotics and the culture was incubated for approximately 8 hours at 37^{o}C (with shaking). A 250 ml conical flask containing 50 ml LB and appropriate antibiotics was then inoculated with 500 µl of this culture and grown at 30^{o}C overnight. Two 12.5 ml aliquots of this overnight culture were used to inoculate two 2 L conical flasks containing 1 L LB supplemented with kanamycin and chloramphenicol and grown at 30^{o}C (with shaking) until OD_{600} reached 0.6.

Isopropyl-β-*D*-thiogalactopyranoside (IPTG) was then added to a final concentration of 1 mM to induce protein expression and cultures were incubated under agitation for a further 2 hours. Cells were then harvested by centrifugation for 10 minutes at 5000 rpm (in a Sorvall RC 5B plus centrifuge fitted with a SLA-3000 rotor). Cell pellets were resuspended in 30 ml nickel affinity loading buffer and stored at -20^{o}C until required.

2.7.3 Protein purification

Cells from overexpression (see above) were thawed mixed with 0.3 ml protease inhibitor cocktail set II (Calbiochem) and disrupted at 10000 psi in a French pressure cell. Broken cells were then centrifuged at 18000 rpm (Sorvall RC 5B plus centrifuge with SS-34 rotor) for 30 minutes at 4^{o}C to remove the cell debris. The clarified cell extract was loaded onto a HiTrap chelating column (GE healthcare) which had been primed with nickel chloride and equilibrated with 6 column volumes of nickel affinity loading buffer. The flow rate used in all purification steps was constant at 1.5 ml min^{-1}. The loaded column was washed with approximately 20 ml nickel affinity loading buffer to remove non-specifically bound protein. An imidazole gradient of 20 mM - 500 mM was then applied to elute the hexahis-tagged protein. Recombinant NifL proteins typically eluted in 20% - 40% nickel affinity elution buffer (~100 mM - 200 mM imidazole). The protein content of fractions recovered from affinity chromatography was analysed by SDS-PAGE. Fractions containing the protein of interest at high concentration and purity were pooled and dialysed into storage buffer. After dialysis, protein samples were stored at -20^{o}C until required.

2.7.4 Bradford assay for protein concentration

Bradford assays are widely used for determination of protein concentration (Bradford, 1976). Coomassie Plus™ protein assay reagent and BSA standard were purchased from Thermo Scientific and the assay was carried out according to the manufacturer's instructions.

2.7.5 Protein buffer exchange

When the buffer used for storage of protein samples was not appropriate for a particular experiment, an aliquot was taken from the stored sample and exchanged into a germane buffer using a Zeba™ desalt spin column (Thermo scientific) as directed by the manufacturer.

2.7.6 Size exclusion chromatography (SEC)

Size exclusion chromatography is a technique commonly used to estimate the molecular mass of proteins in solution. Purified protein samples were run at 0.4 ml min^{-1} over a Superose 12 10/300 GL column (GE healthcare) equilibrated with at least six column volumes of analytical gel filtration buffer. Proteins were injected at a concentration of 104 μM (based on a monomer) unless otherwise stated. Bio-Rad gel filtration standards (thyroglobulin (bovine), γ-globulin (bovine), ovalbumin (chicken), myoglobin (horse) and vitamin B12) were used for calibration as directed by the manufacturer.

2.7.7 Dynamic light scattering (DLS)

Dynamic light scattering is commonly used to analyse the size distribution of proteins in solution. Purified protein samples were buffer exchanged (see section 2.7.5) into 50 mM Tris-HCl (pH 8.0), 100 mM NaCl. Samples were then centrifuged (13000 rpm for 30 seconds in a Thermo Scientific Heraeus Pico microfuge) through a 0.1 μm Ultrafree filter (Millipore) to remove particulate material and a 13 μl aliquot was pipetted into a microsampling cell (Wyatt technologies). Measurements were taken at 293 K using a Dynapro Titan DLS instrument (Protein Solutions Inc.). At least 15 scattering measurements were taken for each sample and the resulting data were analysed using the DYNAMICS 6.9.2.11 software package (Protein Solutions Inc.).

2.7.8 Chemical cross-linking

Chemical cross-linking can be used to study subunit stoichiometry in multimeric proteins. Samples of purified protein (52 μM based on a monomer) were incubated in 50 mM Tris-HCl (pH 8.0), 50 mM NaCl and 0.25% glutaraldehyde for 10 mins at 30^{o}C. The reaction volume was 10 μL. After 10 minutes, 10 μL of SDS-PAGE sample buffer was added and samples were immediately heated to 100 ^{o}C for 4 mins and analysed by SDS-PAGE. Densitometric analysis was performed using SynGene GeneTools software (version 3.06.04) from Synoptics Ltd.

2.7.9 Cysteine cross-linking

Cysteine cross-linking is a technique routinely used to analyse the tertiary and quaternary structures of proteins. Pairs of native or substituted cysteine residues

located in close proximity to one another in a folded protein structure can be oxidised to form disulphide bridges using catalysts such as copper (II) o-phenanthroline (Cu-Phe) or iodine solution. This can provide information regarding the relative positions of pairs of amino acid residues in a three-dimensional protein structure. For cysteine cross-linking experiments on the isolated PAS2 domain of NifL (and its variants), protein samples (52 μM, based on a monomer) were incubated in 17 mM Tris-HCl (pH 8.0), 17 mM NaCl, 17% (v/v) glycerol and 5 μM Cu-Phe for 10 minutes at 37^{o}C. Reactions were stopped by addition *N*-ethylmaleimide (NEM) to a final concentration of 66 mM. NEM irreversibly alkylates thiol groups and thereby prevents further disulphide bond formation. This step prevents non-specific disulphide bridge formation in the denatured protein samples used for SDS-PAGE analysis. Two aliquots were removed from each "stopped" reaction and added to either non-reducing SDS-PAGE loading buffer (Expedeon) or loading buffer containing 25% (v/v) β-mercaptoethanol. Samples were then analysed by SDS-PAGE on pre-cast (12%) polyacrylamide gels (Expedeon) as directed by the manufacturer.

For cysteine cross-linking experiments on the full length NifL protein and its variants, protein samples (8.2 μM based on a monomer) were incubated in 17 mM Tris-HCl (pH 8.0), 17 mM NaCl, 17% (v/v) glycerol and Cu-Phe catalyst at the final concentrations indicated in Figure 5.6 (i.e. 0 μM, 2.5 μM or 5 μM) for 10 minutes at 37^{o}C. Reactions were stopped by addition of NEM to a final concentration of 50 mM. Additional control experiments in which NEM was added prior to catalysis were carried out to ensure the NEM concentration used was sufficient to fully prevent non-specific disulphide bond formation. Samples were analysed by SDS-PAGE as described above.

2.7.10 Analytical ultracentrifugation (AUC)

AUC is commonly used to analyse the molecular mass of macromolecules in solution. Sedimentation equilibrium experiments were performed in a Beckman Optima XL-I analytical ultracentrifuge equipped with absorbance optics and an An50Ti rotor. Purified protein samples were diluted to a concentration of 100 μM and then buffer exchanged (see section 2.7.5) into 50 mM KH_2PO_4 (pH 8.0), 100 mM NaCl. A series of dilutions were prepared for each protein (10-fold, 20-fold and 100-fold dilutions were prepared). To ensure that the freshest possible protein samples were analyzed in the AUC, equilibrium ultracentrifugation experiments were performed immediately after the sample preparation; 110 μL of each sample was loaded into the sample sector of charcoal-filled Epon double sector cells fitted with quartz windows, while 120 μL of buffer was loaded into the reference sector. Samples were centrifuged at speeds of 16,000 and 23,000 rpm and the absorbance was recorded at 275 nm for the higher concentrations, and 230 nm for the lower protein concentrations. The precise concentration of each sample was calculated retrospectively using absorbance measured by the AUC and an experimentally (and independently) determined absorbance co-efficient. Data analysis was performed using Ultrascan II (Demeler, 2005) where profiles of individual samples were initially analysed at single speeds using an ideal, single component model. The parameters for buffer density and partial specific volume were determined using SEDNTERP (Horan et al., 1995).

2.7.11 Spectroscopic analysis of the FAD content of NifL

The FAD absorbance spectrum has characteristic peaks at 450 nm and 378 nm. Incorporation of the FAD molecule into a protein results in alteration of the

absorbance spectrum such that these characteristic peaks obtain "shoulders". It has previously been demonstrated that, under oxidising conditions, the absorbance spectrum of the FAD-bound NifL protein contains peaks at 358 nm, 362 nm and 446 nm (Macheroux et al., 1998). The molar absorption co-efficient of the protein was calculated to be 12250 M^{-1} cm^{-1} at 446 nm (Macheroux et al., 1998). To analyse the FAD content of the NifL protein and its variants, absorption spectra of protein samples of known concentration (calculated by Bradford assay, see section 2.7.4) were recorded over the 300 nm to 700 nm range of wavelengths using a Perkin-Elmer lambda 35 spectrophotometer with a 1 cm path length. The FAD concentration of each sample was calculated using the following equation.

$A = \varepsilon cl$

Where A = absorbance at 446 nm (absorbance units)

ε = molar absorption co-efficient (M^{-1} cm^{-1})

c = FAD concentration (M)

l = path length (cm)

The ratio of protein concentration to FAD concentration in each sample was used to calculate the number of FAD molecules per NifL dimer.

2.7.12 Limited proteolysis

Limited proteolysis is routinely used to study conformational change in proteins. Trypsin and chymotrypsin proteolysis were performed in TA buffer at 25^{o}C. Samples were incubated for 1 hour before initiating digestion with α-chymotrypsin type I-S (Sigma, from bovine pancreas) or trypsin type III (Sigma, from bovine pancreas). The protease was diluted from a 0.5 mg/mL stock solution to a final protease:NifL weight ratio of 1:60. NifL samples were diluted from a 50 μM

(based on a dimer) stock solution to a final concentration of 5 μM. The total reaction volume was 120 μL. After 0, 2, 5, 10, 20, 30 and 60 minutes, a 15 μL aliquot of the proteolysis reaction was withdrawn and added to microcentrifuge tubes containing 0.35 μg of trypsin/chymotrypsin inhibitor (Roche, from soybean). An equal volume of gel loading buffer (Expedeon 4x sample buffer) was added and samples were analysed by SDS-PAGE on 12% pre-cast polyacrylamide gels (Expedeon) as directed by the manufacturer. For reducing conditions, all samples and stock solutions were prepared in sealed glass bijou tubes and sparged with oxygen-free nitrogen for 3 mins before being transferred to a Belle anaerobic chamber (in which the oxygen level was maintained below 3.5 ppm). The stock solution of $NifL_{(1-284)}$ was reduced with a 100-fold excess of dithionite (5 mM) and a sample removed for spectroscopic analysis to confirm that the flavin co-factor was fully reduced. The proteolysis experiment was then performed as described above. Where appropriate, densitometry analysis was performed using SynGene GeneTools software (version 3.06.04) from Synoptics Ltd.

2.8 Western blotting and immunodetection

To obtain protein extracts, cultures of *E. coli* strain ET8000 containing the plasmid of interest were grown as for the β-galactosidase assays (section 2.9). To ensure that cell numbers were equivalent between samples, the volume taken from each culture was adjusted according to differences in OD_{600}. The normalised cell samples were then centrifuged (13000 rpm for 30 seconds in a Heraeus Pico microfuge) and the pellet resuspended in 50 μl SDS-PAGE sample buffer. The resuspended samples were then boiled for 3 minutes at 100^{o}C and subjected to SDS-PAGE. After electrophoresis, proteins were electrotransferred onto nitrocellulose

membranes (Amersham Biosciences) using an XCell II™ blot module (Invitrogen) as directed by the manufacturer. After blotting, membranes were washed in TBS buffer (two 10 minute washes) and blocked by incubating overnight in 20 ml blocking buffer at 4°C. A series of three wash steps was then performed (two 10 minute washes in TBS-Tween/Triton buffer were followed by a 10 minute wash in TBS buffer). Nitrocellulose membranes were then probed with polyclonal antisera against NifL. In order to titrate non-specifically binding antibodies out of the rabbit antiserum, a 10000-fold dilution was prepared in TBS containing 3% BSA and broken ET8000 cells that lacked the plasmid of interest. Membranes were incubated for 1 hour in 25 ml of this solution and subsequently washed as after blocking. Primary antibodies were detected with alkaline phosphatase-conjugated anti-rabbit secondary antibodies raised in goat (Sigma). The membranes were incubated for 45 minutes in 25 ml TBS containing 3% BSA and a 5000-fold dilution of secondary antibody and then washed four times in TBS-Tween/Triton (each wash was 10 minutes). Finally, secondary antibodies were detected by staining with SigmaFAST™ 5-bromo-4-chloro-3-indolylphosphate and nitroblue tetrazolium (Sigma) as directed by the manufacturer.

2.9 Experimental assays

2.9.1 Assay of NifL activity *in vivo*

NifL activity was assayed as the ability of the NifA protein to activate transcription of a *nifH-lacZ* promoter fusion in the presence of NifL under four distinct growth conditions. Chemically competent *E. coli* ET8000 cells carrying the reporter plasmid pRT22 (containing the *nifH-lacZ* fusion) were transformed with pPR34 (carrying *nifL* co-expressed with *nifA*) or a mutant derivative and spread onto

LB agar supplemented with 100 μg ml^{-1} carbenicillin and 35 μg ml^{-1} chloramphenicol (the pRT22 and pPR34 plasmids carry chloramphenicol and carbenicillin resistance markers respectively). Individual colonies were picked and grown at 37^{o}C for 8 hours in LB liquid media containing the same antibiotics. These cultures were used to inoculate "assay cultures" grown in NFDM media supplemented with 5% (v/v) Hino and Wilson buffer, appropriate antibiotics and either 200 μg ml^{-1} casein hydrolysate (for nitrogen limiting conditions) or 1 mg ml^{-1} $(NH_4)_2SO_4$ (for conditions rich in fixed nitrogen). These cultures were grown at 30^{o}C overnight either aerobically in 50 ml conical flasks containing 5 ml of medium with vigorous shaking (250 rpm), or anaerobically in tightly sealed 7 ml bijou tubes containing 7 ml medium. Assays for β-galactosidase activity were then performed as described in section 2.9.3.

2.9.2 Bacterial adenylate cyclase two-hybrid analysis

The bacterial adenylate cyclase two-hybrid (BACTH) system is commonly used to investigate interactions between proteins or protein domains. It relies on functional complementation between subunits of the adenylate cyclase (AC) enzyme expressed *in trans* in an AC-deficient *E. coli* reporter strain. The N-terminal (T25) and C-terminal (T18) domains of the hetero-dimeric AC protein are encoded on plasmids pT25 and pUT18 respectively (Karimova et al., 1998). These plasmids each contain an antibiotic resistance marker and a multiple cloning site to aid the creation of two hybrid proteins containing an AC subunit transcriptionally (and translationally) fused to the protein of interest. Interaction between the proteins of interest results in co-localisation of the AC subunits and functional complementation. Thus, protein-protein interactions result in a substantial increase in AC activity

(Karimova et al., 1998; Karimova et al., 2000). This, in turn, triggers transcription of various reporter genes, including *lacZ*.

The plasmids pT25 and pUT18, or derivatives containing the desired *nifL* sequence, were co-transformed into chemically competent *E. coli* strain BTH101 (Karimova et al., 2000) and spread onto LB agar supplemented with X-gal (40 μg ml^{-1}), chloramphenicol (35 μg ml^{-1}), carbenicillin (100 μg ml^{-1}) and IPTG (0.5 mM). Agar plates were incubated at 30°C for 72 hours. Colonies were then picked and cultured in universal tubes containing 5 ml LB supplemented with chloramphenicol and carbenicillin for approximately 8 hours at 30°C with shaking (250 rpm). 50 μl aliquots of these cultures were used to inoculate 7 ml bijou tubes containing 7 ml LB supplemented with 1% (v/v) glucose, 0.5 mM IPTG and appropriate antibiotics. Bijou tubes were tightly sealed and grown overnight at 30°C. Cultures were then assayed for β-galactosidase activity as described in section 2.9.3.

2.9.3 β-galactosidase assays

After overnight growth of assay cultures (see sections 2.9.1 and 2.9.2) the OD_{600} was recorded. 30 μl of each culture was then added to 970 μl lysis buffer containing 2% (v/v) chloroform, vortexed and incubated at 30°C for at least 10 minutes. Assay reactions were started by addition of 200 μl start buffer (4 mg ml^{-1} 2-nitrophenyl-β-galactopyranoside) and, after a carefully recorded period of incubation at 30°C, each reaction was stopped by adding 500 μl stop buffer (0.5 M $NaCO_3$). The OD_{420} and OD_{550} of each reaction tube was then measured and the β-galactosidase activity was calculated using the following equation:

$$\text{β-galactosidase activity (Miller Units)} = \frac{OD_{420} - (1.75OD_{550})}{vtOD_{600}}$$

Where v = volume (ml) of lysed culture

OD_{420} = optical density at 420 nm

OD_{550} = optical density at 550 nm

OD_{600} = optical density at 600 nm

t = total reaction time (minutes).

Chapter 3 - Influence of the PAS2 domain on NifL function *in vivo*

3.1 Introduction

As discussed in Chapter 1, NifL is a transcriptional anti-activator that regulates the expression of genes required for nitrogenase biosynthesis via interaction with its partner protein, NifA. NifL controls transcriptional activation by NifA in response to cellular levels of oxygen and fixed nitrogen (Martinez-Argudo et al., 2004c). NifL is a modular protein that contains four discrete domains: a C-terminal GHKL domain, an H domain and two N-terminal PAS domains (Figure 1.22). The C-terminal (H and GHKL) domains are important for inhibition of NifA and nitrogen sensing. The first PAS domain, PAS1, is located between NifL residues 1-140 and senses changes in redox potential via a FAD co-factor (Söderbäck et al., 1998). The SMART and Pfam databases recognise the region of NifL between residues 151-268 as a second distinct PAS domain, called PAS2. Secondary structure predictions using the PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred/) and Jpred (http://www.compbio.dundee.ac.uk/www-jpred/) servers and alignments with known PAS structures indicate that the PAS2 domain is likely to contain structural elements found in other PAS domains (Figure 3.1). This domain has no apparent co-factor and prior to this work its function was unknown. As mentioned in Chapter 1, it is common for signalling proteins to contain multiple PAS domains in tandem. Despite the abundance of duplicate PAS domains, relatively few have been characterised. In the studied examples DcuS and KinA, the biological function of the second PAS domain within the tandem pair is often unclear due to the apparent lack of any co-factor or ligand binding pocket that might be indicative of a role signal perception (Etzkorn et al., 2008; Lee et al., 2008). The N-terminal PAS domains of NifL are typical in this sense; the PAS1 domain has a role in signal perception whilst the function of PAS2 is unclear. This work aimed to elucidate the role of the NifL PAS2

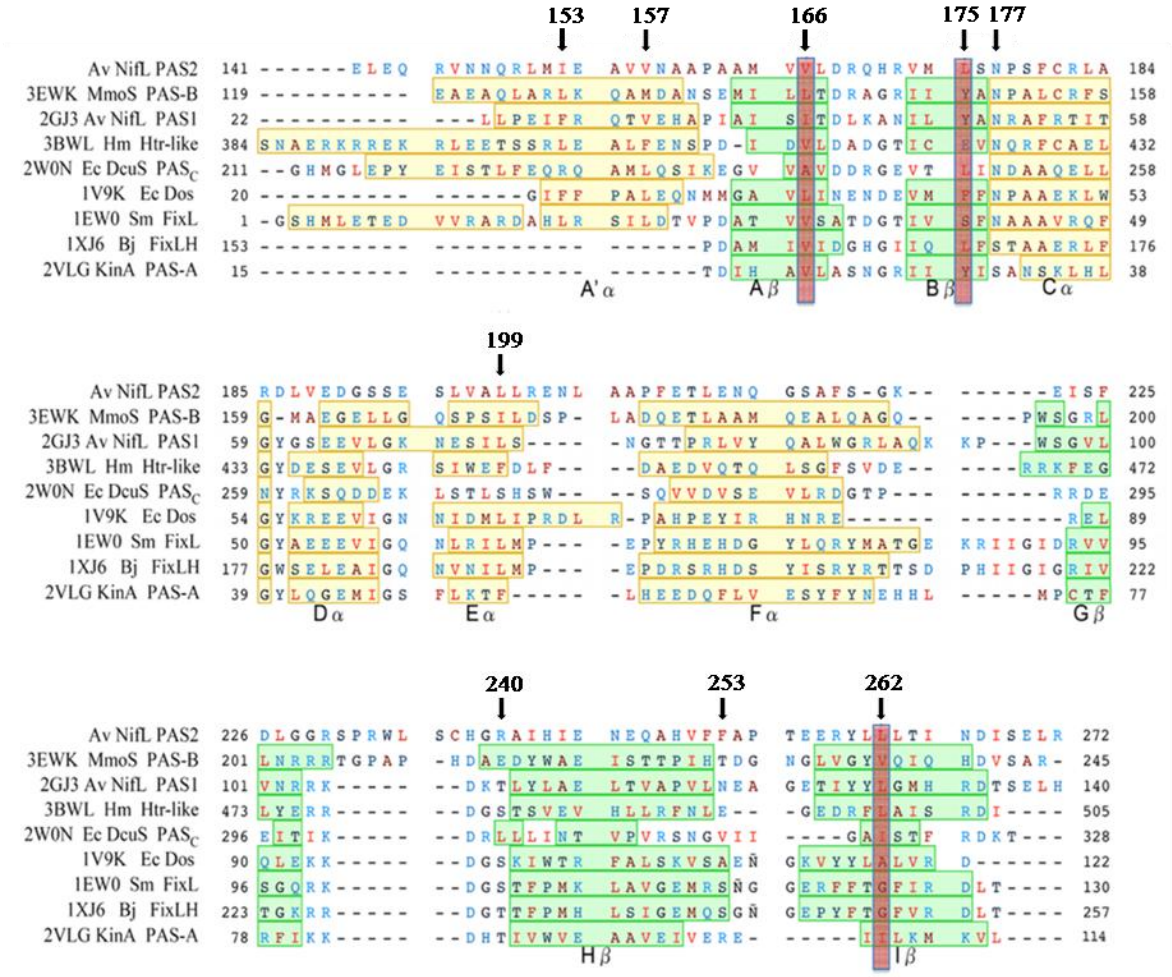


Figure 3.1. Sequence alignment of the *A. vinelandii* NifL PAS2 domain with PAS domains of known structure (PDB code and protein designation are listed to the left of each sequence). Amino acids are coloured according to their level of Kyte-Doolittle hydrophobicity (Kyte and Doolittle, 1982) . The group I (most hydrophobic) residues are coloured light red, group II are dark red, group III are dark blue and groups IV and V (the least hydrophobic) are light blue. The α-helices are highlighted in yellow and β-strands in green. Residues in the conserved PAS dimer interface (Ayers and Moffat, 2008) are highlighted in red. The positions of amino acid substitutions in the NifL PAS2 domain analysed in this thesis are indicated by arrows above the text.

domain in signalling using genetic and biochemical approaches. The first step in this analysis was to mutagenise the region of *nifL* encoding the PAS2 domain to examine whether/how substitutions in this domain might influence signalling in NifL.

3.2 Mutagenesis of the NifL PAS2 domain

The DNA sequence encoding the PAS2 domain was randomly mutagenised using error-prone PCR and the mutant library was inserted into *nifL* to replace the wild-type sequence. The resultant NifL mutants were screened for differences in their ability to inhibit NifA activity using the two-plasmid heterologous reporter system described in section 2.9. In this system, β-galactosidase activity is measured as an indicator of NifA-mediated transcriptional activation from a *nifH-lacZ* reporter. Colonies expressing wild-type NifL (co-transcribed with NifA) are pale blue on Xgal indicator plates containing limiting fixed nitrogen when incubated under aerobic conditions, due to low levels of NifA activity. Mutants that appeared white (indicating enhanced repression of NifA activity by NifL) or dark blue (suggesting the NifL mutant protein was deficient in its ability to inhibit NifA) were selected and the plasmid DNA was recovered and sequenced (see Materials and Methods Chapter 2.5.10). Additionally, site-directed mutagenesis was used to investigate the role of residues predicted to contribute to a conserved dimerisation interface found in many PAS domains of known structure (see section 1.2.6). Single codon changes were generated using a two-step PCR method and the presence of the desired mutation was then confirmed by DNA sequencing. NifL mutants of known sequence were assayed for their ability to inhibit NifA activity in response to redox and fixed nitrogen signals *in vivo*. NifA-mediated transcriptional activation from a reporter plasmid containing a *nifH-lacZ* fusion was measured in the presence of NifL or mutant derivatives using the two-plasmid heterologous system mentioned above. However, unlike the indicator plates used in the

mutant screens, these are quantitative assays for β-galactosidase expression. Assay cultures were grown under four conditions to assess the response of the NifL protein to different environmental cues. Cells were grown either aerobically or anaerobically with casein hydrolysate (to simulate nitrogen limiting conditions) or ammonium sulphate (for excess fixed nitrogen) as the sole nitrogen source (see section 2.9.1 for details). In other words, cultures were grown under nitrogen fixing conditions or in the presence of excess oxygen or excess fixed nitrogen or a combination of both. Three distinct phenotypes emerged from both the random and sire-directed approaches: (i) "locked-on" mutants that constitutively inhibited NifA activity, (ii) "redox signalling" mutants that failed to respond to the presence of oxygen but inhibited NifA activity in the presence of high levels of fixed nitrogen, and (iii) "aerobically inactive" mutants that failed to inhibit NifA under oxidising conditions irrespective of fixed nitrogen availability but retained some ability to respond to fixed nitrogen under anaerobic conditions (Figure 3.2).

(i) "Locked-on" mutants

As demonstrated previously, the wild-type NifL protein represses NifA activity in discrete responses to oxygen (Figure 3.2A, compare the blue and red bars) and fixed nitrogen (Figure 3.2A, compare blue and yellow bars) or a combination of both (Figure 3.2A, green bars). Control experiments demonstrated that there is no transcription from the reporter fusion in the absence of NifA (Figure 3.2A, bars marked "Reporter only") and that NifA is active across all four conditions in the absence of regulation by its partner protein, NifL (Figure 3.2A, bars marked "NifA"). The amino acid substitutions V157A, V166M, L175A, N177S, L199R, R240W, and L262A apparently lock the NifL protein in the inhibitory conformer, causing NifL to inhibit NifA activity in the absence of oxygen and

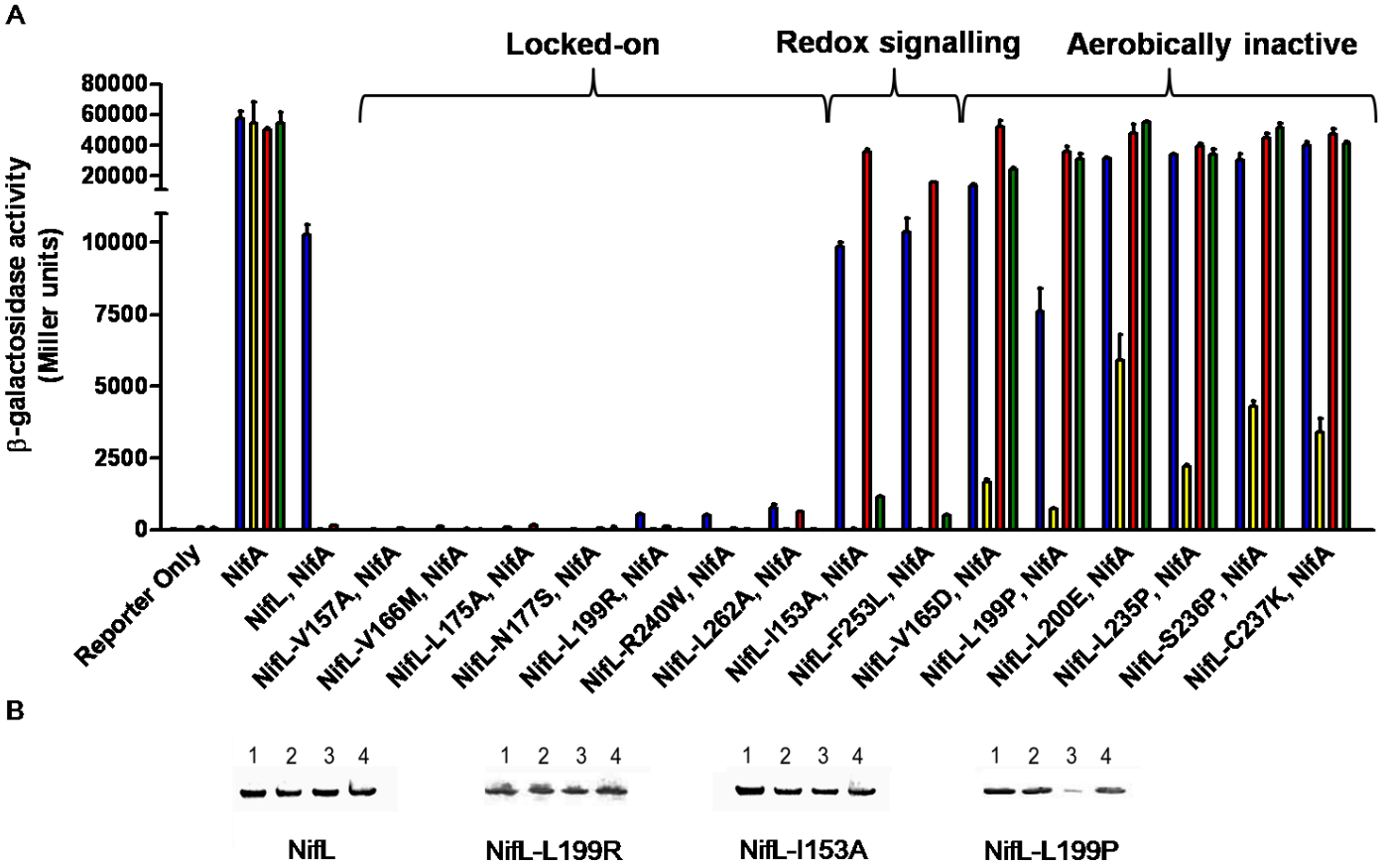


Figure 3.2. Activity and stability of mutant NifL proteins *in vivo*. (A) Influence of mutant NifL proteins on transcriptional activation by NifA *in vivo*. Cultures were grown under the following conditions; (1) anaerobically, under nitrogen-limiting conditions (blue bars), (2) anaerobically with excess fixed nitrogen (yellow bars), (3) aerobically with limiting fixed nitrogen (red bars) and (4) aerobically when fixed nitrogen was replete (green bars). Cultures were assayed for β-galactosidase activity as a reporter of NifA-mediated transcriptional activation from a *nifH-lacZ* fusion. Further experimental detail concerning assays and culture conditions can be found in section 2.9. All experiments were performed at least in duplicate with error bars denoting the standard error of the mean. (B) Anti-NifL western analysis of the wild-type and mutant NifL proteins in strains grown under the same four conditions (labelled 1-4) as the β-galactosidase assay cultures.

fixed nitrogen (Figure 3.2A, blue bars). NifL variants of this sort were termed “locked-on” mutants.

Western analysis indicated that these NifL variants were stable across the four assay conditions and a representative example (NifL-L199R) is shown in Figure 3.2B. The identification of amino acid substitutions in the PAS2 domain that lock NifL in the inhibitory conformer suggests that the PAS2 domain is important to the conformational changes that occur when the NifL protein switches between the inhibitory and non-inhibitory signalling states.

(ii) “Redox signalling” mutants

Two amino acid substitutions (L153A and F253L) in the PAS2 domain gave rise to a form of the NifL protein that is insensitive to redox signals but responds normally to fixed nitrogen. In contrast to wild-type NifL, which strongly represses NifA activity in response to oxygen, these variants allowed high levels of NifA activity under oxidising conditions (Figure 3.2A, red bars). However, NifL-L153A and NifL-F253L retained the ability to respond to fixed nitrogen and addition of ammonium to the media results in strong inhibition of NifA activity (Figure 3.2A, yellow and green bars). Western blotting experiments demonstrated that both mutant proteins were stable under all conditions tested and data for NifL-I153A is shown in Figure 3.2B. Thus, the inability of these NifL variants to respond to oxygen is not due to a change in stability caused by the substitutions but instead indicates a defect in the redox signalling mechanism. Therefore, NifL variants of this type were termed “redox signalling” mutants. The identification of this class of mutant in the PAS2 domain of NifL suggests that PAS2 is involved in redox signalling. As mentioned above, the PAS2 domain has no apparent co-factor and previous studies have demonstrated that a truncated form of NifL lacking the PAS1 domain (i.e. containing

PAS2 and the C-terminal output domains) is not responsive to changes in redox status. Taken together, the available evidence implies that PAS2 may have a role in relaying signals from the redox-sensing PAS1 domain to the C-terminal domains of NifL.

(iii)"Aerobically inactive" mutants

Six NifL variants that were unable to inhibit NifA activity under oxidising conditions were also obtained. These where named "aerobically inactive" mutants. NifL-V165D, NifL-L199P, NifL-L200E, NifL-L235P, NifL-S236P and NifL-C237K allowed high levels of NifA activity in the presence of oxygen and little or no reduction in NifA activity was observed when fixed nitrogen was added to aerobic cultures (Figure 3.2A, red and green bars). In other words, these NifL variants were unable to respond to fixed nitrogen under oxidising conditions. However, they retained some nitrogen sensitivity under anaerobic conditions as addition of ammonium to anaerobic cultures resulted in lower NifA activity (Figure 3.2A, compare blue and yellow bars). Even under these growth conditions (when fixed nitrogen is replete and oxygen is limiting) NifA activity was not fully repressed (Figure 3.2A yellow bars) and, in most cases, the mutant NifL proteins allowed greater NifA activity under nitrogen fixing conditions than the wild-type protein (Figure 3.2A, blue bars). Thus, the "aerobically inactive" NifL variants exhibit an impaired ability to inhibit NifA. Western analysis indicated that the mutant NifL proteins were stable under three of the four test conditions but were relatively unstable under oxidising conditions when fixed nitrogen was absent (data for NifL-L199P is shown in Figure 3.2B as a representative example). This instability may contribute to the failure of the mutant NifL proteins to inhibit NifA activity in the presence of oxygen but cannot account for the deficient response to fixed nitrogen under oxidising conditions. Despite the stability of the variant proteins when oxygen and fixed nitrogen were in excess, neither stimulus resulted

in inhibition of NifA activity (Figure 3.2A, green bars). This implies that the redox response is deficient regardless of changes in protein stability. It is worth noting that all mutations that result in an "aerobically inactive" phenotype encode substitutions that are likely to disrupt the structure of the protein, for example two leucine residues are substituted for proline (L199P and L235P). One possible interpretation of this data is that the "aerobically inactive" phenotype derives from structural perturbation of the PAS2 domain. The isolation of "aerobically inactive" mutants in PAS2 provides further evidence that this domain is involved in redox signalling and implies that a "mis-functioning" PAS2 domain can disrupt the nitrogen response. Further, it suggests that the PAS2 domain can influence the conformation of the C-terminal domains of NifL. Overall, these data suggest that the "aerobically inactive" NifL mutants are competent to interact with GlnK (which conveys the nitrogen signal, see section 1.4.3) and undergo the conformational changes necessary to inhibit NifA under anaerobic conditions. However, the presence of oxygen triggers a deficient redox signalling pathway that somehow perturbs the otherwise functional nitrogen response.

3.2.1 Site-directed mutagenesis at positions 199 and 166

In order to better understand the phenotypes of mutants isolated by random mutagenesis, the importance of some residues was investigated by site-directed mutagenesis. It had previously been observed that introduction of charged glutamate or arginine residues at position 199 both resulted in a "locked-on" phenotype, despite the opposite polarity of their charges (see section 1.5). Therefore, it was postulated that the loss of hydrophobicity at this position was responsible for this phenotype. To test this hypothesis, site-directed substitutions of L199 were generated using a two-step PCR technique. The ability of the resulting NifL variants to inhibit transcriptional activation by

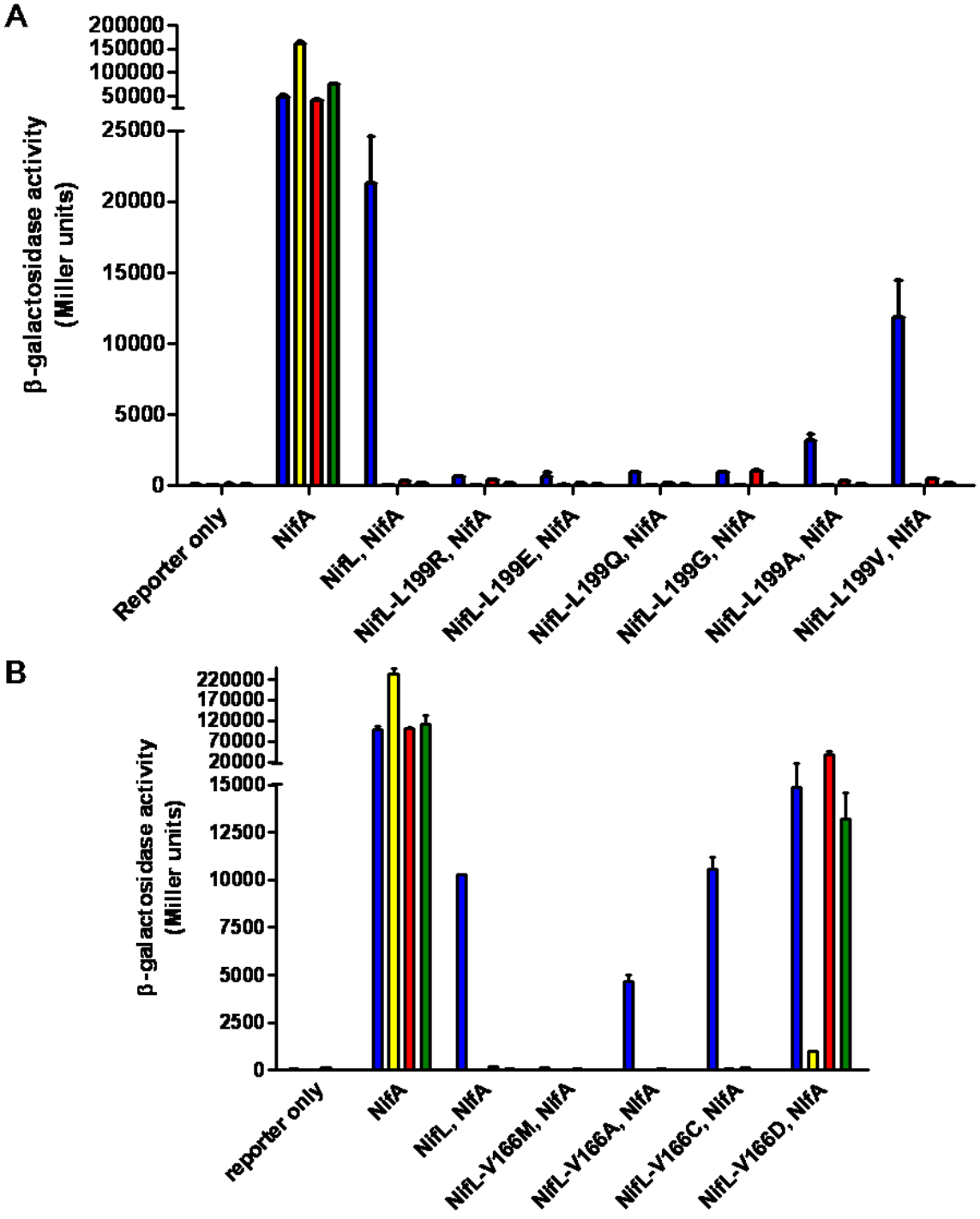


Figure 3.3. (A) NifA activity in the presence of mutant NifL proteins with substitutions of L199 for residues of varying hydrophobicity. (B) NifA activity in the presence of NifL variants with substitutions of V166. Assays and culture conditions in parts A and B are as described in Figure 3.2.

NifA was tested (Figure 3.3A). The L199Q substitution exhibited a "locked-on" phenotype similar to that of NifL-L199E and NifL-L199R (Figure 3.3A). Complete removal of the hydrophobic side chain of residue 199 by substitution of the native leucine for glycine also gave rise to a "locked-on" form of NifL, indicating that the absence of hydrophobicity at this position is sufficient to lock the NifL protein in the inhibitory conformer (Figure 3.3A). When some hydrophobicity was restored by introducing an alanine or valine residue at position 199, a significant increase in NifA activity was detected under nitrogen-fixing conditions (Figure 3.3A, blue bars). Moreover, the extent to which NifA activity increased (relative to NifA activity in the presence of NifL-L199G, NifL-L199R or NifL-L199E) appeared to be proportional to the hydrophobicity of the residue introduced (i.e. NifA activity is greater in the presence of NifL-L199V than in the presence of NifL-L199A). This suggests that, under nitrogen fixing conditions, L199 participates in a hydrophobic interaction which is required in order for NifL to adopt the non-inhibitory conformation.

As mentioned above, substitution of V166 for methionine gives rise to a "locked-on" form of the NifL protein. To investigate the influence of the side chain at this position several substitutions were generated using a two-step PCR technique. The activity of NifA in the presence of NifL-V166M, NifL-V166A, NifL-V166C and NifL-V166D was then determined (Figure 3.3B). As demonstrated previously, NifL-V166M inhibited NifA activity under all four conditions, even when oxygen and fixed nitrogen were limiting (Figure 3.3B, blue bars). The same phenotype was not observed for any of the other NifL variants. NifL-V166C was indistinguishable from the wild-type NifL protein under all conditions tested (Figure 3.3B, compare bars marked "NifL-V166C, NifA" with those marked "NifL-V166M, NifA"). NifL-V166A responded normally to oxygen and fixed nitrogen but allowed approximately 2-fold less NifA activity than the wild-type protein under nitrogen fixing conditions (Figure 3.3B, blue bars). The less conservative

substitution of V166 for aspartate resulted in an "aerobically inactive" phenotype (Figure 3.3B), perhaps indicating that this substitution perturbs the structure of the PAS2 domain. It is interesting that the two sulphur-containing substitutions at position 166 (V166C and V166M) have strikingly different phenotypes. NifL-V166C appears wild type whilst NifL-V166M is "locked-on". Since the side chains of methionine and the native valine are both hydrophobic, it seems likely that the "locked-on" phenotype of the V166M variant is due to steric hindrance when the bulkier methionine side chain is present. This implies that V166 (or C166 in the NifL-V166C variant) is tightly packed against another residue when NifL is in the non-inhibitory conformation but not when the protein is in the inhibitory state. Thus, substitution of V166 for methionine forces the protein into the inhibitory conformer, resulting in a "locked-on" phenotype.

3.2.2 Mutagenesis of the Eα helix

As mentioned in section 3.1, secondary structure predictions indicate that the NifL PAS2 domain contains structural elements found in other PAS domains of known structure. One such structural element is the Eα helix (Figures 3.1 and 1.4B), which forms part of a conserved cleft that accommodates a co-factor in some PAS domains (Möglich et al., 2009b). This helix is amongst the most variable features of the PAS superfamily and its structure and amino acid sequence are often adapted to suite the biological function of specific PAS domains. In the NifL PAS2 domain, Eα is a predicted amphipathic α-helix that contains at least one residue important in signalling, L199 (discussed above). It was decided to further investigate the function of this helix by site-directed mutagenesis. Structural predictions and alignments with PAS domains of known structure indicate that the PAS2 Eα helix is likely to extend from residue 196 to residue 201 (Figure 3.1). However, in the absence of any direct structural information it is difficult to predict its

precise length and position. Additionally, the Eα and Dα helices are merged in some PAS domains (e.g. NifL PAS1) to form one extended α-helix. Therefore, substitutions were generated throughout the region of PAS2 predicted to form the amphipathic Eα helix and extending outwards to its flanking residues (amino acids 192-202). The ability of the resulting NifL variants (NifL-S192G, NifL-S193G, NifL-E194K, NifL-S195G, NifL-L196A, NifL-L199A, NifL-L200A, NifL-L200E, NifL-R201A, NifL-E202A and NifL-E202K) to inhibit transcriptional activation by NifA was examined *in vivo* (Figure 3.4).

Substitution of residues S192, S193 or S195 for glycine did not influence NifL activity *in vivo* (Figure 3.4A). Likewise, the L196A, E202A and E202K substitutions did not significantly alter the ability of NifL to inhibit transcriptional activation by NifA. As demonstrated previously, NifL-L199A responded normally to changes in redox potential and nitrogen status but allowed approximately 5-fold less NifA activity under nitrogen fixing conditions when compared to the wild type (Figure 3.4A). NifL-L199A can be described as intermediate between the "locked-on" and wild-type proteins. Substitutions of two charged residues (R201 and E194) located on the opposite side of the Eα helix (to L199) result in a similar phenotype. The NifL-R201A and NifL-E194K proteins are both impaired in their ability to release NifA from inhibition when oxygen and fixed nitrogen are limiting (Figure 3.4A). Thus, these results cannot be rationalised simply in terms of the amphipathic nature of the Eα helix as substitutions on both the hydrophilic and hydrophobic sides result in similar phenotypes. However, this mutagenesis yielded an interesting NifL variant worthy of further study. NifL-L200A is an "aerobically inactive" mutant that appears relatively stable under oxidising conditions when fixed nitrogen is limiting (compared to other NifL variants of the same phenotype) (Figure 3.4B). It was not certain whether the inability of the "aerobically inactive" mutants to respond to fixed nitrogen under oxidising conditions was directly due to a change in the signalling state of

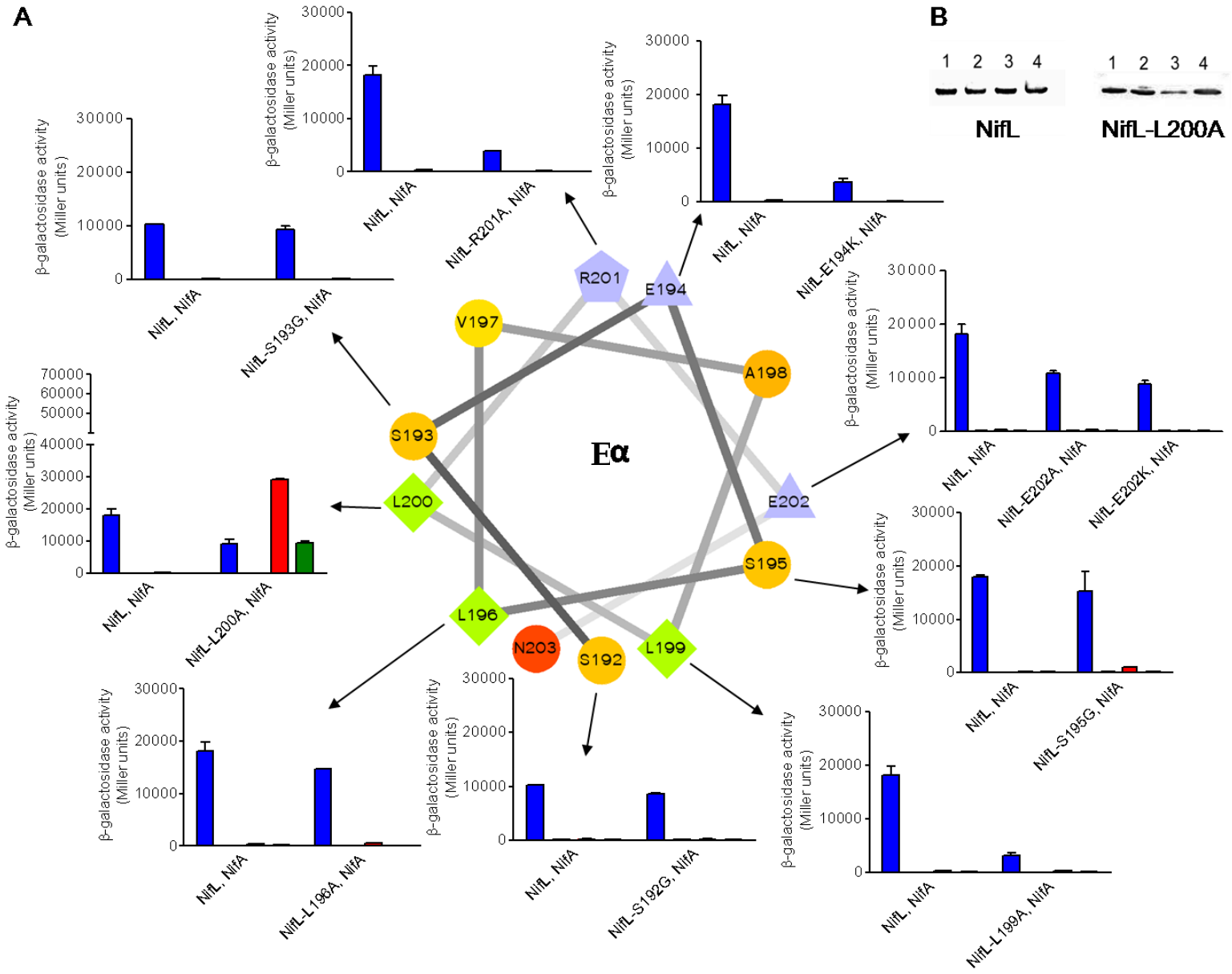


Figure 3.4. Mutagenesis of the Eα helix in the PAS2 domain of NifL. (A) Helical wheel projection of the Eα helix with surrounding graphs showing NifA activity in the presence of mutant NifL proteins. Arrows indicate the appropriate graph for each residue in the helical wheel. Graph legends are as in Figure 3.2. Control experiments measuring NifA activity in the absence of NifL and transcription from the reporter plasmid in the absence of NifA were performed as previously but omitted from the figure for simplicity. (B) Anti-NifL western analysis of NifL and NifL-L200A. Lanes are labelled as in Figure 3.2.

the PAS1 domain or was indirectly caused by physiological differences between cells grown under conditions of differing oxygen availability. In order to discern between these two possibilities, experiments were performed in which the NifL-L200A protein was truncated to remove the redox sensing PAS1 domain (to give $NifL_{(143-519)}$-L200A) or combined with a secondary mutation (E70A) that disrupts redox signalling by PAS1 (see section 1.2.3). Removal of the PAS1 domain or introduction of E70A as a secondary substitution restored the aerobic nitrogen response of NifL-L200A (data not shown). In other words, disruption of redox sensing by PAS1 allowed NifL-L200A to respond normally to fixed nitrogen under oxidising conditions. This demonstrates that perception/transmission of the redox signal by the PAS1 domain influences the conformation of the C-terminal domains in the NifL-L200A variant. As the L200A substitution is located in the PAS2 domain, this result implies that the signalling state of the PAS1 domain influences the C-terminal region of NifL via conformational changes in PAS2.

3.2.3 PAS2 deletions

In addition to studying single amino acid substitutions in PAS2, the effect of removing the PAS2 domain on the ability of NifL to inhibit NifA activity was also investigated. Three plasmids, each encoding a variant form of the NifL protein lacking the PAS2 domain (NifLΔ148-270, NifLΔ146-272 and NifLΔ146-275), were constructed as described in section 2.6.1. The stability of these NifL variants and their ability to inhibit transcriptional activation by NifA *in vivo* were examined (Figure 3.5). Western analysis indicated that the mutant proteins were stable under three of the four conditions tested (Figure, 3.5B). However, a reduction in the stability of the NifLΔ146-272 and NifLΔ146-275 deletions was observed under aerobic conditions when fixed nitrogen was limiting.

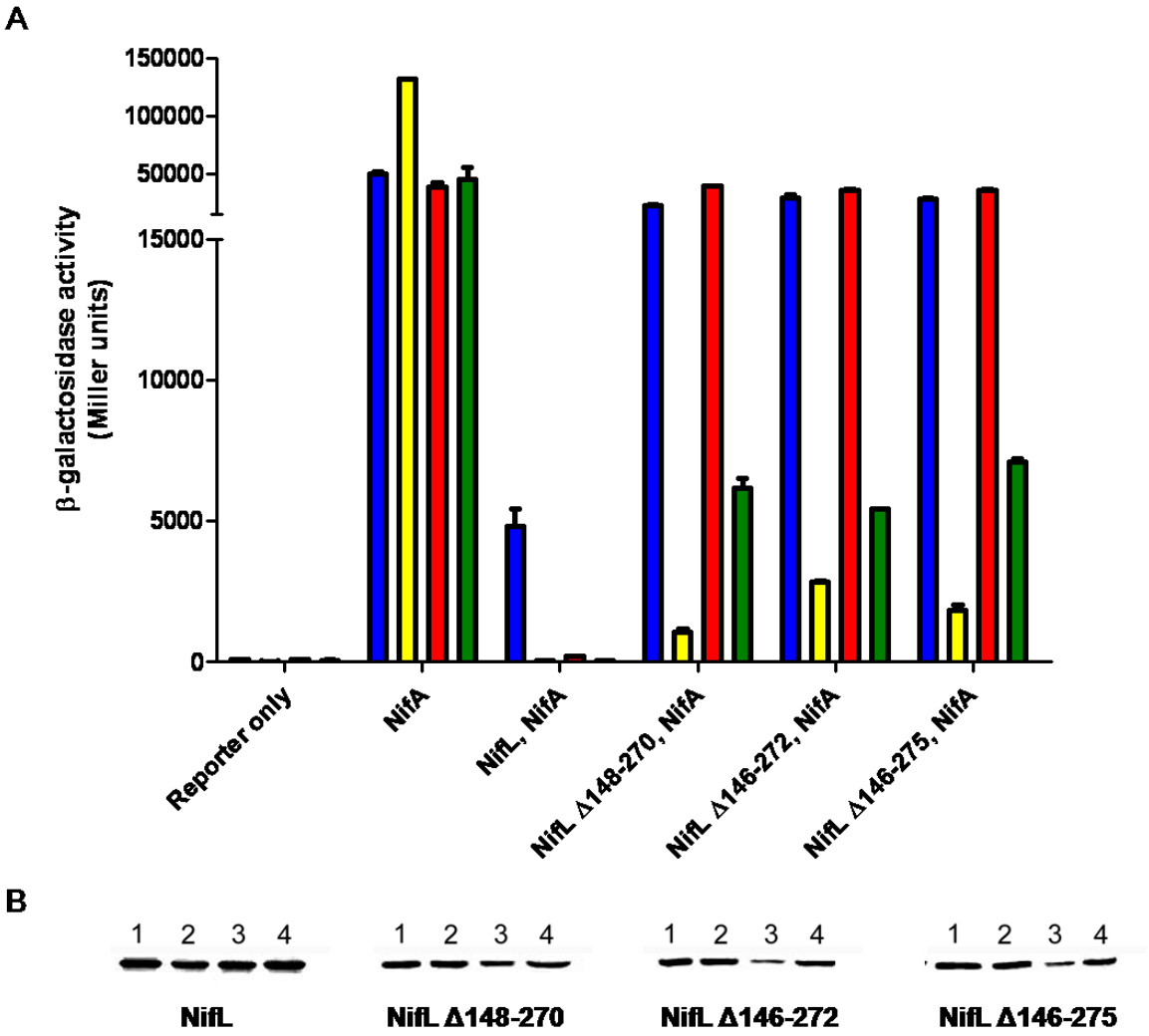


Figure 3.5. Activity and stability of PAS2 deletion mutants *in vivo*. (A) Influence of mutant NifL proteins on transcriptional activation by NifA. Assay and culture conditions are as described in the legend of Figure 3.2. (B) Anti-NifL western analysis of the wild-type and mutant NifL proteins in strains grown under the same four conditions as used for the β-galactosidase assays. Lanes are labelled as in Figure 3.2.

NifLΔ148-270 appeared more stable under these conditions (Figure 3.5B). In all cases, the truncated proteins where unable to inhibit NifA activity in the absence of fixed nitrogen (Figure 3.5A, blue and red bars) but retained some ability to respond to changes in nitrogen status (Figure 3.5A, compare the blue and yellow bars or compare the red and green bars). Under nitrogen limiting conditions, NifA activity in the presence of the NifL variants was similar to that observed when the NifL protein was absent altogether (Figure 3.5A, blue and red bars). This strongly implies that, despite slight differences in protein stability under the relevant conditions, the PAS2 deletion mutants were not sensitive to changes in redox status. Moreover, significant β-galactosidase activity was evident under all conditions, suggesting that none of the variant NifL proteins were able to fully repress NifA activity *in vivo*. The response of the PAS2 deletion mutants to fixed nitrogen was slightly more efficient under anaerobic conditions than under aerobic conditions (Figure 3.5A). Overall, the phenotypes of the PAS2 deletion mutants resemble those of the "anaerobically inactive" NifL variants in that both are unable to inhibit NifA in response to oxygen and are impaired in their ability to respond to fixed nitrogen. Taken together, these data emphasise the importance of PAS2 in redox signalling and imply that the absence of the PAS2 domain from the NifL protein results in a conformation of the C-terminal domains that is not optimal for inhibition of NifA. These results also demonstrate that the NifL protein can sense fixed nitrogen without a functional PAS2 domain, although the efficacy of the response is impaired. The ability of the PAS2 deletion mutants to sense fixed nitrogen is not surprising given that the nitrogen response is mediated by interaction of the GHKL domain with GlnK whilst the reduced efficiency of the response may be symptomatic of an altered conformation of the C-terminal domains in the absence of PAS2.

3.3 Properties of the mutant NifL proteins *in vivo*

3.3.1 "Locked-on" mutants require a functional nucleotide binding domain

It has previously been demonstrated that NifL requires nucleotide binding to the C-terminal GHKL domain in order to inhibit NifA activity. NifL is incompetent to bind NifA *in vitro* in the absence of nucleotide and the binding of ADP to the GHKL domain has been shown to stabilise the NifL-NifA binary complex (Eydmann et al., 1995; Money et al., 1999). Mutations in the GHKL domain that prevent binding of adenosine nucleotides result in the inability of NifL to inhibit NifA *in vivo* (Martinez-Argudo et al., 2004c; Perry et al., 2005). In order to indirectly assess whether the "locked-on" phenotype of the PAS2 mutations requires nucleotide binding, "locked-on" mutants were combined with a secondary amino acid substitution in the GHKL domain (G480A), that has been fully characterised in previous studies and is known to disrupt ADP binding (Perry et al., 2005). Plasmid DNA encoding the L199R or V166M substitutions was cleaved at two unique restriction sites and transferred into a plasmid carrying G480A (pNLG480A), which had been digested with the same enzymes. The new double mutants (encoded on plasmids pPS2 and pPS124) were then assayed for their ability to inhibit NifA activity *in vivo* (Table 3.1). As demonstrated previously, NifL-G480A failed to inhibit transcriptional activation by NifA (Table 3.1, row 4), whilst NifL-V166M and NifL-L199R constitutively inhibited NifA activity (Table 3.1, rows 5 and 6 respectively). When either of these PAS2 mutants were combined with G480A, the G480A phenotype was dominant, since the NifL-V166M, G480A and NifL-L199R, G480A double mutants were severely compromised in their ability to inhibit NifA activity under all conditions tested (Table 3.1, rows 7 and 8). Western analysis confirmed that all of the mutant proteins were stable under the four test conditions (data not shown). Overall, these results suggest that disruption of nucleotide binding nullifies the "locked-on" phenotype of V166M and L199R and that the

			β-Galactosidase activity in Miller Units (+/- SE)			
			Anaerobic		Aerobic	
Row	Plasmid	Protein(s)	N-	N+	N-	N+
1	-	reporter only	90 (+/- 3)	37 (+/- 14)	154 (+/- 2)	30 (+/-4)
2	pPR39	NifA	85217 (+/- 5551)	203320 (+/- 35830)	80003 (+/- 2503)	81818 (+/- 6504)
3	pPR34	NifA, NifL	6115 (+/- 833)	97 (+/- 3)	279 (+/- 0)	51 (+/- 2)
4	pNLG480A	NifA, NifL-G480A	86563 (+/- 634)	90943 (+/- 328)	58257 (+/- 9595)	41418 (+/- 3754)
5	pPS20	NifA, NifL-V166M	304 (+/- 2)	18 (+/- 18)	163 (+/- 7)	42 (+/- 2)
6	pNSK2	NifA, NifL-L199R	181 (+/-0)	14 (+/- 4)	138 (+/- 8)	87 (+/- 1)
7	pPS124	NifA, NifL-V166M, G480A	93644 (+/- 3153)	33592 (+/- 2757)	66683 (+/- 4086)	8108 (+/- 462)
8	pPS2	NifA, NifL-L199R, G480A	33426 (+/- 4634)	20303 (+/- 66)	40315 (+/- 4372)	28531 (+/- 1517)

Table 3.1. The "locked-on" phenotype of mutations in the PAS2 domain requires a functional nucleotide-binding (GHKL) domain. The data presented in Tables 3.1 and 3.2 are derived from at least two independent replicates.

			β-Galactosidase activity in Miller Units (+/- SE)			
			Anaerobic		Aerobic	
Row	Plasmid	Protein(s)	N-	N+	N-	N+
1	-	reporter only	15 (+/- 0)	19 (+/- 2)	124 (+/- 15)	99 (+/- 8)
2	pPR39	NifA	25004 (+/- 9693)	150911 (+/- 3301)	48843 (+/- 2037)	59216 (+/- 2837)
3	pPR34	NifA, NifL	12583 (+/- 546)	28 (+/- 2)	221 (+/- 10)	93 (+/- 15)
4	pPS54	NifA, $NifL_{(143-519)}$	12423 (+/- 471)	93 (+/- 5)	7347 (+/- 286)	565 (+/- 12)
5	pPR54	NifA, $NifL_{(147-519)}$	12199 (+/- 2234)	42 (+/- 0)	6515 (+/- 46)	229 (+/- 9)
6	pPS20	NifA, NifL-V166M	547 (+/- 24)	16 (+/- 1)	114 (+/- 24)	21 (+/- 8)
7	pNSK2	NifA, NifL-L199R	607 (+/- 63)	18 (+/- 2)	383 (+/- 1)	179 (+/- 7)
8	pPS77	NifA, $NifL_{(143-519)}$-V166M	1166 (+/- 57)	17 (+/- 0.3)	609 (+/- 32)	47 (+/- 6)
9	pPS4	NifA, $NifL_{(147-519)}$-L199R	460 (+/- 93)	14 (+/- 1)	484 (+/- 49)	165 (+/- 2)

Table 3.2. The PAS1 domain is not required for the "locked-on" phenotype of mutations in the PAS2 domain.

conformation of the GHKL domain is therefore important for the constitutive inhibition of NifA activity by the "locked-on" PAS2 mutants. This implies that the "locked-on" PAS2 mutants are similar to the wild type in their requirement for nucleotide binding to the GHKL domain in order to inhibit NifA activity. In contrast, substitutions in the H domain of NifL that give rise to a "locked-on" phenotype exhibit a decreased requirement for nucleotide binding *in vitro* and are dominant to the G480A substitution *in vivo* (Martinez-Argudo et al., 2004a). Thus, these data highlight differences between the "locked-on" PAS2 mutants studied in this thesis and previously identified "locked-on" mutants in other domains of the NifL protein.

3.3.2 The PAS1 domain is not required for the "locked-on" phenotype

As mentioned in sections 1.2.3 and 1.4.3, the PAS1 domain of NifL is required for the inhibition of NifA activity in response to oxygen (Söderbäck et al., 1998). In order to investigate whether the phenotype of the "locked-on" PAS2 mutants is influenced by the redox sensing PAS1 domain, the V166M and L199R substitutions were introduced to truncated forms of the NifL protein that lack the PAS1 domain. The various truncated NifL proteins and variant forms of the truncated proteins containing "locked-on" substitutions in the PAS2 domain were assayed for their ability to inhibit transcriptional activation by NifA (Table 3.2). As expected, forms of the NifL protein that lack the PAS1 domain ($NifL_{(143-519)}$ and $NifL_{(147-519)}$) did not inhibit NifA in response to oxygen, but responded normally to fixed nitrogen (Table 3.2, rows 4 and 5) and NifL-V166M and NifL-L199R inhibited NifA activity under all four conditions tested (Table 3.2, rows 6 and 7). However, when the V166M or L199R substitution was present in the truncated proteins, constitutive inhibition of NifA activity was retained (Table 3.2, rows 8 and 9). These data suggest that PAS1 is not required for inhibition of NifA by the "locked-on" PAS2 mutants. This conclusion is

supported by further experiments demonstrating that when the V166M substitution is combined with a secondary substitution in the PAS1 domain (E70A, see section 1.2.3) that blocks redox sensing (resulting in a form of NifL that does not inhibit NifA activity under oxidising conditions), the "locked-on" phenotype of the PAS2 mutant is dominant (data not shown). Taken together, the available information clearly shows that the "locked-on" phenotype of the PAS2 mutations is independent of the PAS1 domain. This does not eliminate the possibility that the PAS2 and PAS1 domains interact during signalling, rather, it indicates that the PAS2 mutants interfere with signal relay downstream of PAS1.

3.4 Discussion

In order to understand the role of the PAS2 domain of NifL in signal transduction, this domain was extensively mutagenised using site-directed and random approaches. All mutant NifL proteins were tested for their ability to inhibit NifA activity in response to redox and fixed nitrogen signals *in vivo*. The data presented in this Chapter suggest that the PAS2 domain can exist in at least two discrete signalling states, as exemplified by mutations that stabilise NifL in either the "on" (inhibitory) or the "off" (non-inhibitory) conformation. The "locked-on" mutations in PAS2 result in a form of NifL that is competent to inhibit NifA, irrespective of the redox state of the FAD co-factor in the PAS1 domain. In contrast, the "redox signalling" mutants apparently fail to communicate the redox state of PAS1 to the C-terminal domains of NifL, but remain responsive to the fixed nitrogen signal. Two NifL variants of this class were identified, NifL-I153A and NifL-F253L. I153 is likely to be positioned in the A'α helix of PAS2 and structural modelling of the PAS2 domain suggests that the F253 may point outwards from the central β-sheet. However, the location of these residues does not yield an obvious explanation for the "redox-signalling" phenotypes observed. "Redox signalling" mutations may act by

stabilising/mimicking the "off" conformation of the PAS2 domain or they may simply disrupt PAS2 function resulting in defunct relay of redox signals in NifL. Whatever the mechanism, these results suggest an important role for PAS2 in redox signal relay from PAS1 to influence the interaction of the C-terminal domains of NifL with NifA.

In addition to the "locked-on" and "redox signalling" mutants, a third class of mutation in the PAS2 domain was identified (called "aerobically inactive" mutants). These mutations result in impairment of the redox response and the aerobic fixed nitrogen response. This lends further credence to the idea that PAS2 can influence the conformation of the C-terminal domains of NifL in a signal dependant manner. PAS2 deletion experiments demonstrate that inhibition of NifA in response to redox signals requires a functional PAS2 domain whereas nitrogen sensing does not. Therefore, the ability of the GHKL domain of NifL to bind GlnK and the subsequent conformational changes in the H and GHKL domains that promote NifA inhibition do not strictly require the PAS2 domain. However, NifL variants in which the PAS2 domain is absent or is not functional (as is likely to be the case for the "aerobically inactive" mutants) often exhibit an impaired ability to inhibit NifA activity, suggesting that the PAS2 domain may stabilise the C-terminal domains in a conformation that is optimal for inhibition of NifA.

Substitutions that give rise to a "locked-on" phenotype are distributed throughout the PAS2 domain (Figure 3.1). However, structural predictions indicate that four of the seven "locked-on" substitutions identified (V157A, V166M, L175A and L262A) are located in a dimerisation interface that includes the A'α helix and extends throughout the central β-sheet in many PAS domains of known structure (Figure 3.1). The V166M variant was obtained via random mutagenesis of PAS2 whilst the V157A, L175A and L262A variants were subsequently generated using the site-directed approach. V166, L175 and L262 correspond to residues in the conserved β-sheet interface and V157 is likely to be

located in the A'α helix (Figure 3.1). The "locked-on" phenotype of substitutions at these positions implies a connection between the quaternary structure of the PAS2 domain and the signalling state of NifL. Given the importance of PAS2 in transduction of the redox signal, it is possible that the "locked-on" variants identified here simulate the oxidised (inhibitory) conformer of the wild-type NifL protein, particularly as nucleotide binding to the GHKL domain is required for inhibition of NifA in both cases. However, the evidence for these hypotheses is highly speculative and a full biochemical analysis is required. Thus, biochemical experiments on the NifL PAS2 domain are the focus of the next chapter in this thesis.

Chapter 4 - Oligomerisation states of the PAS2 domain of NifL

4.1 Introduction

At the time of writing, the structures of 36 PAS domains from prokaryotic organisms had been deposited in the protein data bank (PDB). Of these, 27 form homo-dimers in the crystal structure. PAS domains contain multiple surfaces that can mediate interaction between subunits and, as a result, they can pack together to form dimers in several different ways (Möglich et al., 2009b). Some PAS domains have even been shown to adopt multiple quaternary structures within a single crystal lattice (Ayers and Moffat, 2008; Lee et al., 2008). Despite this plasticity with respect to quaternary arrangement, the residues that comprise the dimer interface are found in conserved structural locations in most PAS domains. Specifically, this dimerisation interface involves the central β-sheet, which can pack against either the central β-sheet of the opposite protomer or its flanking helices. However, a conserved patch of hydrophobic residues on the outer surface of the β-sheet provides inter-subunit contacts in both scenarios (Möglich et al., 2009b). A common structural arrangement of prokaryotic PAS dimers involves the N-terminal α-helix (the A'α helix) that flanks the PAS core (see Chapter 1). In these structures, the N-terminal helices (one from each subunit) associate to form α-helical bundles and pack against the central β-sheet of the opposing protomer (Key et al., 2007a; Key and Moffat, 2005; Kurokawa et al., 2004; Lee et al., 2008; Ma et al., 2008; Park et al., 2004; Verger et al., 2007). In this quaternary arrangement, residues at conserved positions in both the A'α helix and the central β-sheet mediate dimerisation (Möglich et al., 2009b). As mentioned in section 3.4, seven substitutions were identified in the PAS2 domain of NifL that lock the protein in the active/inhibitory conformation. Structural predictions indicated that four of these substitutions are positioned in regions of the central β-sheet and the A'α helix that contribute to the conserved dimerisation interface discussed above. Therefore, the

oligomerisation state of the isolated PAS2 domain and the influence of various substitutions on oligomerisation were investigated.

4.2 Effect of substitutions on the quaternary structure of the PAS2 domain

4.2.1 Bacterial adenylate cyclase two-hybrid analysis of oligomerisation of the PAS2 domain

The bacterial adenylate cyclase two-hybrid (BACTH) system is commonly used to analyse interactions between proteins or protein domains. The Adenylate cyclase (AC) enzyme contains two discretely folded domains and the BACTH system relies on functional complementation between AC subunits expressed *in trans* (in an AC deficient *E. coli* reporter strain). Each subunit is transcriptionally fused to a protein of interest to create two hybrid proteins. Interaction between the proteins of interest results in co-localisation of the AC subunits, enabling functional complementation. Thus, protein-protein interactions result in a substantial increase in AC activity (Karimova et al., 1998; Karimova et al., 2000). The AC enzyme catalyses the conversion of ATP to cAMP. Production of the second messenger, cAMP, triggers transcription of various reporter genes, including *lacZ*. This enables detection of protein-protein interactions via assays of β-galactosidase activity.

The BACTH system was used to investigate interactions between subunits of the isolated PAS2 domain of NifL (Figure 4.1). A strong interaction was detected between subunits of the wild-type PAS2 domain (Figure 4.1, bars marked “WT”), suggesting that the domain is multimeric. This interaction was perturbed in variant forms of the PAS2 domain containing “locked-on” substitutions; each of the seven “locked-on” substitutions indentified by the mutagenic studies described in Chapter 3 was tested and, in every case, the interaction between PAS2 protomers was impaired (Figure 4.1, compare

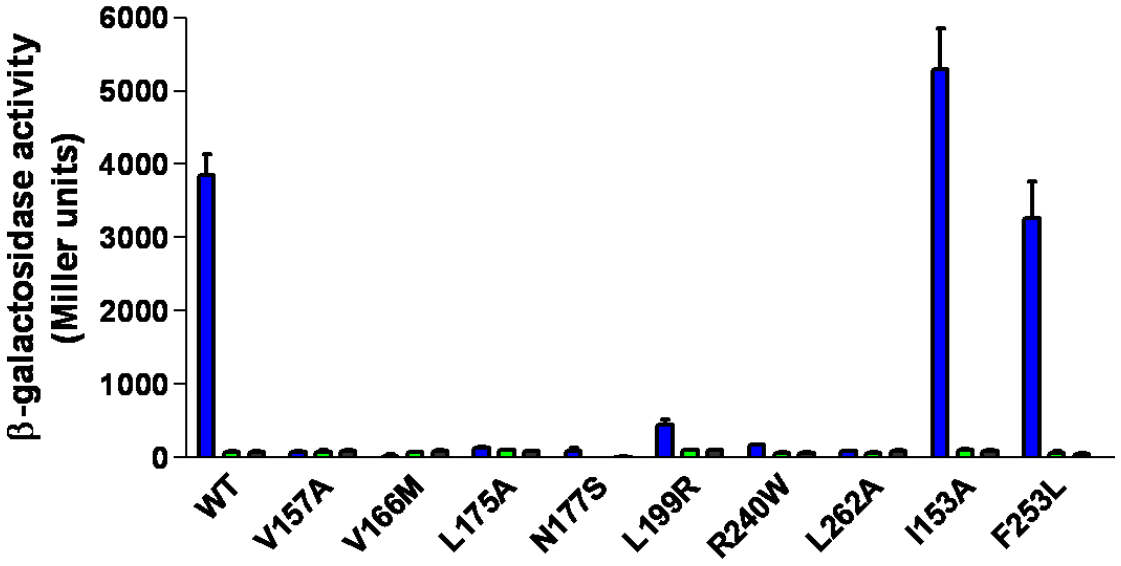


Figure 4.1. Bacterial adenylate cyclase two-hybrid (BACTH) analysis of PAS2 oligomerisation. Hybrid proteins containing the T18 or T25 subunit of adenylate cyclase fused to the NifL PAS2 domain ($NifL_{(147-284)}$) or variants of the PAS2 domain were constructed and expressed as described in sections 2.6.2 and 2.9.2. Data for each NifL variant is shown as a block of 3 bars; the blue bars represent interactions between fusion proteins while the green and black bars are controls in which the T18-PAS2 fusion protein is expressed with the T25 subunit only (green bars) or the T25-PAS2 fusion protein is expressed with the T18 subunit only (black bars). Graphs show an average of at least two replicates and error bars indicate the standard error of the mean.

blue bars marked “WT” with those marked “V157A”, “V166M”, “L175A”, “N177S”, “L199R”, “R240W” and “L262A”). By contrast, variant forms of the PAS2 domain containing the “redox signalling” substitutions (I153A and F253L) were competent to maintain the interaction. Overall, this data demonstrates a correlation between the “locked-on” phenotype and perturbation of the interaction between PAS2 subunits, suggesting that the “locked-on” substitutions may influence NifL activity by disrupting oligomerisation of the PAS2 domain.

4.2.2 Biochemical analysis of oligomerisation of the PAS2 domain

In order to conduct biochemical investigations into the oligomeric state of the PAS2 domain, protein preparations of high purity and concentration were needed. To this end, a DNA fragment encoding NifL residues 143 - 284 (or its mutant derivatives) was cloned into the plasmid pETM11 to create a hexahistidine tagged PAS2 fusion protein for overexpression using the pET expression system. This system utilises the evolved ability of T7 bacteriophage to produce high levels of viral protein in host cells and requires *E. coli* strains such as BL21 (DE3) which contain a chromosomal copy of the gene for T7 RNA polymerase. The presence of a *lac* operator site upstream of the T7 promoter (and on the pET plasmids) allows IPTG (or lactose) inducible expression of target genes (Studier et al., 1990). The His-tagged PAS2 domain was overexpressed and purified by nickel affinity chromatography (see section 2.7). Two variants were selected from each phenotypic class (i.e. two “locked-on” variants and two “redox signalling” variants) and the oligomeric state of the wild-type and variant proteins was analysed using size exclusion chromatography, dynamic light scattering, chemical cross-linking and analytical ultracentrifugation.

(i) Size exclusion chromatography

Size exclusion chromatography (SEC) is a technique commonly used to separate macromolecules on the basis of their hydrodynamic volume. Briefly, the molecules are passed through a column containing a porous matrix. Smaller molecules can penetrate the pores whilst larger molecules cannot. Thus, the smaller molecules are exposed to a greater percentage of the total column volume (i.e. they take a longer route through the matrix) than the larger molecules. As a result, elution of the macromolecules from the column is dependent on their size (and shape). This enables analysis of the molecular weight and oligomeric state of a purified protein sample via comparison with protein standards of known molecular weight.

Size exclusion chromatography was used to analyse the oligomeric state of the PAS2 domain and its mutants (Figure 4.2). When loaded onto the column at various concentrations, the wild-type PAS2 domain eluted in a single peak with an apparent molecular mass somewhat lower than that predicted for a spherical dimeric species (38.56 kDa). Retention volumes were clearly concentration dependent within the range 26 - 519 µM (Figure 4.2A, chromatogram marked "WT") with apparent molecular weights ranging from 32.3 kDa to 37.2 kDa (Table 4.1). The concentration dependence of the elution profile suggests rapid inter-conversion between the monomeric and dimeric forms during the timescale of chromatography. Variant PAS2 domains containing the F253L substitution, which gives rise to a "redox signalling" phenotype in the full-length protein, also eluted in a concentration dependent manner (Figure 4.2A, chromatogram marked "F253L"). However, the elution volumes and apparent molecular weights observed where shifted slightly (relative to wild-type PAS2) towards the value expected for a dimer. The apparent molecular weight of samples injected at concentrations of 26 - 519 µM ranged

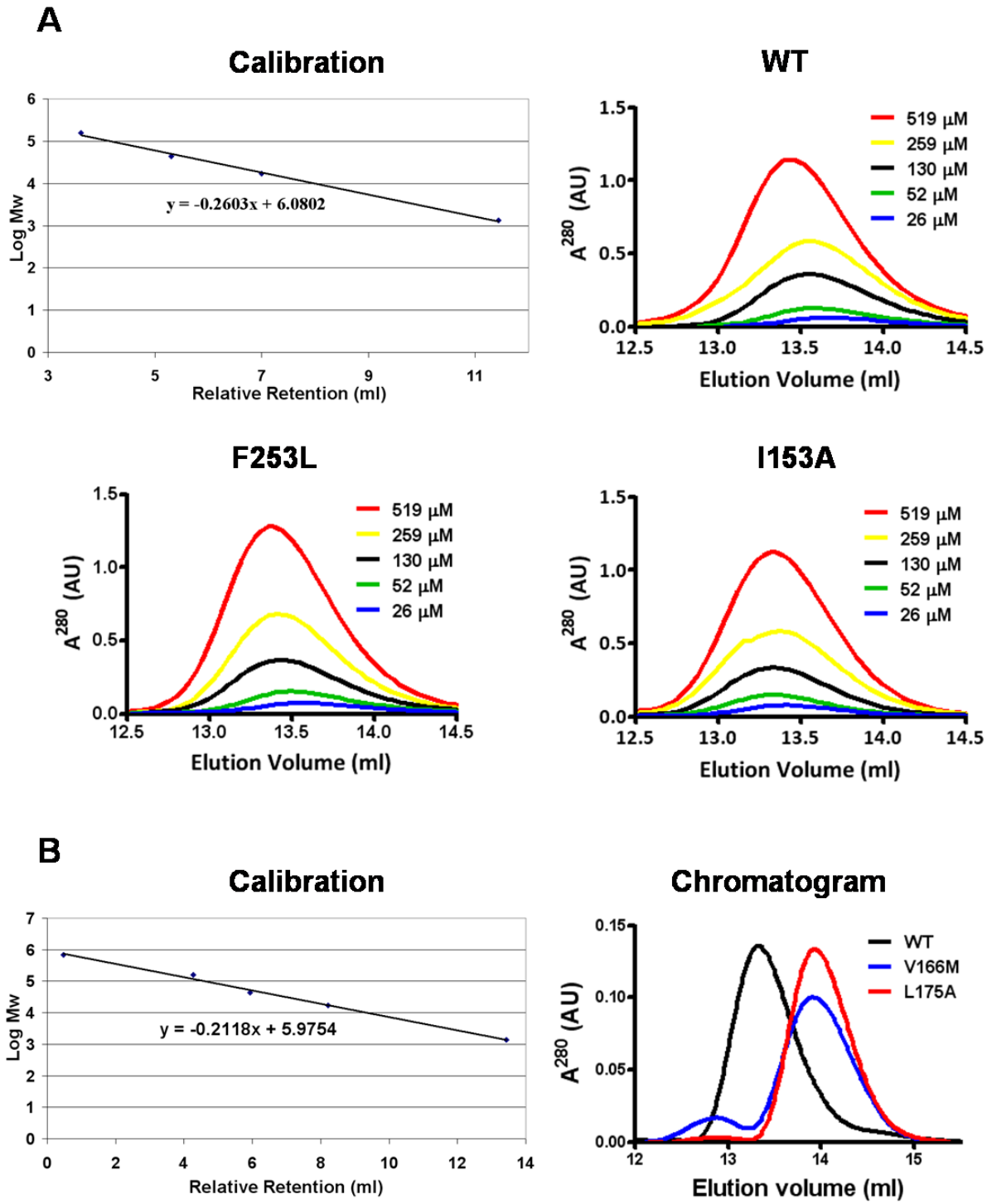


Figure 4.2. Analysis of PAS2 dimerisation by size exclusion chromatography. (A) Calibration and elution profiles for the wild-type PAS2 domain (chromatogram labelled "WT") and the "redox signalling" variants PAS2-I153A (chromatogram labelled "I153A") and PAS2-F253L (chromatogram labelled "F253L") when injected at five different concentrations as indicated in the legend. (B) Calibration and elution profiles of the wild-type PAS2 domain (black line on the chromatogram) and the "locked-on" variants PAS2-V166M (blue line) and PAS2-L175A (red line) injected at a concentration of 104µM. The elution volumes and apparent molecular weights of all species in this figure are tabulated in Table 4.1.

Experiment	Protein & concentration	Elution Volume (ml)	Apparent Mw (kDa)
A	WT 26 μM	13.68	32.3
	WT 52 μM	13.58	34.2
	WT 130 μM	13.56	34.7
	WT 259 μM	13.55	34.9
	WT 519 μM	13.44	37.2
B	I153A 26 μM	13.41	37.7
	I153A 52 μM	13.33	39.6
	I153A 130 μM	13.33	39.6
	I153A 259 μM	13.36	39.1
	I153A 519 μM	13.33	39.6
C	F253L 26 μM	13.56	34.7
	F253L 52 μM	13.5	35.8
	F253L 130 μM	13.44	37.2
	F253L 259 μM	13.42	37.6
	F253L 519 μM	13.38	38.6
D[a]	WT 104 μM	13.30	33.4
	WT 519 μM	13.18	39.3
	L175A 104 μM	13.84	23.9
	L175A 519 μM	13.58	31.0
	V166M 104 μM (major peak)	13.90	23.1
	V166M 104 μM (minor peak)	12.83	48.5

[a] Note that the data shown in part D were obtained separately and using a different calibration to that shown in the other sections.

Table 4.1. Size exclusion chromatography of the NifL PAS2 domain ($NifL_{(143-284)}$) and variant PAS2 domains.

from 34.7 kDa to 38.6 kDa, compared to the spread of 32.3 kDa to 37.2 kDa observed for the wild-type PAS2 domain (Table 4.1, compare experiments A and C). However, this relatively small difference may not be significant given the resolution of SEC; the data may reflect a small change in shape caused by the F253L substitution rather then altered stability of the PAS2 dimer. The second "redox signalling" variant tested, PAS2-I153A, also eluted as a dimer on gel filtration (Figure 4.2A, chromatogram marked "I153A"), but the profile was less concentration dependent than the wild-type PAS2 domain and above 26 μM, the retention volume remained constant with an apparent molecular weight of 39.6 kDa (Table 4.1, experiment B). This suggests that the I153A substitution may shift the monomer-dimer equilibrium towards the dimeric state.

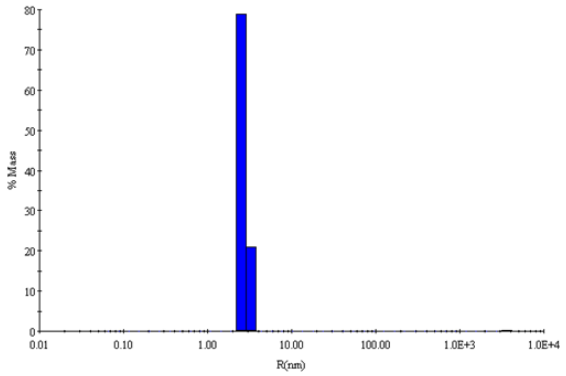
In contrast to the behaviour of wild-type PAS2 and the "redox signalling" mutants, the variant form of the PAS2 domain containing the L175A substitution, which gives rise to a "locked-on" phenotype in the full-length protein, eluted with an apparent molecular weight of 23.9 kDa (when injected at 104 μM), similar to that expected for a monomer (Figure 4.2B, red line). The wild-type PAS2 domain eluted as a species of 33.4 kDa when injected at the same concentration (Table 4.1, experiment D and Figure 4.2B, black line). The elution profile of PAS2-L175A was concentration dependent but did not approach that of the dimeric form when injected at a higher concentration. The apparent molecular weight of PAS2-L175A shifted from 23.9 kDa when injected at 104 μM to 31 kDa when injected at 519 μM, compared to a shift from 33.4 kDa to 39.3 kDa for the wild-type domain when injected at the same concentrations (Table 4.1, experiment D). The second "locked-on" variant tested, PAS2-V166M, sieved as a mixture of two oligomeric species of 23.1 kDa and 48.5 kDa (Figure 4.2B, blue line), which are likely to represent the monomeric and dimeric forms respectively of the variant PAS2 domain. Overall, these data suggest that the monomer-dimer equilibrium is shifted towards the monomeric state in the

"locked-on" variants, consistent with the observation that the subunits of these mutant PAS2 domains fail to interact in the bacterial two-hybrid system.

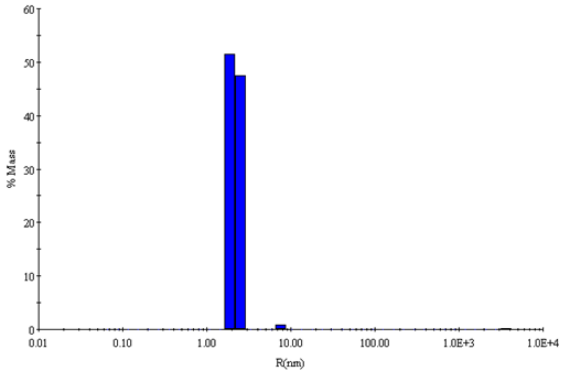
(ii) Dynamic light scattering

Dynamic light scattering (DLS), also known as photon correlation spectroscopy, is often used to analyse the homogeneity and the approximate size of macromolecules in solution. This technique involves shining a beam of polarised light through a sample solution and measuring scattering of the light upon collision with the solute. The solute particles are in Brownian motion and the speed of Brownian motion is dependent on the size of the particles (larger particles move more slowly than smaller ones). Scattering of the monochromatic light upon contact with the dissolved macromolecules (which are assumed to be spherical) is dependent on the speed of Brownian motion and thus DLS can be used to approximate the hydrodynamic radius, molecular weight and size distribution of the macromolecules.

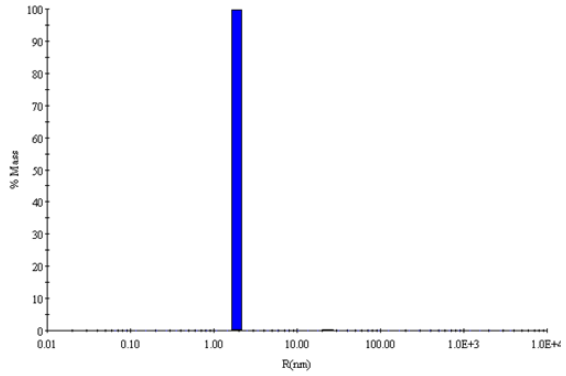
DLS was used to investigate oligomerisation of the NifL PAS2 domain and its mutants (Figure 4.3). The DLS analysis demonstrated a high level of purity in all protein samples as a single species contributed 98.9 - 99.9% of the total mass of each sample (Figure 4.3, inset). Analysis of the wild-type PAS2 domain at a concentration of 104 μM indicated a single species with a hydrodynamic radius of 2.8 nm, which corresponds to a molecular weight of approximately 37 kDa, accounting for 99.8% of the sample mass. This species is likely to represent the PAS2 dimer. Variant forms of the PAS2 domain containing the "redox signalling" substitutions I153A and F253L appeared fully dimeric at the same concentration, with 99.9% of the sample mass forming a single species in both cases (Figure 4.3, inset). PAS2-I153A and PAS2-F253L had hydrodynamic radii of 3.3 nm and 3.1 nm respectively. Given that both the "redox signalling" variants and the wild-type



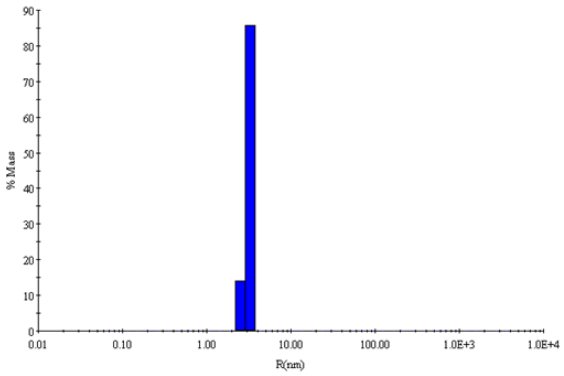
WT



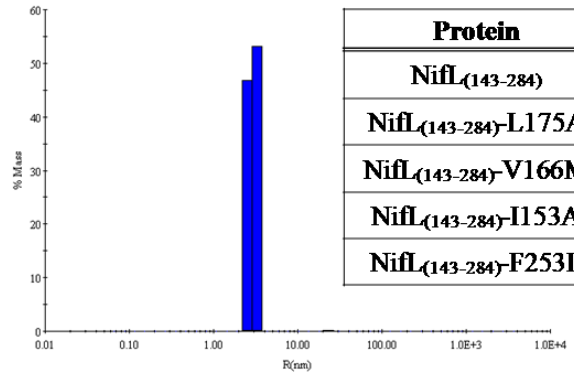
L175A



V166M



I153A



F253L

Protein	R (nm)	%Pd	MW-R (kDa)	%Mass
$NifL_{(143-284)}$	2.8	13.7	37	99.8
$NifL_{(143-284)}$-L175A	2.3	12	24	98.9
$NifL_{(143-284)}$-V166M	1.9	0	15	99.7
$NifL_{(143-284)}$-I153A	3.3	6.3	54	99.9
$NifL_{(143-284)}$-F253L	3.1	11.7	48	99.9

Figure 4.3. Dynamic light scattering of the NifL PAS2 domain and selected PAS2 variants. Graphs show the percentage mass of the sample (%Mass) on the y-axis versus the hydrodynamic radius (R) on the x-axis. Data is given for the wild-type PAS2 domain (labelled "WT"), the "locked-on" variants L175A and V166M and the "redox signalling" variants I153A and F253L. All protein samples were analysed as described in section 2.7.7 at a concentration of 104 μM. Data for the major peak in each sample is tabulated in the inset (%Pd = % polydiversity and MW-R = molecular weight).

domain appeared to be fully dimeric at the concentration tested, the increased hydrodynamic radius of the variant domains may indicate a less compact conformation.

The "locked-on" variants PAS2-L175A and PAS2-V166M both exhibited a reduced hydrodynamic radius and molecular weight compared to the wild-type domain (Figure 4.3, inset). Approximately 98.9% of the total mass of the PAS2-L175A sample formed a species with a hydrodynamic radius of 2.3 nm and a molecular weight of 24 kDa. Data for PAS2-V166M indicated that 99.7% of the sample mass was consistent with spherical particles with a radius of 1.9 nm and a molecular weight of roughly 15 kDa. These results suggest that, at a concentration of 104 μM, the L175A and V166M substitutions both cause an almost complete dissociation of PAS2 subunits. Overall, the DLS data concur with results obtained from the SEC and BACTH analyses.

(iii) Chemical cross-linking

Chemical cross-linking involves the use of a chemical reagent to covalently link two or more macromolecules. This process has many biological applications including the fixation of biological samples and the study of protein-protein or protein-DNA interactions. There are many chemical reagents (known as cross-linking reagents) available for covalent linkage of biological molecules and these reagents can have differing specificities and modes of action. Commonly used cross-linking reagents include formaldehyde, which is able to form protein-protein and protein-nucleic acid cross-links, and glutaraldehyde, which efficiently cross-links amine groups (Baumert and Fasold, 1989). In addition to its application in the study of interactions between macromolecules and complex formation, chemical cross-linking can also be used to investigate oligomerisation and protomer interactions in multi-subunit proteins. Purified protein

samples can be exposed to a cross-linking reagent and the formation of covalent links between subunits can be determined.

To investigate the subunit stoichiometry of the isolated PAS2 domain and its variants, protein samples were chemically cross-linked with glutaraldehyde as described in section 2.7.8 and the products were analysed by SDS-PAGE (Figure 4.4). The amount of protein in each band was quantified by densitometry and the relative amounts of monomer and dimer were calculated as a percentage of the total protein in each lane. These experiments indicated that the isolated PAS2 domain and all the variants tested can be cross-linked in the dimeric form (Figure 4.4). No cross-linked species corresponding to higher order oligomers could be detected by SDS-PAGE. It was evident from the appearance of duplicate bands upon addition of the cross-linking reagent that, in addition to the inter-molecular cross-links, glutaraldehyde was also catalysing the formation of intra-molecular cross-links. However, for the purposes of quantitation, the protein densities of these duplicate bands were pooled to give the total density of protein in either the monomeric or dimeric state. The "locked-on" PAS2 variants exhibited a reduction in the percentage of cross-linked protein, suggesting a shift in the monomer-dimer equilibrium towards the monomeric form. The V166M and L175A substitutions caused a reduction in the percentage of the cross-linked (dimeric) species from 64% in the wild type to 33% and 40% respectively in the "locked-on" variants (Figure 4.4, compare lanes marked "WT", "V166M" and "L175A"). The PAS2-I153A variant showed increased cross-linking (78% cross-linked) relative to the wild-type domain, implying that this substitution may shift the equilibrium towards the dimeric state. However, the F253L variant did not differ greatly from the wild-type domain (60% of the total protein was cross-linked). This difference between the I153A and F253L substitutions, in terms of their influence on PAS2

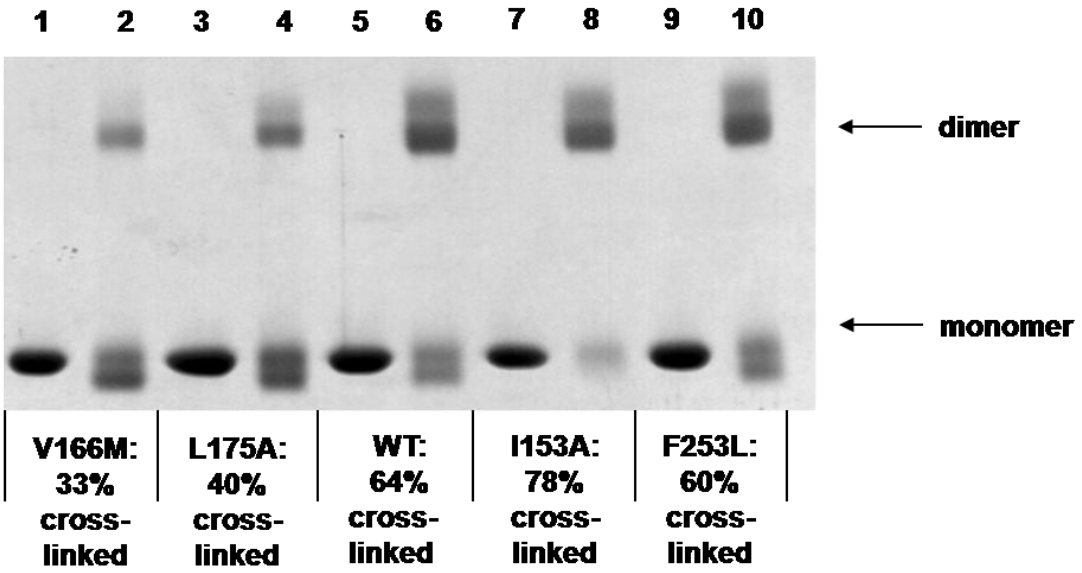


Figure 4.4. Chemical cross-linking of the PAS2 domain ($NifL_{(143-284)}$) and the V166M, L175A, I153A and F253L variant domains. Protein samples (52 μM, based on a monomer) were cross-linked with glutaraldehyde as described in Chapter 2.7.8 and analysed by SDS-PAGE. Lanes are grouped into pairs whereby the cross-linking reactions (even numbered lanes) are shown adjacent to controls where glutaraldehyde was absent from the reaction mixture (odd numbered lanes). Each band was quantified using SynGene densitometry software. For each reaction the amount of dimeric (cross-linked) protein is shown as a percentage of the total protein in the lane.

dimerisation, conforms to the trends identified by SEC. Overall, the results from the chemical cross-linking experiments are in concurrence with results from SEC, DLS and BACTH analysis; the data obtained using all of these techniques indicate that the "locked-on" variants disrupt PAS2 dimerisation while the "redox signalling" variants do not.

(iv) Analytical ultracentrifugation

Analytical ultracentrifugation (AUC) is a technique commonly used to analyse the molecular mass (and thus the association state) of macromolecules in solution. The analytical ultracentrifuge contains an optical detection system that allows the user to observe the distribution of sample concentration over time, upon application of a centrifugal force. The migration of macromolecules through the centrifugal field is proportional to their molecular mass and AUC is broadly considered to be the "gold standard" for molecular weight determination. The experiments performed using the analytical ultracentrifuge fall into two categories: sedimentation velocity and sedimentation equilibrium experiments. Sedimentation velocity experiments involve the application of a large centrifugal force, sufficient to generate relatively rapid sedimentation of the sample. The rate at which this sedimenting "band" of solute moves through the cell can be measured and used to determine the sedimentation co-efficient of the sample. The sedimentation co-efficient provides information regarding the mass and shape of macromolecules in solution. Sedimentation equilibrium experiments require the application of a slightly weaker centrifugal force. In these experiments, sedimentation of the sample results in increasing solute concentration towards the bottom of the cell but sedimentation is antagonised by the force of diffusion "pushing" against the concentration gradient. When the system reaches equilibrium, there is no net movement of the sample over time (i.e. the solute distribution is constant), and these opposing forces are balanced.

Spectroscopic determination of the sample concentration at different points in the equilibrated cell enable precise determination of the molecular weight and, if applicable, the change in molecular weight as a function of solute concentration. Thus, AUC is a valuable tool for analysing the kinetics of subunit association in multimeric proteins and protein complexes.

AUC was used to investigate the dynamic equilibrium between the monomeric and dimeric forms of the NifL PAS2 domain (Figure 4.5). Sedimentation equilibrium profiles of PAS2 indicated that the solution molecular mass varied between 24 kDa and 35 kDa over a concentration range of 10 - 100 μM. In contrast, equilibrium profiles of PAS2-L175A showed a variation of 22 - 28 kDa over a concentration range of 7 - 70 μM. Plotting the sedimentation equilibrium profiles in terms of log absorbance versus radius^2/2 is expected to give a straight line, where the gradient of the line is proportional to the molecular mass of the protein in solution (Horan et al., 1995). The difference in apparent molecular mass of the two forms can be observed in Figure 4.5, where the larger gradient of the data corresponding to the wild-type protein demonstrates a shift towards the dimeric species at the higher protein concentration (Figure 4.5A, triangles), in contrast to the predominance of the monomeric form at the lower protein concentration (Figure 4.5B, triangles). The sedimentation data for each species fitted best to a monomer-dimer model with a dissociation constant (K_d) of 34 μM for the wild type, and 120 μM for the L175A variant. Thus, substitution of L175 for alanine results in a 3.5-fold reduction in the affinity between PAS2 subunits. This reflects the difference between the L175A variant and wild-type forms of the PAS2 domain observed in the BACTH, SEC, DLS and chemical cross-linking experiments. It is noteworthy that the dissociation constants derived from the AUC analysis correlate quite precisely with the results from DLS; at the protein concentration of 104 μM used in the DLS experiments we would expect the wild-type PAS2 domain to be

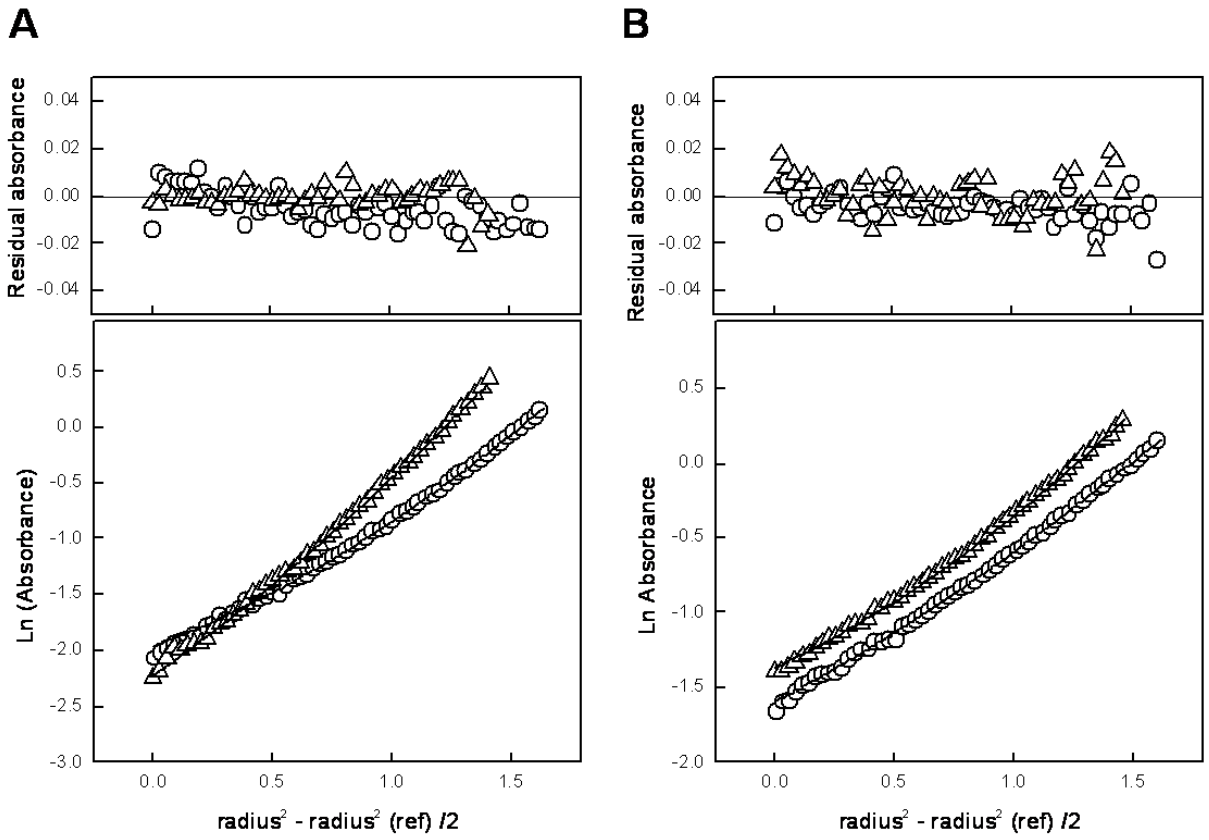


Figure 4.5. Analytical ultracentrifugation analysis of the NifL PAS2 domain. Sedimentation equilibrium profiles of the wild-type PAS2 domain (triangles) and the "locked-on" variant PAS2-L175A (circles) at a rotor speed of 23,000 rpm. *Lower panels*: (A) 100 μM PAS2 and 70 μM PAS2-L175A measured at 275 nm. (B) 10 μM PAS2 and 7 μM PAS2-L175A measured at 230 nm. The lines represent a fit to both data sets for each sample using a 20 kDa monomer-dimer equilibrium model and K_d values of 34 and 120 μM for the wild type domain and L175A variant respectively. *Upper panels*: residual absorbance between the experimental data and the fitted lines. This Figure was kindly provided by Dr. Thomas A. Clarke, UEA.

predominantly dimeric and the L175A variant to be predominantly monomeric based on dissociation constants of 34 μM and 120 μM respectively. The congruence between these techniques, in combination with results form BACTH, SEC and chemical cross-linking experiments, strongly supports the assertion that the "locked-on" PAS2 substitutions act by disrupting dimerisation of the PAS2 domain.

4.3 Substitutions in the PAS2 domain do not influence the overall oligomerisation state of NifL

Given that the "locked-on" substitutions apparently alter the quaternary structure of the isolated PAS2 domain, it was questioned whether these substitutions influence oligomerisation of the full length NifL protein. In other words, it is important to establish whether the PAS2 domain is an oligomerisation determinant of NifL or whether the oligomerisation state of this domain is important for relaying structural signals between domains. Assessing this experimentally is complicated by the presence of two additional oligomerisation interfaces in the NifL protein, since the PAS1 and H domains both contain dimerisation surfaces. The crystal structure of the NifL PAS1 domain indicates that this domain is dimeric (Key et al., 2007a) and the H domain is predicted to form a coiled-coil structure homologous to the dimerisation interface of the histidine protein kinases (Little et al., 2007). Therefore, the oligomeric states of the "locked-on" PAS2 variants were analysed in the context of the full length NifL protein as well as various truncated forms of the protein containing different domain combinations. SEC and BACTH analysis were used to investigate the influence of "locked-on" substitutions on the oligomerisation of these constructs.

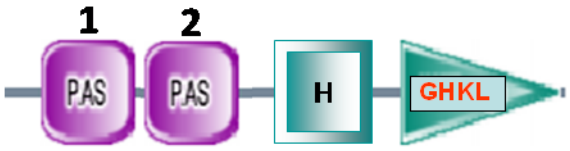
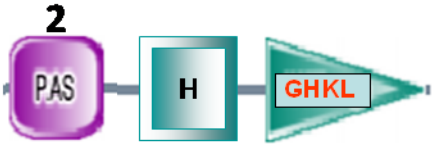
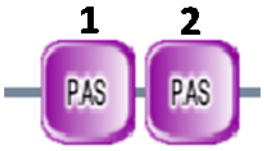
Construct	Domain Architecture
NifL	
$NifL_{(143-519)}$	
$NifL_{(1-284)}$	

Figure 4.6. Domain architectures of the three NifL constructs for SEC analysis.

4.3.1 Chromatographic analysis of NifL domain combinations

SEC was used to analyse the oligomeric state of three different NifL domain combinations: (i) constructs containing only the two N-terminal PAS domains (lacking the H and GHKL domains), (ii) constructs containing the PAS2, H and GHKL domains (lacking the PAS1 domain) and (iii) the full length NifL protein (the domain architectures of each of these constructs are illustrated in Figure 4.6). For each domain combination, the behaviour of the "locked-on" variant V166M on SEC was compared to that of the wild-type protein (Table 4.2). The V166M substitution had no effect on oligomerisation in any of the constructs tested. The full length proteins, NifL and NifL-V166M, both eluted as trimers with apparent molecular weights of 197.7 kDa and 196.7 kDa respectively (Table 4.2, rows 1 and 2). Similarly, when only the PAS2, H and GHKL domains were present (i.e. when the dimerisation interface in the PAS1 domain was absent), the V166M substitution did not influence oligomerisation; the behaviour of $NifL_{(143-519)}$ and $NifL_{(143-519)}$-V166M on SEC were consistent with spherical particles of 108.1 kDa and 110.4 kDa respectively (Table 4.2, rows 3 and 4). These molecular weights are closer to that predicted for a dimer (90.6 kDa on the basis of sequence) then to that predicted for a trimer (135.9 kDa). $NifL_{(1-284)}$, which contains only the two N-terminal PAS domains (also referred to as the "PAS1-PAS2 fragment") and lacks the predicted dimerisation interface present in the H domain, eluted as a trimer with an apparent molecular mass of 96.3 kDa (Table 4.2, row 5). The variant proteins $NifL_{(1-284)}$-I153A (which belongs to the "redox signalling" class of mutants) and $NifL_{(1-284)}$-V166M ("locked-on" class) also sieved as trimers with apparent molecular masses of 92.5 kDa and 92.8 kDa respectively (Table 4.2, rows 6 and 7). Thus, the presence of the "locked-on" V166M substitution does not appear to influence oligomerisation of the PAS1-PAS2 fragment of NifL. Although most of the constructs tested appear trimeric, the true association state is likely to be dimer or tetramer

Protein Construct	Domains Present	Apparent Mw (kDa)	Expected Monomeric Mw (kDa)	Apparent Oligomerisation state
NifL	PAS1, PAS2, H, GHKL	197.7	61.1	Trimer
NifL-V166M	PAS1, PAS2, H, GHKL	196.7	61.1	Trimer
$NifL_{(143-519)}$	PAS2, H, GHKL	108.1	45.3	Dimer
$NifL_{(143-519)}$-V166M	PAS2, H, GHKL	110.4	45.3	Dimer
$NifL_{(1-284)}$	PAS1, PAS2	96.3	35.0	Trimer
$NifL_{(1-284)}$-V166M	PAS1, PAS2	92.5	35.0	Trimer
$NifL_{(1-284)}$-I153A	PAS1, PAS2	92.8	35.0	Trimer

Table 4.2. Size exclusion chromatography of NifL domain combinations.

(Söderbäck et al., 1998) and elution of this these proteins from the SEC column may be aberrant, since sedimentation velocity experiments on the PAS1-PAS2 fragment suggest that this protein is dimeric, irrespective of its redox state (Little and Dixon, unpublished data). Overall, the SEC experiments clearly demonstrate that in the presence of either the PAS1 dimerisation surface or the predicted interface in the H domain (or a combination of both), substitutions that disrupt dimerisation of the isolated PAS2 domain no longer influence the association state. That is, the PAS2 dimerisation interface, which is apparently disrupted by the "locked-on" substitutions, is not important for oligomerisation in the full length NifL protein.

4.3.2 BACTH analysis of oligomerisation of the PAS1-PAS2 fragment

In order to corroborate the results obtained by SEC analysis of the PAS1-PAS2 fragment, a second independent technique was used to analyse oligomerisation of the tandem PAS domains. Using the bacterial two-hybrid system, self-association of the isolated PAS1 and PAS2 domains was detected as anticipated (Figure 4.7, bars marked "PAS1" and "PAS2"). Oligomerisation of the longer construct in which both domains are present in tandem was also detectable (Figure 4.7, bars marked "PAS1, PAS2"). As shown previously (in section 4.2.1), the "locked-on" substitutions V157A and V166M disrupt oligomerisation of the isolated PAS2 domain (Figure 4.7, bars marked "PAS2", "PAS2-V157A" and "PAS2-V166M"). However, when the PAS1 domain was also present, the effect of these substitutions on oligomerisation was nullified and the variant proteins showed the same level of interaction as the wild-type PAS1-PAS2 fragment (Figure 4.7, compare bars marked "PAS1, PAS2", "PAS1, PAS2-V157A" and "PAS1, PAS2-V166M"). This demonstrates that the dimerisation interface present in PAS1 (Ayers and Moffat, 2008; Key et al., 2007a) is sufficient to maintain oligomerisation of the PAS1-

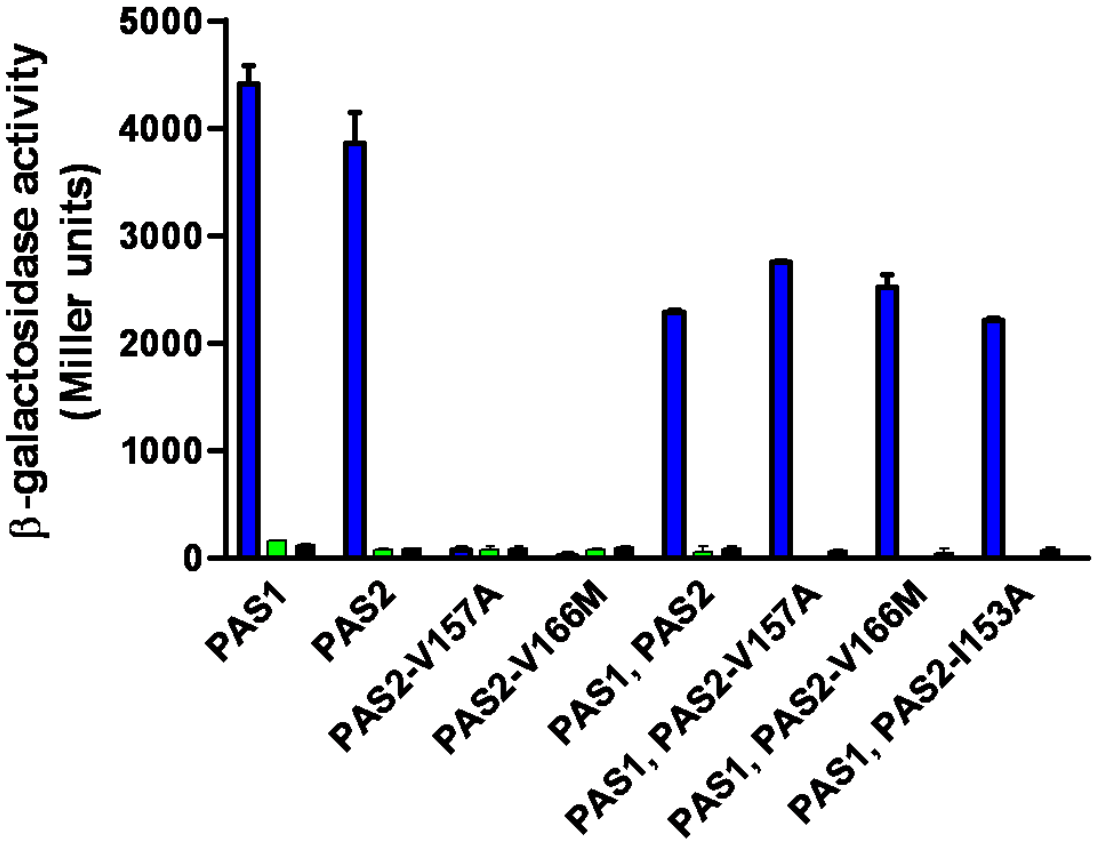


Figure 4.7. BACTH analysis of the influence of substitutions in the PAS2 domain on oligomerisation of the PAS1-PAS2 fragment of NifL. Hybrid proteins containing the PAS1 ($NifL_{(1-146)}$), PAS2 ($NifL_{(147-284)}$) and PAS1-PAS2 ($NifL_{(1-284)}$) fragments of NifL fused to either subunit of adenylate cyclase were constructed and expressed as described in Chapter 2. The graph legend is as in Figure 4.1.

PAS2 construct when PAS2 dimerisation is impaired. The BACTH analysis was performed under oxygen-limiting conditions *in vivo*, unlike the SEC experiments, which were carried out under aerobic conditions. As the PAS1 domain was sufficient to maintain oligomerisation of the PAS1-PAS2 fragment in both experiments, it appears that the association state of this protein is maintained irrespective of redox status (and regardless of the integrity of the PAS2 dimerisation interface). The PAS2 dimerisation interface, therefore, is not important in maintaining the oligomeric state of NifL but instead seems to function in intra-molecular signal transduction.

4.4 Discussion

Taken together, data from BACTH, SEC, DLS, chemical cross-linking and AUC experiments indicate that the isolated PAS2 domain and variant forms of this domain containing "redox signalling" substitutions are dimeric in solution. The data presented in this chapter demonstrate that substitutions in the PAS2 domain that give rise to a "locked-on" phenotype in the full length NifL protein disrupt dimerisation of the isolated PAS2 domain. In other words, when dimerisation of the PAS2 domain is impaired NifL is apparently locked in the inhibitory conformer. This implies that the PAS2 domain is dimeric when NifL adopts the non-inhibitory (or "off") conformer and monomeric when NifL is in the inhibitory (or "on") conformation.

Figure 4.8 shows the positions of all substitutions identified in Chapter 3 on a structural model of the NifL PAS2 domain. The model presented in Figure 4.8 is based on the NifL PAS1 domain (2GJ3) but similar results were obtained using the several other PAS structures as templates for modelling. Residues I153 and F253, which give rise to a "redox signalling" phenotype when substituted for alanine and leucine respectively, are located on opposite ends of the molecule. The position of these residues does not provide

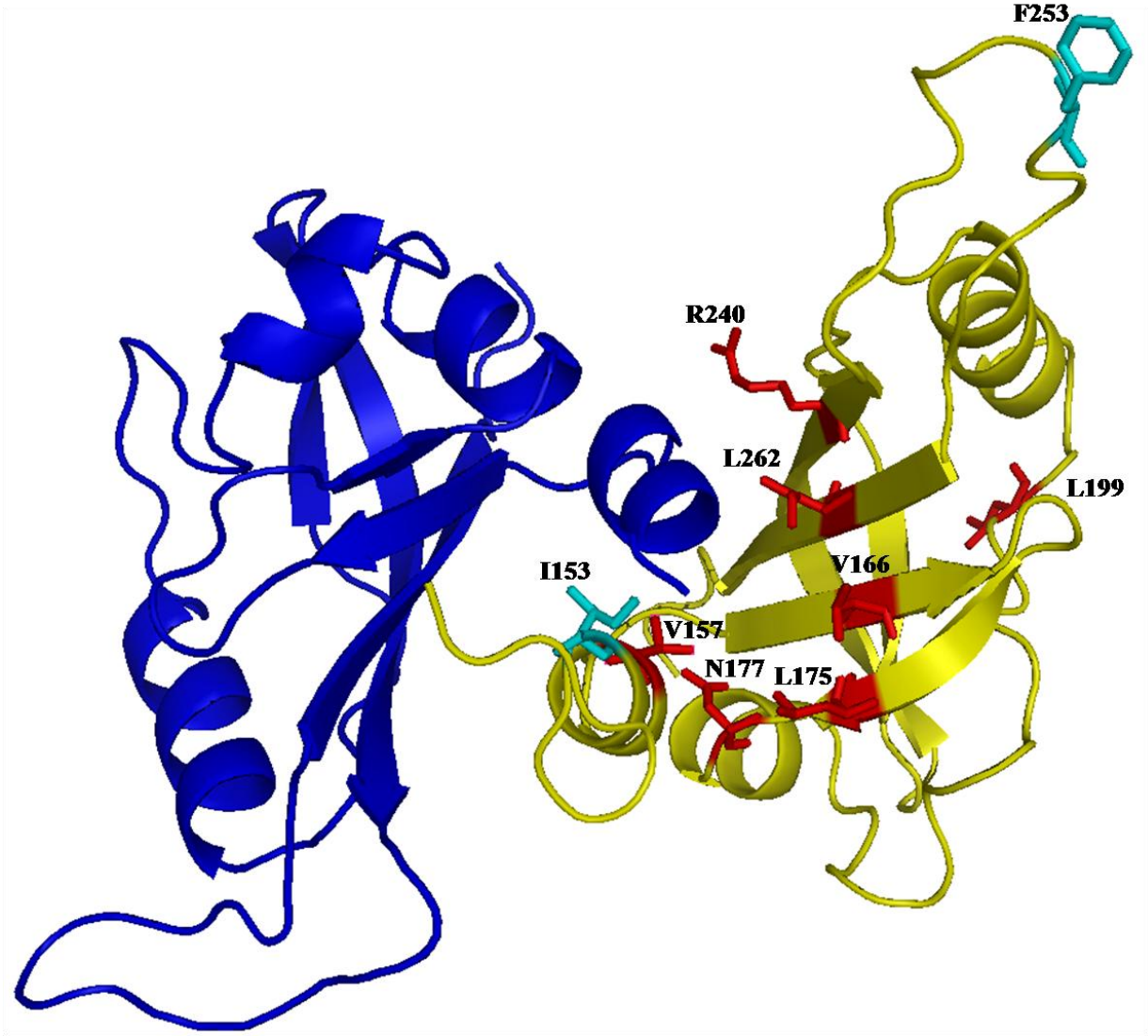


Figure 4.8. Structural model of the dimeric NifL PAS2 domain. The ribbon diagram shows the NifL PAS2 domain modelled on the NifL PAS1 domain. Similar results can be obtained by modelling on other PAS structures (e.g. *Ec*DOS PASA). Subunits are coloured blue and yellow and amino acids substituted in Chapter 3 are shown as sticks on the yellow subunit; residues that give rise to a "locked-on" phenotype when substituted are coloured red and residues that give rise to a "redox signalling" phenotype when substituted are coloured cyan. Note that all "locked-on" substitutions apart from L199 cluster around the putative dimerisation interface.

an obvious explanation of their phenotype. However, SEC and chemical cross-linking analyses suggest that the I153A substitution may stabilise dimerisation of the isolated PAS2 domain and the close proximity of this residue to the putative dimerisation interface in the structural model of the PAS2 domain is consistent with this experimental evidence. F253 is oriented outward from the globular domain and is located in a loop connecting two β-strands in the central β-sheet (Hβ and Iβ). It is possible that this residue forms a contact between PAS2 and another domain of NifL, especially given that SEC and chemical cross-linking experiments indicate that the F253L substitution does not significantly influence dimerisation of the isolated PAS2 domain. That is, F253 may be involved in inter-domain communication in NifL. As mentioned in section 3.4, structural predictions indicate that at least four of the seven "locked-on" substitutions identified are positioned in a dimerisation interface that is conserved in many PAS domains and are therefore likely to directly disrupt interaction between PAS2 protomers (Figure 4.8). One of the remaining three "locked-on" substitutions (N177S), whilst not located in this conserved interface, is positioned close to it and is likely to directly influence the stability of the PAS2 dimer. By contrast, one of the remaining two "locked-on" substitutions, L199R, is distal to the putative dimerisation interface and is therefore unlikely to disrupt subunit interactions directly (Figure 4.8) (the position of the other substitution, R240W, cannot be predicted with a high level of confidence). Despite this, the L199R substitution impairs the interaction between PAS2 subunits as measured by the BACTH system. These seemingly contradictory data can be reconciled if we hypothesise that a global change in conformation accompanies switching of the PAS2 domain from the dimeric "off" state to the monomeric "on" state. According to this scenario, amino acid changes at positions that are remote to the dimerisation interface (such as L199 and possibly R240) could favour the "on" conformation and thereby influence the interaction between PAS2 subunits. Tenuous support for this

assertion can be derived from the BACTH data; in contrast to the other "locked-on" substitutions which completely eliminate interaction between subunits of the PAS2 domain, L199R exhibits a low level of interaction (Figure 4.1). These results are consistent with (but do not demonstrate) an indirect influence of the L199R substitution on PAS2 oligomerisation compared to a direct effect exerted by the other substitutions. Whatever the mechanism, it is clear that changes in the quaternary structure of the PAS2 domain are important for signal transduction in NifL.

Further SEC and BACTH experiments demonstrate that, despite its importance in signalling, the PAS2 dimerisation interface is not an oligomerisation determinant for the full length NifL protein. Hence, changes in the association state of the PAS2 domain do not control the assembly of NifL subunits but instead facilitate switching between alternative quaternary arrangements. As these alternative arrangements represent the inhibitory and non-inhibitory signalling states, it is likely that switching between them is responsive to environmental cues. Given that the PAS2 domain itself does not appear to have a role in sensing, we may speculate that the signalling state of the PAS2 domain (and thus its association state) is responsive to signal perception by other domains. Taken together, the above postulation and the importance of the PAS2 domain in relaying redox signals from the PAS1 domain to the C-terminal domains of NifL (see Chapter 3) imply that PAS2 may be sensitive to signal perception by PAS1. To investigate this possibility further it is necessary to probe the signal dependent conformational changes that occur in the N-terminal domains of NifL and investigate the influence of substitutions in the PAS2 domain on these changes. These issues are the focus of the next Chapter of this thesis.

Chapter 5 - Redox signal relay between the NifL PAS domains

5.1 Introduction

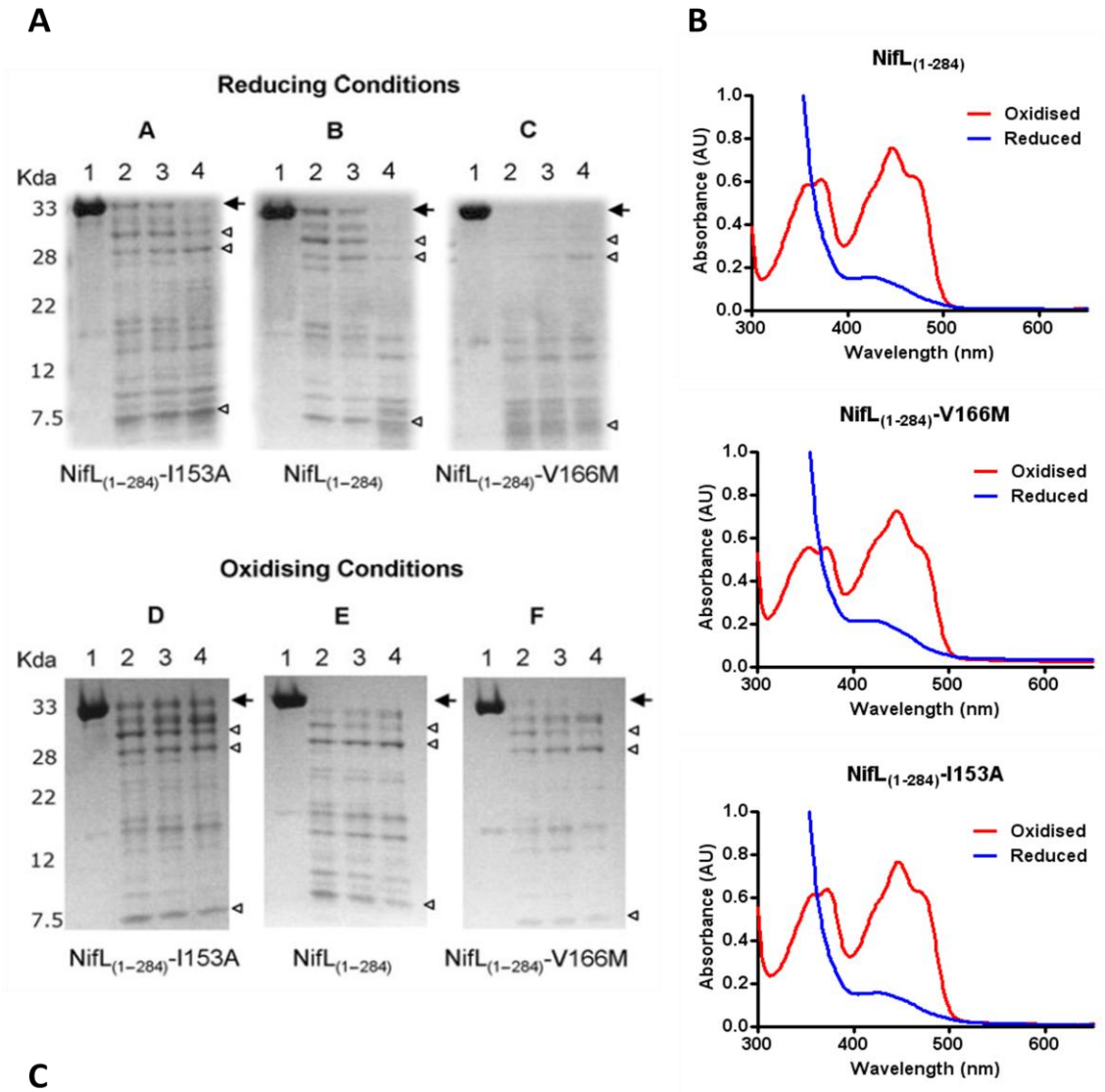
The data presented in the previous two Chapters imply that the PAS2 domain of NifL undergoes a change in quaternary structure in response to the perception of redox signals by the PAS1 domain. That is, the signalling state of the PAS1 domain appears to be communicated to the PAS2 domain, resulting in the movement of PAS2 subunits relative to one another. It appears that the "locked-on" substitutions in the PAS2 domain lock the NifL protein in the oxidised conformer by impairing PAS2 dimerisation. In order to verify these hypotheses, it is necessary to examine the influence of redox signals on the quaternary structure of the PAS2 domain in NifL constructs in which the PAS1 domain is also present. As the association state of the PAS1 domain remains constant irrespective of redox conditions, it is not possible to achieve this by studying changes in oligomerisation. Therefore, it was necessary to assess the quaternary structure of the PAS2 domain indirectly, via analysis of the redox dependent conformational changes that occur in NifL and the influence of substitutions in PAS2 on these changes. Several experimental techniques were employed to this end, including limited proteolysis, BACTH analysis and cysteine cross-linking. Further, the inter-domain region linking the NifL PAS domains was mutagenised in order to investigate the transmission of redox signals between PAS1 and PAS2. Probing conformational changes in NifL also enables comparison between the oxidised wild-type protein and the "locked-on" variant proteins. Thus, such experiments can provide evidence to support or counter the assertion that the "locked-on" substitutions in the PAS2 domain lock NifL in the oxidised conformation.

5.2 Analysis of conformational changes in NifL using limited proteolysis

Limited proteolysis is a technique commonly used to examine conformational change in proteins. When a protein is exposed to a small amount of peptidase, the pattern and rate of proteolytic digestion depend upon the accessibility of the various cleavage sites within the protein. Hence, changes in the conformation of a protein will alter the rate and pattern of its digestion (also known as its proteolytic footprint). Peptidases such as trypsin and chymotrypsin are routinely used for proteolytic footprinting to probe conformational changes in proteins.

5.2.1 Redox dependent conformational changes in the N-terminal PAS domains of NifL

The analysis of the variant proteins and domains presented thus far suggests that a large conformational change, involving a shift in the quaternary structure of the PAS2 domain, accompanies redox signal transduction in NifL. It was desirable to investigate this conformational change in a wild-type context and check for congruence with the findings obtained using variant proteins. To this end, limited chymotrypsin proteolysis was used to analyse the conformation of the PAS1-PAS2 fragment of NifL (Figure 4.6, $NifL_{(1-284)}$) and its variants under oxidising and reducing conditions *in vitro* (Figure 5.1A). These experiments were carried out under anaerobic conditions in a glove box using sodium dithionite to reduce the FAD co-factor in PAS1 (Hill et al., 1996) where appropriate (see section 2.7.12). Protein samples were incubated with chymotrypsin for time periods of 0, 2, 5 and 10 minutes and the progress of the proteolysis reaction at each time point was analysed by SDS-PAGE (Figure 5.1A). When the FAD co-factor was reduced with sodium dithionite, the wild-type protein fragment, $NifL_{(1-284)}$, showed a similar rate and pattern of digestion to the "redox signalling" variant, $NifL_{(1-284)}$-I153A (Figure 5.1A, compare panels A and B). In both cases the undigested protein (indicated by arrows to the right of each



C

Protein	Absorbance at 446 nm	FAD concentration (μM)	Number of FAD molecules per NifL dimer
$NifL_{(1-284)}$	0.755	61.7	1.23
$NifL_{(1-284)}$-V166M	0.726	59.2	1.18
$NifL_{(1-284)}$-I153A	0.765	62.5	1.25

Figure 5.1. Limited chymotrypsin proteolysis and spectroscopic analysis of the PAS1-PAS2 fragment of NifL. (A) $NifL_{(1-284)}$ and two variants were digested with chymotrypsin under oxidising conditions and after reduction with dithionite as described in section 2.7.12. The progress of the proteolysis reaction after 0, 2, 5 and 10 minutes (Lanes 1, 2, 3 and 4 respectively) was analysed by SDS-PAGE. Arrows indicate the uncleaved protein and empty arrow-heads mark putative cleavage products. Data shown is representative of at least three independent replicates. (B) Spectroscopic analysis of NifL fragments before and after reduction. The spectroscopic data were used to calculate the FAD concentration and thus the FAD incorporation for each sample (as described in section 2.7.11) and the results are tabulated in part C.

panel in Figure 5.1A) persisted after a 5 minute incubation with protease (Figure 5.1A, compare lane 3 of panels A and B) but was no longer present after 10 minutes incubation (Figure 5.1A, lane 4 in panels A and B). In contrast, a different proteolysis pattern was observed with $NifL_{(1-284)}$-V166M which was largely digested within 2 minutes of exposure to chymotrypsin (Figure 5.1A, the undigested band is absent from lanes 2, 3 and 4 in panel C). This suggests that, under reducing conditions, $NifL_{(1-284)}$ adopts a similar conformation to $NifL_{(1-284)}$-I153A, whilst differing in conformation to $NifL_{(1-284)}$-V166M. However, when the FAD co-factor was oxidised, there was a change in the proteolytic footprint of $NifL_{(1-284)}$ (Figure 5.1A, the undigested band is present in lanes 2 and 3 of panel B but absent in the same lanes of panel E). In contrast to the digestion pattern observed under reducing conditions, the proteolytic footprint of the oxidised $NifL_{(1-284)}$ protein closely resembled that of $NifL_{(1-284)}$-V166M (Figure 5.1A, compare panels E and F). Similar results were obtained from limited proteolysis analysis of a second "locked-on" variant, $NifL_{(1-284)}$-L175A (data not shown). For both $NifL_{(1-284)}$ and the $NifL_{(1-284)}$-V166M variant, the band corresponding to the uncleaved protein was digested within 2 minutes of addition of the protease under oxidising conditions (Figure 5.1A, lanes 2, 3 and 4 of panels E and F). The shift in the proteolysis pattern of wild-type $NifL_{(1-284)}$, when comparing the results obtained under oxidising and reducing conditions, implies a redox dependent change in protein conformation (Figure 5.1A, compare panels B and E). Moreover, the proteolytic footprint of $NifL_{(1-284)}$ resembles that of the "redox signalling" variant under reducing conditions and that of the "locked-on" variant protein under oxidising conditions. A similar redox dependent conformational change was not evident in either of the variant proteins (Figure 5.1A, compare panels A and D or panels C and F). As an additional control, the oxidation state of the FAD group was monitored spectroscopically to ensure that the dithionite concentration used was sufficient to fully reduce the co-factor in all of the

protein constructs (Figure 5.1B). In each case, protein samples were fully reduced after addition of dithionite, as determined by quenching of the spectral features characteristic of the oxidised flavin group (peaks at 360 nm and 445 nm and shoulders at 420 nm and 470 nm). To eliminate the possibility that the PAS2 substitutions influence the incorporation of FAD into PAS1, the FAD content of each construct was determined (Figure 5.1C). All proteins exhibited 59 - 63% FAD incorporation, indicating that there were no significant differences in folding between the wild-type and variant proteins. Taken together, these data indicate that oxidation-reduction of the FAD group in PAS1 triggers a shift between two distinct conformations of the PAS1-PAS2 construct and that substitutions in the PAS2 domain cause the protein to favour one of these conformers over the other (regardless of signal perception by the PAS1 domain). It is also worth noting that the rate of digestion of $NifL_{(1\text{-}284)}$ was faster under oxidising conditions than under reducing conditions (Figure 5.1, panels B and E). This is consistent with the hypothesis that PAS2 subunits dissociate in response to oxidation of the PAS1 co-factor, leading to a more open conformation. Overall, the results from limited proteolysis of the PAS1-PAS2 fragment of NifL support the hypothesis that the "locked-on" and "redox signalling" substitutions in the PAS2 domain lock the NifL protein in either the oxidised or reduced conformer.

5.2.2 Conformational changes in longer NifL constructs

In order to discern whether the "locked-on" NifL variants adopt the *bona fide* oxidised conformer or promote inhibition of NifA via some alternative mechanism (as is the case for some H domain variants (Martinez-Argudo et al., 2004a)), it was important to analyse the influence of the "locked-on" substitutions on the conformation of the C-terminal domains of the protein. It has previously been shown that nucleotide binding strongly influences the conformation of the GHKL domain of NifL (see section 1.4.3). The

addition of ADP to the reaction buffer in limited trypsin proteolysis experiments results in a conformational change in the C-terminal domains of the full length NifL protein (Söderbäck et al., 1998) and variant forms of NifL that are unable to bind nucleotide, and/or undergo the shift in conformation that accompanies nucleotide binding, fail to inhibit NifA activity *in vivo* (Perry et al., 2005). Therefore, limited trypsin proteolysis was used to probe conformational changes associated with ADP binding in wild-type NifL and a variant form of the protein containing the "locked-on" substitution, V166M (Figure 5.2). Protein samples were incubated with trypsin as described in section 2.7.12 and aliquots were removed after 0, 2, 5, 10, 20, 30 and 60 minutes of exposure to the protease. Aliquots were added to eppendorf tubes containing a trypsin inhibitor and subsequently analysed by SDS-PAGE. The response of two NifL constructs to nucleotide was analysed; proteolysis reactions were performed using both the full length NifL protein and $NifL_{(143\text{-}519)}$ (which lacks the redox sensing PAS1 domain, see Figure 4.6), either in the absence of nucleotide or in the presence of 2 mM ADP. All reactions were carried out under aerobic conditions, when the FAD co-factor in the PAS1 domain of NifL is assumed to be oxidised (Söderbäck et al., 1998). The proteolytic footprint of the full length NifL protein is shown in Figure 5.2A. The percentage of protein that remained uncleaved at each time point was quantified using Syngene densitometry software and plots of percent undigested NifL protein versus time are shown below the SDS-PAGE analysis in Figure 5.2 as an indication of the rate of proteolytic digestion in each reaction. In the absence of nucleotide, different rates of proteolysis were observed for the wild-type and variant proteins. The band representing the full length protein (indicated by an arrow on the right of each panel in Figure 5.2) is degraded more rapidly when the V166M substitution is present, although proteolysis of both proteins yields similar cleavage products (Figure 5.2A, compare panels A and B). This difference in the rate of digestion is also apparent in

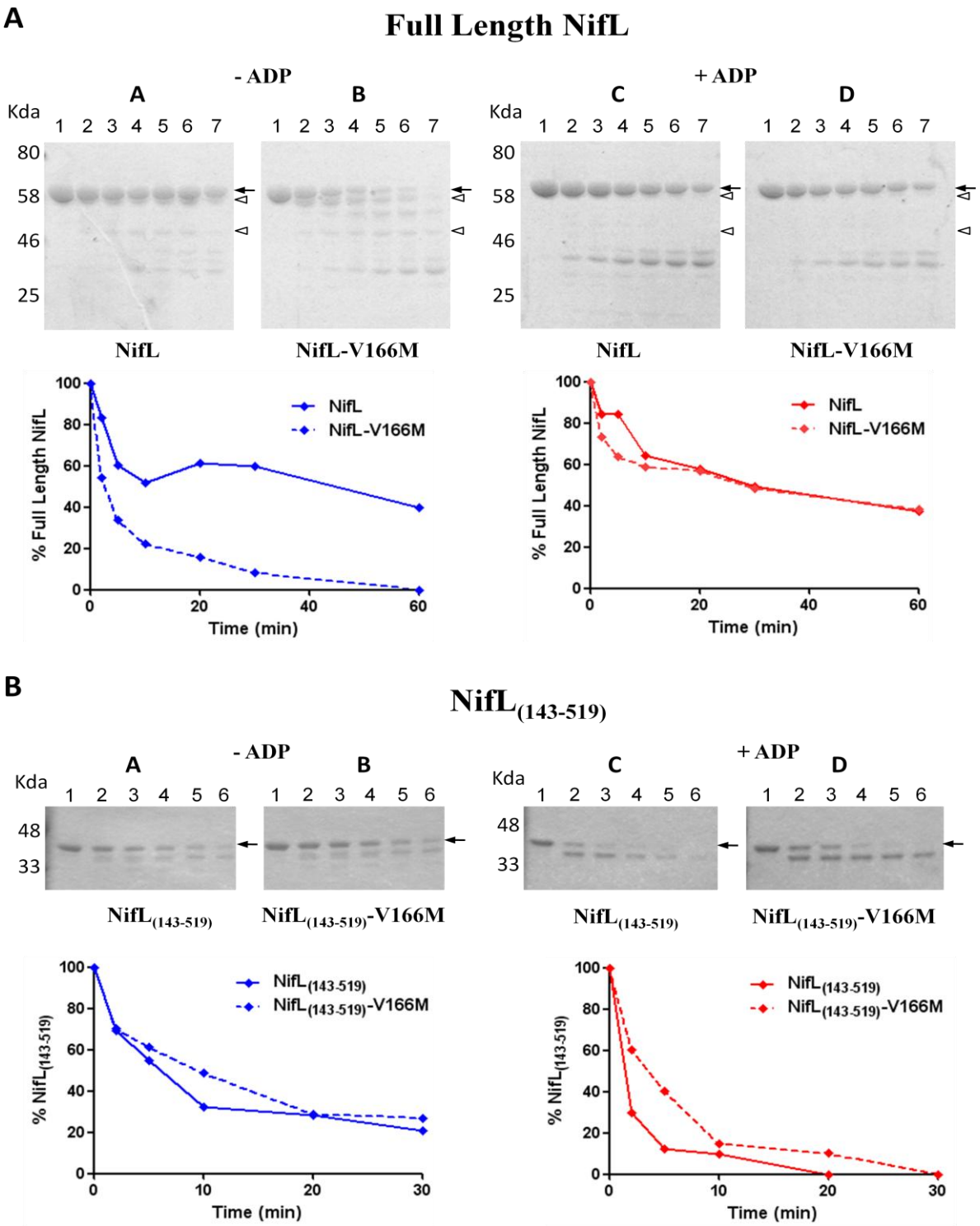


Figure 5.2. Limited trypsin proteolysis of (A) NifL and (B) NifL$_{(143-519)}$. Proteolytic digestion was analysed in either the presence of 2 mM ADP or the absence of nucleotide (as indicated). Samples were removed from the reaction mixture after 0, 2, 5, 10, 20, 30 and 60 minutes (Lanes 1 - 7 respectively) and analysed by SDS-PAGE. Arrows indicate the uncleaved protein and empty arrow-heads indicate putative cleavage products. The rate of proteolysis of each sample was studied using densitometry analysis as described in section 2.7.12 and plots of percent uncleaved protein (relative to t = 0) versus time are shown below the appropriate gel. The data shown is representative of at least two independent replicates.

the densitometry analysis (Figure 5.2A, compare the dashed blue line and the full blue line), suggesting a difference in conformation between NifL and NifL-V166M when nucleotide is absent. However, a nucleotide dependent conformational change is evident in both the wild-type and variant proteins. As shown previously (Perry et al., 2005; Söderbäck et al., 1998), addition of ADP to the reaction buffer results in increased protection of the C-terminal domains from proteolysis in the wild-type NifL protein (Figure 5.2A compare panels A and C). A similar conformational change was observed for the V166M variant protein (Figure 5.2A compare panels B and D). In both cases, two putative cleavage products generated when the proteolysis reaction was performed in the absence of nucleotide (marked by open arrowheads in Figure 5.2A) were not apparent when 2 mM ADP was present in the reaction mixture (Figure 5.2A compare panels A and B with panels C and D). In contrast to results obtained in the absence of nucleotide, the wild-type and variant proteins exhibit a similar pattern and rate of digestion when ADP is present (Figure 5.2A, compare panels C and D). The densitometry analysis indicated that the rate of digestion of NifL and NifL-V166M was similar (Figure 5.2A, compare the dashed and full red lines). These data suggest not only that NifL and NifL-V166M both undergo a nucleotide dependent change in conformation and but also that, in the presence of ADP, the wild-type and "locked-on" variant proteins adopt a similar conformational state.

Limited trypsin proteolysis was also used to analyse the conformation of $NifL_{(143-519)}$ and $NifL_{(143-519)}$-V166M (which lack the redox sensing PAS1 domain) and the response of these proteins to adenosine nucleotides (Figure 5.2B). A similar proteolytic footprint was observed for the wild-type and "locked-on" variant proteins in the absence of nucleotide (Figure 5.2B, compare panels A and B). Densitometry analysis indicated that the V166M substitution did not significantly influence the rate of proteolysis (Figure 5.2B,

dashed and full blue lines). Thus, $NifL_{(143-519)}$ and $NifL_{(143-519)}$-V166M appear to adopt similar conformational states when nucleotide is absent. Addition of ADP to the reaction buffer apparently triggers a conformational change in both proteins, thereby increasing their susceptibility to trypsin proteolysis (Figure 5.2B, compare panels A and B to panels C and D). For example, $NifL_{(143-519)}$ is fully digested after a 20 minute incubation with trypsin in the presence of ADP (Figure 5.2B, panel C, lane 5) whilst a small proportion (~20%) remains undigested after 30 minutes incubation when nucleotide is absent (Figure 5.2B, panel A, lane 6). When compared to $NifL_{(143-519)}$, the V166M variant protein appears more resistant to proteolysis in the presence of ADP (Figure 5.2B, compare panels C and D). This difference is apparent in the densitometry analysis (Figure 5.2B, compare the full and dashed red lines). For example, approximately 60% of the $NifL_{(143-519)}$-V166M protein remains undigested after a 2 minute incubation with trypsin whereas approximately 30% of the $NifL_{(143-519)}$ protein remains undigested after the same time. Overall, limited proteolysis analysis of NifL constructs lacking the PAS1 domain indicates that the wild-type and "locked-on" variant protein fragments adopt a similar conformation in the absence of nucleotide and that ADP binding induces a conformational change in both proteins. However, the resulting ADP-bound conformers of these proteins are not equivalent. That is, the V166M substitution results in an altered conformation of the truncated NifL protein (lacking PAS1) provided ADP is present. For completeness, the limited trypsin proteolysis experiments described in this section were repeated using chymotrypsin and similar results were obtained (data not shown).

Results obtained from the limited proteolysis experiments, particularly those conducted in the presence of nucleotide, correlate well with the available data concerning the behaviour of the various protein fragments and variants *in vivo*. The full length, nucleotide-bound form of NifL exhibits an inhibitory signalling state under oxidising

conditions *in vivo*, as does the V166M variant (Figure 3.2). Thus, we might expect these two forms of the NifL protein to adopt similar conformations when ADP is present and the FAD co-factor is oxidised, as observed in the limited proteolysis experiments. By contrast, the $NifL_{(143-519)}$ and $NifL_{(143-519)}$-V166M proteins adopt different signalling states *in vivo*; truncated forms of the NifL protein, lacking the PAS1 domain, are unable to sense the cellular redox state and fail to inhibit NifA activity under oxidising conditions whereas removal of the PAS1 domain from the V166M variant protein does not alter its "locked-on" phenotype and $NifL_{(143-519)}$-V166M strongly inhibits NifA activity under oxidising conditions (Table 3.2). Based on the *in vivo* data, $NifL_{(143-519)}$ is expected to adopt the non-inhibitory conformer and $NifL_{(143-519)}$-V166M is expected to adopt the inhibitory conformer under the 2 mM ADP condition. Hence, the difference in conformation between the two truncated forms of NifL apparent in the proteolysis experiments reflects known differences in phenotype.

The wild-type and "locked-on" variant forms of the NifL protein both fail to inhibit NifA activity *in vivo* when nucleotide binding is impaired (Table 3.1). Therefore, both proteins are expected to adopt a non-inhibitory conformation in the absence of adenosine nucleotides. However, it is important to note that the conformer adopted in the absence of nucleotide is not equivalent to that adopted by the nucleotide-bound form of NifL under reducing conditions. Thus, despite the data from limited proteolysis clearly indicating that NifL and NifL-V166M are in different conformations when ADP is absent, it is not clear what the physiological relevance of this might be. Interestingly, the difference in conformation between the wild-type protein and the V166M variant in absence of adenosine nucleotides is not apparent in truncated constructs lacking the PAS1 domain. Previous studies have suggested that the PAS1-PAS2 fragment is resistant to cleavage by trypsin (Söderbäck et al., 1998). Taken together, these data imply that the oxidised PAS1

domain exerts a different influence on the conformation of the C-terminal domains in the wild-type protein compared to the V166M variant, provided adenosine nucleotides are absent.

5.3 Influence of signals from PAS1 on the PAS2 dimerisation interface

5.3.1 Cysteine cross-linking analysis

Cysteine cross-linking is a technique routinely used to analyse the tertiary and quaternary structures of proteins (Bass et al., 2007). The side chains of cysteine residues located in close proximity to each other in a folded protein can be oxidised to form a disulphide bridge and the presence of covalent disulphide bonds can then be determined using SDS-PAGE. Thus, cysteine cross-linking analysis has the potential to indentify contacts between pairs of positions in a three-dimensional protein structure. These positional pairs can consist of a single cysteine residue from each subunit in a multimeric protein or sets of two cysteine residues within a single polypeptide chain. In order to perform cysteine cross-linking analysis, it is first necessary to generate a functional "cysteine-free" form of the protein of interest, in which the native cysteines have been removed via site-directed mutagenesis of the coding sequence (known as cysteine replacement mutagenesis). Cysteine residues can then be introduced at positions of interest, informed by structural data or modelling, and contacts between these positions can be indentified.

Using structural models of the PAS2 domain in combination with the results from mutagenic analysis, it was intended to substitute residues in the putative PAS2 dimerisation interface for cysteines and analyse disulphide bond formation between PAS2 subunits in the isolated PAS2 domain and the full length NifL protein. Further, it was intended to examine the influence of the signalling state of the PAS1 domain on disulphide

bond formation between PAS2 subunits. The wild-type NifL protein contains four cysteine residues (C181, C237, C380 and C507). Colleagues in the Dixon laboratory systematically substituted these cysteines for other residues to create a "cysteine-free" form of the NifL protein that is similar to the wild type in its response to oxygen and fixed nitrogen signals (Figure 5.3). The cysteine-free form of the NifL protein, NifL-C181S, C237F, C380S, C507T, will be referred to as $NifL_{(cys\text{-}free)}$ for the rest of this Chapter. Although $NifL_{(cys\text{-}free)}$ allowed greater NifA activity than the wild type under nitrogen fixing conditions (Figure 5.3, blue bars), the variant protein strongly inhibited NifA activity in response to excess fixed nitrogen (Figure 5.3, yellow bars), oxygen (Figure 5.3, red bars) or a combination of both (Figure 5.3, green bars). Western analysis indicated that the "cysteine-free" form of the NifL protein was stable under the four assay conditions (data not shown). In addition to the full length NifL protein, a cysteine-free form of the isolated PAS2 domain ($NifL_{(143\text{-}284)}$-C181S, C237F) was generated. This construct will be referred to as $PAS2_{(cys\text{-}free)}$. Individual cysteine substitutions were then generated at positions 157, 166 and 240 in $PAS2_{(cys\text{-}free)}$ in order to examine inter-subunit disulphide bond formation in the isolated PAS2 dimer. Cysteine cross-liking experiments were then performed as described in section 2.7.9. Variant forms of the PAS2 domain containing either the V157C or the R240C substitution formed disulphide bonds in the presence of 5 μM copper phenanthroline, suggesting that these residues are located in close proximity to their counterparts in the opposite protomer in the assembled PAS2 dimer (Figure 5.4). Alternatively, residues that are surface exposed can be cross-linked when molecules collide in solution. It may also be possible for cysteine substitutions in the dimerisation interface to disrupt PAS2 dimerisation and still form disulphide bridges when PAS2 monomers

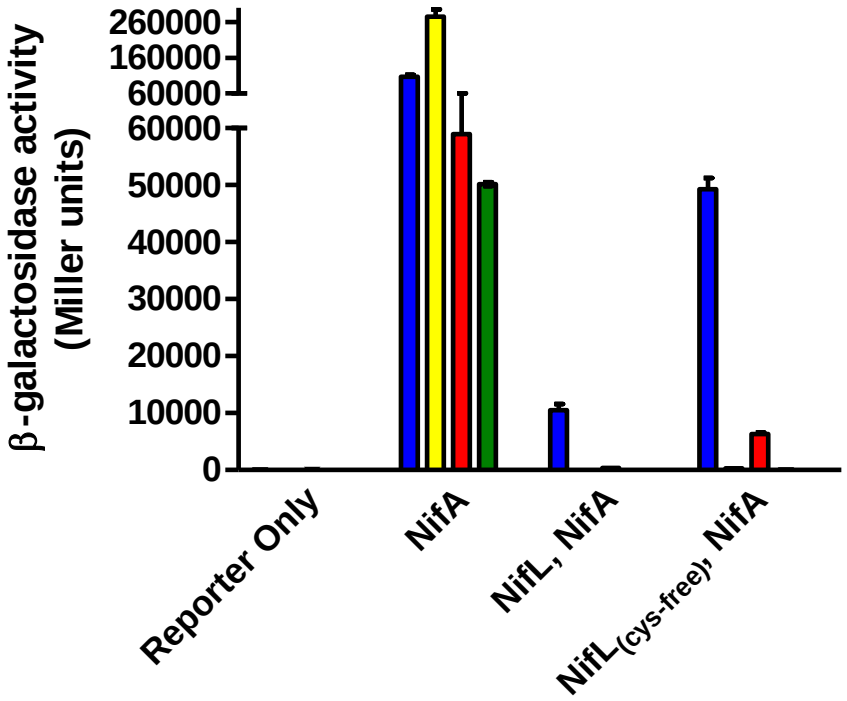


Figure 5.3. Influence of $NifL_{(cys\text{-}free)}$ on NifA activity *in vivo*. Cultures were grown under the following conditions; (1) anaerobically, under nitrogen-limiting conditions (blue bars), (2) anaerobically with excess fixed nitrogen (yellow bars), (3) aerobically with limiting fixed nitrogen (red bars) and (4) aerobically when fixed nitrogen was replete (green bars). Cultures were assayed for β-galactosidase activity as a reporter of NifA-mediated transcriptional activation from a *nifH-lacZ* fusion. All experiments were performed at least in duplicate with error bars denoting the standard error of the mean.

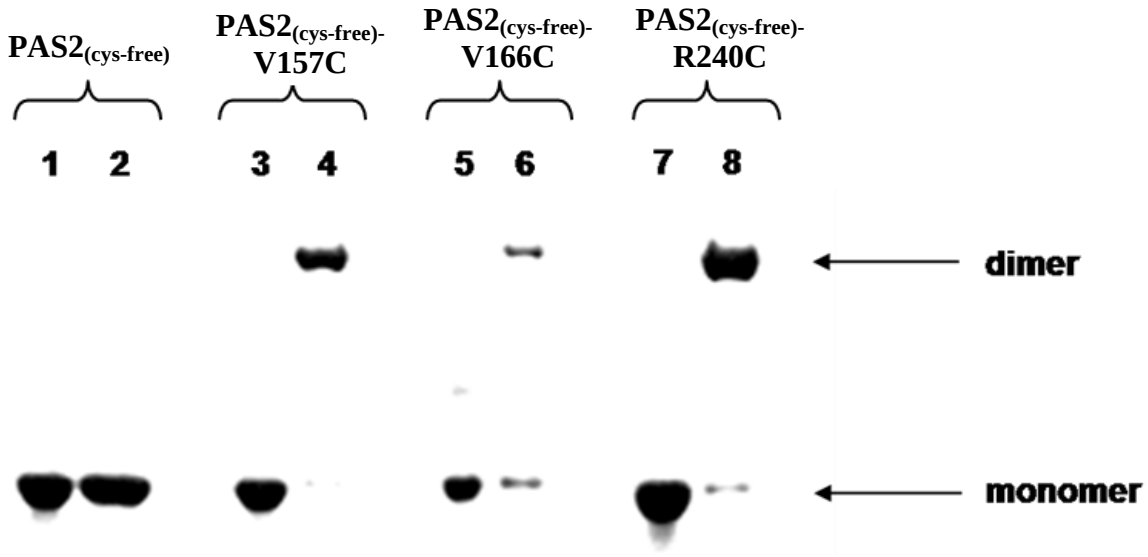


Figure 5.4. Cysteine cross-linking of the PAS2 domain of NifL. A "cysteine-free" variant of the PAS2 domain ($NifL_{(143-284)}$-C181S, C237F) was generated and cysteine residues were then introduced at several positions (157, 166 and 240). Copper phenanthroline catalysed disulphide bridge formation in each variant protein was analysed as described in Chapter 2.7.9. Products from the cross-linking reactions were then added to either reducing SDS-PAGE sample buffer (odd numbered lanes) or non-reducing sample buffer (even numbered lanes) and analysed by electrophoresis. Lanes were loaded as follows: the cysteine-free PAS2 domain in lanes 1 and 2, the V157C variant in lanes 3 and 4, the V166C variant in lanes 5 and 6 and the R240C variant in lanes 7 and 8.

collide. Overall, the results from cysteine cross-linking analysis of the isolated PAS2 domain imply, in congruence with the data presented in Chapter 4, that residues 157 and 240 are positioned in the vicinity of the PAS2 dimerisation interface.

In order to analyse the influence of signals from the PAS1 domain on the cross-linking of PAS2 subunits, it was first necessary to perform phenotypic analysis of the variant proteins. When the V157C, V166C and R240C substitutions were introduced into the full-length $NifL_{(cys\text{-}free)}$ protein and assessed for their ability to inhibit NifA-mediated transcriptional activation *in vivo*, only the V157C substitution had a phenotype similar to the wild-type protein (Figure 5.5). $NifL_{(cys\text{-}free)}$-V157C allowed high NifA activity under nitrogen fixing conditions (Figure 5.5, blue bars) but inhibited NifA in discrete responses to oxygen (Figure 5.5, red bars) and fixed nitrogen (Figure 5.5, yellow bars), or a combination of both (Figure 5.5, green bars). That is, the phenotype of $NifL_{(cys\text{-}free)}$-V157C appeared similar to that of $NifL_{(cys\text{-}free)}$. By contrast, the V166C and R240C substitutions apparently influenced the activity of $NifL_{(cys\text{-}free)}$. $NifL_{(cys\text{-}free)}$-V166C failed to inhibit NifA activity under all conditions tested (Figure 5.5, bars marked "$NifL_{(cys\text{-}free)}$-V166C, NifA"). The null phenotype of V166C may be a consequence of instability of the variant protein, but this possibility was not investigated further. $NifL_{(cys\text{-}free)}$-R240C exhibited a "redox signalling" phenotype and failed to inhibit NifA activity in the presence of excess oxygen but responded normally to fixed nitrogen (Figure 5.5, bars marked "$NifL_{(cys\text{-}free)}$-R240C, NifA"). As $NifL_{(cys\text{-}free)}$-V157C responded to oxygen and fixed nitrogen, this variant was selected for use in further experiments aimed at investigating cysteine cross-linking between PAS2 subunits in the full length NifL protein. As mentioned above, it was intended to examine the influence of the redox signalling state of the PAS1 domain on disulphide bond formation between PAS2 subunits. However, it was not possible to compare oxidising and reducing conditions directly because the presence of a reductant

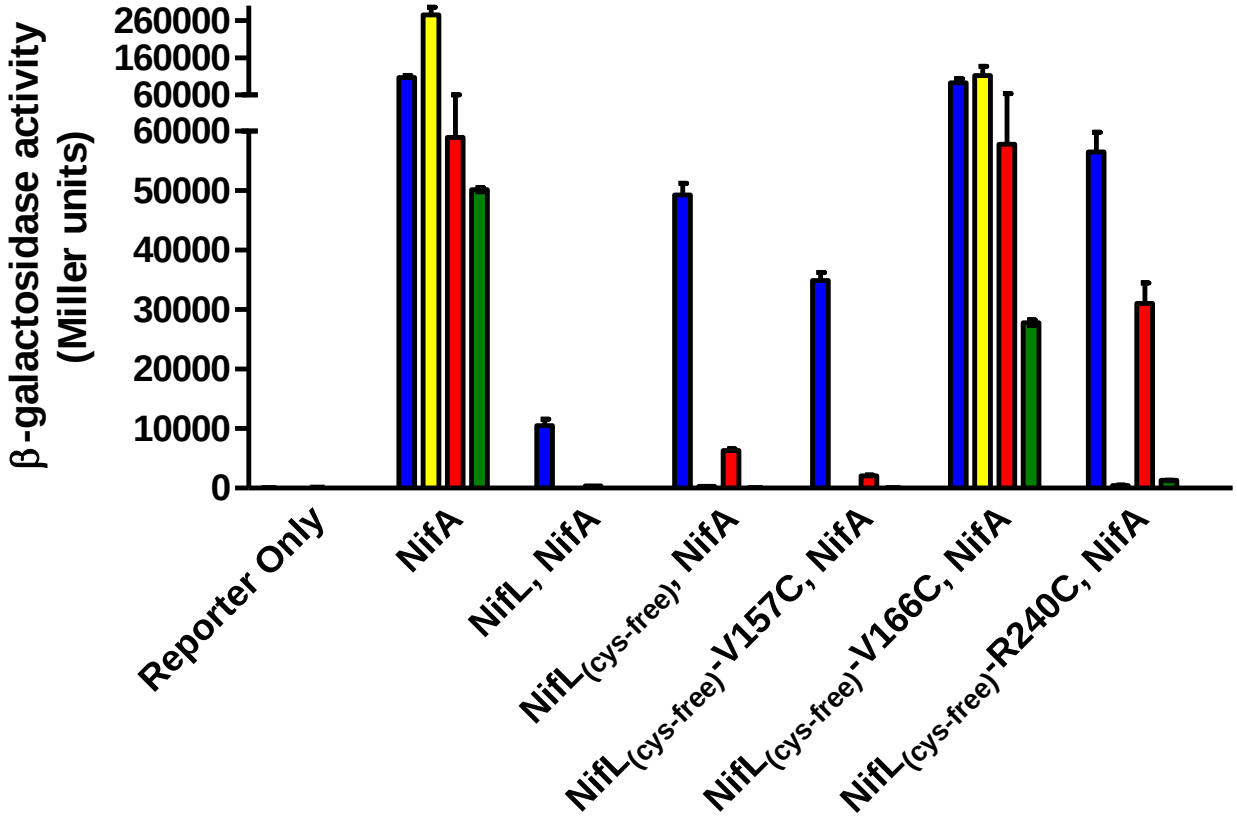


Figure 5.5. Influence of cysteine substitutions in the PAS2 domain on the ability of the $NifL_{(cys\text{-}free)}$ protein to inhibit transcriptional activation by NifA *in vivo*. Cultures were grown under the following conditions; (1) anaerobically, under nitrogen-limiting conditions (blue bars), (2) anaerobically with excess fixed nitrogen (yellow bars), (3) aerobically with limiting fixed nitrogen (red bars) and (4) aerobically when fixed nitrogen was replete (green bars). Cultures were assayed for β-galactosidase activity as a reporter of NifA-mediated transcriptional activation from a *nifH-lacZ* fusion. All experiments were performed at least in duplicate with error bars denoting the standard error of the mean.

would influence disulphide bond formation. Therefore, a substitution in the PAS1 domain that prevents transmission of the redox signal was utilised. The crystal structure of NifL PAS1 suggests that E70 might be involved in the initial conformational changes associated with oxidation of the FAD co-factor (Key et al., 2007a). When this glutamate residue was substituted for alanine, the resultant NifL variant failed to inhibit NifA activity in response to oxygen *in vivo* (Salinas, Little and Dixon, unpublished data). Purified NifL-E70A contains a normal complement of FAD that can be reduced by sodium dithionite (Little and Dixon, unpublished data) suggesting that the phenotype of this mutation arises from a defect in structural propagation of the redox signal rather than a defect in redox chemistry. The ability of V157C to participate in disulphide bond formation between PAS2 subunits under oxidising conditions was investigated, in either the presence or absence of the E70A substitution in PAS1. Appropriately positioned cysteine side chains are oxidised by ambient dissolved oxygen to form disulphide bonds (Bass et al., 2007). However, this reaction often proceeds slowly unless stimulated by the addition of a redox catalyst, such as copper phenanthroline. The presence of disulphide bridges can then be detected as a change in the apparent molecular mass of the protein when analysed by SDS-PAGE. As an additional control, a fraction of the cross-linked sample is often incubated with a reductant, such as β-mercaptoethanol, to demonstrate that this change in molecular mass is reversed upon reduction (and therefore must be due to disulphide bond formation). Protein samples were exposed to varying levels of copper phenanthroline and the formation of covalent cross-links was analysed by SDS-PAGE (Figure 5.6). In the absence of the E70A substitution (i.e. when $NifL_{(cys\text{-}free)}$-V157C was in the oxidised conformer) only small traces of the dimeric cross-linked species could be resolved and there was no obvious decrease in the amount of the monomeric (non cross-linked) protein as the oxidant concentration was

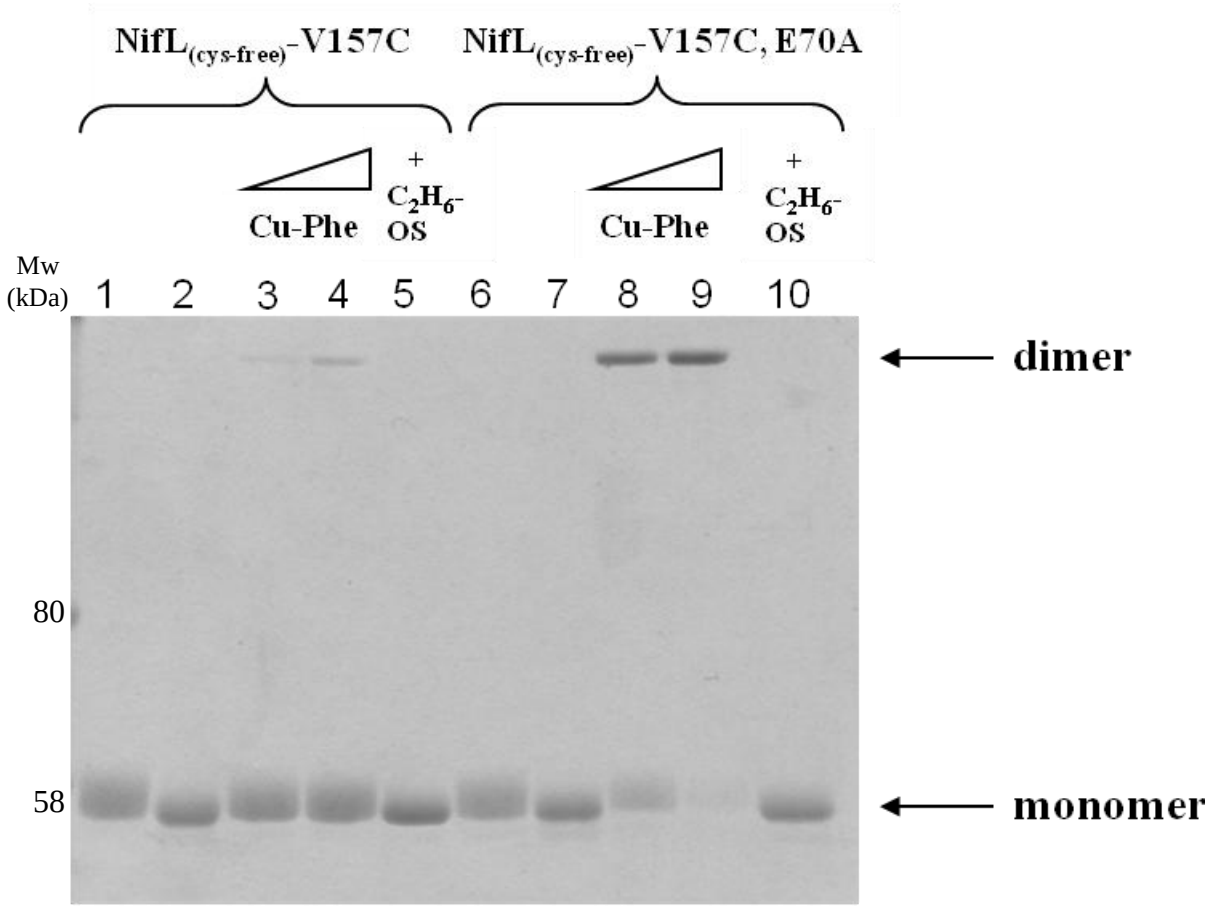


Figure 5.6. Cysteine cross-linking analysis of $NifL_{(cys\text{-}free)}$-V157C and $NifL_{(cys\text{-}free)}$-V157C, E70A. Cysteine cross-linking reactions were performed as described in section 2.7.9 and samples were analysed by SDS-PAGE. Lanes were loaded as follows: 1 = $NifL_{(cys\text{-}free)}$-V157C with NEM added prior to oxidation with 5 µM copper phenanthroline, 2 = $NifL_{(cys\text{-}free)}$-V157C, 3 = $NifL_{(cys\text{-}free)}$-V157C with 2.5 µM copper phenanthroline, 4 = $NifL_{(cys\text{-}free)}$-V157C with 5 µM copper phenanthroline, 5 = $NifL_{(cys\text{-}free)}$-V157C with 5 µM copper phenanthroline and β-mercaptoethanol in the loading dye, 6 = $NifL_{(cys\text{-}free)}$-V157C, E70A with NEM added prior to 5 µM copper phenanthroline, 7 = $NifL_{(cys\text{-}free)}$-V157C, E70A, 8 = $NifL_{(cys\text{-}free)}$-V157C, E70A with 2.5 µM copper phenanthroline, 9 = $NifL_{(cys\text{-}free)}$-V157C, E70A with 5 µM copper phenanthroline, 10 = $NifL_{(cys\text{-}free)}$-V157C, E70A with 5 µM copper phenanthroline and β-mercaptoethanol in the loading dye.

increased (Figure 5.6, lanes 3 and 4). In contrast, introduction of the secondary E70A substitution in PAS1 resulted in a large increase in disulphide bond formation and, in the 5 μM copper phenanthroline condition, the protein was almost entirely in the cross-linked form (Figure 5.6, lanes 8 and 9). This disulphide bridge could be fully removed by addition of β-mercaptoethanol to the SDS sample dye (Figure 5.6, lane 10) and control experiments in which *N*-ethylmaleimide (NEM) reagent (which irreversibly alkylates thiol groups) was added to samples prior to the copper-phenanthroline catalysed oxidation indicated that no non-specific cross-linking occurred between denatured polypeptides during preparation of samples for SDS-PAGE analysis (Figure 5.6, lanes 1 and 6). These experiments were repeated using the PAS1-PAS2 fragment rather then the full length NifL protein and similar results were obtained (data not shown). Hence, $NifL_{(cys\text{-}free)}$-V157C cross-links efficiently only in the presence of a secondary substitution in the PAS1 domain that blocks transduction of the oxygen signal. These results indicate that signals from the PAS1 domain influence the efficiency of disulphide bridge formation between subunits of the PAS2 domain in the context of the full length NifL protein.

Overall, the data appears to indicate that oxidation of the FAD co-factor in the PAS1 domain induces a conformational change in PAS1 that, in turn, triggers a change in the quaternary structure of the PAS2 domain. However, when the ability of the $NifL_{(cys\text{-}free)}$-V157C, E70A protein to inhibit NifA activity in response to the oxygen and fixed nitrogen signals *in vivo* was checked, the protein was indistinguishable from $NifL_{(cys\text{-}free)}$-V157C, which lacks the E70A substitution (Figure 5.7A). That is, contrary to expectation, substitution of E70 for alanine did not result in increased NifA activity under oxidising conditions when V157C and the four cysteine replacement substitutions in NifL were present. This puzzling result prompted investigation of the influence of the V157C substitution (and the V157C, E70A double substitution) on NifL activity in the absence of

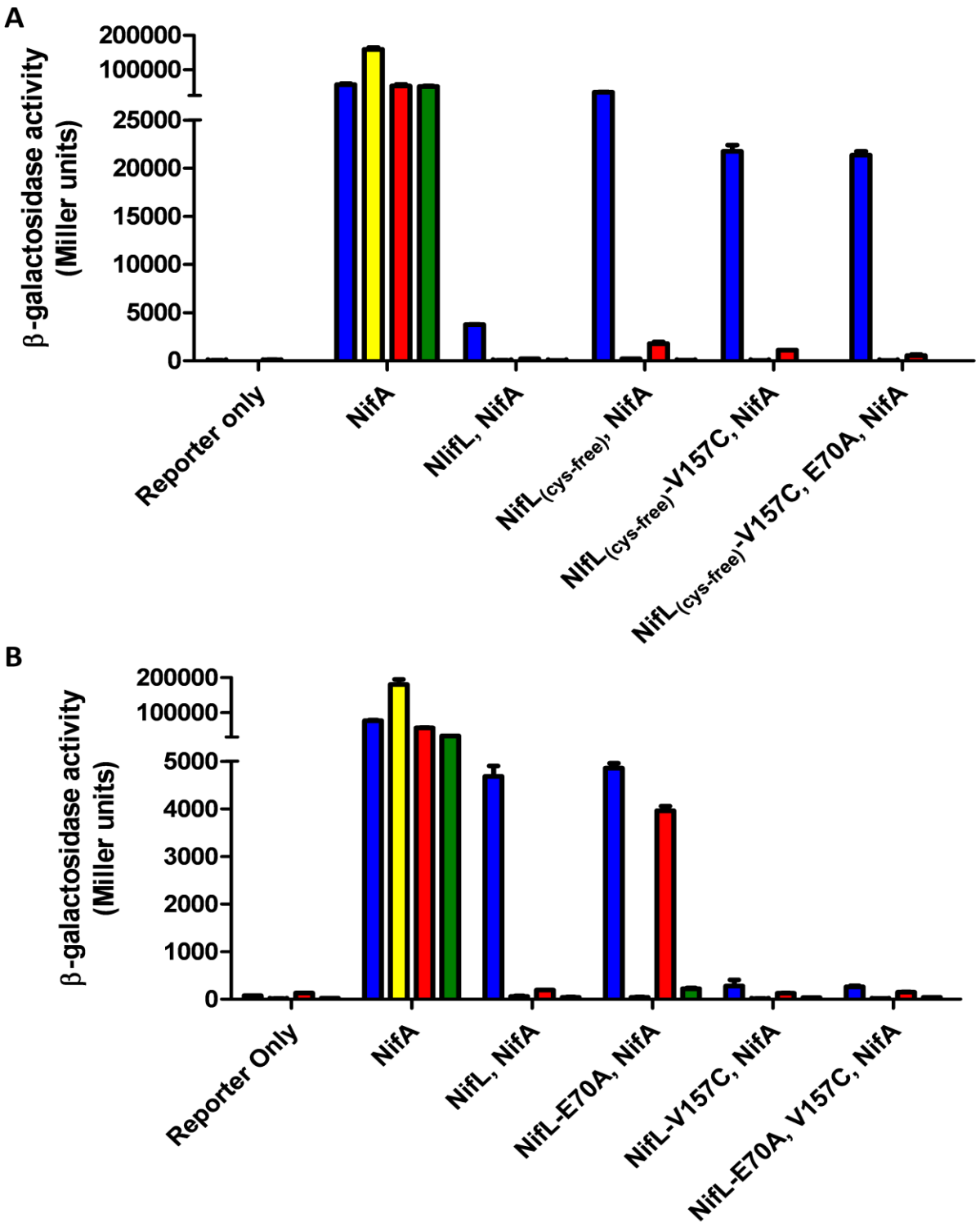


Figure 5.7. Influence of the V157C and E70A substitutions on the ability of (A) $NifL_{(cys-free)}$ and (B) NifL to inhibit NifA activity *in vivo*. Cultures were grown under the following conditions; (1) anaerobically, under nitrogen-limiting conditions (blue bars), (2) anaerobically with excess fixed nitrogen (yellow bars), (3) aerobically with limiting fixed nitrogen (red bars) and (4) aerobically when fixed nitrogen was replete (green bars). Cultures were assayed for β-galactosidase activity as a reporter of NifA-mediated transcriptional activation from a *nifH-lacZ* fusion. All experiments were performed at least in duplicate with error bars denoting the standard error of the mean.

the four cysteine replacements (i.e. in a wild-type background) (Figure 5.7B). As demonstrated previously, NifL-E70A failed to inhibit NifA activity in response to excess oxygen (Figure 5.7B, bars marked "NifL-E70A, NifA"). In contrast to results obtained in the cysteine-free background, NifL-V157C and NifL-V157C, E70A both exhibited a "locked-on" phenotype (Figure 5.7B). That is, the V157C substitution results in a "locked-on" form of the NifL protein and is dominant over the E70A substitution. Although this result provides some explanation as to why the $NifL_{(cys\text{-}free)}$-V157C and $NifL_{(cys\text{-}free)}$-V157C, E70A variants exhibit the same phenotype (Figure 5.7A), it appears to contradict the biochemical data and raises further questions regarding the influence of the cysteine replacements on NifL function. For example, why does the V157C substitution not give rise to a "locked-on" phenotype in the cysteine-free background? It is clear, however, that the $NifL_{(cys\text{-}free)}$ protein does not behave in the same manner as the wild-type protein and that differences between the two forms are likely to be responsible for the discrepancies discussed above. In other words, the cysteine replacements somehow alter the conformation of NifL and thus substitution of V157 for cysteine has different effects in the cysteine-free background compared with the wild type. Given these uncertainties, it is difficult to draw any firm conclusions from the cysteine cross-linking experiments. Nevertheless, despite the lack of congruence between the phenotypes of the variant proteins *in vivo* and the biochemical data, substitutions in the PAS1 domain can influence disulphide bridge formation between PAS2 subunits *in vitro*.

5.3.2 BACTH analysis

The bacterial two-hybrid system was used to investigate the influence of signals from the PAS1 domain on the interaction between subunits of the PAS2 domain. The results presented in Chapter 4 demonstrate that the interaction between PAS2 subunits is

detectable using BACTH analysis. Preliminary experiments demonstrated that an interaction between the isolated PAS2 domain and a longer construct containing the PAS1 and PAS2 domains in tandem (the PAS1-PAS2 fragment) could also be detected using BACTH analysis (data not shown). Moreover, control experiments suggested that this interaction was due to PAS2 oligomerisation, rather than interaction between the PAS1 and PAS2 domains (data not shown). We questioned whether signals from the PAS1 domain could influence interaction of the PAS1-PAS2 fragment with the isolated PAS2 domain, thereby indicating that the signalling state of PAS1 can influence PAS2 dimerisation. Of course, the context of this experiment differs substantially from signalling events in the wild-type NifL protein as only one PAS2 subunit is receiving a signal from PAS1. However, appropriate controls can be performed by introducing a "locked-on" PAS2 substitution to the PAS1-PAS2 fragment to perturb the interaction in just one of the two interacting PAS2 subunits; preliminary experiments demonstrated that this substantially reduced, but did not eliminate, the interaction between PAS2 subunits (data not shown). This implies that conformational changes in a single PAS2 subunit can elicit a measurable difference in the interaction and provides a suitable control by simulating the "on state" of the PAS1-PAS2 fragment.

Initially, an attempt was made to investigate the PAS1-PAS2 versus PAS2 interaction under both oxidising and reducing conditions. However, western analysis using anti-NifL anti-sera indicated that the fusion proteins were unstable in cells grown under oxidising conditions (data not shown). Therefore, a substitution that locks the PAS1 domain in the "on" state was utilised. Substitution of M132 for alanine gives rise to a form of the NifL protein that inhibits NifA activity under reducing conditions *in vivo* (Little and Dixon, unpublished data). In other words, the M132A substitution locks the PAS1 domain in the oxidised (or "on") signalling state and thus causes NifL to inhibit NifA activity in

the absence of an oxygen signal. The interaction between the PAS1-PAS2 fragment and the isolated PAS2 domain was examined under reducing conditions and the influence of the "locked-on" substitutions M132A (in PAS1) and V166M (in PAS2) were analysed (Figure 5.8). As observed previously in the preliminary experiments, interaction between the wild-type PAS1-PAS2 fragment and the isolated PAS2 domain was detected. The strength of this interaction, as quantified by β-galactosidase assays, was 1,536 Miller units (Figure 5.8, experiment A). Introduction of the M132A substitution resulted in a decrease in the strength of this interaction by approximately 500 Miller units (or one third) to 1050 Miller units (Figure 5.8, experiment B). Control experiments indicated that introduction of the V166M substitution (which inhibits PAS2 oligomerisation, see Figure 4.1), into the PAS1-PAS2 fragment (but not the isolated PAS2 domain) resulted in a slightly larger decrease in interaction strength of approximately 700 Miller units (Figure 5.8, experiment C). When both constructs contained the V166M substitution, the interaction was reduced to approximately 250 Miller units (Figure 5.8, experiment D). However, negative controls measuring the interaction of each fusion protein with the opposing AC subunit suggested a background interaction of 273 Miller units in experiments utilising the T18:NifL$_{(1\text{-}284)}$-V166M fusion protein (Figure 5.8, controls column for experiments C and D). Therefore, the true level of interaction in experiments C and D is likely to be lower than the measured interaction. Nevertheless, these results suggest that substitutions in the PAS1 domain can influence the affinity of PAS2 association. As mentioned above, the data presented in the previous Chapters suggests that the transition of the PAS1 domain from the reduced (or "off") state to the oxidised (or "on") state is communicated to the PAS2 domain, resulting in a decrease in the affinity between PAS2 subunits. Based on this hypothesis, it was expected that the interaction between the PAS1-PAS2 fragment and the isolated PAS2 domain would be stronger when the PAS1 domain is in the "off" state compared to the

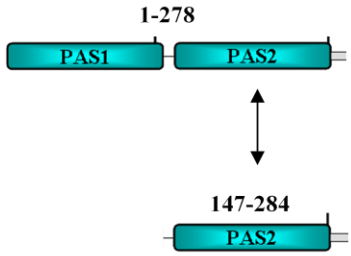
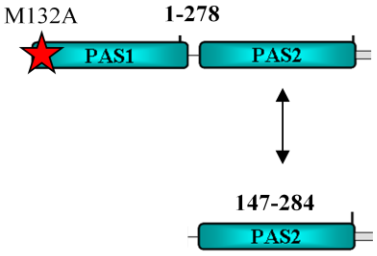
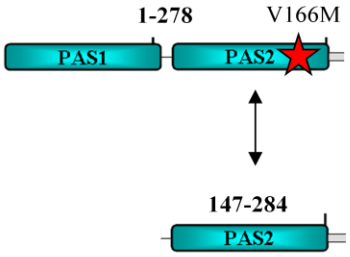
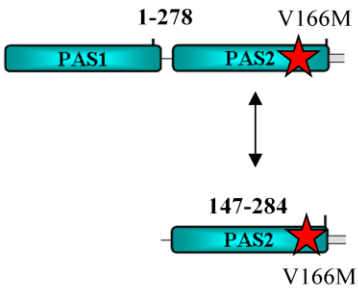
Experiment	Constructs	Interaction (+/- SE) in Miller units	Controls (+/- SE) in Miller units
A		**1536 (+/- 25)**	T18 fusion vs T25 97 (+/- 6) T18 vs T25 fusion 78 (+/- 0.2)
B		**1050 (+/- 34)**	T18 fusion vs T25 69 (+/- 5) T18 vs T25 fusion 78 (+/- 0.2)
C		**806 (+/- 34)**	T18 fusion vs T25 273 (+/- 24) T18 vs T25 fusion 78 (+/- 0.2)
D		**251 (+/-12)**	T18 fusion vs T25 273 (+/- 24) T18 vs T25 fusion 76 (+/- 1)

Figure 5.8. BACTH analysis of the influence of signals from the PAS1 domain on the association of PAS2 subunits. Cells were grown under anaerobic conditions and interactions between hybrid proteins were measured as described in section 2.9.2. The data shown are based on at least three independent replicates.

"on" state. The BACTH results support this prediction. The ability of the PAS1-PAS2 fragment containing the M132A substitution in PAS1 to interact with the isolated PAS2 domain is intermediate between that of the wild-type PAS1-PAS2 fragment and the fragment containing the V166M substitution in PAS2. That is, when the PAS1 domain is locked in the "on" state there is a reduction in the affinity between PAS2 subunits although this reduction is less substantial than that induced by "locked-on" substitutions in the PAS2 domain. Overall, data from the BACTH analysis supports the hypothesis that redox sensing by the PAS1 domain impacts upon the stability of the PAS2 dimer. It should be remembered that clear interpretation of these results is hindered by a lack of information regarding the oligomerisation state of the hybrid proteins; each of the fusion proteins studied here can presumably form a homodimer (as the PAS1 and PAS1-PAS2 fragments of NifL both dimerise) whereas heterologous association between the hybrid proteins must take place in order to yield a measureable interaction. Thus, the data obtained may represent the gross output from several competing dynamic equalibria. Additionally, the possibility that hybrid protein homodimers interact to form heterologous higher order oligomers cannot be eliminated.

5.4 Mutagenesis of the α-helix linking the NifL PAS domains

Secondary structure predictions using the PSIPRED server indicate that the region of NifL between the PAS1 and PAS2 domains forms an α-helix. This helix could be considered a C-terminal extension of the PAS1 domain, an N-terminal extension of the PAS2 domain or a helical linker joining the two PAS domains. As mentioned in Chapter 1, PAS domains often have helical extensions protruding outward from, or flanking, the core α/β fold (Möglich et al., 2009b). However, a recent study investigating signalling in chimeric PAS-based sensor proteins indicated that tandem PAS domains are commonly

linked by short amphipathic α-helices (Möglich et al., 2010) and, in this section, the helix connecting the NifL PAS domains will be referred to as a helical linker. Taken together, secondary structural predictions and the crystal structure of the NifL PAS1 domain (Key et al., 2007a) suggest that this linker helix is likely to start between residues 137 and 139. In the absence of structural data concerning the other NifL domains, it is difficult to predict with confidence where the C-terminal end of the linker may be. However, predictions using the COILS server (http://www.ch.embnet.org) indicate that the region of NifL between residues 139 and 159 may form an α-helical coiled-coil in the NifL dimer (a helical wheel projection of this is shown in Figure 5.9A). This region overlaps with the proposed A'α helix in the PAS2 domain (discussed in sections 3.4 and 4.1) and without detailed structural information it is not possible to discern whether there are (i) two distinct helices, (ii) one extended helix or (iii) the structural predictions and modelling are incorrect. Given that the results presented in this thesis suggest that redox signals are communicated between the NifL PAS domains, it is possible that the inter-domain region may have a role in signal relay. Therefore, alanine scanning mutagenesis was used to analyse the function of the putative helical linker.

5.4.1 Alanine Scanning

Twelve residues in the putative helical linker connecting the NifL PAS domains were substituted for alanine and the ability of the resultant NifL variants to inhibit transcriptional activation by NifA in response to oxygen and fixed nitrogen signals was analysed *in vivo* (Figure 5.9B). Three previously studied alanine substitutions in this region (L151A, I153A and V157A) were included for comparison. Western analysis indicated that all of the variant NifL proteins were stable under the four reaction conditions (data not shown). As demonstrated previously, NifA is active under all conditions in the absence of

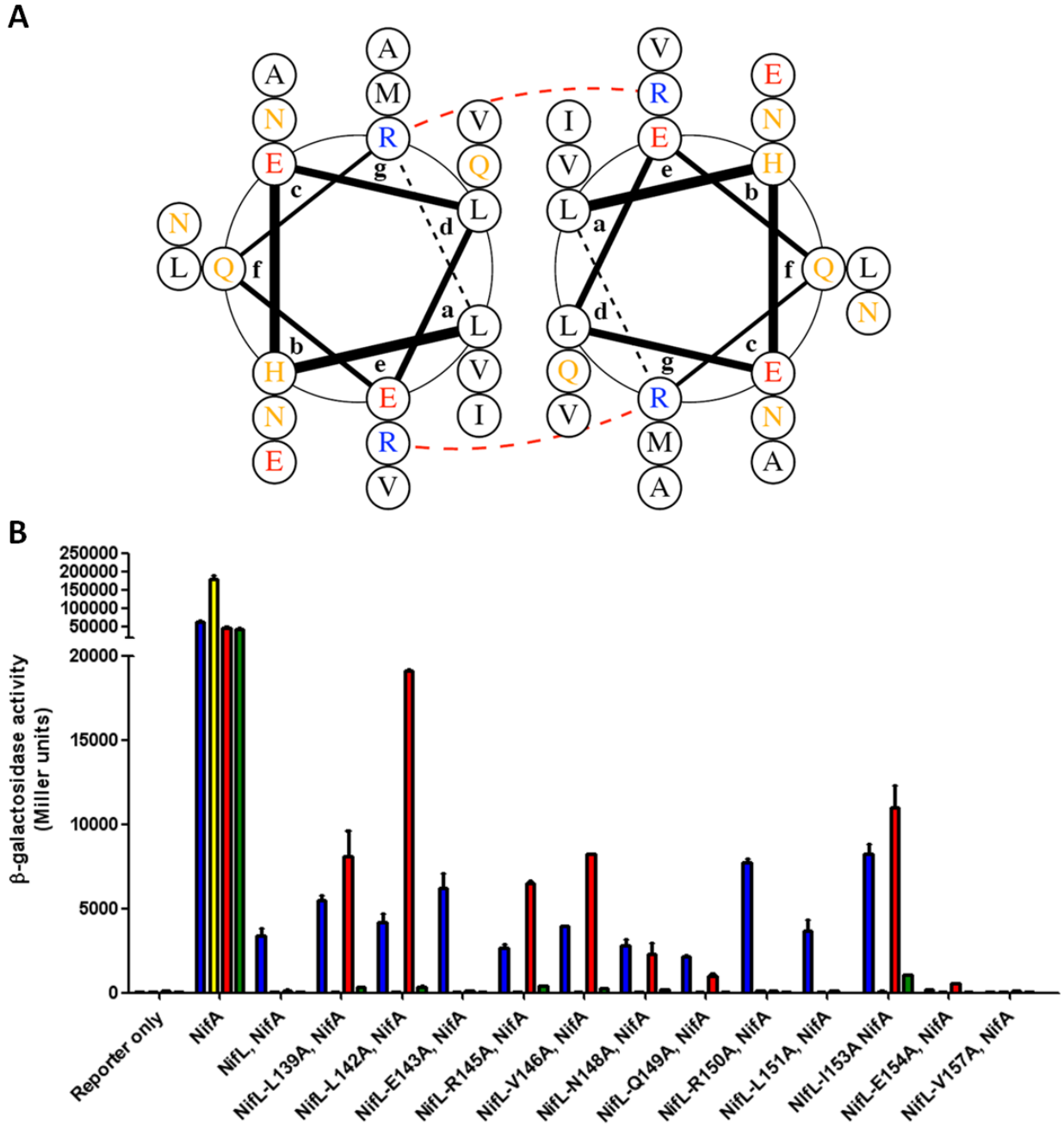


Figure 5.9. Alanine scanning of the linker helix that connects the PAS1 and PAS2 domains of NifL. (A) Helical wheel projection of NifL residues 139-159. (B) Influence of alanine substitutions in the linker helix on the ability of NifL to inhibit NifA-mediated transcriptional activation from a *nifH-lacZ* reporter fusion *in vivo*. The graph legend is as in Figure 5.7.

the NifL regulatory protein (Figure 5.9B, bars marked "NifA") and wild-type NifL inhibits NifA activity in discrete responses to oxygen and fixed nitrogen (Figure 5.9B, bars marked "NifL, NifA"). Seven of the twelve alanine substitutions (L139A, L142A, R145A, V146A, N148A, Q149A and I153A) gave rise to a form of the NifL protein that failed to inhibit NifA activity under oxidising conditions (Figure 5.9B, red bars). That is, over half of the NifL variants examined exhibited a "redox signalling" phenotype. Three of the twelve substitutions (E143A, R150A and L151A) exhibited wild-type phenotypes whilst the remaining two substitutions (E154A and V157A) inhibited NifA activity under all four conditions tested (Figure 5.9B, bars marked "NifL-E154A, NifA" and "NifL-V157A, NifA"). Interestingly, these two "locked-on" substitutions are positioned at the C-terminal end of the inter-domain region and could be considered part of the A'α helix in the PAS2 domain rather than the helical linker, particularly as the V157A substitution is known to influence dimerisation of the $NifL_{(147-284)}$ fragment (i.e. the isolated PAS2 domain) (Figure 4.1). Helical wheel projections for residues 139-159 suggested an electrostatic repulsion between residues R145 and R150 (Figure 5.9A, red dashed lines). However, NifL-R145A was a "redox signalling" variant and NifL-R150A was wild type. Thus, the alanine scanning mutagenesis indicated that this repulsion, if present, is not important in signalling as the R145A and R150A substitutions are phenotypically different and one might anticipate that breaking an important interaction from either side would result in the same phenotype. One plausible explanation of the above data is that the region of NifL mutagenised in this section includes two helices; the helical linker could be located between residues 139 and 149-151 whilst the A'α helix of PAS2 could start between residues 151 and 153. In this scenario, there would be no electrostatic repulsion between R145 and R150 as these residues would be located in different helices and, more importantly, V157 and I153 would be located in the PAS2 domain, explaining the

influence of the V157A and I153A substitutions on oligomerisation of the isolated PAS2 domain (in NifL fragments containing residues 143-284 or 147-284) (see Chapter 4). This postulation is also consistent with the structural model of the NifL PAS2 domain presented in chapter 4 (Figure 4.8).

Overall, the frequency at which "redox signalling" variants were obtained by alanine scanning of the putative linker helix connecting the PAS1 and PAS2 domains of NifL (six of the eight substitutions generated between residues 139 and 150 or seven of the twelve residues substituted between 139 and 159, depending on the true length of the helix) suggests that this region is important in redox signal transduction.

5.4.2 Deletion mutants

A recent bioinformatic analysis examined the linker sequences that connect tandem PAS domains in 5002 proteins (Möglich et al., 2010). As mentioned above, this study indicated that the linkers adopt a well-defined, largely α-helical structure. Moreover, it was found that approximately half (48%) of the putative linkers analysed were 28 residues long and that linkers that deviate from this length commonly differ by sets of three or four residues. For example, the shorter helical linkers were often 24, 21, 17 or 14 residues long. This pattern implies that the periodicity of PAS-PAS helical linkers may be important for their function. That is, the length and conformation of the linker between PAS domains is likely to define their relative orientation and thus influence PAS-to-PAS signalling. These findings are reminiscent of results obtained from the study of signal transmission between sensory PAS domains and effector domains via the connecting amphipathic (Jα) helix (discussed in section 1.2.4.). Experiments examining the influence of various amino acid deletions and insertions in the Jα helices of chimeric PAS sensor proteins have

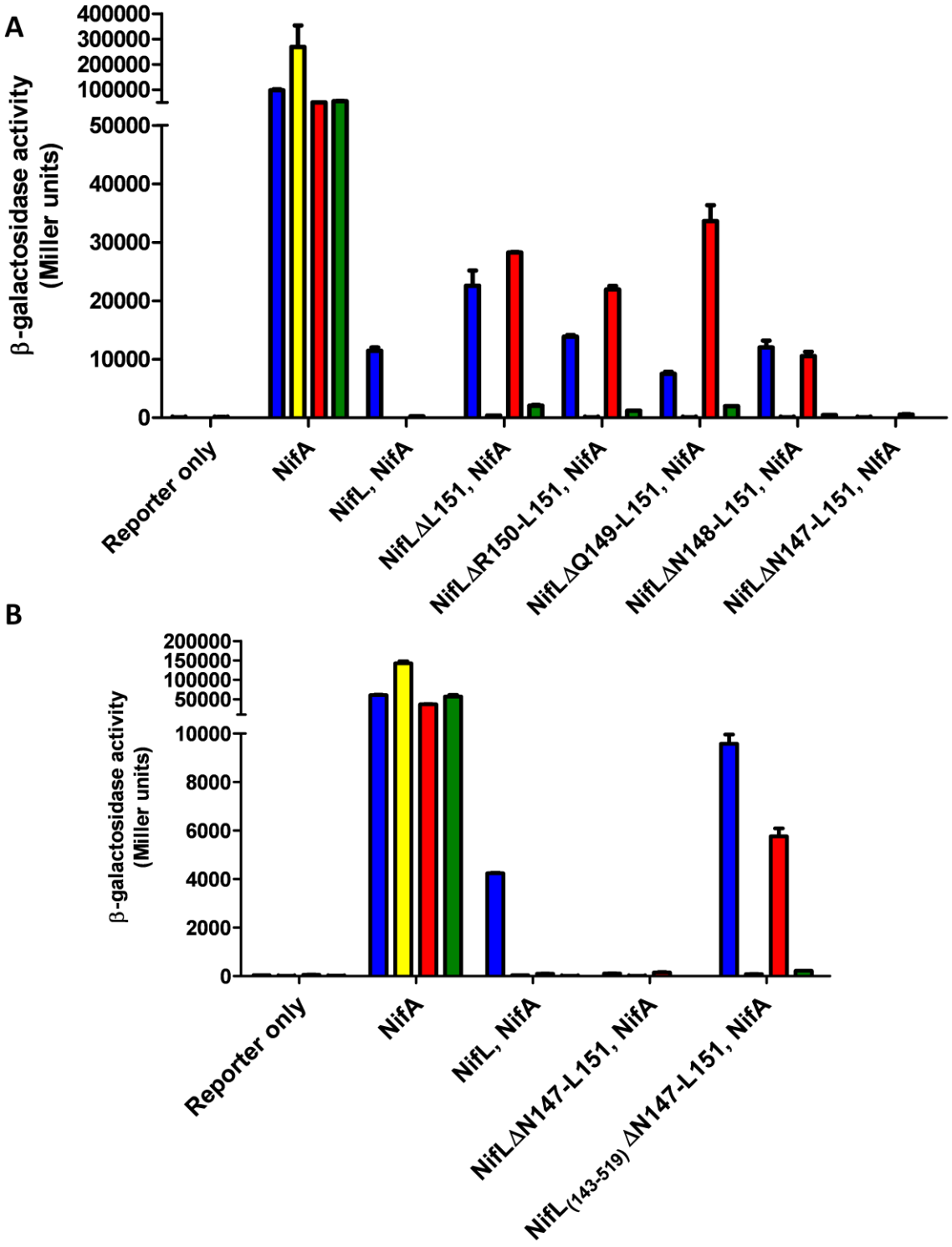


Figure 5.10. (A) Influence of deletions in the linker helix that connects the PAS1 and PAS2 domains of NifL on the ability of the NifL protein to inhibit transcriptional activation by NifA *in vivo*. (B) Requirement of the N-terminal region of NifL for the "locked-on" phenotype of the ΔN147-L151 deletion. Graph legends are as in Figure 5.7.

demonstrated a precise correlation between the phenotype (or signalling state) and the helical length (Möglich et al., 2009b).

To investigate the importance of the length of the putative helical linker connecting the N-terminal NifL PAS domains in redox signal transduction, a series of amino acid deletions was generated and the influence of these deletions on the ability of NifL to inhibit NifA activity in response to oxygen was analysed *in vivo* (Figure 5.10A). Five variant forms of the NifL protein were created: NifLΔL151, NifLΔR150-L151, NifLΔQ149-L151, NifLΔN148-L151 and NifLΔN147-L151 (deletions of 1-5 residues). Western analysis indicated that all proteins were stable under the four assay conditions (data not shown). Four of the five NifL variants failed to inhibit NifA activity in the presence of excess oxygen (Figure 5.10A, red bars) but responded normally to fixed nitrogen (Figure 5.10A, yellow bars). That is, four of the deletions (NifLΔL151, NifLΔR150-L151, NifLΔQ149-L151 and NifLΔN148-L151) gave rise to a "redox signalling" phenotype. As mentioned above, NifL-L151A and NifL-R150A respond normally to oxygen and fixed nitrogen (Figure 5.9B, bars marked "NifL-L151A, NifA" and "NifL-R150A, NifA"). This indicates that the side chains of L151 and R150 are not required for redox signal transduction in NifL. Nevertheless, deletion of either L151, or L151 and R150, resulted in a form of NifL that failed to inhibit NifA activity under oxidising conditions (Figure 5.10A, bars marked "NifLΔL151, NifA" and "NifLΔR150-L151, NifA"). Taken together, these data imply that a shortening of the helical linker, rather than the removal of key amino acid side chains, is responsible for the "redox signalling" phenotype of NifLΔL151 and NifLΔR150-L151. These results emphasise the importance of the inter-domain linker to redox signalling in NifL. However, it is difficult to distinguish between mutations that result in a loss of function (i.e. block redox signalling via perturbation of protein structure) and those that give rise to a form of NifL that is

locked in the "off state" (i.e. favour the reduced conformation). It is possible that the "redox signalling" variants obtained here act by either of these mechanisms. Interestingly, the NifLΔN147-L151 variant inhibited NifA activity under all four conditions, even when oxygen and fixed nitrogen were limiting (Figure 5.10, bars marked "NifLΔN147-L151, NifA"). In other words, the deletion of five residues (N147-L151) from the PAS-PAS linker helix in NifL gave rise to a "locked-on" phenotype. Structural damage in this region of NifL would be expected to result in a "redox signalling" phenotype and thus the identification of this "locked-on" variant cannot be explained by structural perturbation of the linker helix. Hence, the deletion appears to simulate a conformational state normally induced by oxygen signals from the PAS1 domain. This implies that conformational changes in the helical linker are important for transmission of redox signals between the NifL PAS domains.

All "locked-on" NifL variants identified to date, containing substitutions in the PAS2 or H domains, do not require the PAS1 domain in order to inhibit NifA activity (see section 3.3.2). In other words, they retain a "locked-on" phenotype when the N-terminal region of NifL (residues 1-142) is removed. To further investigate the properties of the newly identified "locked-on" variant, NifLΔN147-L151, the N147-L151 deletion was introduced to a truncated form of the NifL protein that lacked the first 142 amino acids. The ability of the variant proteins to inhibit NifA-mediated transcriptional activation was then analysed *in vivo* (Figure 5.10B). As shown previously, NifLΔN147-L151 inhibited NifA activity under all four assay conditions (Figure 5.10B, bars marked "NifLΔN147-L151, NifA"). By contrast, $NifL_{(143-519)}$ΔN147-L151 responded normally to fixed nitrogen (Figure 5.10B, yellow and green bars) but failed to inhibit transcriptional activation by NifA in response to excess oxygen (Figure 5.10B, red bars) or when oxygen and fixed nitrogen were limiting (Figure 5.10B, blue bars). The $NifL_{(143-519)}$ΔN147-L151 protein

exhibited a "redox signalling" phenotype, indicating that the N-terminal truncation overrides the effect of the deletion in the linker helix. This demonstrates that, unlike "locked-on" substitutions in the PAS2 domain, the "locked-on" phenotype of the N147-L151 deletion is context dependent and requires the N-terminal region of the NifL protein. Overall, the data presented in this section indicate that the properties (i.e. the phasing, conformation or energetics) of the helical linker connecting the PAS1 and PAS2 domains are important to the redox signalling mechanism in NifL and that signal relay between the PAS domains is likely to depend upon structural alterations in this region.

5.5 Discussion

The limited proteolysis experiments presented in this Chapter demonstrate that a redox-dependent conformational change occurs in the N-terminal PAS domains of NifL. Moreover, these experiments suggest that substitutions in the PAS2 domain are able to lock the NifL protein in either the oxidised or the reduced conformer. Data from previous Chapters imply that changes in the quaternary arrangement of the PAS2 domain modulate NifL activity and that the quaternary structure of the PAS2 domain is responsive to redox signal perception by the PAS1 domain. Several approaches were taken to test this hypothesis. These included cysteine cross-linking studies and BACTH analysis of domain interactions in NifL. Although the results from each set of experiments were not entirely conclusive, taken together, all approaches provide a body of evidence that is consistent with the hypothesis that the signalling state of the PAS1 domain influences the quaternary structure of PAS2. The cysteine cross-linking experiments provide strong biochemical evidence that substitutions in the PAS1 domain can influence disulphide bridge formation between PAS2 subunits. However, the cysteine replacement substitutions appeared to interfere with NifL activity when measured *in vivo*. The BACTH analysis suggests that the

signalling state of the PAS1 domain influences the affinity of the interaction between PAS2 subunits but the assay conditions were artificial. Nevertheless, when viewed in conjunction with the limited proteolysis experiments and the mutational and biochemical analyses of the PAS2 domain presented in this and previous Chapters, there is strong evidence to support the proposed mechanism of redox signal relay in NifL.

Mutagenic analysis of the linker between the NifL PAS domains indicated that this region of the NifL protein is important in redox signal transduction. Alanine scanning mutagenesis yielded numerous "redox signalling" variants. Moreover, a series of five sequential amino acid deletions gave rise to four "redox signalling" variants and one "locked-on" variant protein. Each deletion results in a specific change in the helix angle and these were plotted against NifL activity (measured as NifA activity in the presence of NifL). The results are shown in Figure 5.11. The deletion of five residues resulted in a helix angle of -140^{o} and gave rise to a form of the NifL protein that adopts an inhibitory conformation irrespective of the signalling state of the PAS1 domain. At helix angles between -150^{o} and -50^{o}, oxygen availability had little influence on NifL activity. In the presence of oxygen, there appeared to be some periodicity in the relationship between helix angle and NifL activity (Figure 5.11, blue line). Taken together, these results indicate that the phasing and conformation of the PAS-PAS linker helix are important in inter-domain redox signal relay. However, the precise mechanism underpinning the transmission of redox signals between the NifL PAS domains remains unclear. Based on the available information from this work and other studies of PAS-containing proteins, several potential mechanisms can be envisaged.

Firstly, the linker helices may act as rigid "spacers" between tandem PAS domains. These "spacers" could ensure that the PAS domains are properly orientated with respect to one another, facilitating direct PAS-PAS interactions. Additionally, quaternary structural

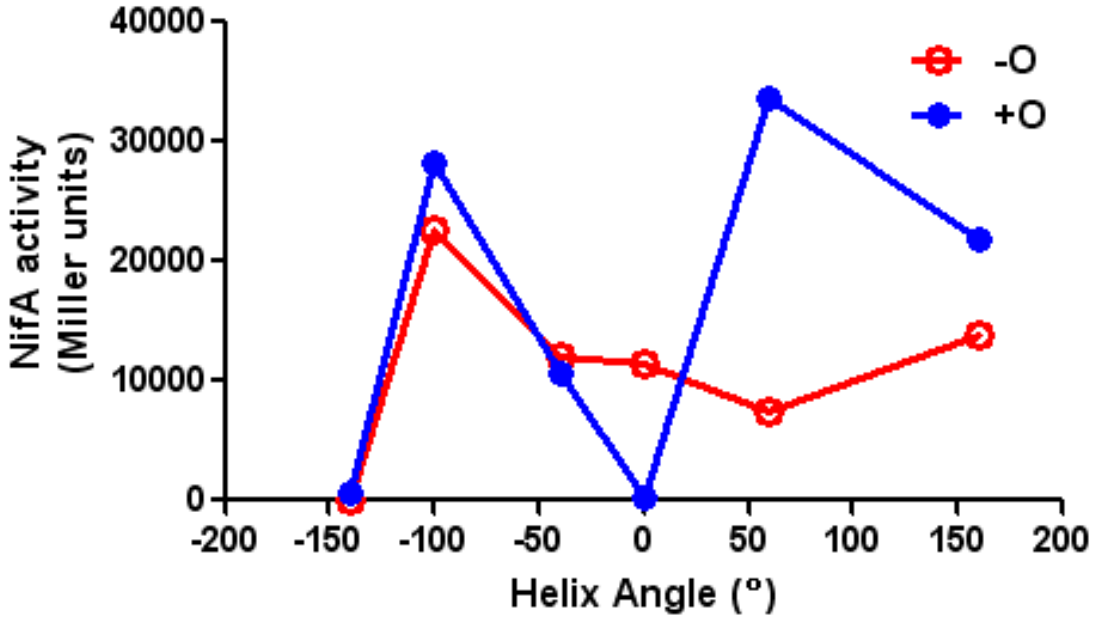


Figure 5.11. Influence of changes in helix angle in the PAS1-PAS2 linker on the ability of NifL to inhibit NifA activity *in vivo*. The data shown here is derived from the results presented in Figure 5.10. Cultures were assayed for β-galactosidase activity as a reporter of NifA-mediated transcriptional activation from a *nifH-lacZ* fusion under aerobic (blue line) or anaerobic (red line) conditions.

changes in the first PAS domain may generate a movement of the two linker helices relative to each other within a dimeric protein. In other words, rigid helical linkers could enable signal transmission either by "tugging" the PAS2 domain in response to changes in the signalling state of PAS1 or by facilitating a direct interaction between the NifL PAS domains which is required for signal relay.

Alternative mechanisms of signal transduction involve conformational and/or energetic changes in the linker region. These may take the form of a helical rotation within a coiled-coil linker or changes in the association state of the linker helix. For example, the helices may form a "zipper" that "zips" or "unzips" depending on the signalling state of the PAS1 domain. However, the available data, although limited, perhaps favour a model whereby signals are transmitted along a coiled-coil linker in the form of torque or helical rotation. Such movements may be triggered, for example, by a partial unwinding of the linker helix in response to signals form the sensory domain. This mechanism has been proposed for several PAS-containing proteins (Möglich et al., 2010; Möglich and Moffat, 2007b; Taylor, 2007). In a recent study that combined experiments on chimeric PAS sensor proteins with the available structural data on tandem PAS domains, Moffat and colleagues observed that multiple PAS domains are commonly arranged along a linear axis such that the N-terminus of the second domain is adjacent to the C-terminus of the first (i.e. they are oriented "head-to-tail"). The authors postulate that addition/reduction of torques along this axis provides a means of integrated signal output from multiple sensory PAS domains (Möglich et al., 2010). It has been proposed that the C-terminal Jα helices protruding from the sensory PAS domains of the YtvA and FixL proteins relay signals to effector domains via a similar mechanism (Möglich et al., 2009b; Möglich and Moffat, 2007). Thus, the torque (or helical rotation) hypothesis conveniently couples signal transmission between PAS domains to the regulation of effector domain activity in these

systems. However, some PAS-containing proteins lack Jα helices. For example, the N-terminal PAS domains and the output domains of the NifL protein are connected by a glutamine rich linker. Structural predictions indicate that this region is likely to be disordered (http://prdos.hgc.jp/). Hence, signalling between the PAS domains and C-terminal domains of NifL is unlikely to occur via helical rotation. This does not eliminate the possibility that signals from the PAS1 domain generate torque in the linker helix to impact the signalling state of PAS2, but instead limits the potential of the helical rotation hypothesis to explain how signals from PAS1 influence NifL activity.

The length of the helical linker connecting the NifL PAS domains is clearly important in signalling as amino acid deletions in this region influence NifL activity, presumably via affects on the relay of redox signals to the PAS2 domain. However, the available information regarding the PAS1-PAS2 linker region in NifL is not sufficient to discriminate between the models of signal transduction discussed above. Nevertheless, each of the proposed mechanisms of PAS-to-PAS signal relay satisfies an important criterion; they are responsive to (or utilise) the quaternary structural changes that characterise PAS signalling.

Chapter 6 - General Discussion

The data presented here demonstrate that the PAS2 domain of NifL can exist in two discrete states, as exemplified by substitutions that stabilise NifL in either the "on" or the "off" conformation. The "on" substitutions in PAS2 result in a form of NifL that is competent to inhibit NifA, irrespective of the redox state of the FAD co-factor in the PAS1 domain. Limited proteolysis experiments suggest that these substitutions lock NifL in a conformation similar to that of the oxidised form. By contrast, the "off" mutants in PAS2 apparently fail to communicate the redox state of PAS1 to the C-terminal domains of NifL, but the variant proteins remain responsive to the fixed nitrogen signal conveyed by interaction with the signal transduction protein, GlnK (Little et al., 2002; Rudnick et al., 2002). The "off" or "redox signalling" variants might influence redox signal transduction in NifL by several different mechanisms including: (i) disrupting interactions between the PAS1 and PAS2 domains, (ii) perturbing interactions between PAS2 and the C-terminal domains of NifL or (iii) stabilising the reduced conformation relative to the oxidised conformer. Evidence from SEC and chemical cross-linking experiments suggests that one of the two "redox signalling" substitutions identified (I153A) acts by stabilising the PAS2 dimer whilst the other substitution (F253L) influences NifL activity via an alternative mechanism. Further evidence for the involvement of the PAS2 domain in redox signalling was obtained from *in vivo* analysis of variant forms of the NifL protein that lack this domain; removal of the PAS2 domain gives rise to a form of NifL that is not competent to respond to changes in redox potential. These results directly demonstrate an important role for PAS2 in redox signal relay from PAS1 to influence the interaction of the C-terminal domains of NifL with NifA.

The quaternary arrangement of the PAS2 subunits within NifL is apparently an important component in redox signal transduction. The isolated PAS2 domain is a dimer in

solution and the "off" state variants are also dimeric, whereas all of the "on" state variants analysed appear to influence the association state of the isolated PAS2 domain towards the monomeric form. However, these changes in association state are not apparent when PAS2 is combined with other domains, suggesting that PAS2 does not contribute to oligomerisation of NifL, but instead provides an interface for alternative quaternary arrangements. Most of the "locked-on" substitutions identified are apparently located within a conserved dimerisation interface recently recognised in PAS domains of known structure (Ayers and Moffat, 2008). Although no structural data is available for the NifL PAS2 domain, structural modelling indicates that four of the seven "locked-on" substitutions indentified in this work are located in the proposed interface and at least one of the remaining three substituted residues are likely to stabilise the PAS2 dimer (Figures 3.1 and 4.8). BACTH analysis demonstrates that all seven of the "locked-on" substitutions disrupt the interaction between PAS2 subunits (Figure 4.1). Thus, all of the "locked-on" substitutions identified in this work perturb dimerisation of the PAS2 domain, regardless of their proximity to the putative dimerisation interface.

Taken together, the data presented in this thesis suggest a model of signal transduction in NifL whereby changes in the quaternary structure of the PAS2 domain mediate the transmission of redox signals from the PAS1 domain to the C-terminal domains of NifL (Figure 6.1). In this model, oxidation of the FAD co-factor induces a conformational change in the PAS1 domain that is communicated to the PAS2 domain via the inter-domain linker helix (see Chapter 5.4). This, in turn, triggers a shift in the monomer-dimer equilibrium to favour the dissociation of PAS2 subunits. Movement of the PAS2 protomers is likely to generate re-organisation of the H and GHKL domains of NifL to promote binding of NifL to NifA. However, there are several aspects of the redox signal transduction pathway in NifL that remain unclear. For example, how do changes in the

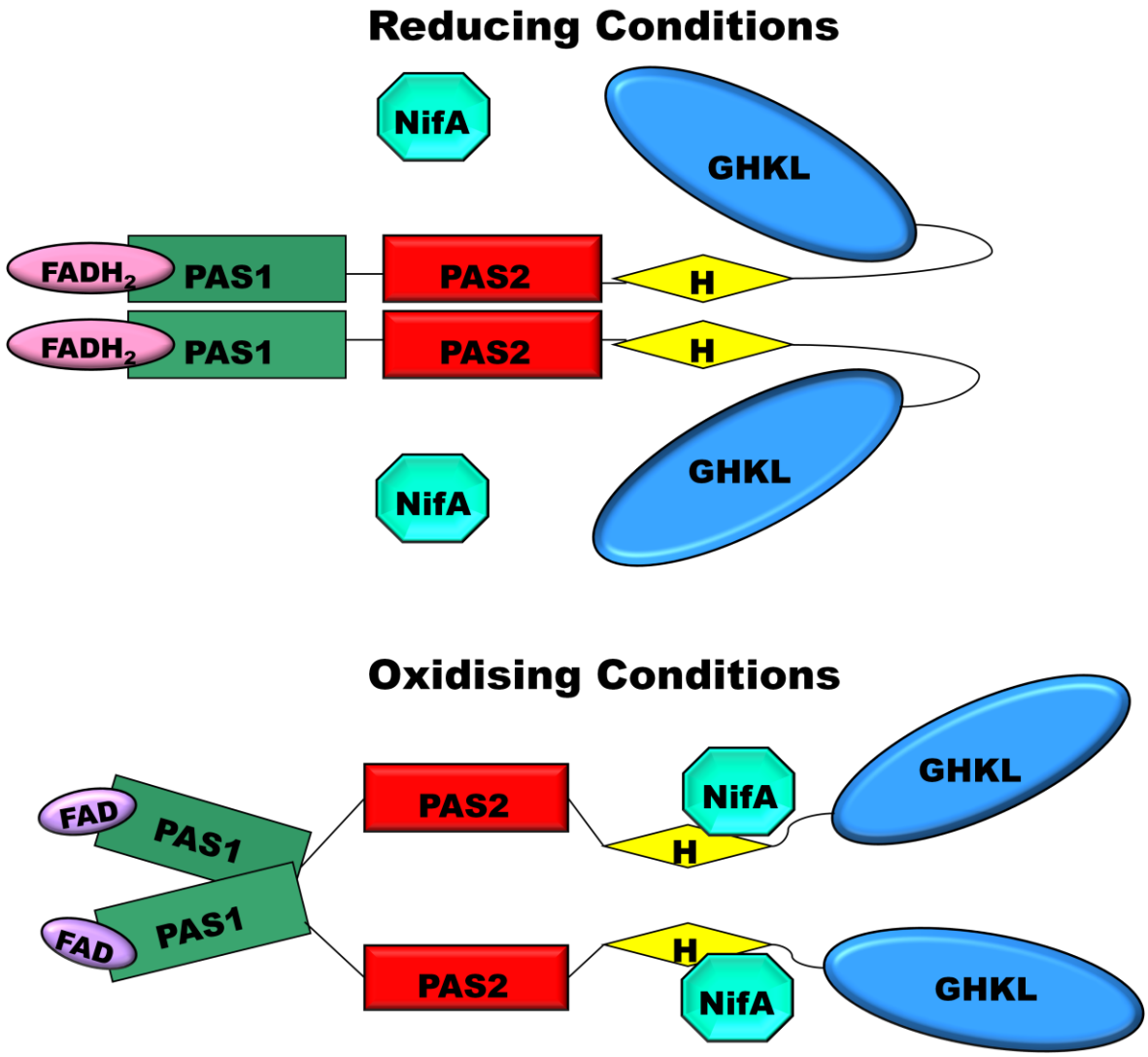


Figure 6.1. Model of redox signal transduction in NifL. Under reducing conditions, when the FAD co-factor is fully protonated, the PAS2 domain is maintained in the dimeric state and the H and GHKL domains are in a conformation that prevents NifA from accessing the surfaces of NifL that mediate the interaction with NifA. Under these conditions, NifA escapes inhibition by NifL. Oxidation of the FAD co-factor generates a conformational change in the PAS1 domain, which is communicated to the PAS2 domain, triggering a movement of PAS2 protomers. This shift in the quaternary structural arrangement of the PAS2 domain results in a re-organisation of the H and GHKL domains of NifL to promote inhibition of NifA activity.

association state of the PAS2 domain influence the activity of the C-terminal domains? The nature of the interaction between PAS2 and the output domains of NifL is ill-defined and although it is clear that changes in the signalling state of the PAS2 domain must impact upon the conformation and/or quaternary arrangement of the H and GHKL domains, the mechanism requires elucidation. The transmission of signals between the NifL PAS domains is another aspect of signal transduction that requires further investigation. Mutagenic analysis indicates that the inter-domain region is important in signal relay but the precise mechanism is not known. Finally, it is likely that switching of the PAS2 domain between the "on" (monomeric) and "off" (dimeric) signalling states involves a conformational change, particularly given that one of the "on" substitutions identified in this domain (L199R) apparently influences oligomerisation despite being located some distance from the dimerisation interface (see Chapter 4.4). However, the nature of this conformational change remains unclear. An improved understanding of each aspect of signal transduction discussed here might, perhaps, arise from further structural data on the NifL protein.

PAS domains have been shown to undergo signal-dependant conformational changes, particularly in the C-terminal beta sheet regions (Card et al., 2005; Erbel et al., 2003; Evans et al., 2009), that may provoke alterations in quaternary structure (see Chapter 1). The *Bacilus subtilis* YtvA protein provides a well-studied example of a PAS domain that exhibits a stimulus dependant quaternary structural change. YtvA regulates responses to blue-light illumination and contains an N-terminal PAS domain that binds an FMN co-factor. This domain forms a stable dimer, the subunits of which rotate by 4-5^{o} relative to one another in response to blue light illumination (Möglich and Moffat, 2007). Similarly, oxidation of the heme iron in the dimeric PAS-A domain from the *E. coli* direct oxygen sensor (*Ec*DOS) protein leads to a 3^{o} rotation of the subunits with respect to each other

(Kurokawa et al., 2004) and ligand binding to the heme co-factor in the sensory PAS domain from *bj*FixL results in a ~ 2^{o} rotation of the protomers (Ayers and Moffat, 2008). Each of these proteins is discussed in detail in Chapter 1.2. A similar rotation may enable the subunits of the NifL PAS1 domain to undergo a "scissor-like" movement with respect to one another in response to redox changes, to influence in turn the quaternary structure of PAS2 (Figure 6.1). Since redox signal transduction by PAS1 appears to alter the oligomerisation state of PAS2, resulting in dissociation of the PAS2 dimer under oxidising conditions, it is possible that this provides a mechanism for amplifying the signal to effect the conformational movements necessary to switch the activity of the C-terminal domains of NifL. The importance of the monomer-dimer equilibrium in signal relay by the PAS2 domain mirrors the properties of other PAS domains in which homo or hetero-dimerisation plays an important role in the signalling mechanism. Examples include the light-sensing proteins Vivid and phototropin (Nakasone et al., 2008; Zoltowski et al., 2007), the mammalian transcription factors AhR and ARNT (Perdew, 1988; Reisz-Porszasz et al., 1994) and the *B. subtilis* KinA protein (Lee et al., 2008). Thus, the importance of alterations in quaternary structure has been demonstrated in the signalling mechanisms of evolutionarily distant PAS domains of diverse function. In this respect, the findings presented in this thesis are congruent with previous research and highlight the importance of quaternary structural plasticity in the signalling mechanism of PAS domains.

The properties of the NifL PAS2 domain suggest that it is a representative of an emerging subclass of PAS domains that are apparently involved in signal relay rather than sensing. The presence of multiple PAS domains within a single protein is surprisingly common; the SMART (http://smart.embl.de/) and Pfam (Finn et al., 2006) databases both indicate that a total of over 21,000 PAS domains are present in around 14,000 proteins. Of the relatively few PAS domains characterised to date, it is often the case that no obvious

sensory function can be attributed to the additional PAS domain(s) within a tandem pair (or triplet). There are several examples of well-studied PAS domains that may belong to this subclass. DcuS is a membrane-embedded histidine protein kinase that contains a periplasmic C4-dicarboxylate-sensing PAS domain (PASp), which transmits signals to a cytoplasmic PAS domain (PASc) via two transmembrane helices. The PASc domain has no known role in signal perception but the structural plasticity of this domain is believed to be important for signal transduction to the histidine kinase domains. When substitutions in PASc resulting in ligand-independent (constitutive) activation of DcuS were modelled on the dimeric crystal structure of the NifL PAS1 domain, it was observed that these residues were located close to the A'α helix that forms part of the extended dimerisation interface (Etzkorn et al., 2008). This suggests a model in which signal perception by the PASp domain in DcuS impacts upon the stability of the PASc dimer interface, analogous to the influence of the NifL PAS1 domain on the quaternary structure of NifL PAS2. Another example of the potential role of PAS domains in signal relay is provided by KinA, a cytoplasmic histidine protein kinase that regulates sporulation in *B. subtilis* in response to an unknown signal(s). KinA has an N-terminal sensory region that consists of three PAS domains. The oligomerisation state of the most N-terminal of these PAS domains, PAS-A, is important for histidine kinase function. A combination of biophysical and biochemical experiments indicate that this domain exhibits considerable structural plasticity and substitutions that favour the monomeric state of the isolated PAS-A domain activate autokinase activity in the full length KinA protein (Lee et al., 2008). Intriguingly, structural predictions indicate that the substitution in PAS-A giving rise to the greatest level of kinase activity, Y29A, is located at a position equivalent to L175 in the NifL PAS2 domain (Figure 3.1). The Y29A and L175A substitutions both strongly disrupt dimerisation of their respective PAS domains. Further analogies can be drawn between the

sensory regions of KinA and NifL in that both contain duplicate PAS domains that have no known role in signal perception yet can influence the conformation of downstream domains via changes in quaternary structure. Other examples of this class may include the redox sensing region of MmoS, which contains an N-terminal FAD-binding sensory PAS domain and a more C-terminal PAS domain that has no apparent co-factor or ligand-binding pocket (Ukaegbu and Rosenzweig, 2009) and *Ec*DOS, which contains an N-terminal heme-binding PAS domain in tandem with a second PAS domain of unknown function (Sasakura et al., 2006).

Another pertinent example of a "signal relay" PAS domain has been provided by recent mutagenic and crystallographic studies of the DctB protein from *S. meliloti* (Nan et al., 2010; Zhou et al., 2008). DctB is a dimeric, membrane-bound histidine protein kinase that regulates the transcription of C_4-dicarboxylate transport (*dct*) genes in rhizobia. The histidine kinase output domains are located in the cytoplasm whilst the sensory region of the DctB protein (DctBp) is periplasmic and contains two PAS domains, known as the membrane-proximal (PASp) and membrane-distal PAS (PASd) domains (Figure 6.2). Crystal structures of DctBp in both the *apo* and ligand-bound states indicate that the PASd domain binds C_4-dicarboxylates whereas the membrane-proximal PASp domain is not associated with any co-factor or ligand (Zhou et al., 2008). In the absence of C_4-dicarboxylate ligands, both PAS domains form homo-dimers to maintain DctBp in the dimeric form. Ligand binding results in a large decrease in affinity between DctBp subunits, imparted predominantly by monomerisation of the PASp domain (Nan et al., 2010). That is, ligand binding to the PASd domain induces a conformational change that is relayed to the PASp domain, resulting in dissociation of PASp subunits. Using directed mutational analysis, the authors were able to identify substitutions in the PASp domain that lock the DctB protein in the active conformer, presumably by destabilising the PASp dimer

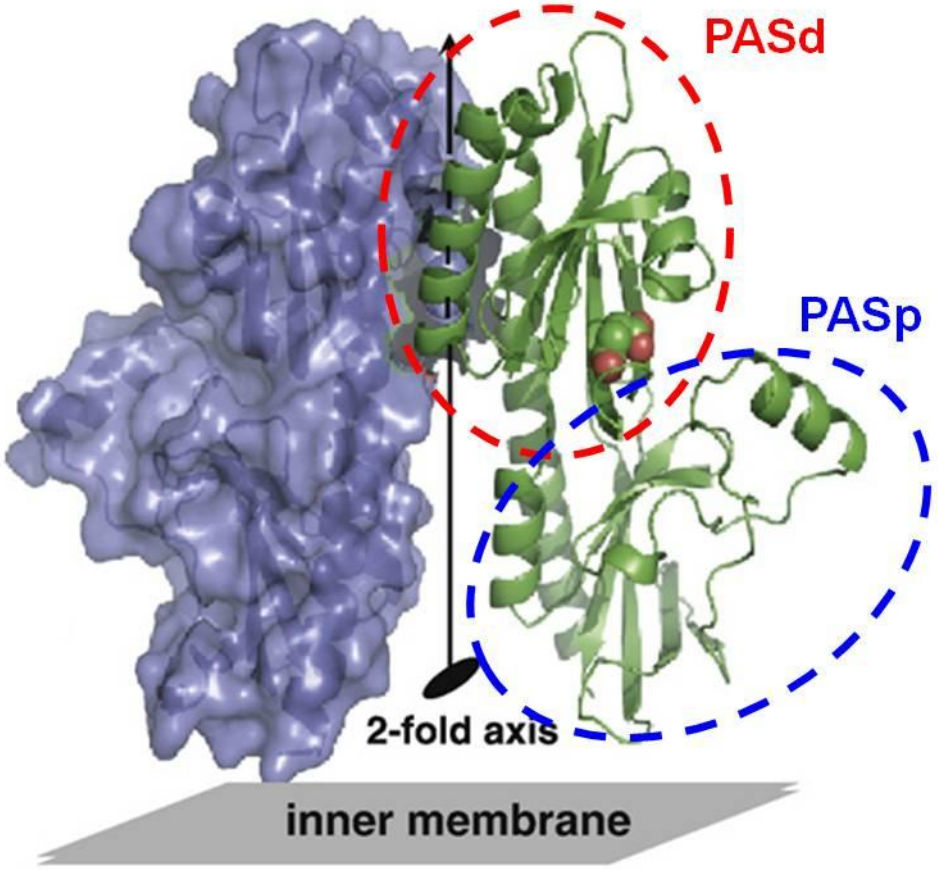


Figure 6.2. Crystal structure of the periplasmic region of DctB (DctBp) in the succinate-bound form (Zhou et al., 2008). One subunit of the DctB dimer is shown as a space-filling model and the other as a ribbon diagram. The membrane-distal (red) and membrane-proximal (blue) PAS domains are circled. The 2-fold crystallographic axis of symmetry and the approximate position of the inner membrane are marked and the succinate molecule is represented as a ball diagram.

(Nan et al., 2010). It has been postulated that changes in the association state of this domain mediate transmission of signals from the ligand-binding periplasmic sensor region to the cytoplasmic histidine kinase effector domains (Nan et al., 2010). The signal transduction mechanism of DctB has clear parallels with that proposed for NifL and, together with the available data regarding the function and abundance of tandem PAS domains, these model systems provide strong evidence to support the existence of an emergent class of "signal relay" PAS domains.

The data presented in this thesis demonstrate that the two PAS domains in NifL function in tandem and that the quaternary structure of PAS2 is responsive to signal perception by PAS1. In addition, substitutions in NifL PAS2 are sufficient to either block the relay of signals from PAS1 or mimic the active state in the absence of a signal. However, for cytoplasmic proteins, the benefits of having an additional PAS domain solely for signal relay are not readily apparent. In the examples discussed above multiple PAS domains might be advantageous for amplification of structural signals, thus driving appropriate conformational changes in the C-terminal DHp (or H in the case of NifL) and GHKL domains. Current models for histidine kinase autophosphorylation and phospho-transfer, based on crystal structures, suggest that large movements of the DHp and GHKL domains relative to one another are required for signalling (Albanesi et al., 2009; Marina et al., 2005). However, not all histidine protein kinases have multiple PAS domains and, for example, experiments with chimeric proteins demonstrate that the single light-sensing PAS domain from YtvA is sufficient to provide input signal specificity and regulate the activity of the C-terminal kinase domains of FixL in response to light (Möglich and Moffat, 2007). Both YtvA and FixL contain an α-helical coiled-coil linker, connecting their input and output domains, comprising the Jα helix often found associated with PAS domains. Signal-dependent unfolding of this coiled-coil sequence may result in rotational movements that

regulate kinase activity (Harper et al., 2004; Harper et al., 2003; Möglich et al., 2009a; Möglich and Moffat, 2007). In proteins such as NifL, DcuS and KinA it is possible that this α-helical coiled-coil mechanism is replaced by the "signal relay" PAS domain, which provides conformational flexibility associated with the inter-conversion between the dimeric and monomeric forms. Additionally, duplicate PAS domains may provide a fulcrum for conformational changes in large, multi-domain proteins. Speculative evidence to support this notion can be found by comparing the FixL proteins from *B. japonicum* and *S. meliloti*. As mentioned in Chapter 1.2, the domain architectures of these proteins are not identical. *S. meliloti* FixL (*Sm*FixL) is a membrane-bound protein containing an N-terminal transmembrane region, a PAS domain and C-terminal histidine kinase effector domains. By contrast, *B. japonicum* FixL (*Bj*FixL) is a cytoplasmic protein that lacks the transmembrane region found in *Sm*FixL but instead contains an additional N-terminal PAS domain. This additional PAS domain may replace the membrane anchor as a fulcrum for conformational changes, particularly as removal of the domain narrows the output range of the protein (Möglich et al., 2010). That is, although the N-terminal PAS domain has no apparent role in signal perception, the presence of this domain facilitates higher activity from the output domains under inducing conditions and lower activity in the absence of a signal. Hence, the efficacy of signal transduction is aided by the presence of a non-sensory PAS domain in *Bj*FixL. However, examples of cytoplasmic histidine kinases containing a single PAS domain can be readily found in the SMART and Pfam databases. Given the abundance of PAS domains and their diversity of function, it seems unlikely that duplicate PAS domains perform identical roles in all proteins in which they are found. Rather, it is probable that these highly adaptable modules are utilised for wide ranging purposes that extend beyond the functions identified to date. Moreover, since the number of proteins containing multiple PAS domains of unknown function vastly exceeds the relatively few

studied examples, there remains a possibility that some of these additional PAS domains respond to as yet undiscovered stimuli and can modulate signal relay accordingly. In other words, there is potential for some "signal relay" PAS domains to act as biological "logic gates" to aid the integration of multiple signals within complex modular proteins.

Chapter 7 - References

Akbar, S., Gaidenko, T.A., Kang, C.M., O'Reilly, M., Devine, K.M., and Price, C.W. (2001). New family of regulators in the environmental signaling pathway which activates the general stress transcription factor sigma(B) of *Bacillus subtilis*. *J Bacteriol* **183**, 1329-1338.

Albanesi, D., Martin, M., Trajtenberg, F., Mansilla, M.C., Haouz, A., Alzari, P.M., de Mendoza, D., and Buschiazzo, A. (2009). Structural plasticity and catalysis regulation of a thermosensor histidine kinase. *PNAS* **106**, 16185-16190.

Arcondeguy, T., Jack, R., and Merrick, M. (2001). P(II) signal transduction proteins, pivotal players in microbial nitrogen control. *Microbiol Mol Biol Rev* **65**, 80-105.

Avila-Perez, M., Hellingwerf, K.J., and Kort, R. (2006). Blue light activates the sigmaB-dependent stress response of *Bacillus subtilis* via YtvA. *J Bacteriol* **188**, 6411-6414.

Avila-Perez, M., Vreede, J., Tang, Y., Bende, O., Losi, A., Gartner, W., and Hellingwerf, K. (2009). *In vivo* mutational analysis of YtvA from *Bacillus subtilis*: mechanism of light activation of the general stress response. *J Biol Chem* **284**, 24958-24964.

Ayers, R.A., and Moffat, K. (2008). Changes in quaternary structure in the signaling mechanisms of PAS domains. *Biochemistry* **47**, 12078-12086.

Baca, M., Borgstahl, G.E., Boissinot, M., Burke, P.M., Williams, D.R., Slater, K.A., and Getzoff, E.D. (1994). Complete chemical structure of photoactive yellow protein: novel thioester-linked 4-hydroxycinnamyl chromophore and photocycle chemistry. *Biochemistry* **33**, 14369-14377.

Ban, C., Junop, M., and Yang, W. (1999). Transformation of MutL by ATP binding and hydrolysis: a switch in DNA mismatch repair. *Cell* **97**, 85-97.

Barrett, J., Ray, P., Sobczyk, A., Little, R., and Dixon, R. (2001). Concerted inhibition of the transcriptional activation functions of the enhancer-binding protein NIFA by the anti-activator NIFL. *Mol Microbiol* **39**, 480-493.

Bass, R.B., Butler, S.L., Chervitz, S.A., Gloor, S.L., and Falke, J.J. (2007). Use of site-directed cysteine and disulfide chemistry to probe protein structure and dynamics: applications to soluble and transmembrane receptors of bacterial chemotaxis. *Methods Enzymol* **423**, 25-51.

Baumert, H.G., and Fasold, H. (1989). Cross-linking techniques. *Methods Enzymol* **172**, 584-609.

Bick, M.J., Lamour, V., Rajashankar, K.R., Gordiyenko, Y., Robinson, C.V., and Darst, S.A. (2009). How to switch off a histidine kinase: crystal structure of *Geobacillus stearothermophilus* KinB with the inhibitor Sda. *J Mol Biol* **386**, 163-177.

Bilwes, A.M., Alex, L.A., Crane, B.R., and Simon, M.I. (1999). Structure of CheA, a signal-transducing histidine kinase. *Cell* **96**, 131-141.

Blanco, G., Drummond, M., Woodley, P., and Kennedy, C. (1993). Sequence and molecular analysis of the *nifL* gene of *Azotobacter vinelandii*. *Mol Microbiol* **9**, 869-879.

Borgstahl, G.E., Williams, D.R., and Getzoff, E.D. (1995). 1.4 A structure of photoactive yellow protein, a cytosolic photoreceptor: unusual fold, active site, and chromophore. *Biochemistry* **34**, 6278-6287.

Bott, M., Meyer, M., and Dimroth, P. (1995). Regulation of anaerobic citrate metabolism in *Klebsiella pneumoniae*. *Mol Microbiol* **18**, 533-546.

Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**, 248-254.

Brudler, R., Gessner, C.R., Li, S., Tyndall, S., Getzoff, E.D., and Woods, V.L., Jr. (2006). PAS domain allostery and light-induced conformational changes in photoactive yellow protein upon I2 intermediate formation, probed with enhanced hydrogen/deuterium exchange mass spectrometry. *J Mol Biol* **363**, 148-160.

Buck, M., Gallegos, M.T., Studholme, D.J., Guo, Y., and Gralla, J.D. (2000). The bacterial enhancer dependant σ^{54} (σ^{N}) transcription factor. *J Bacteriol* **182**, 4129-4136.

Buttani, V., Gartner, W., and Losi, A. (2007). NTP-binding properties of the blue-light receptor YtvA and effects of the E105L mutation. *Eur Biophys J* **36**, 831-839.

Buttani, V., Losi, A., Polverini, E., and Gartner, W. (2006). Blue news: NTP binding properties of the blue-light sensitive YtvA protein from *Bacillus subtilis*. *FEBS Lett* **580**, 3818-3822.

Card, P.B., Erbel, P.J.A., and Gardner, K.H. (2005). Structural Basis of ARNT PAS-B Dimerization: Use of a Common Beta-sheet Interface for Hetero- and Homodimerization. *J Mol Biol* **353**, 664-677.

Casino, P., Rubio, V., and Marina, A. (2009). Structural insight into partner specificity and phosphoryl transfer in two-component signal transduction. *Cell* **139**, 325-336.

Cheung, J., Bingman, C.A., Reyngold, M., Hendrickson, W.A., and Waldburger, C.D. (2008). Crystal structure of a functional dimer of the PhoQ sensor domain. *J Biol Chem* **283**, 13762-13770.

Cho, U.S., Bader, M.W., Amaya, M.F., Daley, M.E., Klevit, R.E., Miller, S.I., and Xu, W. (2006). Metal bridges between the PhoQ sensor domain and the membrane regulate transmembrane signaling. *J Mol Biol* **356**, 1193-1206.

Christen, M., Christen, B., Folcher, M., Schauerte, A., and Jenal, U. (2005). Identification and characterization of a cyclic di-GMP-specific phosphodiesterase and its allosteric control by GTP. *J Biol Chem* **280**, 30829-30837.

Contreras, A., Drummond, M., Bali, A., Blanco, G., Garcia, E., Bush, G., Kennedy, C., and Merrick, M. (1991). The product of the nitrogen fixation regulatory gene *nfrX* of *Azotobacter vinelandii* is functionally and structurally homologous to the uridylyltransferase encoded by *glnD* in enteric bacteria. *J Bacteriol* **173**, 7741-7749.

Csaki, R., Bodrossy, L., Klem, J., Murrell, J.C., and Kovacs, K.L. (2003). Genes involved in the copper-dependent regulation of soluble methane monooxygenase of *Methylococcus capsulatus* (Bath): cloning, sequencing and mutational analysis. *Microbiology* **149**, 1785-1795.

Demeler, B. (2005). UltraScan - A comprehensive data analysis software package for analytical ultracentrifugation experiments. In *Analytical Ultracentrifugation* Scott, D, Harding, S, and Rowe, A (eds) Cambridge: RSC publishing, pp 210-229.

Denison, M.S., Pandini, A., Nagy, S.R., Baldwin, E.P., and Bonati, L. (2002). Ligand binding and activation of the Ah receptor. *Chem Biol Interact* **141**, 3-24.

Dixon, R., and Kahn, D. (2004). Genetic regulation of biological nitrogen fixation. *Nat Rev Microbiol* **2**, 621-631.

Drummond, M.H., and Wootton, J.C. (1987). Sequence of nifL from *Klebsiella pneumoniae*: mode of action and relationship to two families of regulatory proteins. *Mol Microbiol* **1**, 37-44.

Dutta, R., and Inouye, M. (2000). GHKL, an emergent ATPase/kinase superfamily. *Trends Biochem Sci* **25**, 24-28.

El-Mashtoly, S.F., Nakashima, S., Tanaka, A., Shimizu, T., and Kitagawa, T. (2008). Roles of Arg-97 and Phe-113 in regulation of distal ligand binding to heme in the sensor domain of Ec DOS protein. Resonance Raman and mutation study. *J Biol Chem* **283**, 19000-19010.

Erbel, P.J., Card, P.B., Karakuzu, O., Bruick, R.K., and Gardner, K.H. (2003). Structural basis for PAS domain heterodimerization in the basic helix-loop-helix-PAS transcription factor hypoxia-inducible factor. *PNAS* **100**, 15504-15509.

Etzkorn, M., Kneuper, H., Dunnwald, P., Vijayan, V., Kramer, J., Griesinger, C., Becker, S., Unden, G., and Baldus, M. (2008). Plasticity of the PAS domain and a potential role for signal transduction in the histidine kinase DcuS. *Nat Struct Mol Biol* **15**, 1031-1039.

Evans, M.R., Card, P.B., and Gardner, K.H. (2009). ARNT PAS-B has a fragile native state structure with an alternative beta-sheet register nearby in sequence space. *PNAS* **106**, 2617-2622.

Eydmann, T., Soderback, E., Jones, T., Hill, S., Austin, S., and Dixon, R. (1995). Transcriptional activation of the nitrogenase promoter in vitro: adenosine nucleotides are required for inhibition of NIFA activity by NIFL. *J Bacteriol* **177**, 1186-1195.

Fabret, C., Feher, V.A., and Hoch, J.A. (1999). Two-component signal transduction in *Bacillus subtilis*: how one organism sees its world. *J Bacteriol* **181**, 1975-1983.

Fedtke, I., Kamps, A., Krismer, B., and Gotz, F. (2002). The nitrate reductase and nitrite reductase operons and the *narT* gene of *Staphylococcus carnosus* are positively controlled by the novel two-component system NreBC. *J Bacteriol* **184**, 6624-6634.

Finn, R.D., Mistry, J., Schuster-Bockler, B., Griffiths-Jones, S., Hollich, V., Lassmann, T., Moxon, S., Marshall, M., Khanna, A., Durbin, R., Eddy, S.R., Sonnhammer, E.L., and Bateman, A. (2006). Pfam: clans, web tools and services. *Nucleic Acids Res* **34**, D247-251.

Gaidenko, T.A., Kim, T.J., Weigel, A.L., Brody, M.S., and Price, C.W. (2006). The blue-light receptor YtvA acts in the environmental stress signaling pathway of *Bacillus subtilis*. *J Bacteriol* **188**, 6387-6395.

Genick, U.K., Borgstahl, G.E., Ng, K., Ren, Z., Pradervand, C., Burke, P.M., Srajer, V., Teng, T.Y., Schildkamp, W., McRee, D.E., Moffat, K., and Getzoff, E.D. (1997). Structure of a protein photocycle intermediate by millisecond time-resolved crystallography. *Science* **275**, 1471-1475.

Genick, U.K., Soltis, S.M., Kuhn, P., Canestrelli, I.L., and Getzoff, E.D. (1998). Structure at 0.85 Å resolution of an early protein photocycle intermediate. *Nature* **392**, 206-209.

Gilles-Gonzalez, M.A., Caceres, A.I., Sousa, E.H., Tomchick, D.R., Brautigam, C., Gonzalez, C., and Machius, M. (2006). A proximal arginine R206 participates in switching of the *Bradyrhizobium japonicum* FixL oxygen sensor. *J Mol Biol* **360**, 80-89.

Gilles-Gonzalez, M.A., Ditta, G.S., and Helinski, D.R. (1991). A haemoprotein with kinase activity encoded by the oxygen sensor of *Rhizobium meliloti*. *Nature* **350**, 170-172.

Gilles-Gonzalez, M.A., and Gonzalez, G. (2004). Signal transduction by heme-containing PAS-domain proteins. *J Appl Physiol* **96**, 774-783.

Glekas, G.D., Foster, R.M., Cates, J.R., Estrella, J.A., Wawrzyniak, M.J., Rao, C.V., and Ordal, G.W. (2010). A pas domain binds asparagine in the chemotaxis receptor MCPB in *Bacillus subtilis*. *J Biol Chem* **285**, 1870-1878.

Golby, P., Davies, S., Kelly, D.J., Guest, J.R., and Andrews, S.C. (1999). Identification and characterization of a two-component sensor-kinase and response-regulator system (DcuS-DcuR) controlling gene expression in response to C4-dicarboxylates in *Escherichia coli*. *J Bacteriol* **181**, 1238-1248.

Gong, W., Hao, B., Mansy, S.S., Gonzalez, G., Gilles-Gonzalez, M.A., and Chan, M.K. (1998). Structure of a biological oxygen sensor: A new mechanism for heme-driven signal transduction. *PNAS* **95**, 15177-15182.

Grebe, T.W., and Stock, J.B. (1999). The histidine protein kinase superfamily. *Adv Microb Physiol* **41**, 139-227.

Groot, M.L., van Wilderen, L.J., Larsen, D.S., van der Horst, M.A., van Stokkum, I.H., Hellingwerf, K.J., and van Grondelle, R. (2003). Initial steps of signal generation in photoactive yellow protein revealed with femtosecond mid-infrared spectroscopy. *Biochemistry* **42**, 10054-10059.

Hagiwara, D., Yamashino, T., and Mizuno, T. (2004). Genome-wide comparison of the His-to-Asp phosphorelay signaling components of three symbiotic genera of *Rhizobia*. *DNA Res* **11**, 57-65.

Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* **166**, 557-80.

Hankinson, O. (1995). The aryl hydrocarbon receptor complex. *Annu Rev Pharmacol Toxicol* **35**, 307-340.

Harper, S.M., Christie, J.M., and Gardner, K.H. (2004). Disruption of the LOV-Jalpha helix interaction activates phototropin kinase activity. *Biochemistry* **43**, 16184-16192.

Harper, S.M., Neil, L.C., and Gardner, K.H. (2003). Structural basis of a phototropin light switch. *Science* **301**, 1541-1544.

Hefti, M.H., Francoijs, K.J., de Vries, S.C., Dixon, R., and Vervoort, J. (2004). The PAS fold. A redefinition of the PAS domain based upon structural prediction. *Eur J Biochem* **271**, 1198-1208.

Hendriks, J., Gensch, T., Hviid, L., van Der Horst, M.A., Hellingwerf, K.J., and van Thor, J.J. (2002). Transient exposure of hydrophobic surface in the photoactive yellow protein monitored with Nile Red. *Biophys J* **82**, 1632-1643.

Henry, E.C., and Gasiewicz, T.A. (2003). Agonist but not antagonist ligands induce conformational change in the mouse aryl hydrocarbon receptor as detected by partial proteolysis. *Mol Pharmacol* **63**, 392-400.

Hill, S. (1992). Physiology of free-living heterotrophs. In *Biological nitrogen fixation* Stacey, G, Burris, R H, and Evans H J (eds) New York, NY: Chapman and Hall, pp 87-134.

Hill, S., Austin, S., Eydmann, T., Jones, T., and Dixon, R. (1996). *Azotobacter vinelandii* NIFL is a flavoprotein that modulates transcriptional activation of nitrogen-fixation genes via a redox-sensitive switch. *PNAS* **93**, 2143-2148.

Ho, Y.S., Burden, L.M., and Hurley, J.H. (2000). Structure of the GAF domain, a ubiquitous signaling motif and a new class of cyclic GMP receptor. *EMBO J* **19**, 5288-5299.

Hoff, W.D., Xie, A., Van Stokkum, I.H., Tang, X.J., Gural, J., Kroon, A.R., and Hellingwerf, K.J. (1999). Global conformational changes upon receptor stimulation in photoactive yellow protein. *Biochemistry* **38**, 1009-1017.

Horan, T., Wen, J., Arakawa, T., Liu, N., Brankow, D., Hu, S., Ratzkin, B., and Philo, J.S. (1995). Binding of Neu differentiation factor with the extracellular domain of Her2 and Her3. *J Biol Chem* **270**, 24604-24608.

Hord, N.G., and Perdew, G.H. (1994). Physicochemical and immunocytochemical analysis of the aryl hydrocarbon receptor nuclear translocator: characterization of two monoclonal antibodies to the aryl hydrocarbon receptor nuclear translocator. *Mol Pharmacol* **46**, 618-626.

Hu, P., Leighton, T., Ishkhanova, G., and Kustu, S. (1999). Sensing of nitrogen limitation by *Bacillus subtilis*: comparison to enteric bacteria. *J Bacteriol* **181**, 5042-5050.

Huang, L.E., Gu, J., Schau, M., and Bunn, H.F. (1998). Regulation of hypoxia-inducible factor 1alpha is mediated by an O_2-dependent degradation domain via the ubiquitin-proteasome pathway. *PNAS* **95**, 7987-7992.

Huang, Z.J., Edery, I., and Rosbash, M. (1993). PAS is a dimerization domain common to *Drosophila* period and several transcription factors. *Nature* **364**, 259-262.

Hutchings, M.I., Hoskisson, P.A., Chandra, G., and Buttner, M.J. (2004). Sensing and responding to diverse extracellular signals? Analysis of the sensor kinases and response regulators of *Streptomyces coelicolor* A3(2). *Microbiology* **150**, 2795-2806.

Ikeda, T.P., Shauger, A.E., and Kustu, S. (1996). *Salmonella typhimurium* apparently perceives external nitrogen limitation as internal glutamine limitation. *J Mol Biol* **259**, 589-607.

Imamoto, Y., and Kataoka, M. (2007). Structure and photoreaction of photoactive yellow protein, a structural prototype of the PAS domain superfamily. *Photochem Photobiol* **83**, 40-49.

Ishitsuka, Y., Araki, Y., Tanaka, A., Igarashi, J., Ito, O., and Shimizu, T. (2008). Arg97 at the heme-distal side of the isolated heme-bound PAS domain of a heme-based oxygen sensor from *Escherichia coli* (Ec DOS) plays critical roles in autoxidation and binding to gases, particularly O_2. *Biochemistry* **47**, 8874-8884.

Jasaitis, A., Hola, K., Bouzhir-Sima, L., Lambry, J.C., Balland, V., Vos, M.H., and Liebl, U. (2006). Role of distal arginine in early sensing intermediates in the heme domain of the oxygen sensor FixL. *Biochemistry* **45**, 6018-6026.

Jiang, Z., Swem, L.R., Rushing, B.G., Devanathan, S., Tollin, G., and Bauer, C.E. (1999). Bacterial photoreceptor with similarity to photoactive yellow protein and plant phytochromes. *Science* **285**, 406-409.

Kamikubo, H., Koyama, T., Hayashi, M., Shirai, K., Yamazaki, Y., Imamoto, Y., and Kataoka, M. (2008). The photoreaction of the photoactive yellow protein domain in the light sensor histidine kinase Ppr is influenced by the C-terminal domains. *Photochem Photobiol* **84**, 895-902.

Kamps, A., Achebach, S., Fedtke, I., Unden, G., and Gotz, F. (2004). *Staphylococcal* NreB: an O_2-sensing histidine protein kinase with an O_2-labile iron-sulphur cluster of the FNR type. *Mol Microbiol* **52**, 713-723.

Kanamaru, K., Aiba, H., Mizushima, S., and Mizuno, T. (1989). Signal transduction and osmoregulation in *Escherichia coli*. A single amino acid change in the protein kinase, EnvZ, results in loss of its phosphorylation and dephosphorylation abilities with respect to the activator protein, OmpR. *J Biol Chem* **264**, 21633-21637.

Karimova, G., Pidoux, J., Ullmann, A., and Ladant, D. (1998). A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *PNAS* **95**, 5752-5756.

Karimova, G., Ullmann, A., and Ladant, D. (2000). A bacterial two-hybrid system that exploits a cAMP signaling cascade in *Escherichia coli*. *Methods Enzymol* **328**, 59-73.

Kaspar, S., and Bott, M. (2002). The sensor kinase CitA (DpiB) of *Escherichia coli* functions as a high-affinity citrate receptor. *Arch Microbiol* **177**, 313-321.

Kaspar, S., Perozzo, R., Reinelt, S., Meyer, M., Pfister, K., Scapozza, L., and Bott, M. (1999). The periplasmic domain of the histidine autokinase CitA functions as a highly specific citrate receptor. *Mol Microbiol* **33**, 858-872.

Keener, J., and Kustu, S. (1988). Protein kinase and phosphoprotein phosphatase activities of nitrogen regulatory proteins NTRB and NTRC of enteric bacteria: roles of the conserved amino-terminal domain of NTRC. *PNAS* **85**, 4976-4980.

Key, J., Hefti, M., Purcell, E.B., and Moffat, K. (2007a). Structure of the Redox Sensor Domain of *Azotobacter vinelandii* NifL at Atomic Resolution: Signaling, Dimerization, and Mechanism. *Biochemistry* **46**, 3614-3623.

Key, J., and Moffat, K. (2005). Crystal structures of deoxy and CO-bound bjFixLH reveal details of ligand recognition and signaling. *Biochemistry* **44**, 4627-4635.

Key, J., Srajer, V., Pahl, R., and Moffat, K. (2007b). Time-resolved crystallographic studies of the heme domain of the oxygen sensor FixL: structural dynamics of ligand rebinding and their relation to signal transduction. *Biochemistry* **46**, 4706-4715.

Klopprogge, K., Grabbe, R., Hoppert, M., and Schmitz, R.A. (2002). Membrane association of *Klebsiella pneumoniae* NifL is affected by molecular oxygen and combined nitrogen. *Arch Microbiol* **177**, 223-234.

Kneuper, H., Janausch, I.G., Vijayan, V., Zweckstetter, M., Bock, V., Griesinger, C., and Unden, G. (2005). The nature of the stimulus and of the fumarate binding site of the fumarate sensor DcuS of *Escherichia coli*. *J Biol Chem* **280**, 20596-20603.

Kobayashi, K., Ogura, M., Yamaguchi, H., Yoshida, K., Ogasawara, N., Tanaka, T., and Fujita, Y. (2001). Comprehensive DNA microarray analysis of *Bacillus subtilis* two-component regulatory systems. *J Bacteriol* **183**, 7365-7370.

Kort, R., Vonk, H., Xu, X., Hoff, W.D., Crielaard, W., and Hellingwerf, K.J. (1996). Evidence for trans-cis isomerization of the p-coumaric acid chromophore as the photochemical basis of the photocycle of photoactive yellow protein. *FEBS Lett* **382**, 73-78.

Kumar, M., and Chatterji, D. (2008). Cyclic di-GMP: a second messenger required for long-term survival, but not for biofilm formation, in *Mycobacterium smegmatis*. *Microbiology* **154**, 2942-2955.

Kurokawa, H., Lee, D.S., Watanabe, M., Sagami, I., Mikami, B., Raman, C.S., and Shimizu, T. (2004). A redox-controlled molecular switch revealed by the crystal structure of a bacterial heme PAS sensor. *J Biol Chem* **279**, 20186-20193.

Kyte, J., and Doolittle, R. (1982) A Simple Method for Displaying the Hydropathic Character of a Protein. *J Mol Biol* **157**, 105-132.

Kyndt, J.A., Fitch, J.C., Meyer, T.E., and Cusanovich, M.A. (2007). The photoactivated PYP domain of *Rhodospirillum centenum* Ppr accelerates the recovery of the bacteriophytochrome domain after white light illumination. *Biochemistry* **46**, 8256-8262.

Kyndt, J.A., Hurley, J.K., Devreese, B., Meyer, T.E., Cusanovich, M.A., Tollin, G., and Van Beeumen, J.J. (2004a). *Rhodobacter capsulatus* photoactive yellow protein: genetic context, spectral and kinetics characterization, and mutagenesis. *Biochemistry* **43**, 1809-1820.

Kyndt, J.A., Meyer, T.E., and Cusanovich, M.A. (2004b). Photoactive yellow protein, bacteriophytochrome, and sensory rhodopsin in purple phototrophic bacteria. *Photochem Photobiol Sci* **3**, 519-530.

Lee, J., Tomchick, D.R., Brautigam, C.A., Machius, M., Kort, R., Hellingwerf, K.J., and Gardner, K.H. (2008). Changes at the KinA PAS-A dimerization interface influence histidine kinase function. *Biochemistry* **47**, 4051-4064.

Lees, M.J., and Whitelaw, M.L. (1999). Multiple roles of ligand in transforming the dioxin receptor to an active basic helix-loop-helix/PAS transcription factor complex with the nuclear protein Arnt. *Mol Cell Biol* **19**, 5811-5822.

Liebl, U., Bouzhir-Sima, L., Kiger, L., Marden, M.C., Lambry, J.C., Negrerie, M., and Vos, M.H. (2003). Ligand binding dynamics to the heme domain of the oxygen sensor Dos from *Escherichia coli*. *Biochemistry* **42**, 6527-6535.

Lindebro, M.C., Poellinger, L., and Whitelaw, M.L. (1995). Protein-protein interaction via PAS domains: role of the PAS domain in positive and negative regulation of the bHLH/PAS dioxin receptor-Arnt transcription factor complex. *EMBO J* **14**, 3528-3539.

Little, R., Colombo, V., Leech, A., and Dixon, R. (2002). Direct interaction of the NifL regulatory protein with the GlnK signal transducer enables the *Azotobacter vinelandii* NifL-NifA regulatory system to respond to conditions replete for nitrogen. *J Biol Chem* **277**, 15472-15481.

Little, R., and Dixon, R. (2003). The amino-terminal GAF domain of *Azotobacter vinelandii* NifA binds 2-oxoglutarate to resist inhibition by NifL under nitrogen-limiting conditions. *J Biol Chem* **278**, 28711-28718.

Little, R., Martinez-Argudo, I., Perry, S., and Dixon, R. (2007). Role of the H domain of the histidine kinase-like protein NifL in signal transmission. *J Biol Chem* **282**, 13429-13437.

Little, R., Reyes-Ramirez, F., Zhang, Y., van Heeswijk, W.C., and Dixon, R. (2000). Signal transduction to the *Azotobacter vinelandii* NIFL-NIFA regulatory system is influenced directly by interaction with 2-oxoglutarate and the PII regulatory protein. *EMBO J* **19**, 6041-6050.

Little, R., S. Hill, S. Perry, S. Austin, F. Reyes-Ramirez, R. Dixon, and P. Macheroux. (1999). Properties of NifL, a regulatory flavoprotein containing a PAS domain. In *Flavins and flavoproteins* Ghisla, S,Kroneck, P, Macheroux, P and Sund H (eds) Germany, Berlin: Rudolf Weber, pp 737-740.

Ma, X., Sayed, N., Baskaran, P., Beuve, A., and van den Akker, F. (2008). PAS-mediated dimerization of soluble guanylyl cyclase revealed by signal transduction histidine kinase domain crystal structure. *J Biol Chem* **283**, 1167-1178.

Macheroux, P., Hill, S., Austin, S., Eydmann, T., Jones, T., Kim, S.O., Poole, R., and Dixon, R. (1998). Electron donation to the flavoprotein NifL, a redox-sensing transcriptional regulator. *Biochem J* **332 (Pt 2)**, 413-419.

Marina, A., Mott, C., Auyzenberg, A., Hendrickson, W.A., and Waldburger, C.D. (2001). Structural and mutational analysis of the PhoQ histidine kinase catalytic domain. Insight into the reaction mechanism. *J Biol Chem* **276**, 41182-41190.

Marina, A., Waldburger, C.D., and Hendrickson, W.A. (2005). Structure of the entire cytoplasmic portion of a sensor histidine-kinase protein. *EMBO J* **24**, 4247-4259.

Marles-Wright, J., Grant, T., Delumeau, O., van Duinen, G., Firbank, S.J., Lewis, P.J., Murray, J.W., Newman, J.A., Quin, M.B., Race, P.R., Rohou, A., Tichelaar, W., van Heel, M., and Lewis, R.J. (2008). Molecular architecture of the "stressosome," a signal integration and transduction hub. *Science* **322**, 92-96.

Martinez-Argudo, I., Little, R., and Dixon, R. (2004a). A crucial arginine residue is required for a conformational switch in NifL to regulate nitrogen fixation in *Azotobacter vinelandii*. *PNAS* **101**, 16316-16321.

Martinez-Argudo, I., Little, R., and Dixon, R. (2004b). Role of the amino-terminal GAF domain of the NifA activator in controlling the response to the antiactivator protein NifL. *Mol Microbiol* **52**, 1731-1744.

Martinez-Argudo, I., Little, R., Shearer, N., Johnson, P., and Dixon, R. (2004c). The NifL-NifA System: a multidomain transcriptional regulatory complex that integrates environmental signals. *J Bacteriol* **186**, 601-610.

Martinez-Argudo, I., Little, R., Shearer, N., Johnson, P., and Dixon, R. (2005). Nitrogen fixation: key genetic regulatory mechanisms. *Biochem Soc Trans* **33**, 152-156.

Mascher, T., Helmann, J.D., and Unden, G. (2006). Stimulus perception in bacterial signal-transducing histidine kinases. *Microbiol Mol Biol Rev* **70**, 910-938.

Meyer, M., Dimroth, P., and Bott, M. (2001). Catabolite repression of the citrate fermentation genes in *Klebsiella pneumoniae*: evidence for involvement of the cyclic AMP receptor protein. *J Bacteriol* **183**, 5248-5256.

Meyer, T.E. (1985). Isolation and characterization of soluble cytochromes, ferredoxins and other chromophoric proteins from the halophilic phototrophic bacterium *Ectothiorhodospira halophila*. *Biochim Biophys Acta* **806**, 175-183.

Meyer, T.E., Yakali, E., Cusanovich, M.A., and Tollin, G. (1987). Properties of a water-soluble, yellow protein isolated from a halophilic phototrophic bacterium that has photochemical activity analogous to sensory rhodopsin. *Biochemistry* **26**, 418-423.

Mizuno, T. (1997). Compilation of all genes encoding two-component phosphotransfer signal transducers in the genome of *Escherichia coli*. *DNA Res* **4**, 161-168.

Mizuno, T. (1998). His-Asp phosphotransfer signal transduction. *J Biochem (Tokyo)* **123**, 555-563.

Moffett, P., Reece, M., and Pelletier, J. (1997). The murine *Sim-2* gene product inhibits transcription by active repression and functional interference. *Mol Cell Biol* **17**, 4933-4947.

Möglich, A., Ayers, R.A., and Moffat, K. (2009a). Design and signaling mechanism of light-regulated histidine kinases. *J Mol Biol* **385**, 1433-1444.

Möglich, A., Ayers, R.A., and Moffat, K. (2009b). Structure and Signaling Mechanism of Per-ARNT-Sim Domains. *Structure* **17**, 1282-1294.

Möglich, A., Ayers, R.A., and Moffat, K. (2010). Addition at the Molecular Level: Signal Integration in Designed Per-ARNT-Sim Receptor Proteins. *J Mol Biol* **400**, 477-486.

Möglich, A., and Moffat, K. (2007). Structural basis for light-dependent signaling in the dimeric LOV domain of the photosensor YtvA. *J Mol Biol* **373**, 112-126.

Money, T., Jones, T., Dixon, R., and Austin, S. (1999). Isolation and properties of the complex between the enhancer binding protein NIFA and the sensor NIFL. *J Bacteriol* **181**, 4461-4468.

Monsieurs, P., De Keersmaecker, S., Navarre, W.W., Bader, M.W., De Smet, F., McClelland, M., Fang, F.C., De Moor, B., Vanderleyden, J., and Marchal, K. (2005). Comparison of the PhoPQ regulon in *Escherichia coli* and *Salmonella typhimurium*. *J Mol Evol* **60**, 462-474.

Morett, E., and Segovia, L. (1993). The sigma 54 bacterial enhancer-binding protein family: mechanism of action and phylogenetic relationship of their functional domains. *J Bacteriol* **175**, 6067-6074.

Müllner, M., Hammel, O., Mienert, B., Schlag, S., Bill, E., and Unden, G. (2008). A PAS domain with an oxygen labile $[4Fe\text{-}4S]^{2+}$ cluster in the oxygen sensor kinase NreB of *Staphylococcus carnosus*. *Biochemistry* **47**, 13921-13932.

Nakamura, H., Kumita, H., Imai, K., Iizuka, T., and Shiro, Y. (2004). ADP reduces the oxygen-binding affinity of a sensory histidine kinase, FixL: the possibility of an enhanced reciprocating kinase reaction. *PNAS* **101**, 2742-2746.

Nakasako, M., Zikihara, K., Matsuoka, D., Katsura, H., and Tokutomi, S. (2008). Structural basis of the LOV1 dimerization of *Arabidopsis* phototropins 1 and 2. *J Mol Biol* **381**, 718-733.

Nakasone, Y., Eitoku, T., Zikihara, K., Matsuoka, D., Tokutomi, S., and Terazima, M. (2008). Stability of dimer and domain-domain interaction of Arabidopsis phototropin 1 LOV2. *J Mol Biol* **383**, 904-913.

Nambu, J.R., Lewis, J.O., Wharton, K.A., Jr., and Crews, S.T. (1991). The *Drosophila* single-minded gene encodes a helix-loop-helix protein that acts as a master regulator of CNS midline development. *Cell* **67**, 1157-1167.

Nan, B., Liu, X., Zhou, Y., Liu, J., Zhang, L., Wen, J., Zhang, X., Su, X.D., and Wang, Y.P. (2010). From signal perception to signal transduction: ligand-induced dimeric switch of DctB sensory domain in solution. *Mol Microbiol* **75**, 1484-1494.

Neuwald, A.F., Aravind, L., Spouge, J.L., and Koonin, E.V. (1999). AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res* **9**, 27-43.

Ninfa, A.J., and Atkinson, M.R. (2000). PII signal transduction proteins. *Trends Microbiol* **8**, 172-179.

Ninfa, A.J., and Magasanik, B. (1986). Covalent modification of the *glnG* product, NRI, by the *glnL* product, NRII, regulates the transcription of the *glnALG* operon in *Escherichia coli*. *PNAS* **83**, 5909-5913.

Ninfa, E.G., Atkinson, M.R., Kamberov, E.S., and Ninfa, A.J. (1993). Mechanism of autophosphorylation of *Escherichia coli* nitrogen regulator II (NRII or NtrB): trans-phosphorylation between subunits. *J Bacteriol* **175**, 7024-7032.

Nixon, B.T., Ronson, C.W., and Ausubel, F.M. (1986). Two-component regulatory systems responsive to environmental stimuli share strongly conserved domains with the nitrogen assimilation regulatory genes *ntrB* and *ntrC*. *PNAS* **83**, 7850-7854.

Ogura, M., Yamaguchi, H., Yoshida, K., Fujita, Y., and Tanaka, T. (2001). DNA microarray analysis of *Bacillus subtilis* DegU, ComA and PhoP regulons: an approach to comprehensive analysis of *B.subtilis* two-component regulatory systems. *Nucleic Acids Res* **29**, 3804-3813.

Pappalardo, L., Janausch, I.G., Vijayan, V., Zientz, E., Junker, J., Peti, W., Zweckstetter, M., Unden, G., and Griesinger, C. (2003). The NMR structure of the sensory domain of the membranous two-component fumarate sensor (histidine protein kinase) DcuS of *Escherichia coli*. *J Biol Chem* **278**, 39185-39188.

Park, H., Suquet, C., Satterlee, J.D., and Kang, C. (2004). Insights into signal transduction involving PAS domain oxygen-sensing heme proteins from the X-ray crystal structure of *Escherichia coli* Dos heme domain (Ec DosH). *Biochemistry* **43**, 2738-2746.

Parkinson, J.S., and Kofoid, E.C. (1992). Communication modules in bacterial signaling proteins. *Annu Rev Genet* **26**, 71-112.

Pellequer, J.L., Wager-Smith, K.A., Kay, S.A., and Getzoff, E.D. (1998). Photoactive yellow protein: a structural prototype for the three-dimensional fold of the PAS domain superfamily. *PNAS* **95**, 5884-5890.

Perdew, G.H. (1988). Association of the Ah receptor with the 90-kDa heat shock protein. *J Biol Chem* **263**, 13802-13805.

Perry, S., Shearer, N., Little, R., and Dixon, R. (2005). Mutational Analysis of the Nucleotide-binding Domain of the Anti-activator NifL. *J Mol Biol* **346**, 935-949.

Petrulis, J.R., and Perdew, G.H. (2002). The role of chaperone proteins in the aryl hydrocarbon receptor core complex. *Chem Biol Interact* **141**, 25-40.

Pinotsis, N., Petoukhov, M., Lange, S., Svergun, D., Zou, P., Gautel, M., and Wilmanns, M. (2006). Evidence for a dimeric assembly of two titin/telethonin complexes induced by the telethonin C-terminus. *J Struct Biol* **155**, 239-250.

Pioszak, A.A., Jiang, P., and Ninfa, A.J. (2000). The *Escherichia coli* PII signal transduction protein regulates the activities of the two-component system transmitter protein NRII by direct interaction with the kinase domain of the transmitter module. *Biochemistry* **39**, 13450-13461.

Pioszak, A.A., and Ninfa, A.J. (2003). Genetic and biochemical analysis of phosphatase activity of *Escherichia coli* NRII (NtrB) and its regulation by the PII signal transduction protein. *J Bacteriol* **185**, 1299-1315.

Pollenz, R.S., Sattler, C.A., and Poland, A. (1994). The aryl hydrocarbon receptor and aryl hydrocarbon receptor nuclear translocator protein show distinct subcellular localizations in Hepa 1c1c7 cells by immunofluorescence microscopy. *Mol Pharmacol* **45**, 428-438.

Pongratz, I., Antonsson, C., Whitelaw, M.L., and Poellinger, L. (1998). Role of the PAS domain in regulation of dimerization and DNA binding specificity of the dioxin receptor. *Mol Cell Biol* **18**, 4079-4088.

Ponting, C.P., and Aravind, L. (1997). PAS: a multifunctional domain family comes to light. *Curr Biol* **7**, 674-677.

Portnoy, M.E., Rosenzweig, A.C., Rae, T., Huffman, D.L., O'Halloran, T.V., and Culotta, V.C. (1999). Structure-function analyses of the ATX1 metallochaperone. *J Biol Chem* **274**, 15041-15045.

Pos, K.M., and Dimroth, P. (1996). Functional properties of the purified Na(+)-dependent citrate carrier of *Klebsiella pneumoniae*: evidence for asymmetric orientation of the carrier protein in proteoliposomes. *Biochemistry* **35**, 1018-1026.

Puga, A., Ma, C., and Marlowe, J.L. (2009). The aryl hydrocarbon receptor cross-talks with multiple signal transduction pathways. *Biochem Pharmacol* **77**, 713-722.

Radchenko, M.V., Thornton, J., and Merrick, M. (2010). Control of AmtB-GlnK complex formation by intracellular levels of ATP, ADP and 2-oxoglutarate. *J Biol Chem*. Epub. doi: 10.1074/jbc.M110.153908.

Rajagopal, S., Anderson, S., Srajer, V., Schmidt, M., Pahl, R., and Moffat, K. (2005). A structural pathway for signaling in the E46Q mutant of photoactive yellow protein. *Structure* **13**, 55-63.

Rappas, M., Schumacher, J., Beuron, F., Niwa, H., Bordes, P., Wigneshweraraj, S., Keetch, C.A., Robinson, C.V., Buck, M., and Zhang, X. (2005). Structural insights into the activity of enhancer-binding proteins. *Science* **307**, 1972-1975.

Ray, P., Smith, K.J., Parslow, R.A., Dixon, R., and Hyde, E.I. (2002). Secondary structure and DNA binding by the C-terminal domain of the transcriptional activator NifA from *Klebsiella pneumoniae*. *Nucleic Acids Res* **30**, 3972-3980.

Reinelt, S., Hofmann, E., Gerharz, T., Bott, M., and Madden, D.R. (2003). The structure of the periplasmic ligand-binding domain of the sensor kinase CitA reveals the first extracellular PAS domain. *J Biol Chem* **278**, 39189-39196.

Reinhart, F., Huber, A., Thiele, R., and Unden, G. (2009). Response of the oxygen sensor NreB to air in vivo: Fe-S containing and apoNreB in aerobically and anaerobically growing *Staphylococcus carnosus*. *J Bacteriol* **192**, 86-93.

Reisz-Porszasz, S., Probst, M.R., Fukunaga, B.N., and Hankinson, O. (1994). Identification of functional domains of the aryl hydrocarbon receptor nuclear translocator protein (ARNT). *Mol Cell Biol* **14**, 6075-6086.

Reyes-Ramirez, F., Little, R., and Dixon, R. (2001). Role of *Escherichia coli* nitrogen regulatory genes in the nitrogen response of the *Azotobacter vinelandii* NifL-NifA complex. *J Bacteriol* **183**, 3076-3082.

Reyes-Ramirez, F., Little, R., and Dixon, R. (2002). Mutant forms of the *Azotobacter vinelandii* transcriptional activator NifA resistant to inhibition by the NifL regulatory protein. *J Bacteriol* **184**, 6777-6785.

Reynolds, M.F., Ackley, L., Blizman, A., Lutz, Z., Manoff, D., Miles, M., Pace, M., Patterson, J., Pozzessere, N., Saia, K., Sato, R., Smith, D., Tarves, P., Weaver, M., Sieg, K., Lakat-Rofgers, G.S., and Rodgers K.R. (2009). Role of conserved F(alpha)-helix residues in the native fold and stability of the kinase-inhibited oxy state of the oxygen-sensing FixL protein from *Sinorhizobium meliloti*. *Arch Biochem Biophys* **485**, 150-159.

Rodrigue, A., Quentin, Y., Lazdunski, A., Mejean, V., and Foglino, M. (2000). Two-component systems in *Pseudomonas aeruginosa*: why so many? *Trends Microbiol* **8**, 498-504.

Rowlands, J.C., and Gustafsson, J.A. (1997). Aryl hydrocarbon receptor-mediated signal transduction. *Crit Rev Toxicol* **27**, 109-134.

Rudnick, P., Kunz, C., Gunatilaka, M.K., Hines, E.R., and Kennedy, C. (2002). Role of GlnK in NifL-mediated regulation of NifA activity in *Azotobacter vinelandii*. *J Bacteriol* **184**, 812-820.

Saito, K., Ito, E., Hosono, K., Nakamura, K., Imai, K., Iizuka, T., Shiro, Y., and Nakamura, H. (2003). The uncoupling of oxygen sensing, phosphorylation signalling and transcriptional activation in oxygen sensor FixL and FixJ mutants. *Mol Microbiol* **48**, 373-383.

Sanowar, S., and Le Moual, H. (2005). Functional reconstitution of the *Salmonella typhimurium* PhoQ histidine kinase sensor in proteoliposomes. *Biochem J* **390**, 769-776.

Sasakura, Y., Yoshimura-Suzuki, T., Kurokawa, H., and Shimizu, T. (2006). Structure-function relationships of EcDOS, a heme-regulated phosphodiesterase from *Escherichia coli*. *Acc Chem Res* **39**, 37-43.

Schirmer, T., and Jenal, U. (2009). Structural and mechanistic determinants of c-di-GMP signalling. *Nat Rev Microbiol* **7**, 724-735.

Schmidt, A.J., Ryjenkov, D.A., and Gomelsky, M. (2005). The ubiquitous protein domain EAL is a cyclic diguanylate-specific phosphodiesterase: enzymatically active and inactive EAL domains. *J Bacteriol* **187**, 4774-4781.

Schmidt, J.V., and Bradfield, C.A. (1996). Ah receptor signaling pathways. *Annu Rev Cell Dev Biol* **12**, 55-89.

Schmitz, R.A., Klopprogge, K., and Grabbe, R. (2002). Regulation of nitrogen fixation in *Klebsiella pneumoniae* and *Azotobacter vinelandii*: NifL, transducing two environmental signals to the *nif* transcriptional activator NifA. *J Mol Microbiol Biotechnol* **4**, 235-242.

Senior, P.J. (1975). Regulation of nitrogen metabolism in *Escherichia coli* and *Klebsiella aerogenes*: studies with the continuous-culture technique. *J Bacteriol* **123**, 407-418.

Sevvana, M., Vijayan, V., Zweckstetter, M., Reinelt, S., Madden, D.R., Herbst-Irmer, R., Sheldrick, G.M., Bott, M., Griesinger, C., and Becker, S. (2008). A ligand-induced switch in the periplasmic domain of sensor histidine kinase CitA. *J Mol Biol* **377**, 512-523.

Söderbäck, E., Reyes-Ramirez, F., Eydmann, T., Austin, S., Hill, S., and Dixon, R. (1998). The redox- and fixed nitrogen-responsive regulatory protein NIFL from *Azotobacter vinelandii* comprises discrete flavin and nucleotide-binding domains. *Mol Microbiol* **28**, 179-192.

Song, Y., Peisach, D., Pioszak, A.A., Xu, Z., and Ninfa, A.J. (2004). Crystal structure of the C-terminal domain of the two-component system transmitter protein nitrogen regulator II (NRII; NtrB), regulator of nitrogen assimilation in *Escherichia coli*. *Biochemistry* **43**, 6670-6678.

Sprenger, W.W., Hoff, W.D., Armitage, J.P., and Hellingwerf, K.J. (1993). The eubacterium *Ectothiorhodospira halophila* is negatively phototactic, with a wavelength dependence that fits the absorption spectrum of the photoactive yellow protein. *J Bacteriol* **175**, 3096-3104.

Stock, A.M., Robinson, V.L., and Goudreau, P.N. (2000). Two-component signal transduction. *Annu Rev Biochem* **69**, 183-215.

Stock, J. (1999). Signal transduction: Gyrating protein kinases. *Curr Biol* **9**, R364-367.

Stock, J.B., Stock, A.M., and Mottonen, J.M. (1990). Signal transduction in bacteria. *Nature* **344**, 395-400.

Studier, F.W., Rosenberg, A.H., Dunn, J.J., and Dubendorff, J.W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol* **185**, 60-89.

Surette, M.G., Levit, M., Liu, Y., Lukat, G., Ninfa, E.G., Ninfa, A., and Stock, J.B. (1996). Dimerization is required for the activity of the protein histidine kinase CheA that mediates signal transduction in bacterial chemotaxis. *J Biol Chem* **271**, 939-945.

Swanson, R.V., Bourret, R.B., and Simon, M.I. (1993). Intermolecular complementation of the kinase activity of CheA. *Mol Microbiol* **8**, 435-441.

Szurmant, H., White, R.A., and Hoch, J.A. (2007). Sensor complexes regulating two-component signal transduction. *Curr Opin Struct Biol* **17**, 706-715.

Taguchi, S., Matsui, T., Igarashi, J., Sasakura, Y., Araki, Y., Ito, O., Sugiyama, S., Sagami, I., and Shimizu, T. (2004). Binding of oxygen and carbon monoxide to a heme-regulated phosphodiesterase from *Escherichia coli.* Kinetics and infrared spectra of the full-length wild-type enzyme, isolated PAS domain, and Met-95 mutants. *J Biol Chem* **279**, 3340-3347.

Tanaka, A., and Shimizu, T. (2008). Ligand binding to the Fe(III)-protoporphyrin IX complex of phosphodiesterase from *Escherichia coli* (Ec DOS) markedly enhances catalysis of cyclic di-GMP: roles of Met95, Arg97, and Phe113 of the putative heme distal side in catalytic regulation and ligand binding. *Biochemistry* **47**, 13438-13446.

Tanaka, A., Takahashi, H., and Shimizu, T. (2007). Critical role of the heme axial ligand, Met95, in locking catalysis of the phosphodiesterase from *Escherichia coli* (Ec DOS) toward Cyclic di-GMP. *J Biol Chem* **282**, 21301-21307.

Tanaka, T., Saha, S.K., Tomomori, C., Ishima, R., Liu, D., Tong, K.I., Park, H., Dutta, R., Qin, L., Swindells, M.B., Yamazaki T., Ono A.M., Kainosho M., Inouye M., and Ikura M. (1998). NMR structure of the histidine kinase domain of the *E. coli* osmosensor EnvZ. *Nature* **396**, 88-92.

Taylor, B.L. (2007). Aer on the inside looking out: paradigm for a PAS-HAMP role in sensing oxygen, redox and energy. *Mol Microbiol* **65**, 1415-1424.

Thorneley, R.N., and Lowe, D.J. (1983). Nitrogenase of *Klebsiella pneumoniae*. Kinetics of the dissociation of oxidized iron protein from molybdenum-iron protein: identification of the rate-limiting step for substrate reduction. *Biochem J* **215**, 393-403.

Thummer, R., Klimmek, O., and Schmitz, R.A. (2007). Biochemical studies of *Klebsiella pneumoniae* NifL reduction using reconstituted partial anaerobic respiratory chains of *Wolinella succinogenes*. *J Biol Chem* **282**, 12517-12526.

Tomomori, C., Tanaka, T., Dutta, R., Park, H., Saha, S.K., Zhu, Y., Ishima, R., Liu, D., Tong, K.I., Kurokawa, H., Qian, H., Inouye, M., and Ikura, M. (1999). Solution structure of the homodimeric core domain of *Escherichia coli* histidine kinase EnvZ. *Nat Struct Biol* **6**, 729-734.

Tuckerman, J.R., Gonzalez, G., Dioum, E.M., and Gilles-Gonzalez, M.A. (2002). Ligand and oxidation-state specific regulation of the heme-based oxygen sensor FixL from *Sinorhizobium meliloti*. *Biochemistry* **41**, 6170-6177.

Tuckerman, J.R., Gonzalez, G., and Gilles-Gonzalez, M.A. (2001). Complexation precedes phosphorylation for two-component regulatory system FixL/FixJ of *Sinorhizobium meliloti*. *J Mol Biol* **308**, 449-455.

Ukaegbu, U.E., Henery, S., and Rosenzweig, A.C. (2006). Biochemical characterization of MmoS, a sensor protein involved in copper-dependent regulation of soluble methane monooxygenase. *Biochemistry* **45**, 10191-10198.

Ukaegbu, U.E., and Rosenzweig, A.C. (2009). Structure of the Redox Sensor Domain of *Methylococcus capsulatus* (Bath) MmoS. *Biochemistry* **48**, 2207-2215.

Ulrich, L.E., Koonin, E.V., and Zhulin, I.B. (2005). One-component systems dominate signal transduction in prokaryotes. *Trends Microbiol* **13**, 52-56.

van Heeswijk, W.C., Stegeman, B., Hoving, S., Molenaar, D., Kahn, D., and Westerhoff, H.V. (1995). An additional PII in *Escherichia coli*: a new regulatory protein in the glutamine synthetase cascade. *FEMS Microbiology Letters* **132**, 153-157.

Verger, D., Carr, P.D., Kwok, T., and Ollis, D.L. (2007). Crystal structure of the N-terminal domain of the TyrR transcription factor responsible for gene regulation of aromatic amino acid biosynthesis and transport in *Escherichia coli* K12. *J Mol Biol* **367**, 102-112.

Vescovi, E.G., Ayala, Y.M., Di Cera, E., and Groisman, E.A. (1997). Characterization of the bacterial sensor protein PhoQ. Evidence for distinct binding sites for Mg2+ and Ca2+. *J Biol Chem* **272**, 1440-1443.

Vreede, J., van der Horst, M.A., Hellingwerf, K.J., Crielaard, W., and van Aalten, D.M. (2003). PAS domains. Common structure and common flexibility. *J Biol Chem* **278**, 18434-18439.

Woodley, P., and Drummond, M. (1994). Redundancy of the conserved His residue in *Azotobacter vinelandii* NifL, a histidine autokinase homologue which regulates transcription of nitrogen fixation genes. *Mol Microbiol* **13**, 619-626.

Yamamoto, K., Matsumoto, F., Minagawa, S., Oshima, T., Fujita, N., Ogasawara, N., and Ishihama, A. (2009). Characterization of CitA-CitB signal transduction activating genes involved in anaerobic citrate catabolism in *Escherichia coli*. *Biosci Biotechnol Biochem* **73**, 346-350.

Yang, J., Zhang, L., Erbel, P.J., Gardner, K.H., Ding, K., Garcia, J.A., and Bruick, R.K. (2005). Functions of the Per/ARNT/Sim domains of the hypoxia-inducible factor. *J Biol Chem* **280**, 36047-36054.

Yang, Y., and Inouye, M. (1993). Requirement of both kinase and phosphatase activities of an *Escherichia coli* receptor (Taz1) for ligand-dependent signal transduction. *J Mol Biol* **231**, 335-342.

Yoshimura, T., Sagami, I., Sasakura, Y., and Shimizu, T. (2003). Relationships between heme incorporation, tetramer formation, and catalysis of a heme-regulated phosphodiesterase from *Escherichia coli*: a study of deletion and site-directed mutants. *J Biol Chem* **278**, 53105-53111.

Zhang, X., Chaney, M., Wigneshweraraj, S.R., Schumacher, J., Bordes, P., Cannon, W., and Buck, M. (2002). Mechanochemical ATPases and transcriptional activation. *Mol Microbiol* **45**, 895-903.

Zhou, Y.F., Nan, B., Nan, J., Ma, Q., Panjikar, S., Liang, Y.H., Wang, Y., and Su, X.D. (2008). C4-dicarboxylates sensing mechanism revealed by the crystal structures of DctB sensor domain. *J Mol Biol* **383**, 49-61.

Zhulin, I.B., and Taylor, B.L. (1998). Correlation of PAS domains with electron transport-associated proteins in completely sequenced microbial genomes. *Mol Microbiol* **29**, 1522-1523.

Zhulin, I.B., Taylor, B.L., and Dixon, R. (1997). PAS domain S-boxes in Archaea, Bacteria and sensors for oxygen and redox. *Trends Biochem Sci* **22**, 331-333.

Zientz, E., Bongaerts, J., and Unden, G. (1998). Fumarate regulation of gene expression in *Escherichia coli* by the DcuSR (*dcuSR* genes) two-component regulatory system. *J Bacteriol* **180**, 5421-5425.

Zoltowski, B.D., Schwerdtfeger, C., Widom, J., Loros, J.J., Bilwes, A.M., Dunlap, J.C., and Crane, B.R. (2007). Conformational switching in the fungal light sensor Vivid. *Science* **316**, 1054-1057.

Zwir, I., Shin, D., Kato, A., Nishino, K., Latifi, T., Solomon, F., Hare, J.M., Huang, H., and Groisman, E.A. (2005). Dissecting the PhoP regulatory network of *Escherichia coli* and *Salmonella enterica*. *PNAS* **102**, 2862-2867.

Appendix - Publications