STRUCTURAL INSIGHTS INTO SPORULATION IN *BACILLUS SUBTILIS*

Approved by Supervisory Committee

Kevin Gardner, Ph.D., Mentor

Mike Rosen, Ph.D., Committee Chair

Philip Thomas, Ph.D.

Joseph Albanesi, Ph.D.

STRUCTURAL INSIGHTS INTO SPORULATION IN *BACILLUS SUBTILIS*

by

James Lee

Dissertation

Presented to the Faculty of the Graduate School of Biomedical Science
The University of Texas Southwestern Medical Center at Dallas
In Partial Fulfillment of the Requirements
For the Degree of

Doctor of Philosophy

The University of Texas Southwestern Medical Center at Dallas

Dallas, Texas

April, 2010

STRUCTURAL INSIGHTS INTO SPORULATION IN *BACILLUS SUBTILIS*

James Lee, Ph.D.

The University of Texas Southwestern Medical Center at Dallas, 2006

Supervising Mentor: Kevin H. Gardner, Ph.D.

PAS domains are modular domains that providing specific interaction surfaces for a diverse array of ligands, from small organic compounds to intra and intermolecular protein domains. As such, they are ubiquitous throughout signal transduction pathways in all three kingdoms of life: Bacteria, Archae, and Eucarya (Huang, Edery et al. 1993).

The significant role played by PAS domains in *Bacillus subtilis* sporulation is underscored by the finding that the PAS-A domain is necessary for efficient phosphorylation in the sporulation kinase, KinA(Wang, Fabret et al. 2001). This activity is necessary for initiating a phosphorelay that results in the upregulation of genes required for sporulation. In KinA, dimerization is necessary for individual monomers of the histidine kinase domain to transphosphorylate partner subunits. The dimerization of KinA involves interactions in the PAS-A

domain, but molecular details regarding PAS-A dimerization and its importance to KinA activity has not been previous characterized.

To investigate these interactions within the context of the KinA homodimer, we expressed the N-terminal PAS-A domain and solved the X-ray crystal structure. Conformational variability was implicated through the observation of different orientations of the dimerization interface in two distinct structural models found in the asymmetric unit of the crystal lattice. These models were used to identify key interfacial residues and the roles of these were tested in a variety of ways by site-directed mutagenesis.

LIST OF FIGURES

LIST OF TABLES

Chapter 1: Introduction Part I

In order for cells to preserve their physiological homeostasis and proliferate, they need to detect a wide assortment of environmental signals and affect the appropriate response. Exogenous signals often modulate the biochemistry of individual pathways such as to maintain cyclic behaviors, cell metabolism, cell growth patterns and/or differentiation. The environment of the cell holds these mechanisms under strict regulation; deviation from this stringent control leads to disease progression and possibly death. Scientific investigation of these signal transduction pathways affecting multiple hierarchies in biological life benefits greatly from diverse methodologies that require collaborations between scientists working both at the submicroscopic level and with whole organisms. The purpose of my thesis is to investigate the KinA, the master regulator of signal transduction within the sporulation pathway of *Bacillus subtilis* and how its structural foundation can be investigated by a multi-disciplinary approach utilizing many techniques. More specifically, I have chosen to investigate within KinA a special type of protein fold called the PAS domain which is involved in signal transduction for a number of pathways in biological life.

PAS domains

PAS domains are modular domains capable of providing specific interaction surfaces for a diverse array of ligands, from small organic compounds to intra and intermolecular protein domains. Such utility contained within a relatively small framework consisting of approximately 100 residues makes them ubiquitous throughout signal transduction pathways in all three kingdoms of life: Bacteria, Archae, and Eucarya (Huang, Edery et al. 1993). The PAS nomenclature derives from the first three proteins in which they were identified: the *Drosophila* period clock protein (PER) was discovered through the search for proteins involved in regulating circadian rhythms (Hamblen, Zehring et al. 1986; Yu, Jacquier et al. 1987); the vertebrate Aryl Hydrocarbon Nuclear Translocator (ARNT) found to be essential to the xenobiotic response pathway modulated by Aryl Hydrocarbon Receptor (Ahr) (Hoffman, Reyes et al. 1991); finally, the *Drosophila* single-minded protein (SIM) is a regulatory protein controlling midline cell lineage (Jackson, Bargiello et al. 1986; Nambu, Lewis et al. 1991). PAS domains are integrated with many types of effector modules to create a wide variety of signaling proteins that vary in responsiveness from refined sensitivity to robust responses. Many proteins contain multiple PAS domains. KinA, the signaling protein studied in this thesis, has three PAS domains, each more homologous to PAS domains from archaeal and human proteins than to one another—arguing against the idea that these domains were simply duplicated to intensify a signaling

response. Instead, proteins like KinA require multiple signal inputs to regulate very complicated systems (Taylor and Zhulin 1999). Thus when comparing different PAS domains either in the same or different proteins, only 17% sequence identity has been observed in the SMART database of PAS domains (http://smart.embl-heidelberg.de). This low sequence identity is striking especially considering that these domains only have 100 residues, but still adopt virtually identical structures.

PAS domain structure and diversity

In regard to this apparently high sequence diversity, it is difficult to robustly identify PAS domain sequences without using known structural information to align the sequences (Hefti, Francoijs et al. 2004). Even among similar pathways in different species of the same genus such as Bacillus (**Fig. 1-1**), PAS domains are not similar in sequence thus suggesting that these PAS domains have divergent functions. It may be that these organisms may have evolved such that they found that the most efficient means of implementing a particular signal transduction mode was through the common usage of PAS domains. The first computational methods that attempted to align these dissimilar sequences produced results that failed to produce a consensus PAS domain sequence (Altschul, Madden et al. 1997). The additional finding of PAS-A/PAS-B repeat sequences in bHLH/PAS proteins, one of the first PAS domains to be thoroughly

investigated, further confounded these attempts to define the PAS domain. The advent of Gapped and PSI-Blast programs allowed the alignment of PAS domains such that they are now identified as consisting of a five-stranded antiparallel β-sheet flanked by several α-helices, encoded within a sequence of approximately 100 residues (**Fig. 1-2**). Relatively more conserved N-terminal and C-terminal regions are designated as the PAS and PAC boxes (Zhulin, Taylor et al. 1997; Hefti, Francoijs et al. 2004), respectively. The PAS box contains the Aβ-Fα segments, while the PAC box contains the final three strands Gβ-Iβ. These elements are linked by sequences highly variable as to composition and length. (Zhulin, Taylor et al. 1997).

Despite the low sequence identity of seventeen percent between different PAS domains, computational biologists have identified over 6000 PAS domains in the three kingdoms of life through a combination of improved sequence alignment programs and secondary structure prediction algorithms, and analysis of the three-dimensional structures for PYP, FixL, and HERG—now considered the prototypical PAS domains (Rost and Sander 1993; Taylor and Zhulin 1999). Indeed, the first robust and systematic identification of a great number of PAS domains came when the results of PhD secondary structure prediction were compared to the structures of PYP, FixL, and HERG (Rost and Sander 1993; Taylor and Zhulin 1999).

The most important computational device for the identification of PAS domains from genomes came with the incorporation of hidden Markov models (HMMs) (Krogh, Brown et al. 1994). This statistical method is based on the now well-accepted assumption that even when primary sequences differ widely in structurally similar protein domains, there still exist conserved residues that can encode the folding and function of what is being queried. HMMs encode this information by generating a matrix that includes the sequences of representative members of a domain family. This matrix is now able to describe the probability of finding any given amino acid at each position in the domain. As such, a query sequence can be scored as a whole to generate a probability distribution for being a member of the domain family included in the matrix of the HMM. HMM methods have proven to be robust even in PAS domain sequences such as PER that contain large loops inserted between the β-strands (Hefti, Francoijs et al. 2004; Yildiz, Doi et al. 2005).

The two most utilized databases of HMM-derived protein domain searches are Pfam and SMART (Simple Modular Architecture Research Tool). Unlike other computational methodologies, Pfam included the entire span of representative domains in the family to find new members from sequence databases (http://pfam.wustl.edu) (Sonnhammer, Eddy et al. 1998). SMART is disadvantaged compared to Pfam when given incomplete domain sequences. However, because SMART incorporated 667 domain families in its matrix that

were verified by laboratory experimentation to have a certain biological context (http://smart.embl-heidelberg.de) (Schultz, Milpetz et al. 1998) SMART has its advantage in that it may predict a domain function. Nevertheless, the combination of these two powerful techniques has allowed for the rapid identification of many new PAS domains.

PAS domain mechanisms

Although mutagenesis studies have been able to identify common secondary and tertiary structure elements used in signal transduction, a general consensus for the molecular mechanism by which these elements act in signal propagation has remained elusive, perhaps due to the following reasons: (1) PAS domains are often found in tandem with other regulatory domains that also modulate the effector domain/protein such that it is difficult to deconvolute the effects of a single PAS domain and/or (2) despite a strikingly similar overall fold, sequence diversity is such that robust identification of new PAS domains often requires incorporation of tertiary structure and/or the hidden Markov models (see above) (Hefti, Francoijs et al. 2004). In this regard, within a small and malleable protein framework, PAS domains can truly provide a wide assortment of molecular switches suitable for any role in signal transduction from direct signal sensor to protein inhibitor/activator to dimerization partner. Further underscoring this idea is that even in the organism of principal interest in this thesis—*Bacillus*—the

domain architecture and sequence diversity of PAS-containing sporulation kinases vary widely within different species and their corresponding family of sporulation kinases, suggesting that not only do they respond to different signals, but that the way they respond to signals and effect downstream targets also diverges (**Fig. 1-1**).

Shown in **Figure 1-3** is a representative listing of a variety domain architectures found in proteins containing just one special type of PAS domain that senses light, often referred to as the LOV (Light, Oxygen, Voltage) domain. As for other PAS domains, effectors for LOV domains include phosphodiesterases, serine/threonine kinases, histidine kinases, and DNA binding domains. The different types of signals identified have also been variable. CitA binds a simple molecule citrate(Reinelt, Hofmann et al. 2003), while LuxP senses the signal secondarily through an effector protein that binds Autoinducer-2 (AI-2), a ligand signal that accumulates when large populations of *Vibrio harveyi* build up (Neiditch, Federle et al. 2005). In other cases, these PAS domains are assisted by cofactors to help create the conformational change they need to signal to their respective stimuli. For example the PAS domain of *Avena sativa* phototropin1, also referred to as the LOV (light-oxygen-voltage) domain, uses an FMN chromophore to affect the unfolding of a helix that regulates a downstream serine/threonine kinase(Salomon, Christie et al. 2000; Crosson and Moffat 2001; Swartz, Corchnoy et al. 2001; Christie, Swartz et al. 2002; Harper, Christie et al.

2004). Also assisted by extra-protein elements, FixL also uses a non-protein cofactor, heme, to affect a histidine kinase (Gong, Hao et al. 1998; Gong, Hao et al. 2000). Meanwhile, a great number of PAS domains have no identified associated ligands or signaling effectors. Indeed, it may be that a number of these PAS domains do not respond to any ligand, but provide crucial protein interaction surfaces that allow other regulatory domains to modulate the specific effect they require. For instance, ARNT contains PAS domains that serve as dimerization domain for several other transcription factors, including HIFα, and AHR(Carver, Jackiw et al. 1994; Bruick 2000; Bracken, Whitelaw et al. 2003; Erbel, Card et al. 2003; Card, Erbel et al. 2005). The PAS domain of ARNT binds these proteins upon their activation by hypoxia and xenobiotic stress, respectively, to create a fully functionally active transcription factor in a signal-dependent manner.

PAS domains as environmental sensors

Robust and specific responses to extracellular signals are necessary for the proper maintenance of intracellular processes. Post-translational modifications such as hydroxylation, acetylation, methylation, ubiquitination, and phosphorylation have long been investigated with well-established technologies in molecular biology and biochemistry. These chemical modifications made to specific binding surfaces modulate the ability of a signaling protein to interact with other components in signaling networks either reversibly or irreversibly.

However, an alternative mechanism that serves to accomplish time- or signal-dependent modulation of protein structure is the reversible interaction with small molecules, such that it may be possible to achieve both a faster initiation and attenuation of that response. Small molecules often can penetrate the cell membrane and as such, serve as environmental signals. For example, oxygen, electrons, and even light energy are very simple direct signals produced by the environment that can be detected by cellular proteins to monitor conditions of the cell environment.

Amongst proteins that directly monitor small molecules, PAS domains are a very common and robust tool used by cells to monitor changes in redox potential, energy levels, light and oxygen tension in the cell. In this manner, these domains play important roles not only in restoring homeostasis when conditions threaten the balance of physiological equilibria but also in shifting growth conditions when conditions are conducive to cell proliferation.

a. Redox state and oxygen sensing

Because of the crucial association of the electron transport chain with the production of energy equivalents, it is thought that alterations to electron transport are important first indications of energy depletion. As the terminal electron acceptor for oxidative phosphorylation, oxygen is a vital component that many types of cells closely monitor such that they can make the appropriate prompt

response. To this end, *E coli* has evolved the Aer protein to sense conditions of low oxygen and transduce the appropriate signals so that it can move toward oxygen in a process called aerotaxis. Aerotaxis has been traditionally thought to be mediated by methylation of transcription factors that activate aerotaxis genes(Bibikov, Miller et al. 2004), but Aer provides an alternative means for this process. The N-terminal cytoplasmic PAS domain of this protein, in response to oxygen abundance and the ensuing cellular condition of low redox state, oxidizes a cysteinyl side chain such that it forms an adduct with the FAD prosthetic group and changes conformation to affect a downstream histidine kinase (Bibikov, Miller et al. 2004).

An alternative means of redox sensing by PAS domains is exemplified by the *E coli* ArcB sensor kinase, which does not utilize a bound FAD, but instead oxidizes two redox active cysteine residues such that they form cysteinyl adducts with the corresponding cysteine residues in its partner for homodimerization (Malpica, Franco et al. 2004). The formation of these intermolecular covalent bonds changes the conformation of the ArcB dimer such as to inhibit kinase activity. (**Fig. 1-4a**).

Another protein that is tied to cellular oxygen levels is involved in nitrogen fixation. During anaerobic conditions, *Bradyzodium japonica* reorganizes its cellular machinery toward creating energy equivalents in the form of ammonium derivatives such as amino acids or even ammonium ions

themselves, both of which are strong reducing agents for oxidative phosphorylation (Taylor and Zhulin 1999). Furthermore, there is the regulatory advantage that nitrogenases, being oxygen labile, have their protein concentration levels regulated by the same oxygen signaling molecule. FixL contains a PAS domain that utilizes a heme cofactor to directly sense oxygen . The binding of cyanide—considered by the authors to be equivalent to an oxygen signal—drives the planarization of the heme proprionate such that the His214 residue of the FG loop is shifted 1.6 Å away from the heme (Gong, Hao et al. 1998). The authors then propose that this conformational change induces other unspecified interactions between the FG loop and the histidine kinase leading to inhibition of kinase activity.

b. PAS domains as light sensors

In certain halophiles, Photoactive yellow protein (PYP) is a bacterial blue-light photoreceptor that induces events mediating negative phototaxis—the process of swimming away from light (Dux, Rubinstenn et al. 1998; Genick, Soltis et al. 1998). Photocycle-associated conformational changes in the protein are precipitated by light-induced cis-trans isomerization in the *p*-coumaric acid chromophore bound in the hydrophobic core of the protein (Baca, Borgstahl et al. 1994; Hoff, Dux et al. 1994; Xie, Hoff et al. 1996). The association of an N-terminal helix-turn-helix structure with the solvent exposed side of the β-sheet

surface, implicated in many other PAS domains involved in signal transduction, has lead investigators to suggest that this element may be involved in propagating conformational changes in the core to downstream effectors for PYP.

Another example of photosensory PAS domains exists in the system used to entrain the circadian rhythm for the mold organism *Neurospora crassa.* Here, special types of PAS domains called LOV domains, change conformation in response to light with the assistance of FAD or FMN chromophores (discussed in next paragraph). The N-terminal LOV domain of the White-collar 1 (WC-1) transcription factor has been shown to be necessary for all light responses in this organism. Light induction is correlated to the formation of higher order complexes containing WC-1 that are capable of inducing asexual sporulation (conidiation) (Liu 2003).

Recently, phototropins have become a prominent class of LOV-containing proteins thoroughly investigated with both molecular biological and biophysical techniques. These proteins are responsible for essential plant activities such as stomatal opening, phototropism, chloroplast relocation and leaf opening (Briggs and Christie 2002; Sakamoto and Briggs 2002). These PAS domains use flavin mononucleotide (FMN) or flavin adenosyl-dinucleotide (FAD) chromophores to drive photon-induced covalent bond formation with the protein such as to create a new signaling state. Reversible protein cysteinyl adduction to this moiety as a result of this chemical photocycling leads to global conformation changes

throughout the protein that are propagated via a variety of mechanisms. In the *Avena sativa* (oat) Phototropin1, consisting of a serine/threonine type kinase preceded by two LOV domains, LOV1 and LOV2, uses FMN to sense light and undergo a photocycle. LOV2 upon photoactivation forms a covalent cysteine adduct between an absolutely conserved Cys450 and its chromophore (Salomon, Christie et al. 2000; Swartz, Corchnoy et al. 2001). Mutation of the conserved Cys450 to alanine prevents both adduct formation and kinase activation, directly linking these two events in the signal transduction model (Christie, Swartz et al. 2002). A C-terminal helical extension (Jα) beyond the basic PAS domain associates with the solvent exposed side of the central β-sheet. When LOV2 becomes light-induced, dramatic conformational changes occur throughout the domain that cause the unpacking of this helix from the rest of the domain, with its helical character destabilized enough to cause its own unfolding (Harper, Neil et al. 2003; Harper, Christie et al. 2004). Subsequent events eventually lead to kinase activation, thus implicating either activation of the kinase domain by the LOV2 domain or attenuation of inhibitory interactions by the linker region as possible mechanisms for kinase activation. Although the use of extra-PAS domain elements such as the Jα helix is not likely to be a feature found in all PAS-mediated signaling, at least thus far, the solvent-exposed central-β sheet does seem to be this fold's most conserved signaling element perhaps underscoring the adaptability of a robust β-sheet core that is stabilized by a number of helices.

Variability of PAS-containing complexes

For many PAS domains, no known endogenous small molecule ligand has been found. However, in many of these cases, these PAS domains are utilized to transduce environmental signals to other proteins or domains not through the detection of any direct signal, but by providing specific interfaces for time or signal-dependent protein-protein interactions to be formed. For example, LuxQ contains two PAS domains that homodimerize and bind to LuxP forming a 1:1 complex (**Fig. 1-4b)**. Interactions that keep this homodimer together are temporarily broken only to allow the formation of other interactions in the presence of a liganded LuxP (Neiditch, Federle et al. 2005). Other PAS-containing protein have shown the ability to form higher order multimeric complexes. For example, biochemical studies have shown that the isolated N-terminal PAS domain of HERG can form tetramers (Li, Xu et al. 1997), while the PAS domain-containing RmFixL protein was shown to form a 2:1 complex with RmFixJ in solution(Tuckerman, Gonzalez et al. 2001).

PAS domains as components of transcription factors

The regulation of gene expression serves an important role in not only immediate responses, but also longer-term adaptation of an organism to continued environmental pressures. PAS domains first gained prominent scientific recognition because of their roles in the formation of transcription factors that

control such functions (Huang, Edery et al. 1993). PAS-containing transcription factors use a number of different DNA binding domains, of which one of the most common are the basic helix-loop-helix (bHLH) domain containing proteins. This family of PAS-containing proteins controls processes involved in the maintainence of homeostasis. For example, a class of proteins called hypoxia inducible factors is crucial in regulating oxygen homeostasis in mammals. These proteins consist of a bHLH DNA binding domain, two tandem PAS domains (PAS-A and PAS-B), N-terminal and C-terminal transactivation domains (N-TAD and C-TAD), and an oxygen degradation domain (ODD) (Crews 1998; Crews and Fan 1999; Taylor and Zhulin 1999). They dimerize with ARNT through interfaces provided by the bHLH DNA-binding domain and the PAS domains. In response to hypoxia, constitutive degradation after oxygen-dependent hydroxylation of these proteins is abolished leading to the formation of a fully functional HIFα-ARNT transcription factor. Functionally, this complex elicits adaptation to low oxygen by upregulating expression of genes involved in glycolysis, angiogenesis, epithelial permeability and tissue remodeling (Bruick 2000; Bracken, Whitelaw et al. 2003). Other than the maintenance of homeostasis, bHLH PAS transcription factors also are important in developmental and physiological processes. For example, the regulation of neural development in many organisms is modulated by the Single-minded (SIM), a protein in which, again, tandem PAS domains are

required to dimerize with PAS domains in the ubiquitous ARNT adaptor protein to form the fully functional transcription factor (Nambu, Lewis et al. 1991).

Structure/Function Relationships

a. Small molecule binding

Many PAS domains have been shown to bind specific small molecule compounds using their hydrophobic cores. Ligand perturbation of distinct structural features at these sites often leads to large conformational changes in this small domain demonstrating that this domain can undergo extensive dynamics. PYP and CitA have been shown to use a binding cavity formed from the Aβ and Bβ strands and the Cα and Dα helices. However, for the majority of small-molecule binding PAS domains, ligand binding can be generalized to that of the pocket formed by the Gβ, Hβ, and Iβ strands and the Eα and Fα helices. This site has been used not only to bind cofactors such as heme and FMN in FixL and Phy3, respectively (Gong, Hao et al. 1998; Miyatake, Mukai et al. 2000; Crosson and Moffat 2001), but also for diffusible small molecule ligands as with human PAS kinase and HIF2α (**Fig. 1-5)**. In the case of PAS kinase (Amezcua, Harper et al. 2002) and HIF2α (Erbel, Card et al. 2003), a small molecule library was screened and a number of artificial ligands were found to bind with micromolar affinity even though the apo structures revealed no preformed cavity for such an interaction. These deformable cores may very well underscore the flexiblity of the PAS

domain and the ability of side chains in the hydrophobic core to reorient themselves to accommodate a ligand, whether artificial or natural.

In what probably is an example of the most deformable core in a PAS domain, the Aryl Hydrocarbon Receptor has been identified to an even greater number of small diffusible small molecules—hundreds of exogenous ligands have been identified to activate this transcription factor. However, it has been suggested that this promiscuity comes at the cost of the requirement of stabilization by the chaperone Hsp90 (Carver, Jackiw et al. 1994; Meyer, Pray-Grant et al. 1998). In this regard, it appears that the protein core is so deformable that chaperone binding is needed to stabilize exposed hydrophobic residues until they are ready to interact with ligands (Carver, Jackiw et al. 1994; Meyer, Pray-Grant et al. 1998; Denison, Pandini et al. 2002) .

b. Polypeptide binding and signal transduction

From the structural studies of many PAS domains to date, it has been speculated that the β-sheet interface and loop regions between secondary structure elements are important sites for protein binding as a mode for signal transduction (Gong, Hao et al. 1998; Amezcua, Harper et al. 2002; Erbel, Card et al. 2003; Reinelt, Hofmann et al. 2003; Card, Erbel et al. 2005). Often the binding of a small diffusible molecule or relatively minor alterations induced in an internally bound chromophore can cause significant enough changes to these structural elements to

affect their ability bind cognate partners. Both PYP and AsLOV2 exhibit only small structural adjustments to their respective FAD and FMN chromophores with light illumination, but major perturbations are made to residues in several strands of the β-sheet (Hoff, Xie et al. 1999; Harper, Neil et al. 2003). These changes attenuate the binding of α-helices to this interface that are correlated to changes in activity. Similarly, in the case of HIF2α, our lab has found that the binding of small molecules—identified through an in-house-developed small molecule library screen—changes features of the β-sheet interface such that NMR-based titration studies suggest weakened heterodimerization between the HIF2α and its partner ARNT.

Similarly, despite a lack of structure, loop regions have also been found to be able to affect conformational changes within some PAS domains. Mutations to a ligand-binding site in the PAS domain of hPASk lead to large changes in the adjacent FG loop (Amezcua, Harper et al. 2002). When such mutations are also introduced into constructs containing both the PAS domain and the kinase domain increased kinase activity was observed. The authors speculate that the ligand-induced changes at the binding site lead to changes in the FG loop that affect its ability to inhibit the kinase domain. Such a mechanism of inhibition by a loop region has also been speculated for FixL based upon the conformational changes seen when the X-ray crystal structures of the met- and cyanomet-bjFixLH PAS domain(Gong, Hao et al. 1998). However, here the changes begin with a heme

chromophore that is in direct contact with this same FG loop. The binding of oxygen or other ligands to the heme is suggested by the authors to cause conformational changes that modulate the conformation of this FG loop such as to inhibit the downstream kinase either directly or indirectly (Gong, Hao et al. 1998; Gong, Hao et al. 2000).

c. Dimerization in PAS domain structures

As already mentioned, PAS domains also transduce signals to receiver/effector domains. These mechanisms often follow simple dimerization patterns either in heterotypic or homotypic fashion. Indeed homodimerization and heterodimerization with different partners is a robust way for ARNT to modulate different response pathways. One partner, AHR, is found in the cytosol associated with cellular chaperones like Hsp90 (Swanson, Tullis et al. 1993; Carver, Jackiw et al. 1994; Meyer, Pray-Grant et al. 1998; Denison, Pandini et al. 2002). In response to the binding of xenobiotics, AHR translocates to the nucleus where it heterodimerizes with ARNT to form the functional transcription factor and initiate the transcription of target genes related to drug metabolism. Similarly, homodimerization is also an important mechanism for PAS domains to control cellular responses. Despite its ability to bind many different partners, ARNT PAS-B still has enough information in its primary sequence to encode an interaction surface for homodimerization—perhaps underscoring the versatility of

the PAS fold. This mode of binding has been shown to be considerably lower in affinity than binding to its heterodimerization partners (Card, Erbel et al. 2005). It is tempting to speculate that the competition between binding partners and ARNT PAS-B itself for dimerization may establish a threshold for activation of certain responses that serves as a cellular checkpoint control. This would ensure that no pathway is spontaneously activated by transient fluctuations in the environment (Swanson, Tullis et al. 1993; Carver, Jackiw et al. 1994; Meyer, Pray-Grant et al. 1998; Bruick 2000; Denison, Pandini et al. 2002; Bracken, Whitelaw et al. 2003).

Indeed, even PAS domains that are constitutively homodimerized and never dissociated by exogenous signals can serve important roles in signaling. *E. coli* ArcB sensor kinase uses its PAS domain to sense changes in redox potential (Malpica, Franco et al. 2004); in response to oxidation, this PAS domain forms inter-subunit disulfide bonds that constrain the dimer in such a manner that the histidine kinase is allosterically inhibited (**Fig. 1-4a**). Similarly, the binding of a partner protein by the tandem PAS domains of LuxQ in *Vibrio harveyi* allosterically inhibits kinase function such as to actually convert it to a phosphatase (Neiditch, Federle et al. 2005). Neiditch et al. suggested that binding of the partner protein changed one set of interactions in the tandem PAS dimerization interface to another set of interactions, indicating that these PAS domains have a flexible interface ready to be switched from one conformation to

another. One basic question in this regard is how homodimeric PAS domains interact structurally to maintain their regulatory functions. To this end, there are now several crystal structures of homodimeric PAS domains also confirmed by gel-filtration to be dimeric. This group of structures includes those for CitA, Per, FixL, and KinA PAS-A, presented in this thesis (Gong, Hao et al. 1998; Reinelt, Hofmann et al. 2003; Yildiz, Doi et al. 2005). Yet deciphering the structural mechanism remains elusive since it is not always possible to crystallize every conformation of a PAS domain in its different regulatory modes as well as provide the necessary biochemical evidence underlying each conformation's physiological significance. However, the growing number of crystal structures of these homodimeric PAS complexes has begun to shed some light to this regard.

In this regard, homodimeric PAS domains have shown significant differences not only in the orientation in which they assemble, but also in the PAS structural elements utilized. For instance, the homodimeric PAS domain of RmFixLH essentially has a parallel orientation with respect to N and C termini (**Fig. 1-6a)**, while the EcDOS PAS homodimer exhibits an orthogonal interaction with respect to its β-strands. Both of these structures reveal extensive interactions by N-terminal α-helical regions not normally considered to be part of the PAS fold (Miyatake, Mukai et al. 2000). On the other hand, bjFixLH contains a homodimeric PAS domain with a predominantly β-sheet interface that can be oriented with respect to its partner subunit's same interface both in a parallel (**Fig.**

1-6b) or anti-parallel manner depending on the crystallization condition (Gong, Hao et al. 1998). This experimental result suggests that caution must be used in the interpretation of these models as crystallization may artificially induce certain contacts that become interpreted as being part of an interface. No structure exemplifies this principle more than that of the X-ray crystal structure of CitA periplasmic PAS domain. Two completely different dimeric forms could be observed in the asymmetric unit (Reinelt, Hofmann et al. 2003). One dimer form involved mainly the β-sheet interface in an orientation much like the anti-parallel form of the BjFixlH PAS homodimer (**Fig. 1-6c first structure**). The total buried surface for both subunits along this dimer interface is 1798 $Å^2$. The other dimer form has an interface dominated by only short regions of its Cα and Eα helices, thus only burying 383 $Å^2$ (**Fig. 1-6c second structure).** It remains questionable that these drastically different interfaces could both—or even just one—could be physiologically relevant as gel-filtration studies suggest that the protein is mainly monomeric.

Despite the existence of many homodimeric PAS domain structures, there are no heterodimeric PAS domain X-ray structures, possibly reflecting the low affinity of interaction and the difficulties of protein expression and purification. Indeed, this may reflect the idea that PAS-mediated cellular responses need to be transient to preventing the cell from overcompensating for a given signal. However, our laboratory was able to generate a low resolution model for the

HIF2α/ARNT PAS-B heterodimer using nitroxide spin-labels to achieve site-specific paramagnetic broadening of resonances belonging to residues that are in their proximity (Card, Erbel et al. 2005). With this approach, the two domains were shown to interact in an antiparallel orientation utilizing the common β-sheet interface utilized in many other PAS: PAS interactions (**Fig. 1-6d**) (Cusanovich and Meyer 2003; Erbel, Card et al. 2003). In contrast to this, preliminarily models for the ARNT PAS-B homodimer demonstrate a parallel association between the subunits (Card, Erbel et al. 2006 (in review)). Whether these different orientations for homo and heterodimerization and the conformational change between them will have a physiological significance remains to be seen.

Questions addressed in this thesis

Thus, in regard to PAS dimerization and its role in regulating signal transduction mechanisms, my thesis focuses on a PAS-containing histidine kinase called KinA. The ArcB and LuxPQ sensor kinases mentioned above are two PAS-containing histidine kinases where it could be shown that the specific conformation maintained in dimeric PAS domains could allosterically inhibit the histidine kinase domain (Malpica, Franco et al. 2004; Neiditch, Federle et al. 2005). However, these proteins are *membrane*-tethered; it remains a question if such PAS-mediated mechanisms could also exist in a *cytosolic* histidine kinase. To

address this question, we investigate whether dimerization by the PAS-A domain has a regulatory role in controlling KinA activity.

Introduction Part II: Sporulation

Bacillus subtilis has two different life cycles: One is vegetative growth in which cells divide symmetrically into two identical daughter cells and the other is sporulation, in which cells divide asymmetrically into a very large mother cell and a much smaller forespore. In this latter process, the forespore is actively transcribing genes necessary to create a thick outer coat (**Fig. 1-7**). Eventually the mother cell dissipates and the forespore buds off as the spore. With its thick outer coat known as the cortex, the spore is resistant to a number of harsh environmental conditions including high temperature, extreme pH, and physical force or pressure (Stragier and Losick 1996; Stephenson and Lewis 2005). The decision to undergo such a process is radical because it attenuates proliferation in order to rededicate cellular resources toward the production of the cortex. Such a decision is made only under radical conditions that absolutely threaten the cell's ability to propagate. These conditions include drastic temperature and pH changes, but sporulation is most commonly elicited as a reaction to nutrient deprivation (Piggot and Hilbert 2004; Barak, Ricca et al. 2005; Stephenson and Lewis 2005).

After several decades of investigation, the exact trigger for sporulation has eluded scientists. Depriving nutrients in a systematic manner has produced no

compound that can be removed from media to robustly trigger sporulation perhaps underscoring the great number of regulatory controls that prevent these cells from receiving a "false alarm" and halting propagation when conditions do not merit it (Cooney, Whiteman et al. 1977; Nishihara 1979; Sadaie and Kada 1985; Hecker, Schumann et al. 1996). Recently, a series of kinases were identified that have been shown to be essential for initiation of sporulation. Their biochemical purification and characterization has facilitated biophysical measures for identifying the means by these kinases are triggered.

Promoter for sporulation: Spo0IIA

The promoter Spo0IIA was identified through traditional mutagenesis methods that controlled the upregulation of most of the genes required for sporulation. Subsequently, it was shown that SpoIIA the transcription factor that recruited the proper enhancer elements for activation of this promoter. Furthermore, SpoIIA proved to be the final phosphoacceptor in a phosphorelay involving two response regulators SpoIIF and SpoIIB, and an initiator (sporulation) kinase (**Fig. 1-8**). Two sporulation kinases, KinA and KinB, were originally identified that contribute phosphates to the phosphorelay through the observation that deleting the genes that encode these proteins lead to a sporulation deficient cell line(Trach and Hoch 1993). Later, through the demonstration that certain regulons of SpoIIA were active even in a $KinA^-/KinB^-$ deficient cell line, three other Sporulation

kinases—KinC, KinD, and KinE (**Fig. 1-1**)—were also identified as contributors (Jiang, Shao et al. 2000). Finally, the finding that a KinA-deficient cell line could decrease sporulation rates to 5% of wild type (Perego, Cole et al. 1989; Wang, Fabret et al. 2001) has designated KinA as the master regulator for sporulation under laboratory conditions. Because of this importance in initiation of sporulation, my research focused on the mechanism by which KinA controlled its kinase activities.

KinA

At the molecular level, KinA comprises three tandem PAS domains followed by a histidine kinase domain, the domain boundaries for which were identified by V8 limited proteolysis (Wang, Fabret et al. 2001). Of the five *B. subtilis* sporulation kinases, only KinC and KinE also contain a PAS domain, possibly reflecting the divergent nature of signal detection between these proteins. Nevertheless, without the PAS-A domain, the KinA loses 95% of its *initial rate* of autophosphorylation, demonstrating the critical importance of the KinA PAS-A domain in its function. The significant role played by PAS domains in KinA is further underscored by the finding that *all* autophosphorylation is abolished with the truncation of all the PAS domains. In addition, mutations to PAS-A (F77S) and PAS-C (I280T) both produced sporulation-deficient phenotypes, implicating these two domains in signal sensing. The lack of the same effects by these mutants on *in vitro* KinA

phosphorylation tempts speculation that a possible upstream effector molecule is unable to interact with these mutant KinA molecules *in vivo* (Stephenson and Hoch 2001). More recently, it was demonstrated through UV-crosslinking to ATP that ATP was a ligand for this domain (Wang, Fabret et al. 2001). Furthermore, by monitoring the transfer of γ-phosphate to GDP on thin layer chromatography, Wang et al., showed that KinA PAS-A catalyzed this phosphotransfer reaction (Wang, Fabret et al. 2001). This led to speculation that hydrolysis of ATP may initiate additional conformation changes within the context of the full-length KinA such as to increase its activity. These studies suggest that the N-terminal signal-sensing portion of KinA, particularly the PAS-A domain appears essential to both signal sensing and conformational activation of KinA. As such, the emphasis of my research became to investigate the regulation of kinase activity by the PAS A domain.

Kinase structure and function

Bacterial histidine kinases are identified by unique signature sequences called H, N, G1, F and G2 blocks (Parkinson and Kofoid 1992) (**Fig 1-9a)**. Block H, located in the N-terminal half of the kinase, includes the histidine residue that serves as the site for autophosphorylation. Blocks G1 and G2 are often similar to glycine-rich portions of nucleotide-binding domains (Rossmann, Moras et al. 1974). Block N contains a well-conserved asparagine residue and is often found

in β-strand structures. Mutations in blocks N, G1, and G2 eliminate autokinase activity, but some proteins only have G1 and no G2 blocks (Yang and Inouye 1993). On the other hand, little is known about the F block other than that it often serves as a spacer between blocks G1 and G2. Because of these well-conserved sequences, the catalytic domains of histidine kinases often have similar structures (Stock, Ninfa et al. 1989) (**Fig. 1-10**).

Differences in domain organization is the basis for their grouping into two major classes. Almost all histidine kinases, including KinA, identified belong to the Class I histidine kinases, which consist of H-box containing regions that are directly linked to the regions that contain the other four conserved boxes. The other regions have been designated as the catalytic and ATP-binding domain (**Fig. 1-9b)**. Class II histidine kinases, however, have H-box containing regions that are several domains away from the catalytic domain (**Fig. 1-9c)**.

Histidine kinases trans-autophosphorylate the conserved histidine from its partner subunit in the dimer (Dutta, Qin et al. 1999). The substrate domain for phosphotransfer in EnvZ, a Class I histidine kinase, also serves as its dimerization domain and, as such, it is known as the Dimerization and Histidine phosphotransfer domain (DHp domain). The two DHp subunits form a four-helix bundle. It has been speculated that the catalytic domain must flank the four-helix bundle in such a manner that the helix containing the histidine-phosphoacceptor site directly faces the twin subunit's catalytic domain (**Fig. 1-10b&c**). Because

the catalytic and DHp domains of EnvZ were solved separately as NMR structures, this idea remains unproven.

This domain organization is clearly different for CheA, a Class II histidine kinase, where dimerization and phosphotransfer elements exist as distinct domains. As exhibited in the crystal structure, the catalytic core faces away from the dimerization domain to find the exposed Histidine-containing Phosphotransfer domain (HPt) (**Fig. 1-10a)**. Because CheA is the only member of this class identified to date, this classification system is becoming less popular and more scientists are simply considering CheA as an aberrant version of a canonical histidine kinase.

Kinase regulation

Data so far concerning the exact molecular mechanisms by which histidine kinases are regulated has been speculative. Perhaps the most definitive mechanism to date is in regard to KinA histidine kinase regulation by an inhibitor Sda that is upregulated in response to DNA damage such that progression sporulation is halted, thus serving an important DNA damage checkpoint that ensures that only *B. subtilis* with integral genomes continue to propagate in future surviving generations. Mechanistically, Sda binds directly to the DHp domain, sterically occluding the histidine phosphotransfer site preventing exposure to the catalytic domain (Rowland, Burkholder et al. 2004).

Although direct inhibitors like Sda provide simplistic models for kinase regulation, probably most regulation of histidine kinases occur through more intricate allosteric changes that are induced by the activation of sensor domains. The most studied examples of this phenomenon have been in membrane-tethered sensor histidine kinases. In *E. coli*, ArcB sensor kinase modulates the response to redox potential. It was shown through band-mobility changes on Western blot analysis that the N-terminal cytosolic PAS domain responds to oxidants by forming two intermolecular disulfide bonds between preformed dimers (**Fig. 1-4a**). This covalent modification propagates an allosteric change to the histidine kinase domain that abolishes its autophosphorylation activity (Malpica, Franco et al. 2004). Similarly, in order to modulate quorum sensing in Vibrio harveyi, LuxP has N-terminal tandem PAS domains that respond to the binding of quorum-activated LuxP. Mutations that disrupt the crystallographically-defined apoLuxP:LuxQ interface as reconstituted in *Vibrio harveyi* cells increase the sensitivity of these organisms to AI-2, suggesting that certain interactions are disrupted to facilitate a conformational change that exhibits the formation of new interactions. This new signaling state then allosterically affects the histidine kinase such that it reverses its function to that of a phosphatase (**Fig. 1-4b**) (Neiditch, Federle et al. 2005).

Because no sensor domains have been shown to have a direct interaction with their cognate histidine kinase, it has long been speculated that the linker

between the sensor and the kinase domain must have structural importance to allow the transduction of environmental signals. And indeed, the identification and characterization of HAMP domains (domain present in histidine kinases, adenylyl cyclases, methyl-accepting proteins and phosphatases) (Aravind and Ponting 1999) that connect sensors with kinases has lent support to this claim. Existing in many signaling contexts, this domain comprises of an approximately 50 residue helix-turn-helix motif. In EnvZ, mutations to the HAMP domain have been able to block its osmosensing ability (Park, Saha et al. 1998). Probably underscoring the universality of linker signaling function in histidine kinases is the fact that fusing various sensor modules and disparate histidine kinase domains from other systems with HAMP domains maintains ligand responsiveness (Utsumi and Inoue 1991; Appleman, Chen et al. 2003).

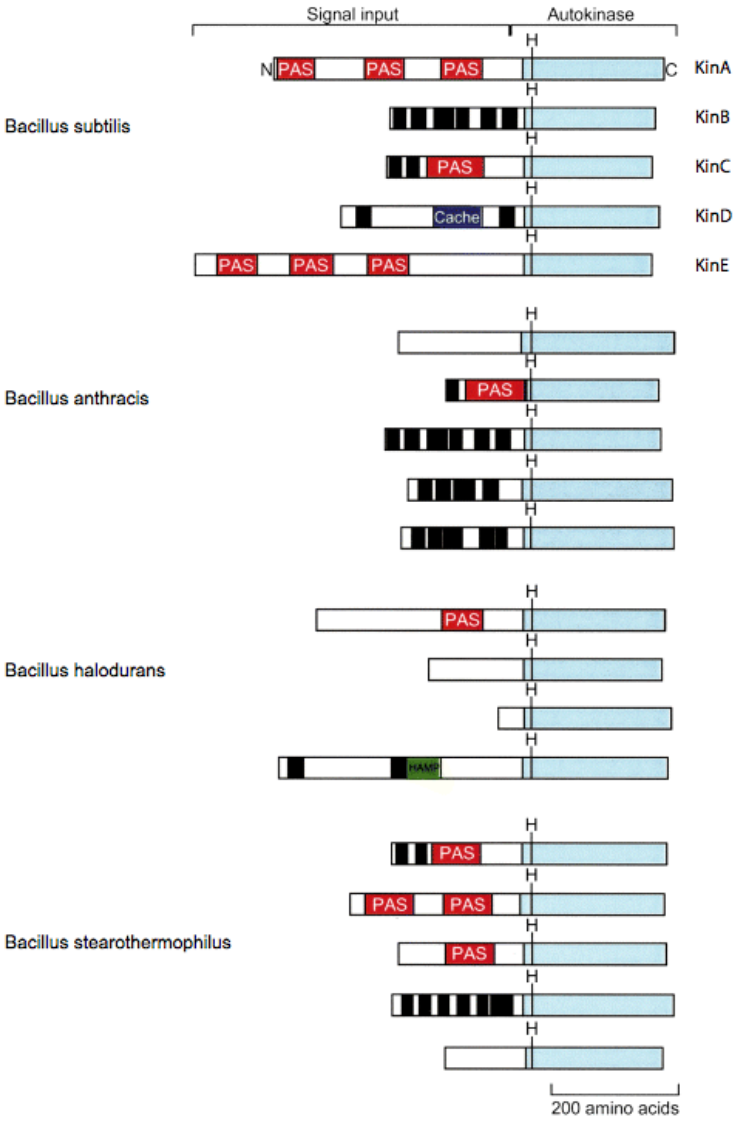


Figure 1-1. Domain architecture of putative *Bacillus* species phosphorelay sensor kinases (Figure adapted from Stephenson and Hoch (Stephenson and Hoch 2002).

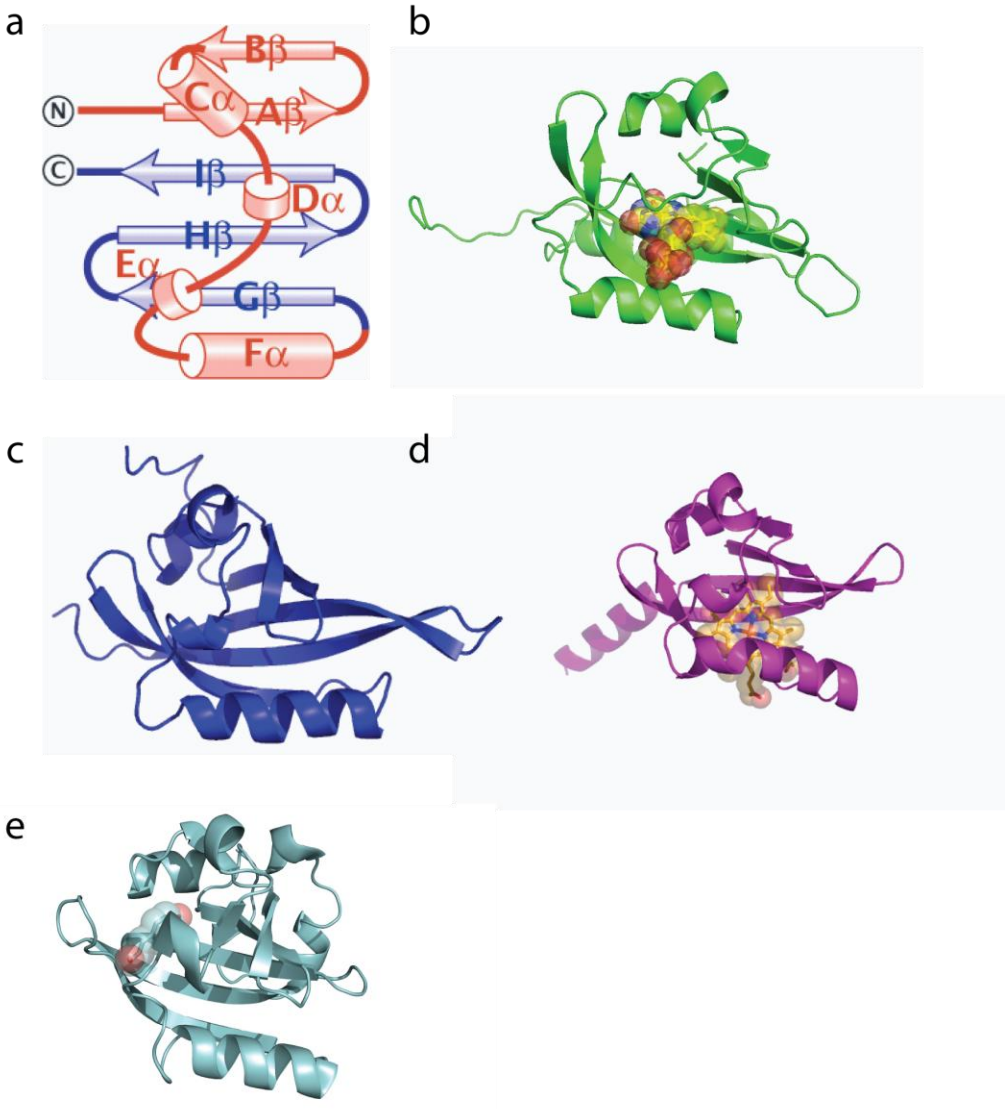


Figure 1-2. PAS domain structures (a) Topology with labeling of PAS (blue) and PAC (red) boxes. (b-d) Representive structures (b) LOV2 of *Avena Sativa* phototropin1 (Protein Data Bank, PDB entry, 1G28 (c) Human ARNT PAS-B (PDB entry 1X0O) (d) bjFixLH (PDB entry 1DRM) (e) PYP from Ectothiorhodospira halophila (PDB entry 2PHY).

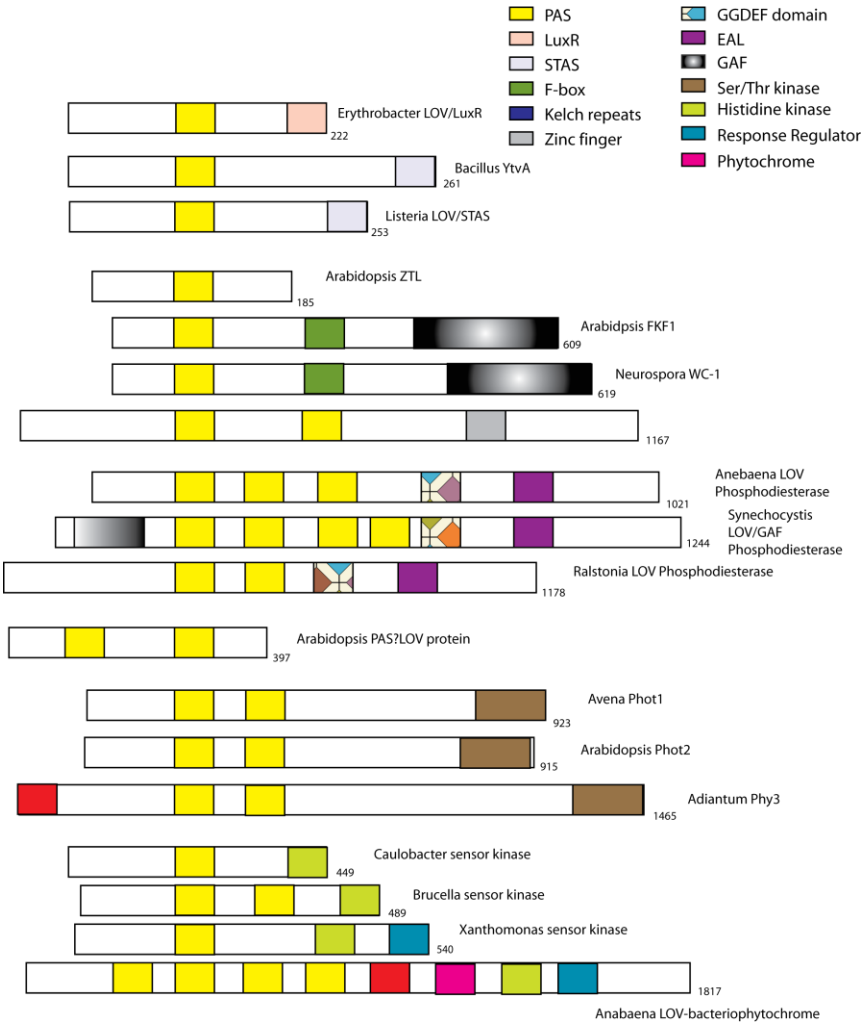


Figure 1-3. Representative domain architectures of proteins containing LOV domains, which represent a unique type of the PAS fold capable of light sensing. Figure adapted from Crosson et al., 2003 (Crosson, Rajagopal et al. 2003).

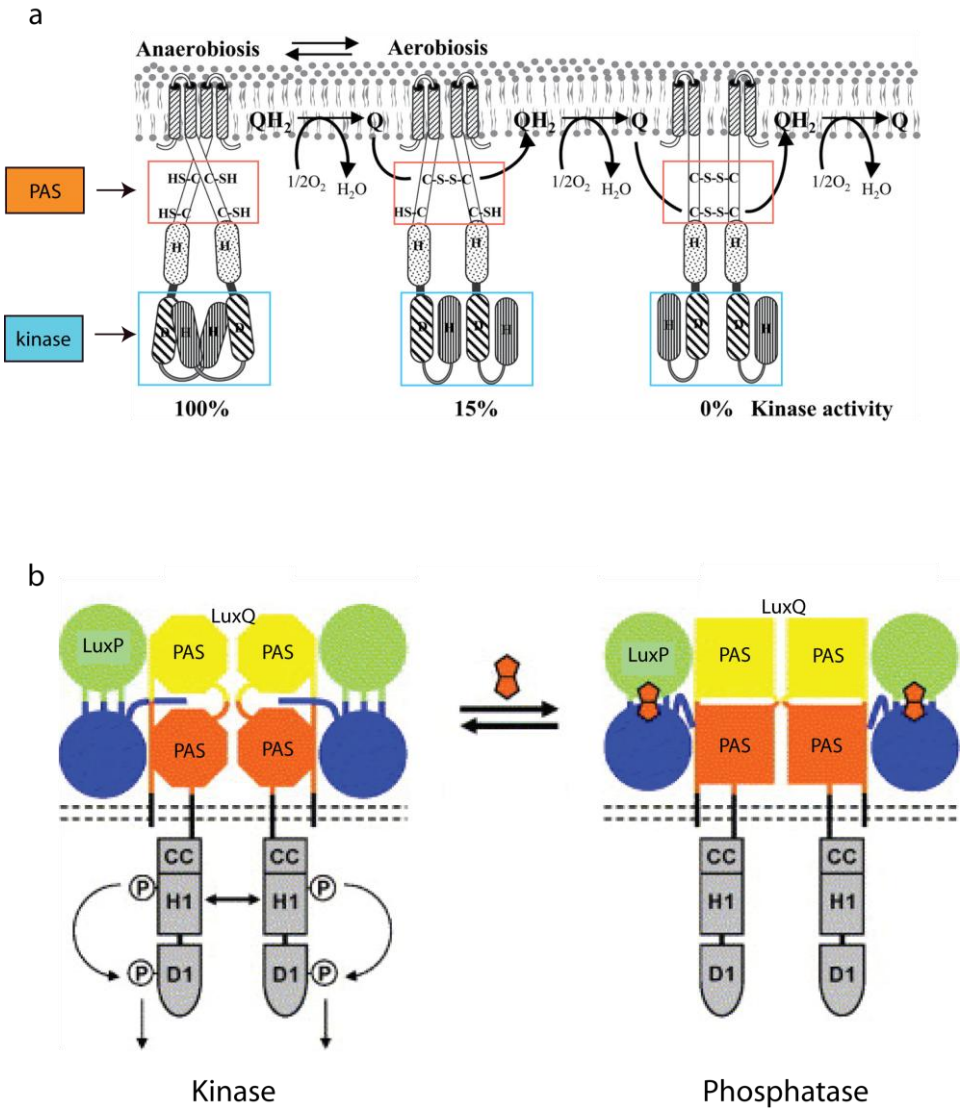


Figure 1-4. Representative models for histidine kinase regulation. (a) The signal molecule autoinducer-2 (AI-2) binds to LuxP, which complexes with the PAS domains of LuxQ. Conformational changes in the PAS domains allosterically converts the histidine kinase to a phosphatase. Figure adapted from Neiditch et al (Neiditch, Federle et al. 2005). (b) ArcB sensor kinase forms covalent bonds in response to low redox potential. Ensuing conformational changes allosterically inhibits the histidine kinase. Figure adapted from Malpica et al. (Malpica, Franco et al. 2004)

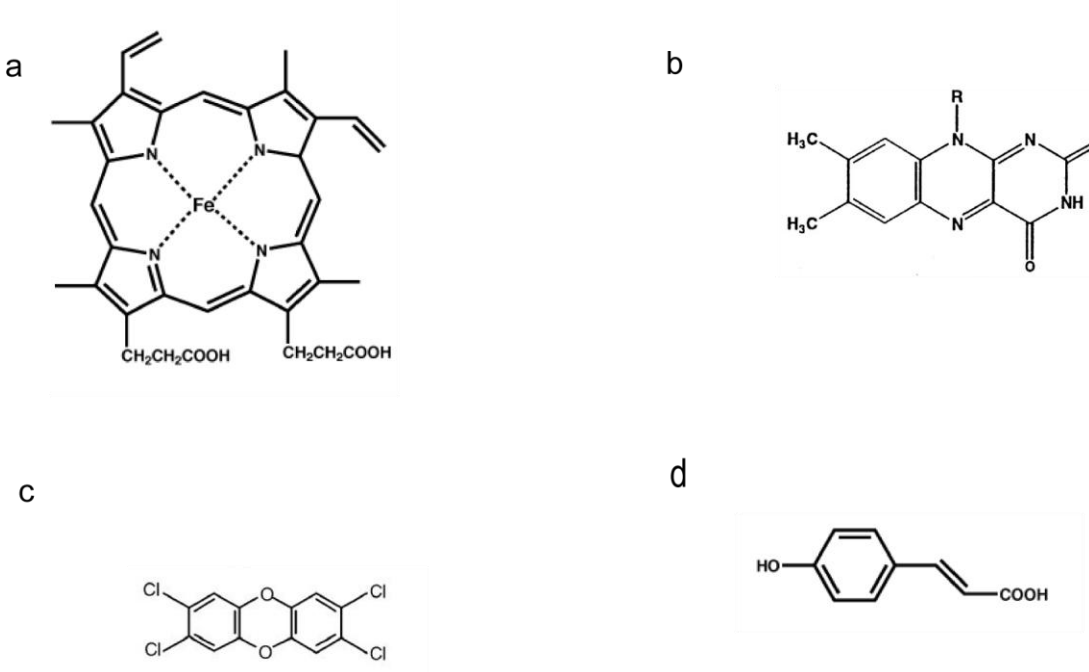


Figure 1-5. Representative PAS ligands (a) Heme used as cofactor in PAS domain of bjFixL to sense oxygen (b) Flavin Mononucleotide (FMN) used by LOV domains to affect photoactivation via covalent bond formation (c) Dioxin activates Aryl Hydrocarbon Receptor to form functional transcription factor. (d) 4-hydrocinnamic acid, a chromophore is PYP, is light-activated to affect cis-trans bond isomerization.

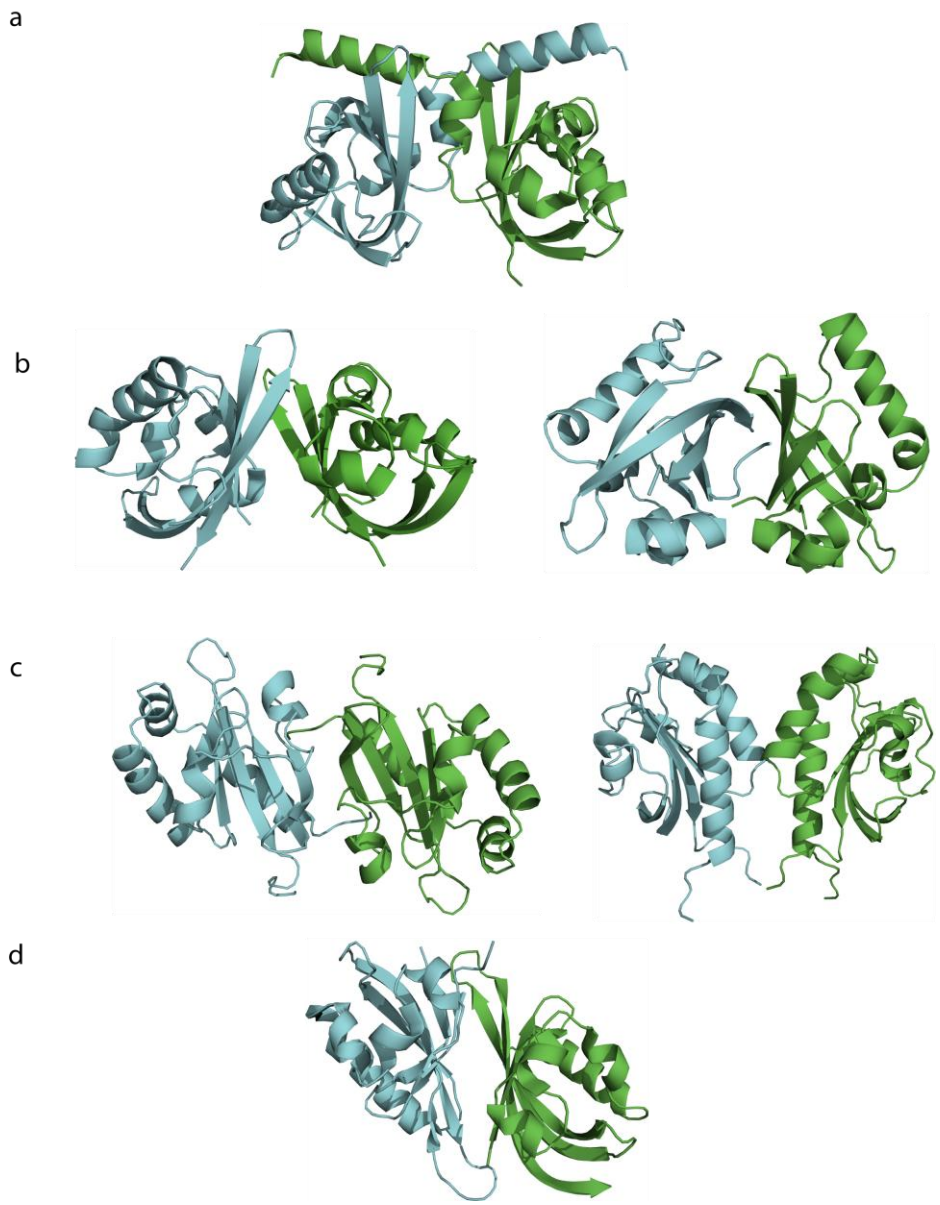


Figure 1-6. PAS dimer structures (a) rmFixLH (PDB entry 1EW0) (b) bjFixLH C32 (PDB entry 1LSW) and R32(PDB entry 1XJ6) forms, respectively.(c) CitA EG and GJ dimers, respectively (PDB entry (1P0Z) (d) HIF/ARNT heterodimer (PDB entry 2A24).

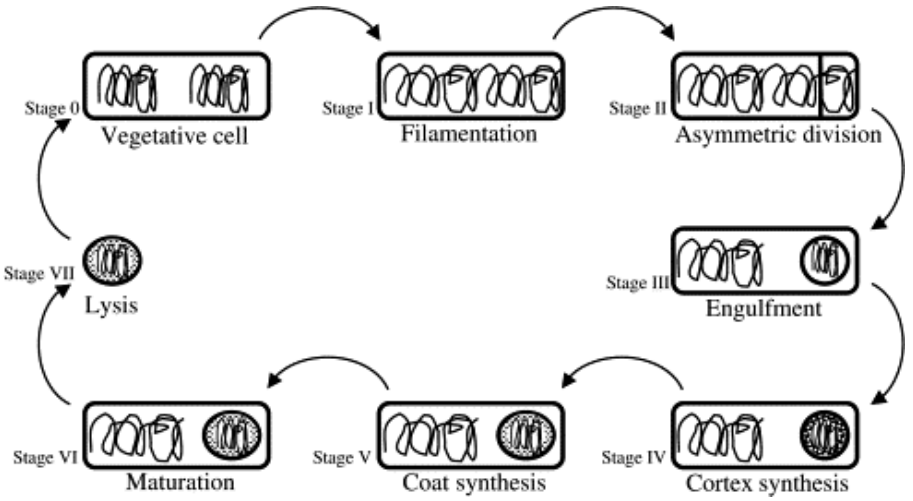


Figure 1-7. Summary of stages of sporulation. Cells in vegetative growth complete replication of the chromosome as shown by squiggly lines. On the other hand, cells just initiating sporulation do not complete DNA replication, but instead form a pair of partial replicated chromosomes in an axial chromatin filament that spans continuously the length of the cell (stage I). By stage II, the mother cell and smaller prespore have divided asymmetrically, trapping a third of the chromosome in the prespore. After the prespore is engulfed by the mother cell, it becomes the independent protoplast completing stage III. Cortex and peptidoglycan synthesis in the prespore occur in stage IV and V, respectively. Spore maturation exhibits few physical changes, but during this time, the spore acquires resistance properties (stage VII). Finally, stage VII completes the cell cycle with the release of the mature spore after lysis of the mother cell. Figure adapted from Stephenson and Lewis (Stephenson and Lewis 2005).

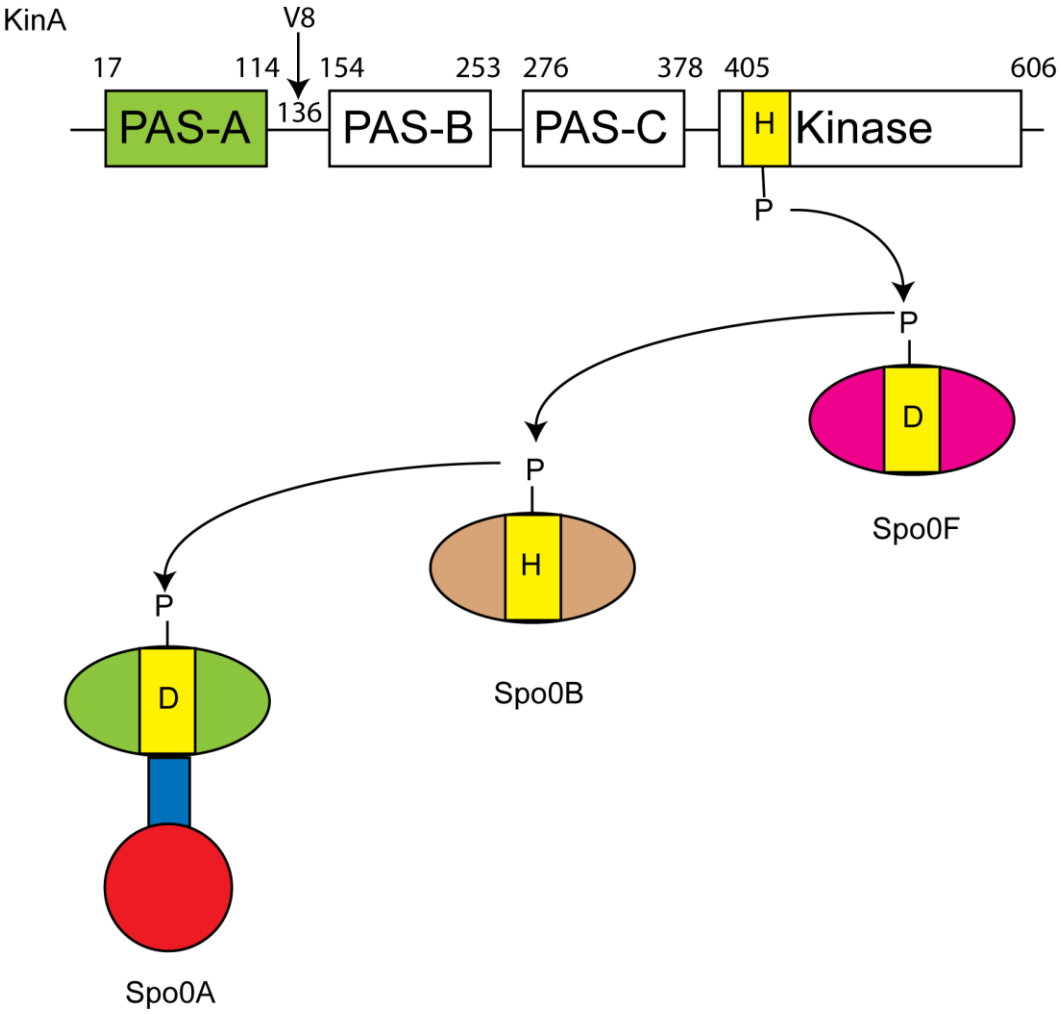


Figure 1-8. Sporulation phosphorelay. KinA controls the checkpoint for sporulation in *B. subtilis* by initiating a phosphorelay, consisting of three other response regulators. The final response regulator in this pathway, Spo0A, activates hundreds of genes involved in phosphorylation. Domain boundaries are from Stephenson and Hoch, 2001 (Stephenson and Hoch 2001). V8 proteolysis site is based on Wang and Hoch, 2001 (Wang, Fabret et al. 2001).

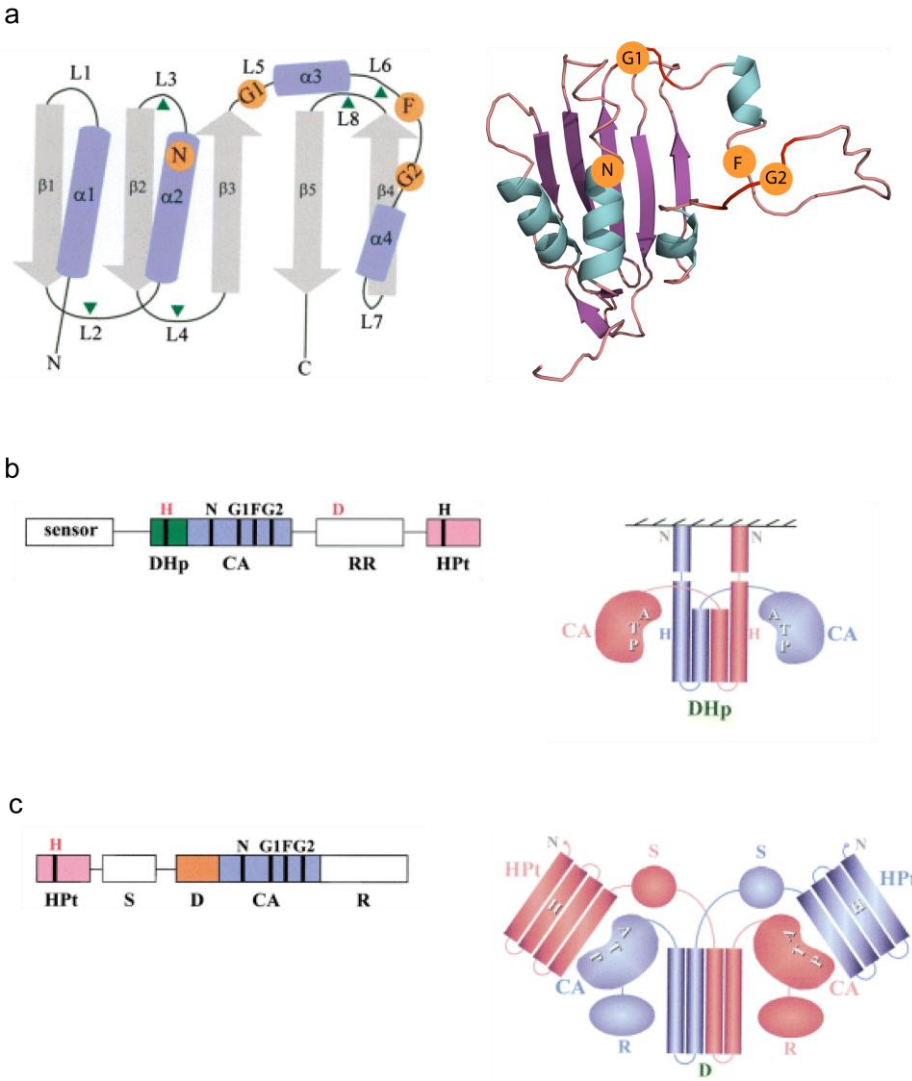


Figure 1-9. Schematic diagrams of histidine kinases (a) Ribbon diagram of EnvZ with conserved elements mapped to the structure (PDB entry 1BXD). (b) Representative domain architecture and organization of a Type I histidine kinase (EnvZ). Dimerization and phosphotransfer elements (DHp) are together in one domain, flanked by a catalytic domain. Not shown in domain organization are response regulators (RR) and sensor domains(c) Representative architecture and organization of a Type II kinase (CheA). Dimerization (D) and phosphotransfer elements (HPt) are in separate domains. A regulatory domain (R) binds upstream effector CheW, while the substrate binding domain (S) binds the phosphorylation target CheB. Figures adapted from Dutta et al. (Dutta, Qin et al. 1999).

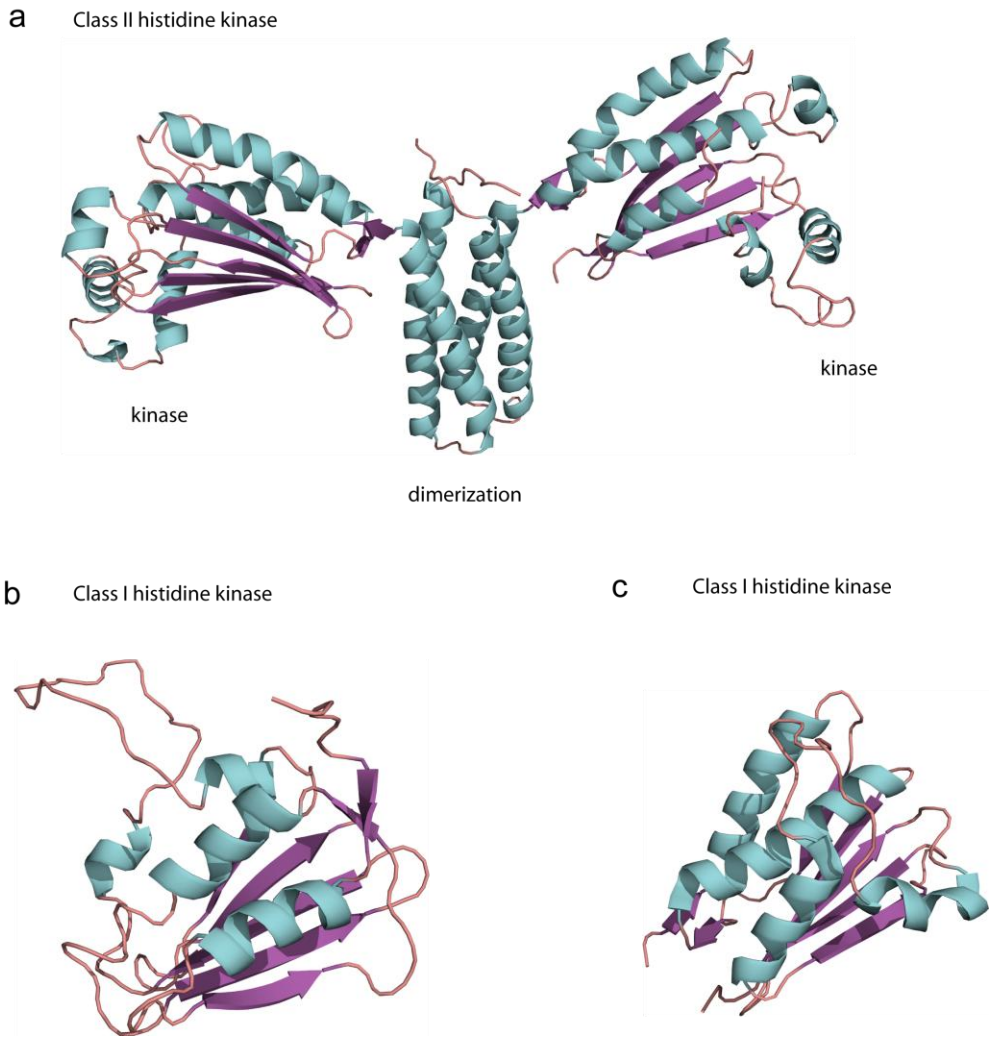


Figure 1-10. Comparison of structures of Histidine kinases, revealing strong conservation of catalytic domain structure. (a) Kinase core (dimerization and kinase domain) of Thermotoga maritima CheA (Protein Data Bank, PDB, entry 1B3Q) [15]. (b) Catalytic domain of *E. coli* EnvZ complexed with ADP (PDB entry 1BXD). (c) Catalytic domain of *E. coli* PhoQ complexed with Mg^{2+} and the nucleotide analog AMPNP (PDB entry 1ID0).

Chapter 2: Initial Characterization

To investigate why KinA requires the PAS-A domain for its function, we began by purifying bacterially-expressed kinase and PAS-A domains and assessing the folding of each individual domain. After successful purification, we wanted to study possible direct interactions between the two domains using NMR. Finally, we sought to validate the previously-reported interaction between PAS-A and ATP.

MATERIALS AND METHODS

Cloning and expression

Routinely in this laboratory, we use the Streptococcus protein G beta 1 (Gβ1) domain an N-terminal fusion tag improving the expression and solubility of a wide range of target proteins. To improve the resistance of this domain to N-terminal proteolytic processing, we use a point mutant (T2Q) variant of this protein (Gronenborn and Clore 1996; Huth, Bewley et al. 1997). Because it is a very stable small protein of 56 amino acids, Gβ1 produces well-dispersed NMR spectra and contributes little to the overall rotational correlation time of derivative fusion proteins, especially since the linker to the target protein is quite flexible. Thus, it can provide also provide a good positive control for protein expression and solubility in whole cell lysate screens (Huth, Bewley et al. 1997; Zhou,

Lugovskoy et al. 2001; Hammarstrom, Hellgren et al. 2002; Card and Gardner 2005). Finally, although it is possible to purify Gβ1-fusion proteins by affinity purification utilizing IgG resins, the use of harsh elution buffers (pH<4) often necessitates an alternative purification strategy. To this end, we added a His_6 sequence on the N-terminus of Gβ1, facilitating standard Ni^{2+}-NTA affinity purification (Harper, Neil et al. 2003).

For high level expression of proteins, we PCR-amplified from chromosomal DNA isolated from *B. subtilis* strain 1A40 (Bacillus Genetic Stock Center, Columbus, OH) and cloned into the parallel vector pHis$_6$Gβ1 (Harper, Neil et al. 2003) using EcoRI and XhoI restrictions sites. Constructs encoding residues for PAS-A (aa 1-117, 10-117, 10-151, and 10-117) and kinase domain (aa 383-606) were overexpressed as TEV-cleavable N-terminal His$_6$Gβ1-tag fusion proteins using *E coli* BL21 (DE3). Cells were grown to OD_{600} =0.6 at 37 °and induced with 0.5 mM IPTG for 16 h at 20 °C. Cells were then centrifuged and the resulting pellet resuspended in buffer containing 25 mM Tris (pH 7.5) and 100 mM NaCl for PAS-A constructs and 25 mM Tris (pH 7.0) and 25 mM NaCl for kinase domain.

Purification of KinA PAS-A and kinase domain

Resuspended cells harboring these overexpressed proteins were lysed by French press or sonication, His$_6$Gβ1-tagged kinase domain (**Fig. 2-1**) or PAS-A (**Fig. 2-**

4) were purified using Ni^{2+}-loaded NTA affinity chromatography columns, eluted using a linear 20 mM-300 mM imidazole gradient, and subsequently exchanged into imidazole-free buffer. The His_6Gβ1-tag was cleaved by incubating overnight with His_6-TEV at 25°C. TEV cleaved proteins were then loaded on another Ni^{2+}-loaded NTA column. TEV-cleaved KinA PAS-A constructs were found to bind this second Ni^{2+}-loaded NTA affinity chromatography column with low affinity (20 mM imidazole). As such, additional separation could be achieved from flow-through contaminants and the cleaved His_6Gβ1-tag bound with higher affinity (300 mM imidazole) (**Fig. 2-5**).

NMR sample preparation

For NMR samples, cells were grown in M9 minimal media containing 1g/L of $^{15}NH_4Cl$ for U-^{15}N samples supplemented with 3g/L of [$^{13}C_6$]glucose for U-$^{15}N/^{13}C$ labeled samples. The same protocol was applied in producing deuterated samples, except cells were grown in 99.9% D_2O M9 minimal media **(***Conditions for spectra are detailed in the figures where they are presented***).**

RESULTS

^{15}N-^{1}H HSQC of the Kinase domain reveals well-folded domain

Using domain boundaries that had previously been established as being able to bind a inhibitor Sda and also be a phosphoacceptor for Spo0F (Rowland, Burkholder et al. 2004), we were able to overexpress a kinase domain that could be purified to homogeneity through the use of Ni^{2+}-NTA affinity chromatography columns and would run as a monodisperse peak on a Superdex 75 prep grade column (Pharmacia) with negligible void peak, indicating the sample contained minimal aggregation (**Fig. 2-1b**). The ^{15}N-^{1}H HSQC spectra of the kinase domain indicates that the kinase domain is well-folded with peak broadening consistent for a protein of this size (**Fig. 2-2a**). The lack of intense peaks with poor proton chemical shift dispersion in the middle of the amide region of the spectra ($^{1}H^{N}$ 7.8-8.5 ppm) as compared to the rest of the spectra led us to believe that N and C-termini flanked just enough of the minimally-required sequence for folding. To resolve any ambiguities in the folding of this protein we improved the spectra through the use of deuteration and increased field strength (**Fig. 2-2b**). These latter improvements produced decently well-dispersed spectra whereby the automatic peak picking option of NMRview could find (Laskowski 1993) 206 resonance peaks out of a possible 224 residues for the kinase domain.

Furthermore, protein samples were generally stable without degradation for at least 2 days. As such, we proceeded to purify the PAS-A domain.

Domain optimization of PAS-A

Initial characterization of the PAS-A domain was performed on a lyophilized protein sample provided by the laboratory of Klaas Hellingwerf (University of Amsterdam) with the PAS-A domain boundaries established through V8 proteolysis (Wang, Fabret et al. 2001). This sample showed extensive degradation when samples were run on SDS-PAGE (**Fig 2-3a**). NMR spectra reflected this result; despite having an absorbance reading at 280 nm that indicated protein concentrations of approximately 300 μM, the spectra required 224 scans per transient (~13 hours total acquisition time) to begin observing signs of folded chemical shifts (**Fig 2-3b**). These spectra showed extensive differential broadening and crowding in the unfolded region of the spectra (^{1}H 7.8-8.4 ppm). However, two facets of this spectra allowed us some optimism at being able to use NMR to characterize this protein: (1) Signals with ^{1}H shifts outside the unfolded regions were observed, indicating some portion of the protein had stable tertiary structure and (2) a signal with a very far downfield chemical shift was observed (^{1}H ~10.7 ppm) that our laboratory believes has found to be diagnostic of a PAS domain fold. Examples of this are found in the C-terminal PAS-B

domains of ARNT and HIF-2α, where a residue in the FG loop is placed in the strong magnetic dipole created by the F helix (**Fig 2-3b**).

To further characterize this previously-lyophilized KinA PAS-A sample, we performed size exclusion chromatography (SEC) using a Pharmacia Superdex 75 prep grade column. Although previously reported as a monomer (Wang, Fabret et al. 2001), our elution profile was such that the calculated molecular weight for the major peak was ~31 kDa which is consistent with the protein being dimeric. Furthermore, other peaks consistent with a tetramer and higher order aggregate were also observed in the elution profile (**Fig. 2-4**). Along with the NMR data, we speculated at this time that the difference in oligomeric status was due to either aggregation or excess unfolded termini as unfolded sequences often run much larger than expected on gel-filtration. Despite this deviation from published results, we were initially satisfied that the protein could run as a monodisperse peak on gel-filtration, an indication of some folded character in this protein.

With these reasons to believe that we could express a folded PAS-A domain, we decided to more stringently map the domain boundaries; the rationale being that our initial construct was either too short at a termini such as to destabilize the overall fold or it was too long such that extraneous unfolded regions led to aggregation and/or protein degradation via proteases. Four constructs were created to investigate the domain boundaries: #1-151, 10-151, 1-117, and 10-117. As discussed in the Material and Methods above, each were

purified via Ni^{2+}-NTA affinity chromatography, cleaved from the His_6GB1 solubility enhancer tag, and repurified by Ni^{2+}-NTA affinity chromatography after buffer exchange to low imidazole. As the cleaved PAS-A domain has low affinity (20 mM imidazole) for Ni^{+2}-NTA affinity chromatography, it was possible to achieve very good purity with such a simple purification scheme (**Fig. 2-5**).

NMR spectra of these constructs were recorded and analyzed. The inclusion of extraneous sequences only led to spectra more crowded in the unfolded region by intense sharp peaks, indicating that unfolded termini were being extended (**Fig 2-6**). As such, we decided to more extensively characterize the construct 10-117. Spectra of this construct were well dispersed and contained much less differential broadening than spectra from the other constructs (**Fig. 2-7**). Furthermore, NMR samples remained stable at 37 º C for over a week as monitored both by monitoring ^{15}N-^{1}H HSQC spectra each day for a week and by running NMR samples on SDS-PAGE. However, only 72 of a possible 108 peaks could be counted by the automatic peak picking option of NMRview (Johnson 2004). Nevertheless, more assured that we had removed excessive unfolded termini and achieved a tight domain boundary, we again performed gel-filtration to confirm the monomeric status reported by Wang et al. However, we continued to achieve results consistent with the PAS-A domain being dimeric (**Chapter 5, Fig. 5-1**) even when the protein was buffer-exchanged into a number of different conditions varying both pH and salt concentrations.

As a final test of the appropriateness of our domain boundaries, we used limited proteolysis (Cohen, Ferre-D'Amare et al. 1995). In this technique, proteins are subjected to small quantities of protease with to remove extraneous sequences at the termini and produce more compact domains. The caveat of such studies is that solvent-exposed loops or extended sequences with partial properties of secondary structure within well-defined domains are sometimes also susceptible to cleavage.

Trypsin, chymotrypsin, and endoproteinase Glu-C (V8 protease) were all used to cleave our NMR optimized KinA PAS-A domain (residues 10-117). Trypsin (bovine pancreas; Sigma) cleaves after lysine and arginine residues; chymotrypsin cleaves mainly after phenylalanines, tyrosines, tryptophans, and some sometimes after large hydrophobic residues such as leucines and methionines; V8 protease (*Staphylococcus aureus*) cleaves after glutamates and aspartates. Chymotrypsin and V8 protease had very modest effects and only after extended incubation times. However, trypsin proteolysis affected a ~2 kDa shift in the KinA PAS-A protein band on SDS-PAGE (**Fig. 2-8a**). To quantitate the exact mass change of this result, we submitted trypsin-proteolyzed KinA PAS-A fragments for ESI-mass spectrometry analysis. The masses found in our sample by this technique were 10,785.88 and 10,530.38 Da (**Fig. 2-8b**). Guided by the program Protein Prospector found on the UCSF Structural Biology web site (http://prospector.ucsf.edu/), we compared calculated masses for possible

fragments to the masses observed by ESI-mass spectrometry. We found that shortened sequences encoding the regions of residues 10-99 and 10-97 had calculated molecular weights with 1 Da of the mass spectrometry-observed fragments of 10,785.88 and 10,530.38 Da, respectively. Another possible fragment encoding the residues 16-112 had a calculated molecular weight within 2 Da of the 10,530.38 Da mass found by mass spectrometry. If we locate residues 97 and 99 on the JPRED secondary structure prediction, the cleavage sites map to the residues located right after the Hβ strand, in the HI loop (**Fig. 2-9a red arrows**). The other possible fragment locates trypsin cleavage sites in the β-turn between the Aβ and Bβ strands and at residue 106 in the Hβ strand (**Fig. 2-9a blue arrows**). Like the possible cleavage at the HI loop, this would also indicate non-specific cleavage in a region between two JPRED-predicted β-strands. We hypothesized that the most likely explanation was that cleavage was occurring in loop regions without affecting the tertiary structure of KinA PAS-A. Nevertheless, to be certain that both our NMR-optimized domain boundaries were appropriate and that Aβ and Iβ strands were in their JPRED-predicted positions, we recorded ^{15}N-^{1}H HSQC of proteolyzed ^{15}N-labeled KinA PAS-A. Because no peaks disappeared from the unfolded region of the spectra, we concluded that proteolysis was indeed occurring not at extraneous unfolded termini, but at solvent-exposed loops such that the domain maintained its tertiary structure despite having an internal cleavage site in its polypeptide backbone (**Fig. 2-10**).

As such, we decided that our domain boundaries were optimal and we could proceed to investigate any putative interaction between the PAS-A domain and kinase domain.

PAS-A does not bind ATP *in vitro*

Previous results had established that KinA PAS could be UV-crosslinked to ATP, thus indicating that ATP was a ligand for this domain (Stephenson and Hoch 2001). Titrating unlabeled ATP into 100μM ^{15}N-labeled KinA PAS-A, we investigated the putative ATP-binding site using NMR spectroscopy. In theory, NMR signals from signals at or near an ATP-binding site should reside in a changed magnetic environment and thus, exhibit a different chemical shift. ATP, buffered at the same pH as that for KinA PAS-A, was first added in equimolar ratios followed by the addition of progressively higher ratios of 1:10, 1:100, and finally 1:1000 (**Fig. 2-11**). No chemical shift changes were observed in ^{15}N-^{1}H HSQC at any ratios suggesting that binding does not occur, at least in the range of 100 millimolar affinities. We concluded that any putative ATP-binding by KinA PAS-A could not be observed by NMR. Thus, we explored other means by which KinA PAS-A regulates KinA activity.

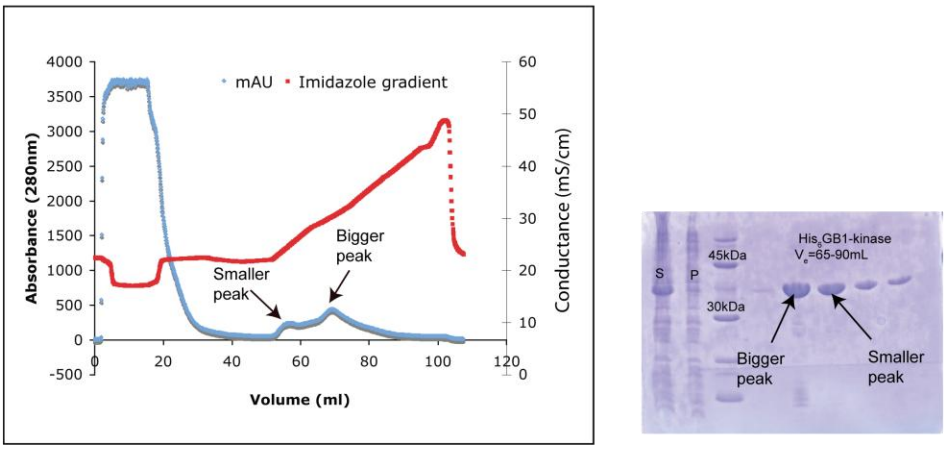
PAS-A does not interact with kinase domain *in vitro*

The aforementioned results suggest that the kinase domain does not require auxillary elements for its folding. The question still remains then why PAS-A is

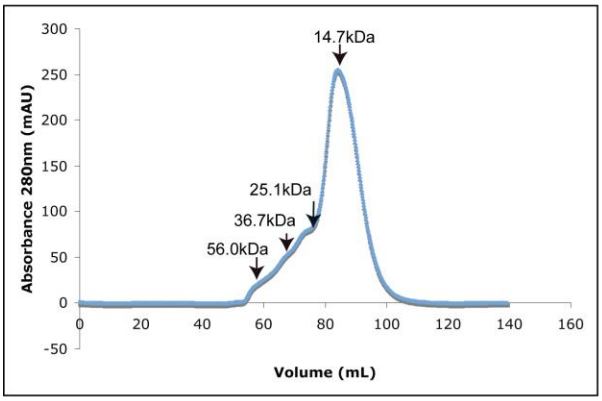
required for its activity? To investigate possible interactions between the two domains, we titrated purified kinase domain and monitored changes in ^{15}N-^{1}H HSQC spectra. We performed this experiment first by titrating unlabeled kinase domain into 100 μM ^{15}N-labeled KinA PAS-A at 1:1, 1:4, and 1:9 ratios and observed no spectral changes (**Fig. 2-12a**). The reverse experiment where unlabeled KinA PAS-A domain was titrated into ^{15}N-labeled kinase domain in ratios of 1:1 and 1:10 was also performed also to result in no chemical shift changes. (**Fig. 2-12b**). Thus, it was concluded that there was no micromolar affinity between the two domains, at least *in vitro*.

These results prompted the reevaluation of our preliminary hypothesis that the PAS-A domain directly interacts with the kinase domain to modulate its activity and that ATP- binding serves as an effector for this interaction. As such, we directed our future experiments toward investigating how more generally a PAS domain could regulate a kinase domain without any direct interaction. We turned our attention now to the large degree of differential broadening seen in the KinA PAS-A spectra that tempts speculation of conformational exchange that could provide the basis for a mechanism for KinA regulation.

a



b



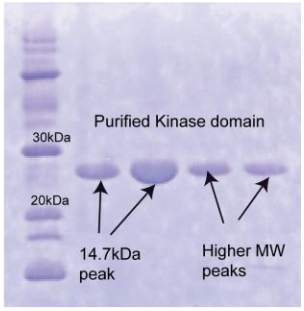


Figure 2-1. Purification of kinase domain from KinA. See text for details. (a) Elution of His_6Gβ1-tagged kinase domain (#383-606) Ni^{2+}-NTA resin with imidazole gradient as measured by salt conductance (mS/cm). SDS-PAGE shows comparison of the soluble (S) and insoluble (P) fractions after french press extrusion. (b) The SEC75 profile of the kinase domain after buffer exchange to low imidazole, TEV-cleavage, and removal of His_6Gβ1-tag by a second Ni^{2+}-NTA resin column. Calculated molecular weights of each peak are indicated with arrows. SDS-PAGE gel shows the elutions for the 14.7kDa peak in one fraction and the higher molecular weight peaks in another fraction.

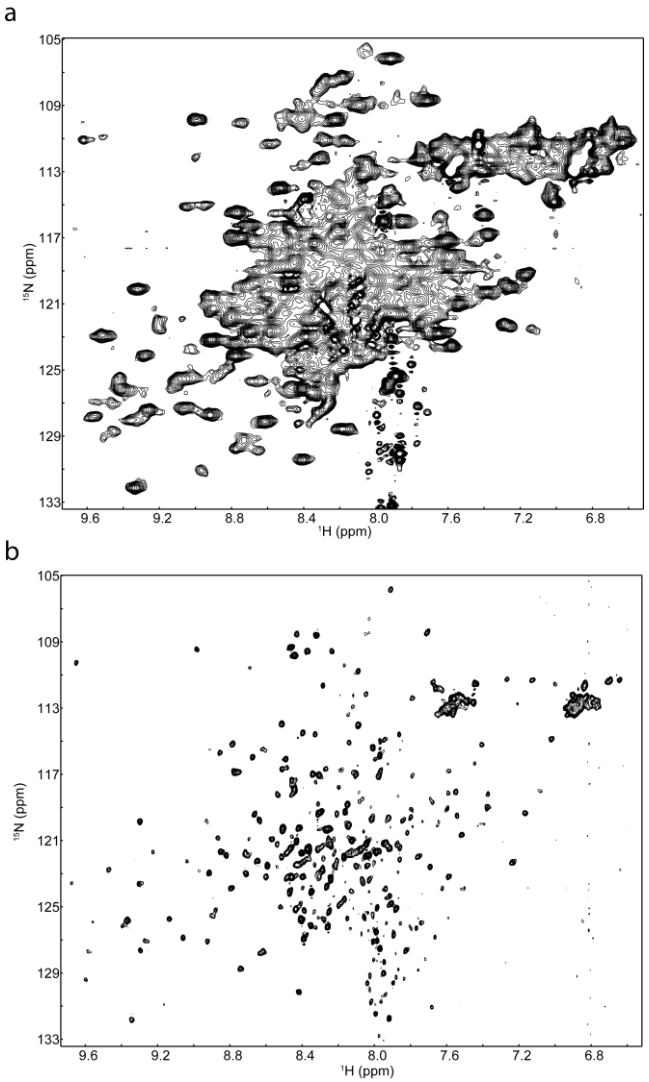


Figure 2-2. ^{15}N-^{1}HSQC spectra of the kinase domain from KinA (a) 250μM ^{15}N-labeled kinase domain (25mM Tris (pH 7.5), 25mM NaCl) sampled on 500 MHz spectrometer at 25°C. Number of transients = 200 (~8 hours). (b) 350μM ^{2}H ^{15}N ^{13}C-labeled kinase domain (25mM Tris (pH 7.5), 100mM NaCl) sampled on 800 MHz spectrometer at 25°C. Number of transients =64 (~4 hours).

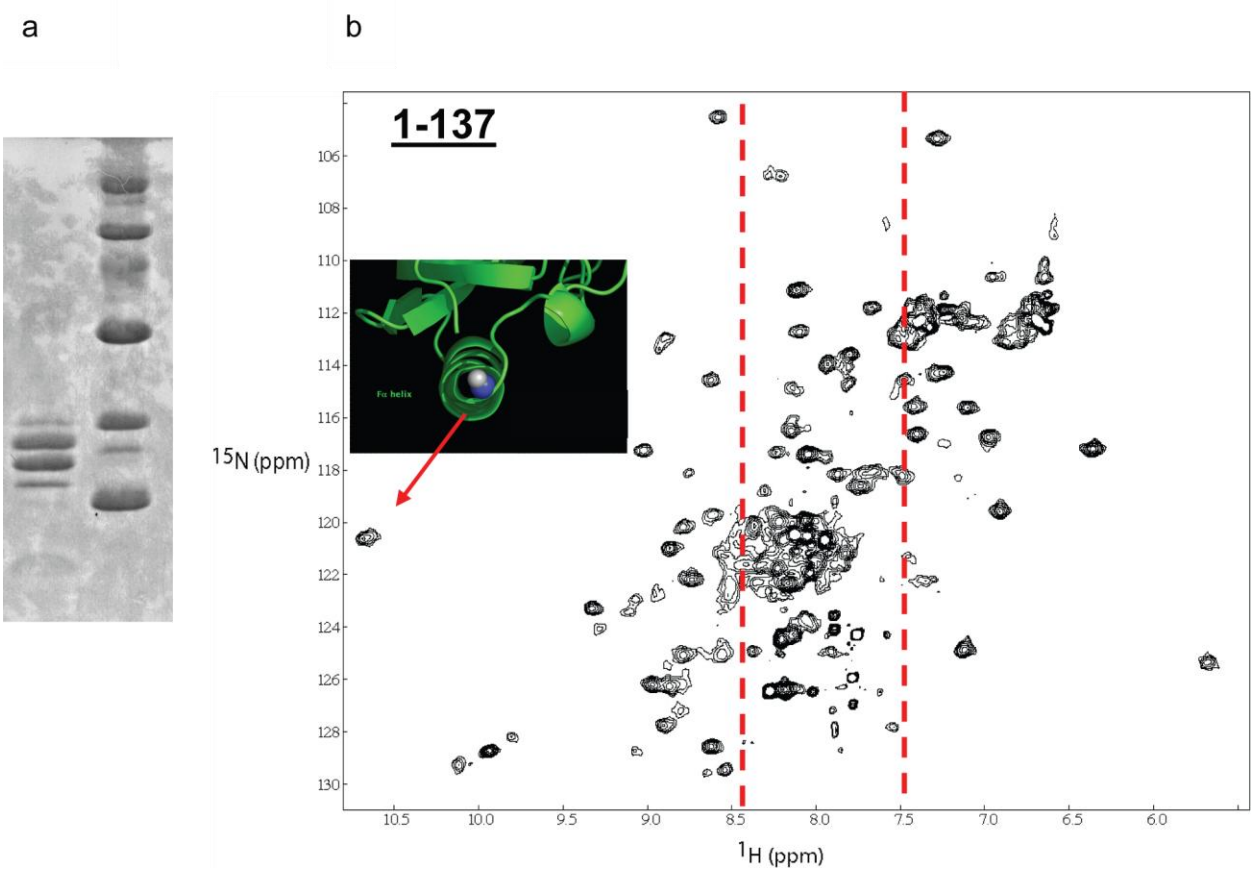


Figure 2-3. Initial characterization of the KinA PAS-A construct 1-137. Sample from lyophilized sample resuspended in buffer of 20 mM Tris (7.0) and 200 mM NaCl. (a) SDS-PAGE (b)^{15}N-^{1}H HSQC with arrow on downfield peak. (Inset) Similar peak from ARNT PAS-B is due to FG loop residue residing in large magnetic dipole of Fα helix (see text). Dashed lines demarcate region of spectra often affiliated with unfolded residues. Spectra recorded with 224 scans per transient for a total acquisition time of approximately 13 hours at 37ºC on a cold probe equipped on a Varian Inova 500 Mhz spectrometer.

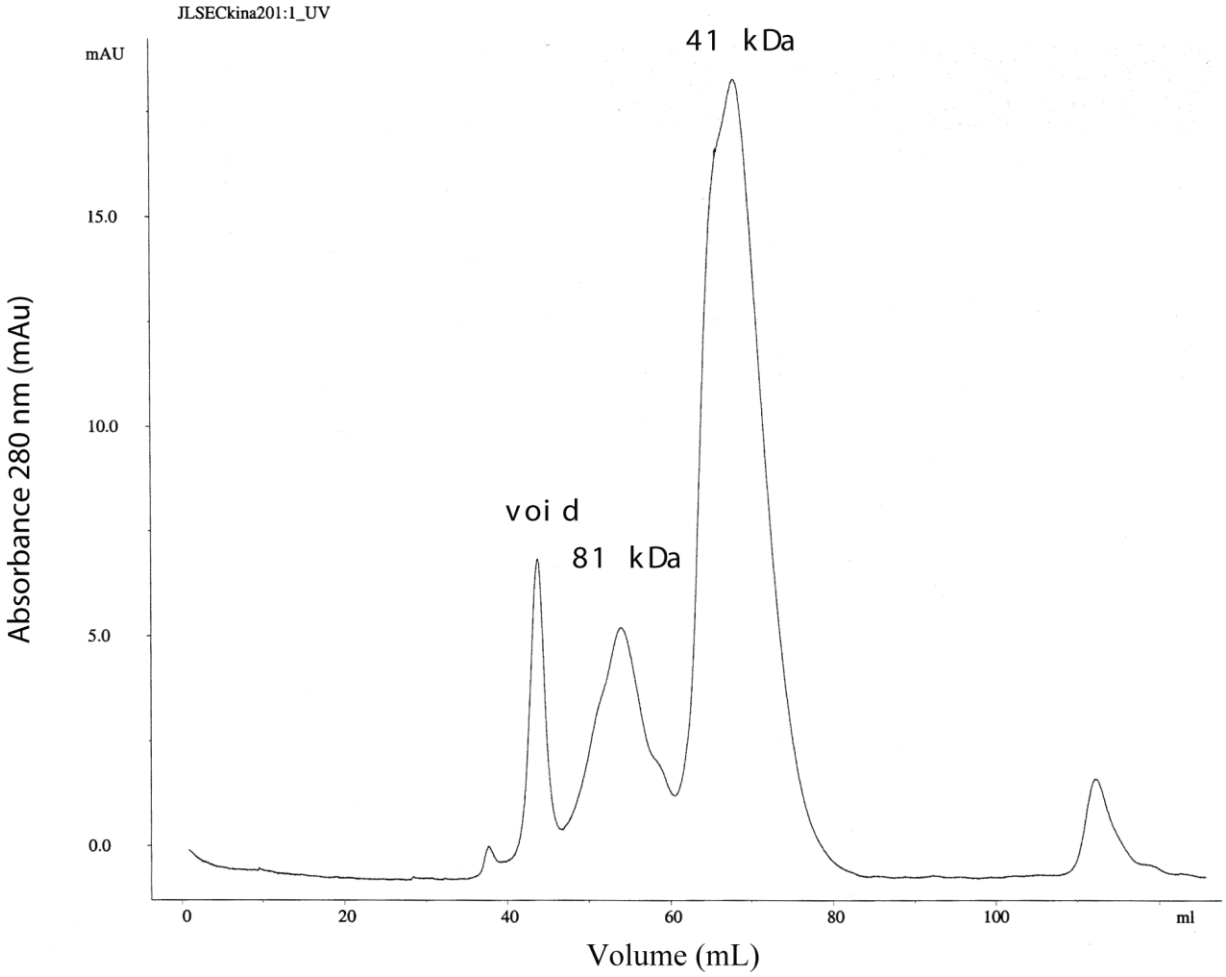
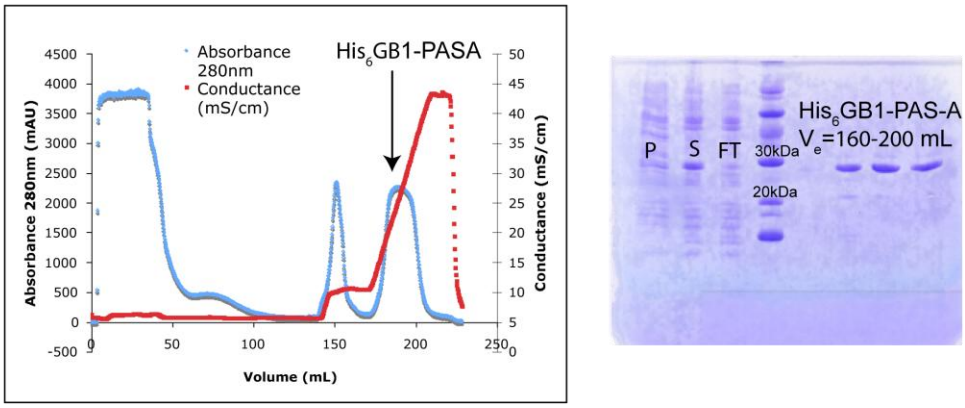


Figure 2-4. Gel-filtration profile of the KinA PAS-A construct 1-137 had three major peaks that would indicating the oligomerization state as being dimer, tetramer, and higher order aggregate.

a



b

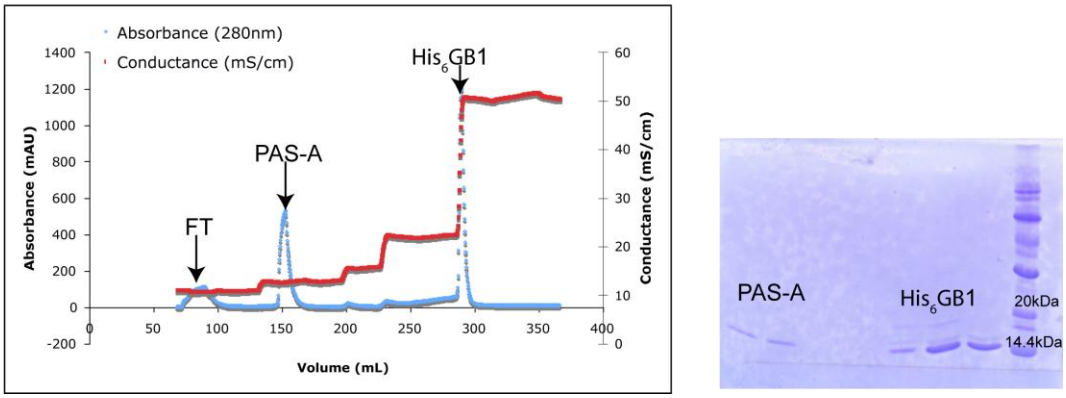


Figure 2-5. Purification of KinA PAS-A domain (aa 10-117). See text for details. (a) Elution of His_6Gβ1-tagged PAS-A domain (aa 10-117) Ni^{2+}-NTA resin with imidazole concentration as measured by salt conductance (mS/cm). As is common with this purification, the first peak at low imidazole concentrations exhibits flow-through of protein impurities commonly found in whole-cell lysates. The second peak (arrow) indicates where kinase domain elutes. SDS-PAGE below chromatogram shows comparison of insoluble (P) and soluble (S) fraction from whole cell lysate, flow-through (FT) non-binding fraction from Ni^{2+}-NTA affinity, and the major elution of His_6Gβ1-PAS-A(b) Passage of cleaved PAS-A through second Ni^{2+}-NTA resin sees PAS-A bind at low imidazole concentration (~20mM) while the His_6Gβ1 elutes at much higher imidazole concentration (250mM imidazole). SDS-PAGE shows the ability to resolve these major species.

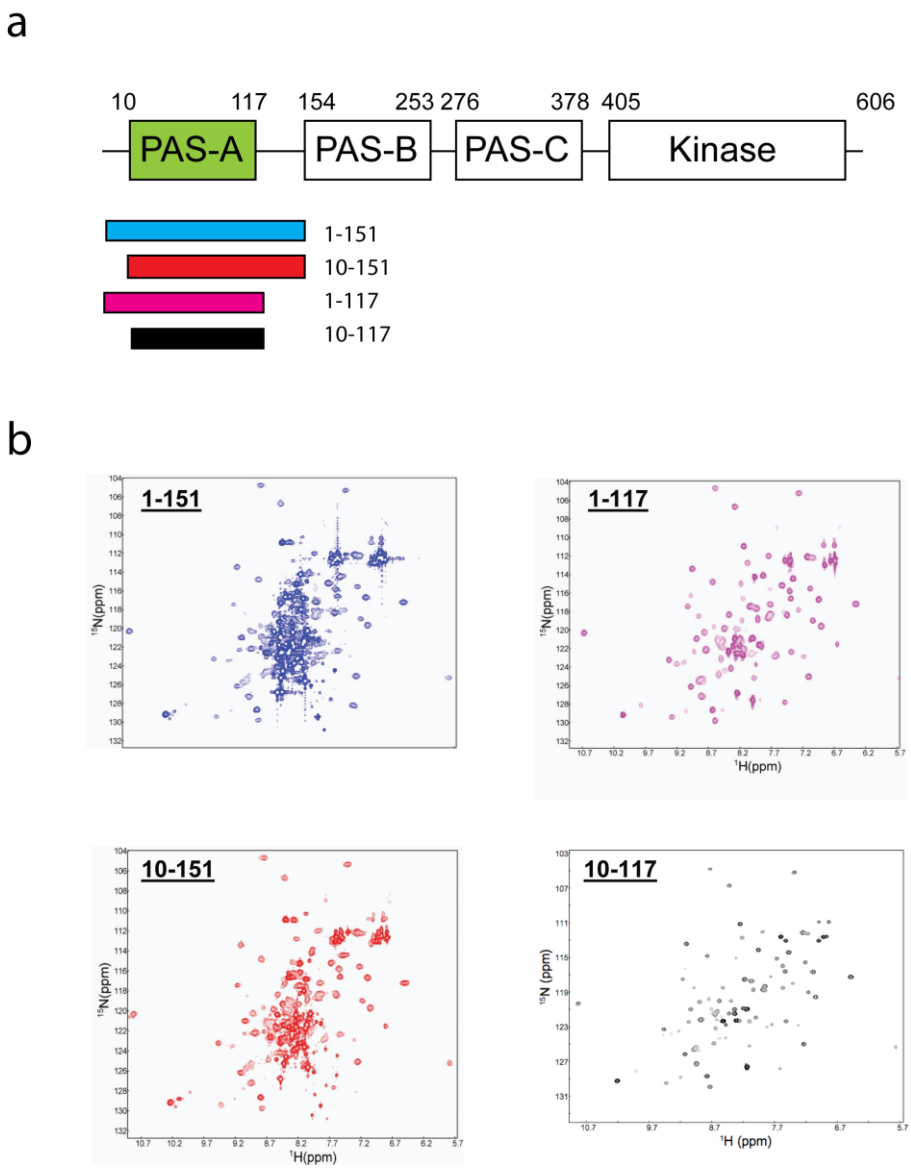


Figure 2-6. Domain mapping of KinA PAS-A. (a) Domain mapping strategy with color legend for corresponding ^{15}N-^{1}H HSQCs in (b). (b) ^{15}N-^{1}H HSQCs of various constructs indicated. All samples are in buffer containing 25 mM Tris (pH 7.5) and 100 mM NaCl and are of 200 μM concentration except construct 10-151 (150 μM). 500 MHz spectrometer recorded all spectra at 25ºC except for construct 10-117 (600 MHz).

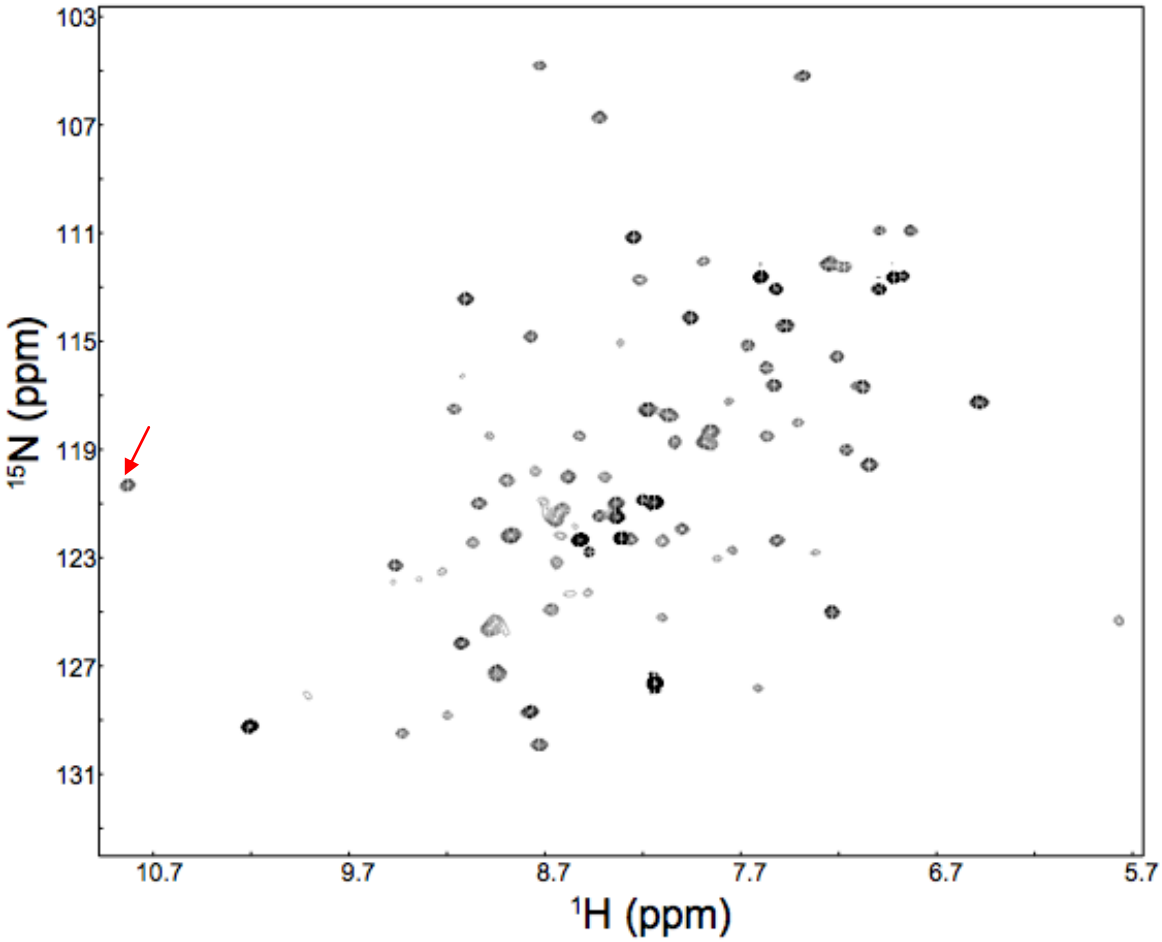


Figure 2-7. ^{15}N-^{1}H HSQC of final KinA PAS-A construct, which encodes residues 10-117. Red arrow indicates peak diagnostic of PAS domain fold. 125 μM sample was in buffer containing 25 mM Tris (pH 7.5) and 100 mM NaCl. Spectra was recorded on 500 MHz spectrometer with nt=32 for approximately 2 hours.

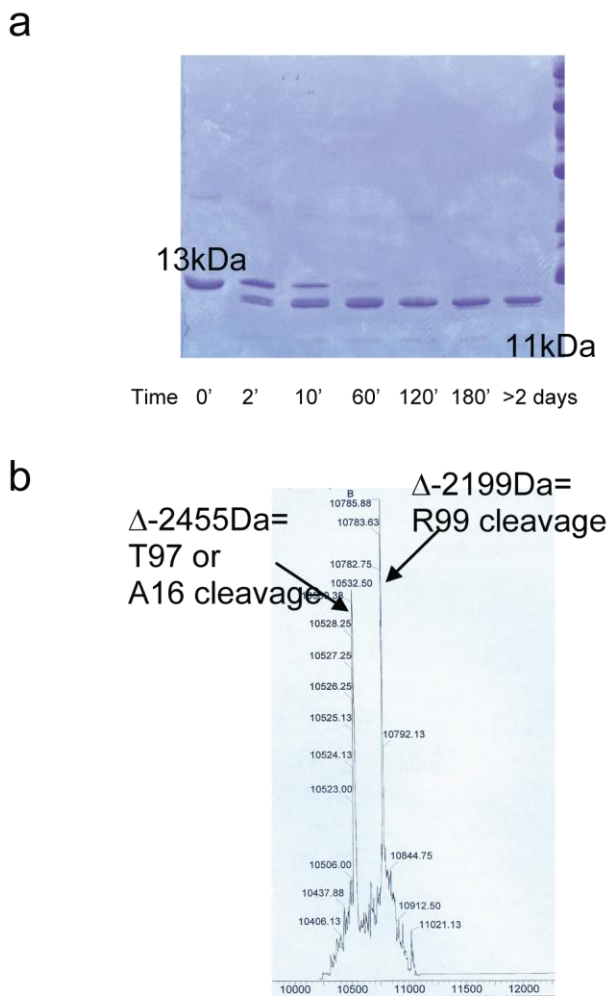


Figure 2-8. Trypsin proteolysis of KinA PAS-A. (a) Monitoring of progression of trypsin proteolysis with time by SDS-PAGE. [Trypsin]=5μg/ml. [PAS-A] =220 μM. Both trypsin and PAS-A are in 25 mM Tris (pH 7.5) and 100 mM NaCl. (b) Electrospray mass spectroscopy profile indicating two major species.

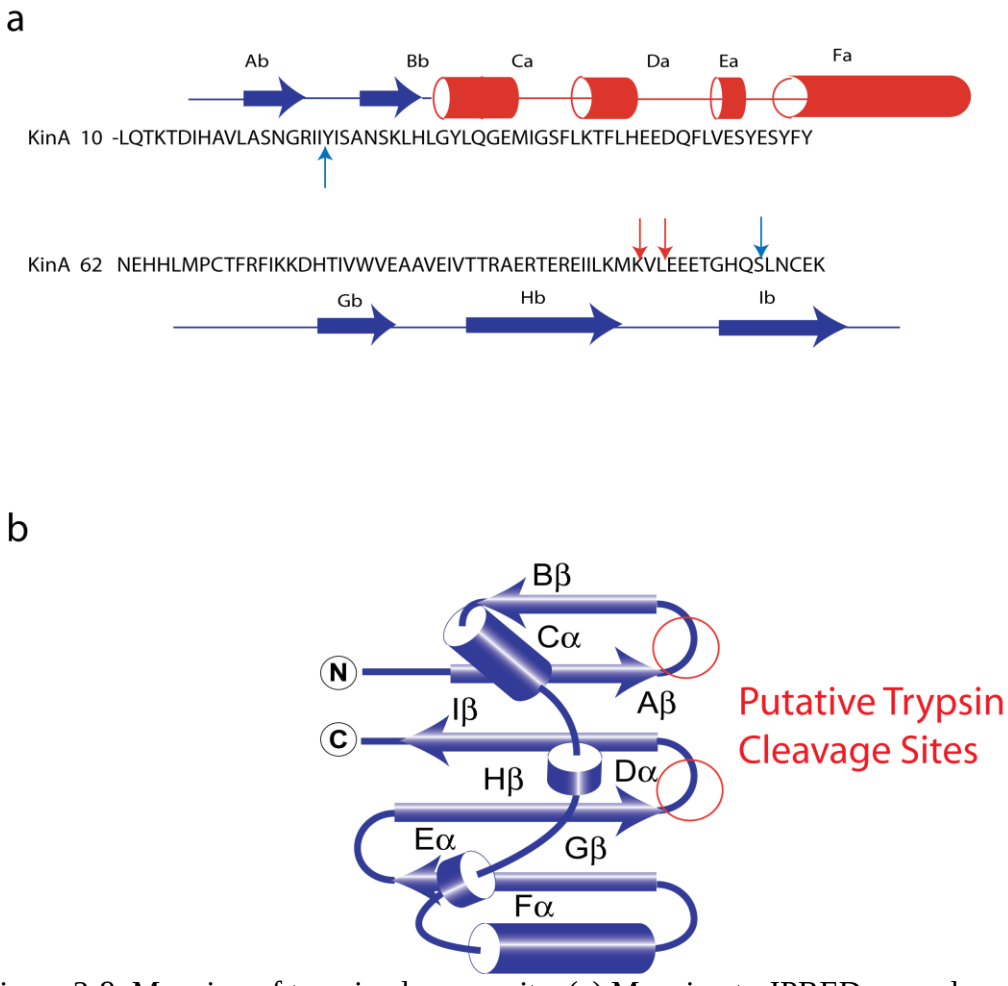


Figure 2-9. Mapping of trypsin cleavage site. (a) Mapping to JPRED secondary structure prediction. Red arrows the C-terminal boundary for a possible cleavage fragment with borders at the N-terminus (residue 10) and at residues 97 or 99 which locates in the HI loop. Blue arrows demarcate another possible cleavage fragment with an N-terminal boundary at residue 22 in the β-turn between the Aβ and Bβ strands and at residue 106 in the Hβ strand. (b) Diagram of PAS-fold and relative trypsin cleavage site.

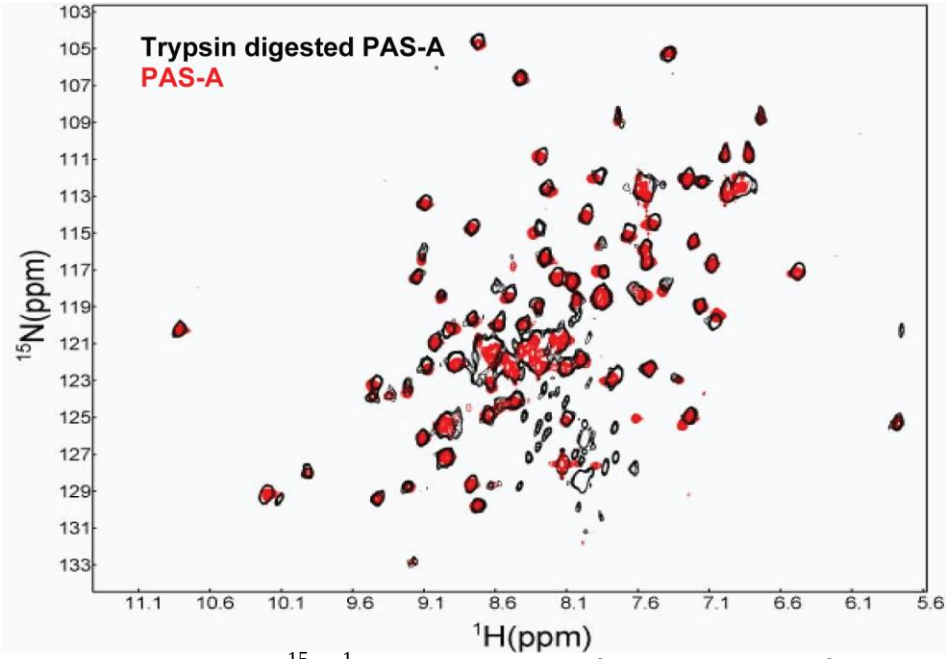


Figure 2-10. Overlaid ^{15}N-^{1}H HSQC spectra of KinA PAS-A before (red, nt=32) and after trypsin proteolysis (black, nt=32). [Trypsin]=5μg/ml. [PAS-A] =220 μM. Both trypsin and PAS-A are in 25 mM Tris (pH 7.5) and 100 mM NaCl. Spectra of trypsinized PAS-A recorded on 500 MHz spectrometer with nt=24 for total acquisition time of 58 minutes.

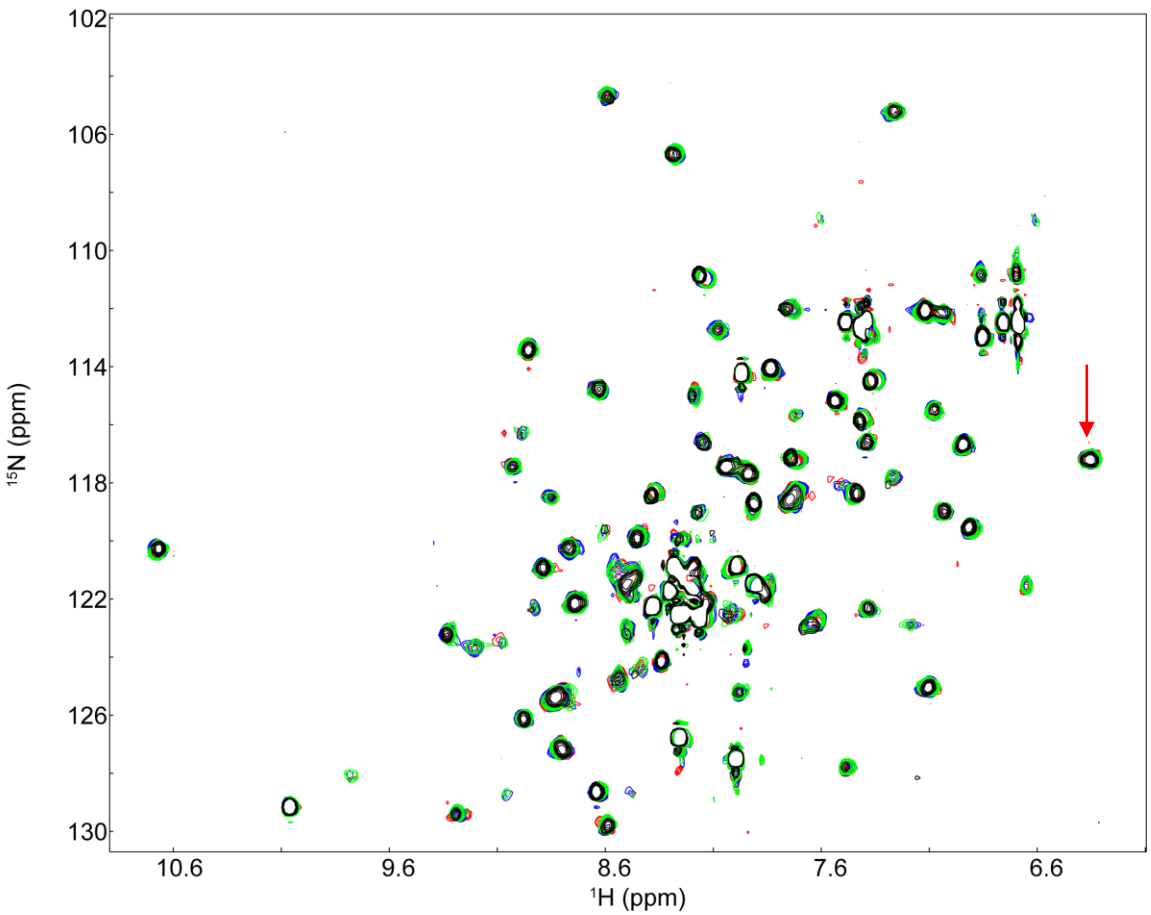


Figure 2-11. Titration analysis of interaction between KinA PAS-A (1-110) and ATP. Both PAS-A and ATP are in buffer containing 25 mM Tris (pH 7.5) and 100 mM NaCl. (a) ATP added incrementally to 100 μM PAS-A (blue spectrum, nt=32) to obtain ratios of 1:10 (green, nt=88), 1:100 (red, nt=16) and 1:1000 (black, nt=56). Red arrow indicates new peak solely due to ATP amine moiety. Spectra recorded on 500 MHz spectrometer at 25ºC.

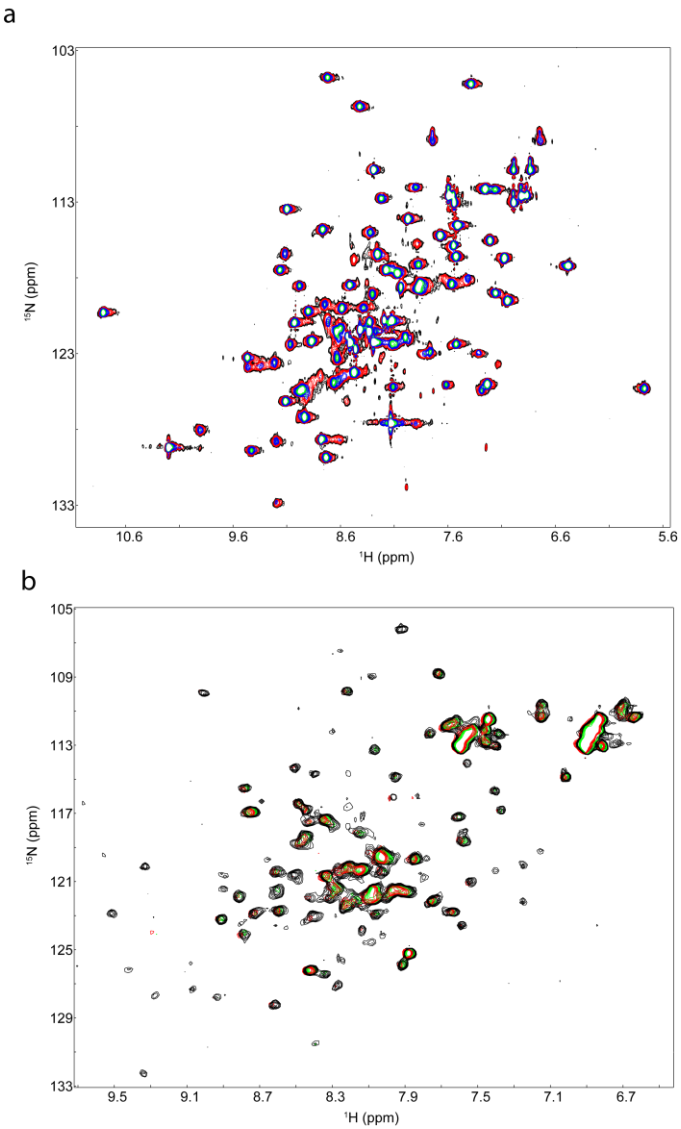


Figure 2-12. Titration analysis of interaction between KinA PAS-A and kinase domain. PAS-A was in buffer containing 25 mM Tris (7.5) and 100 mM NaCl while kinase domain was in buffer containing 25 mM Tris (7.0) and 25 mM NaCl. (a) Kinase domain added incrementally to 125 μM PAS-A (black spectrum) to obtain ratios of 1:1 (red, nt=8), 1:4 (blue, nt=16) and 1:9 (cyan, nt=64). Spectra recorded on 800 MHz spectrometer at 25ºC. (b) PAS-A domain added incrementally to 250μM kinase domain (black, nt =32) to obtain ratios of 1:1 (red, nt=200) and 1:10 (cyan, nt=32 at 37ºC). Spectra recorded on 500 MHz spectrometer at 25ºC unless stated otherwise.

Chapter 3: KinA PAS-A structure

NMR spectra revealed a number of missing or low intensity resonance peaks indicative of differential broadening. Such phenomenon made NMR structural studies difficult. In order to gain insight into the potential structural mechanisms of exchanges processes in KinA PAS-A, we attempted to determine the KinA PAS-A structure by X-ray crystallography (Appendix I contains information on the methodology of the crystallization of KinA PAS-A).

Basic features of the asymmetric unit

We determined the crystal structure of the PAS-A domain of KinA (Fig. 3-1). Crystals of this domain (residues 10-117) contain four monomers in the asymmetric unit, arranged as two dimers of distinctly different quaternary structure, designated as Dimer 1 and Dimer 2 (**Fig. 3-2**). The PAS-A monomer displays the secondary structure and mixed α/β topology commonly observed in other PAS domains, although portions of the Hβ and Iβ strands and the loop connecting them are ordered only in Dimer 2. Inspection of the packing of the PAS-A monomers revealed approximate two-fold non-crystallographic symmetry (NCS) axes between monomers A and B and monomers C and D (Fig.3-2b). The designation of Dimer 1 and Dimer 2 is supported by the buried surface area between monomers A and B (1060 $Å^2$) and monomers C and D (1290 $Å^2$). The

dimer designation is also supported by the large number of specific side chain interactions at the monomer-monomer interface within each dimer. Gel filtration and ^{15}N T_1/T_2 relaxation analyses of rotational correlation times confirms that PAS-A is indeed also dimeric in solution (Chapter 4 and 5). A global view of the ribbon diagrams for these two dimers shows that the interface used for oligomerization is very similar, both employing the same β-sheet interface that is observed in many PAS domains as a specific protein interaction surface (Borgstahl, Williams et al. 1995; Gong, Hao et al. 1998; Chen, Zou et al. 1999; Erbel, Card et al. 2003; Harper, Neil et al. 2003). However, the dimer conformations observed in the crystal are clearly distinct in orientation.

Two different dimer orientations in the crystal

To gain a better perspective on the relative differences in quaternary structure of the two dimers, we aligned monomers A and C along the same axis and compared the relative orientations of monomers B and D in Dimers 1 and 2, respectively (**Fig. 3-2b**). The β-sheet strands in monomers A and B are parallel in Dimer 1, and henceforth this will be termed the parallel conformation. In this dimer, residues Thr91 to Glu98—the HI β-hairpin loop—are solvent-exposed and presumably adopt multiple conformations, as there is no observable electron density in the crystal structure for this region (**Fig. 3-2b**). A β-bulge in strand H formed by residues 88-90 and the resultant disruption of cross-strand hydrogen

bonding may be responsible for the observed flexibility of the HI β-hairpin loop. In Dimer 2 the β-sheet strands in monomers C and D are nearly orthogonal, and henceforth this conformation will be termed the crossed orientation (**Fig 3-2b**). The electron density for the HI β-hairpin loop in Dimer 2 was sufficiently strong and unambiguous to build a model for the residues in this region, which lacks a β-bulge in strand H. A superimposed view of the first subunits of each dimer conformation clearly displays the structure of the HI β-hairpin loop in the crossed conformation (**Fig. 3-2c**). It will also be noted that the crossed conformation has greater electrostatic complementarity (**Fig.3-3**), mostly due to the number of negative charges on the HI loop that are able to interact with the positive residues on its dimerization partner. But despite the major differences in the orientation of respective dimer subunits for the parallel and crossed conformations, both appear to use many similar sets of interactions.

Dimerization rests on a stable base of residues pivoted by Tyr23

We analyzed all the residues at the dimerization interface that could be classified as providing either hydrophobic or polar interactions such that they were less than 4.0 Å and 3.6 Å, respectively, from putative interaction partners. We then classified these residues as either being involved in the dimer interface of both conformations or just one of the two, and listed them in the table below.

Table 3-1. Dimerization residues observed in the crystal structure.

Both conformations		Crossed conformation		Parallel conformation	
His12	Aβ	Thr9	N-terminus	Arg20	Bβ
Val14	Aβ	Ser25	BC loop	Gln36	Cα
Ile22	Bβ	Asn27	Cα	Gly37	Dα
Tyr23	Bβ	Val90	Hβ	Ile40	Dα
Ala26	Cα	Thr92	Hβ	Ile89	HI loop
Glu88	Hβ	Lys104	Iβ	Thr91	HI loop
Glu100	HI loop				
Ile102	Iβ				

Despite the fact that significantly higher buried surface area is calculated for the crossed conformation, each of the individual conformations appear to have equal numbers of residues involved at the interface. Furthermore, it appears that more than half of the total number of interactions are common to the dimer interface for both conformations. This suggests that this set of residues form a stable platform for dimerization. In order to investigate this idea, we examined the superimposition of monomers A and C (**Fig. 3-4).** Residues Val14, Ile22, Ile23, Ala26, Ile102 and Leu103 essentially superimpose, reflecting a lack of movement despite two drastically different dimer conformations. We thus asked how it might

be possible to affect such a conformational change without changing the relative placement of the core of the interface.

In order to investigate this further, we decided to more precisely define principal residues at this interaction surface by first inspecting the center of the interface. Here we observed what seemed like an important residue, Tyr23. This residue exhibits little change within a monomer, as shown in the superimposition of monomers A and C. The Tyr23 of monomer B exhibits a major change in its relative orientation when compared to Tyr23 of monomer D. However, the packing across a dimer interface is radically different, as shown Tyr23 of monomer B undergoing a major change in its relative orientation when compared to Tyr23 of monomer D. In the parallel orientation, Tyr23 is against its respective counterpart from the other subunit such that their hydroxyl moieties are oriented in the same direction (**Fig 3-5a**). In the crossed conformation, these tyrosines pack against each and face in crossed directions. The side chain positions of Tyr23 in the core of each interface reflects the position of the non-crystallographic two-fold axis within each dimer. Viewing both dimer interfaces from their respective centers, we hypothesized that Tyr23 plays a critical role in the binding surface for both conformations. As such, in order to investigate the importance of this residue to the dimerization interface, we generated the point mutant Y23A. On Superdex 75 gel filtration, this Y23A mutant exhibited a shift of approximately 11.0 mL in

elution volume as compared to wild type, consistent with an apparent monomeric molecular weight of 16.1 kDa (**see Chapter 5**). Furthermore, significant spectroscopic changes were observed with solution NMR spectroscopy (**see Chapter 4**) to suggest that intermediate exchange broadening at the dimerization interface was being relieved. From these two latter experiments and its central position in the core of the dimerization interface, we concluded that Tyr23 served a critical role in pivoting the two monomers within a dimeric assembly.

Inspection of the interface just outward from this core Y23A residue, it appeared that Ile102 exhibits a more intricate network of hydrophobic interactions that are related to the polar interaction formed between His12 and Glu88. Just outward from this core Tyr23 residue, Ile102 provides van der Waals stabilizing surfaces for more intricate networks of hydrophobic interactions that are related to a polar interaction between His12 and Glu88. In the crossed conformation, this interaction is inter-molecular but seemingly destabilized by the poor geometry of the near-orthogonal orientation of His12's histidine ring with respect to Glu88's carboxylate. However, Ile102 provides an intra-molecular interactions with His12 that may stabilize its otherwise untenable inter-molecular interaction with Glu88. In the parallel conformation, this network of interactions is not observed as Ile102 only forms an inter-molecular van der Waals interaction with Ile40. This shift at the interface leads to the repositioning of His12 such that its most stable orientation is formed through an intra-molecular polar interaction with Glu88

from the other monomer. This interaction has good geometry as the aromatic ring of His12 is planar with respect to the carboxylate moiety of Glu88 (**Fig. 3-5b**). The importance of Ile102 to hydrophobic networks in the core of the dimerization interface is underscored by the shift in elution volume of an I102A mutant of 7.0 mL from wild type (**see Chapter 5, Fig. 5-1**). From these observations, we conclude that concerted side chain reorientation in monomers B and D of the crossed conformation is the basic mechanism by which the interface is able to adopt different conformations.

Parallel conformation stabilized by a hydrophobic network flanking interface

To understand how such a dynamic interface might be stabilized toward a single conformation, we first examined residues that supplemented the common dimerization interface of the parallel conformation. Thus, we searched for the changes in buried surface area when the parallel conformation is compared to the crossed conformation. The major difference between the two conformations seen in the KinA PAS-A structure is that the interface of the parallel conformation observes the burial of a surface within helix Dα by residues just C-terminal of the Hβ strand (**Fig. 3-6**). Inspection of this region of the parallel structure reveals a network of hydrophobic interactions consisting of van der Waals contacts between Ile40 and Gly37 from one monomer and Ile89 of the other monomer as

well as between Ile40 of one monomer and Ile102 of the other monomer. Noticeably, these interactions are not seen in the crossed conformation as Gly37, Ile40 and Ile89 become partially solvent exposed while Ile102 becomes more deeply buried in the core to form other interactions already mentioned above. Noticeably, these interactions are completely lost in the crossed conformation as Gly37, Ile40 and Ile89 become partially solvent exposed while Ile102 becomes more deeply buried in the core. Interestingly, mutation of Ile40 to alanine does not lead to significant spectroscopic or SEC changes, but mutation of Ile89 to alanine or glutamate do so (**see Chapter 5, Fig. 5-1**), suggesting that other conformations intermediate of our crystallographic structures may lie in equilibrium to help stabilize the parallel conformation interface. These intermediate conformations may have residues in proximity to Ile40 that compensate for the loss of this interaction. As hydrophobic interaction surfaces can be non-specific, it would not be difficult for other hydrophobic sidechains/groups to substitute an interaction.

We next questioned the type of protein contacts that would be needed to be formed to disrupt this general hydrophobic surface and effect a transition to the crossed conformation. In light of the dramatic structural shifts from parallel to crossed conformation, we hypothesized that important new contacts are created in the conformational change to the crossed conformation.

New electrostatic interactions switch parallel conformation to crossed conformation

To test this hypothesis, we looked to the differences in the interactions of Glu88, Lys104, and Asp36 in the two conformations. Interestingly, as mentioned above, His12 and Glu88 form an intramolecular polar interaction in the parallel orientation that spans across the β-sheet. This same interaction becomes intermolecular in Dimer 1 such that two pairs of polar interactions span across the dimerization interface (**Fig 3-5b**). But His12 exhibits essentially no movement from Dimer 1 to Dimer 2 (**Fig. 3-4**). This tempts speculation that Glu88's side chain reorientation and interaction with His12 is a major driving force in inducing the β-bulge for Dimer 1.

Another charged residue that makes a substantial reorientation to provide electrostatic stability to the crossed conformation is Lys104, which shifts 4.71 Å to move between the two dimeric conformations. In the parallel conformation, this residue is solvent exposed. In the switch to the crossed conformation, this residue is partially buried in the core of the interface (**Fig. 3-7a)**. In the present crystal structure the side chain amino group of Lys104 in the crossed conformation is too far (over 5 Å) from the side chain carboxylates of Asp10 or Glu88 in the present crystal structure to form a specific interaction. However, there is sufficient room for these lysines to adopt different side chain rotamer

conformations that may allow the amino groups to interact with these carboxylates in solution, and thus provide charge neutralization.

(**Fig. 3-7a)**.

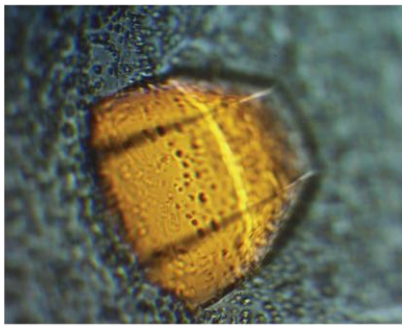
Like Lys104, Gln36 forms a new contact in the crossed conformation. However, its interaction drives a change in the interactions of another residue Arg20. In Dimer 1 the guanadinium group of Arg20 hydrogen bonds to the carboxylate of Glu100 in the symmetry-related monomer. This interaction is impossible in Dimer 2, and instead it adopts a different side chain rotamer and forms an intramolecular hydrogen bond to the main chain carbonyl of Gly41 (**Fig. 3-7b**). This switch in rotamer conformation and interaction partner is driven by the guanadino group of Gln36 forming a hydrogen bond with the carboxylate of the Glu100 from the symmetry related monomer. Thus, three specific electrostatic interactions in the core of the dimerization interface play a major change in dimer orientation.

A key feature distinguishing the crossed conformation is the ordered secondary structure of the HI loop that is outside the core dimerization interface. Thus, we hypothesized that the ordering of the HI loop as it packs against the Cα helix (residues 25-29) of the other monomer must be based on specific contacts. Inspection of this region of the dimer interface revealed three main reasons for the ordering of this loop: (1) the beta-hairpin of the HI loop is formed based on

numerous cross-strand hydrogen bonds; (2) specific hydrogen bonding interactions involving residues Glu36, Glu100, Tyr23 anchor the HI loop to the monomer; (3) positioning of several negatively charged carboxylates possibly provide synergistic, but transient interactions with positively charged residues in the symmetry-related monomer (**Fig. 3-8**). However, possibly reflecting the fragile nature of the packing of HI loop is the finding that NMR resonances for residues in this HI loop also suffer intermediate exchange broadening (**see Chapter 4**), further evidencing of the dynamics of a putative conformational change between the parallel and crossed conformations.

As shown above, dimerization in KinA appears be supported both by a network of static, non-shifting, residues and by other residues that change their hydrogen-bonding interactions from the shift of one conformation to the other. If dynamic shifting among dimeric conformations is indeed physiologically relevant, then it remains to be answered if these structures represent end-points or transition points in a conformational shift. Given that crystallization is typically thought to stabilize the lowest energy conformations, perhaps the former is more relevant here.

a



b

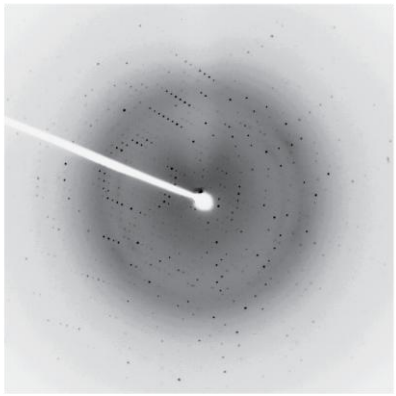


Figure 3-1. Crystallography of KinA PAS-A. (a) Typical crystal produced under conditions outlined in text. (b) Resulting diffraction pattern, showing diffraction out to 1.97 Å.

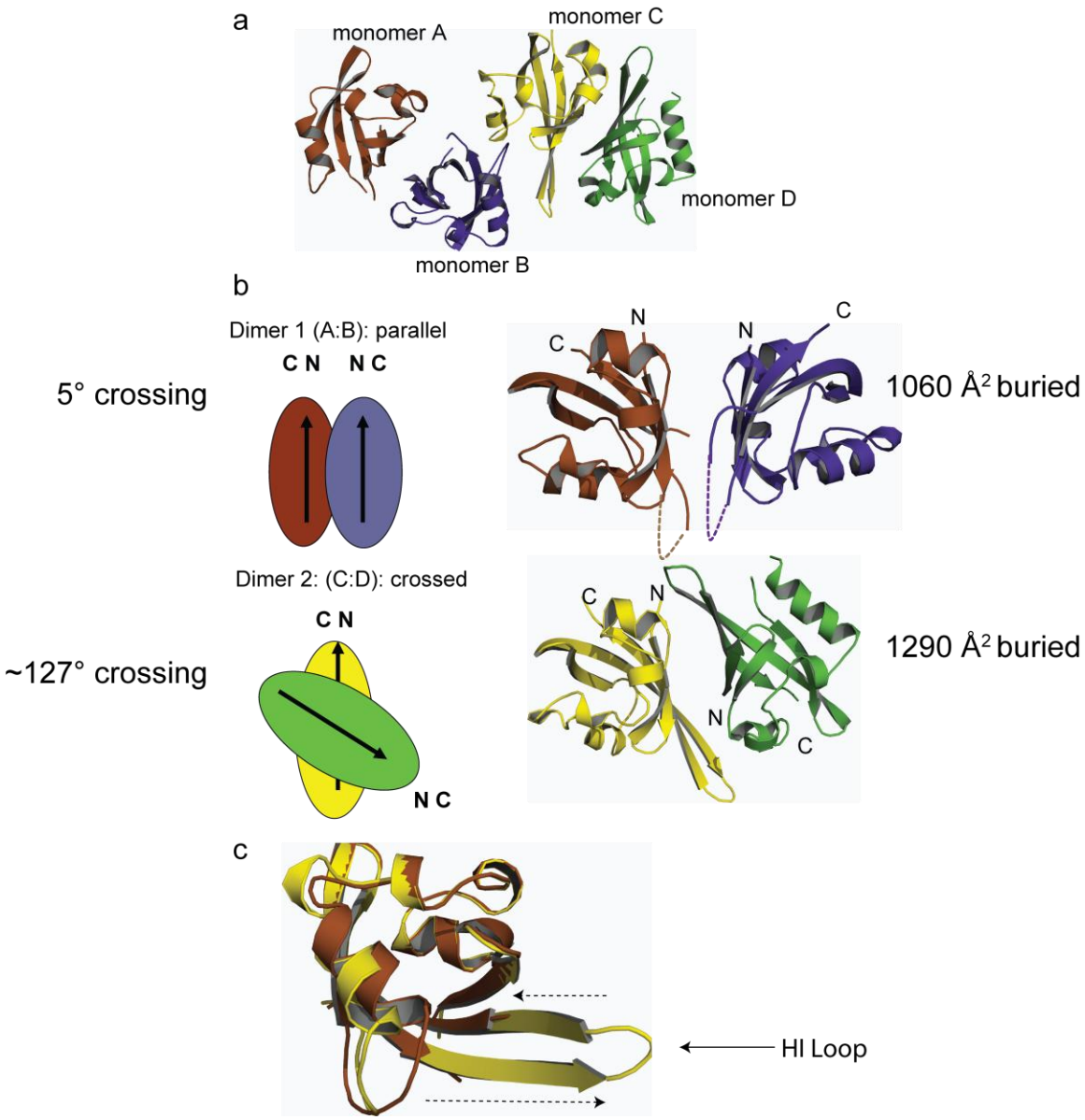


Figure 3-2. KinA PAS-A structure (a) Asymmetric unit of crystal produced as described in text with arbitrary chain designations (b) Each individual dimer as represented through aligning one monomer from each conformation (monomers A and C) along the common axis of the direction of the β-strand. Diagrams of their relative orientation are shown to the left of the structure. (c) Superimposition of chains A and C reveals details of HI loop ordering in Dimer 2.

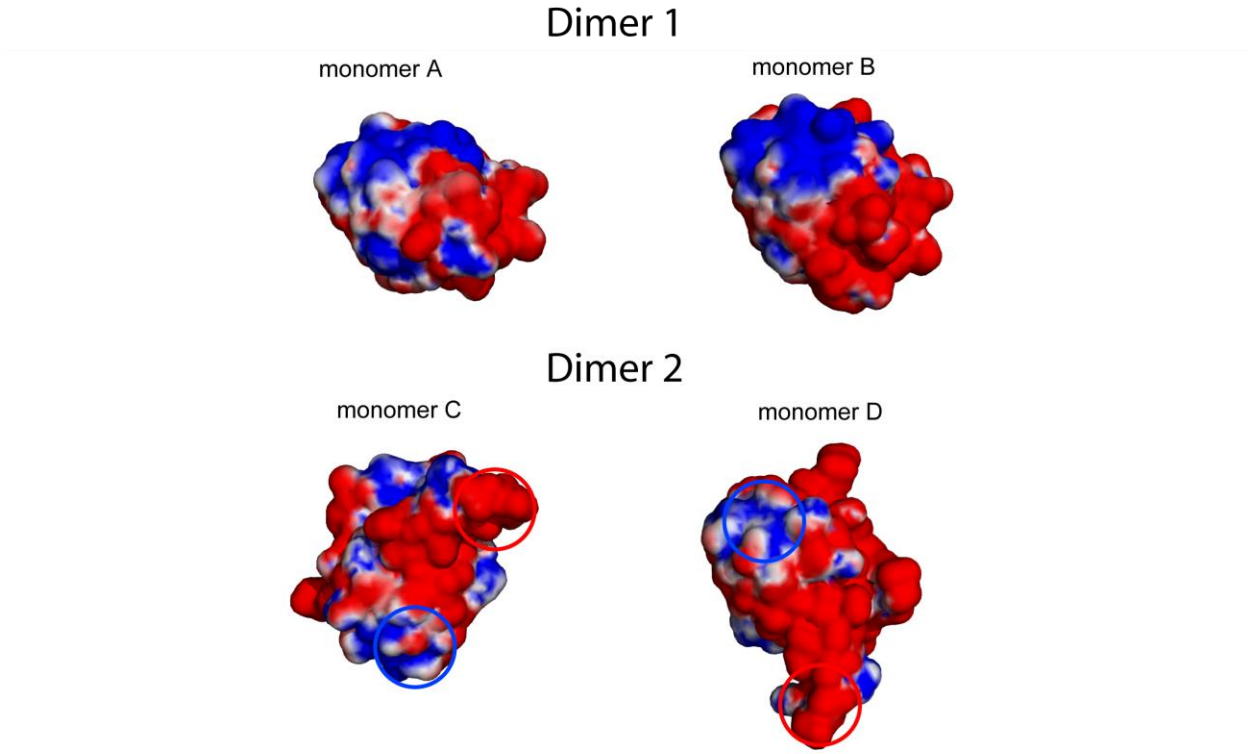


Figure 3-3. Electrostatic potential mapping of KinA PAS-A monomers. Starting from the views of dimers 1 and 2 seen in figure 3-2, the individual monomers were rotated 90° in opposite directions to reveal their interfacial surfaces. Surface electrostatics were calculated using APBS and displayed on each structure using a color gradient ranging from negative (red= -1 kT) to neutral to positive (blue=+1 kT).

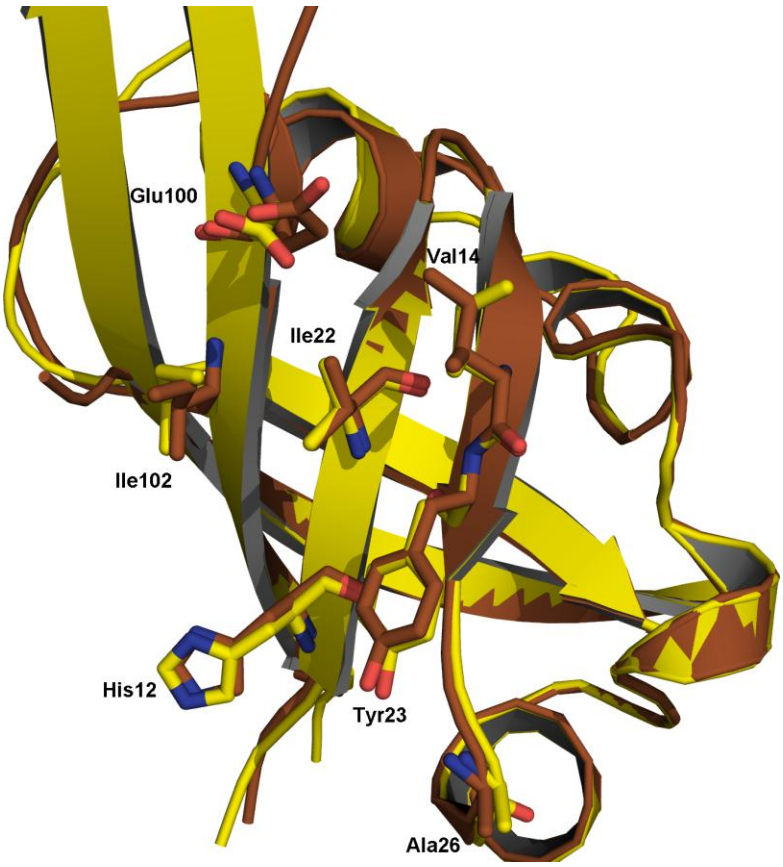


Figure 3-4. Superimposition of monomers A and C reveal similar side chain conformations between Dimer 1 and Dimer 2.

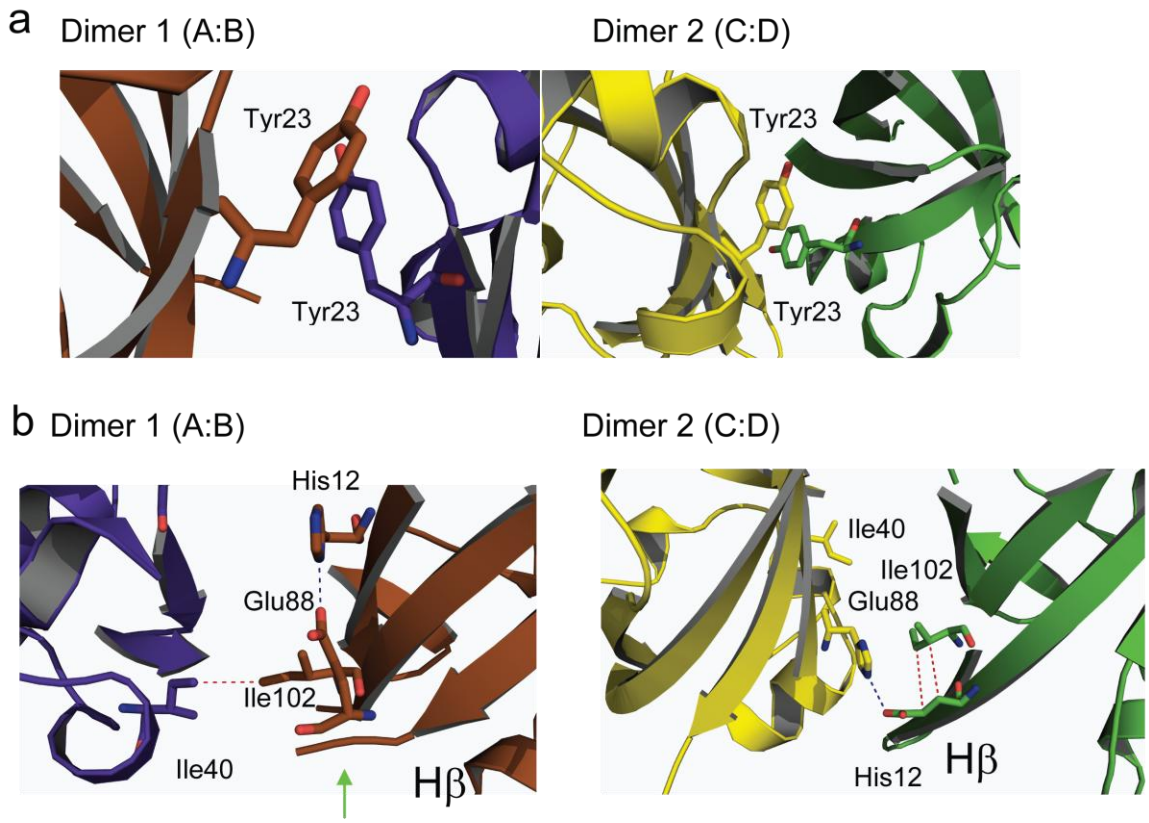


Figure 3-5. Dimerization rests on stable network of residues. (a) Ball and stick representation of the interaction of Tyr23 with its own cognate pair in both parallel and crossed orientations. (b) Ball and stick model of Ile102 and its interaction with the His12/Glu88 salt-bridge in parallel and crossed conformations. Dashed lines indicate putative hydrogen bonds (blue) or van der Waals interactions (red). Green arrow points to the β bulge possibly stabilized by Glu88 polar interaction with His12 in Dimer 1.

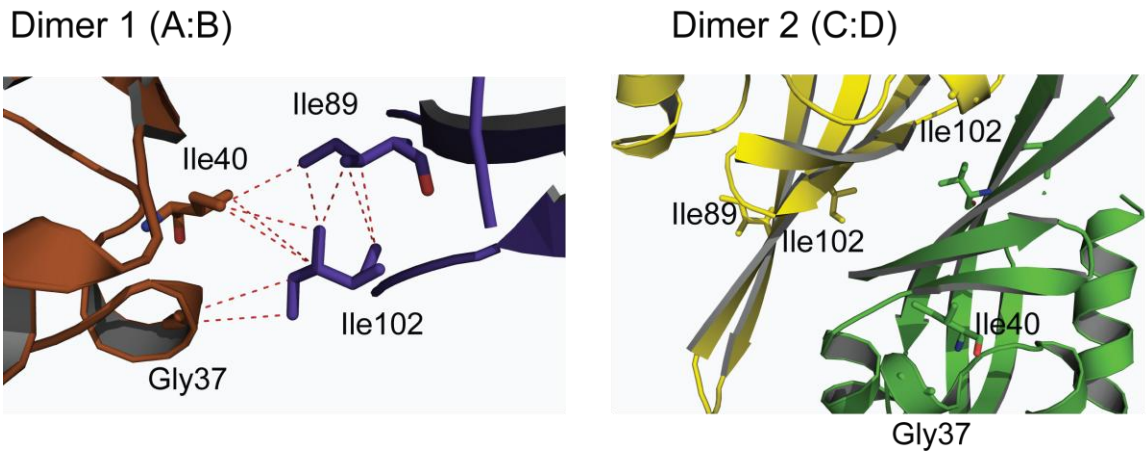


Figure 3-6. Hydrophobic network in parallel conformation broken in switch to crossed conformation. Red dashed lines indicate putative van der Waals interactions.

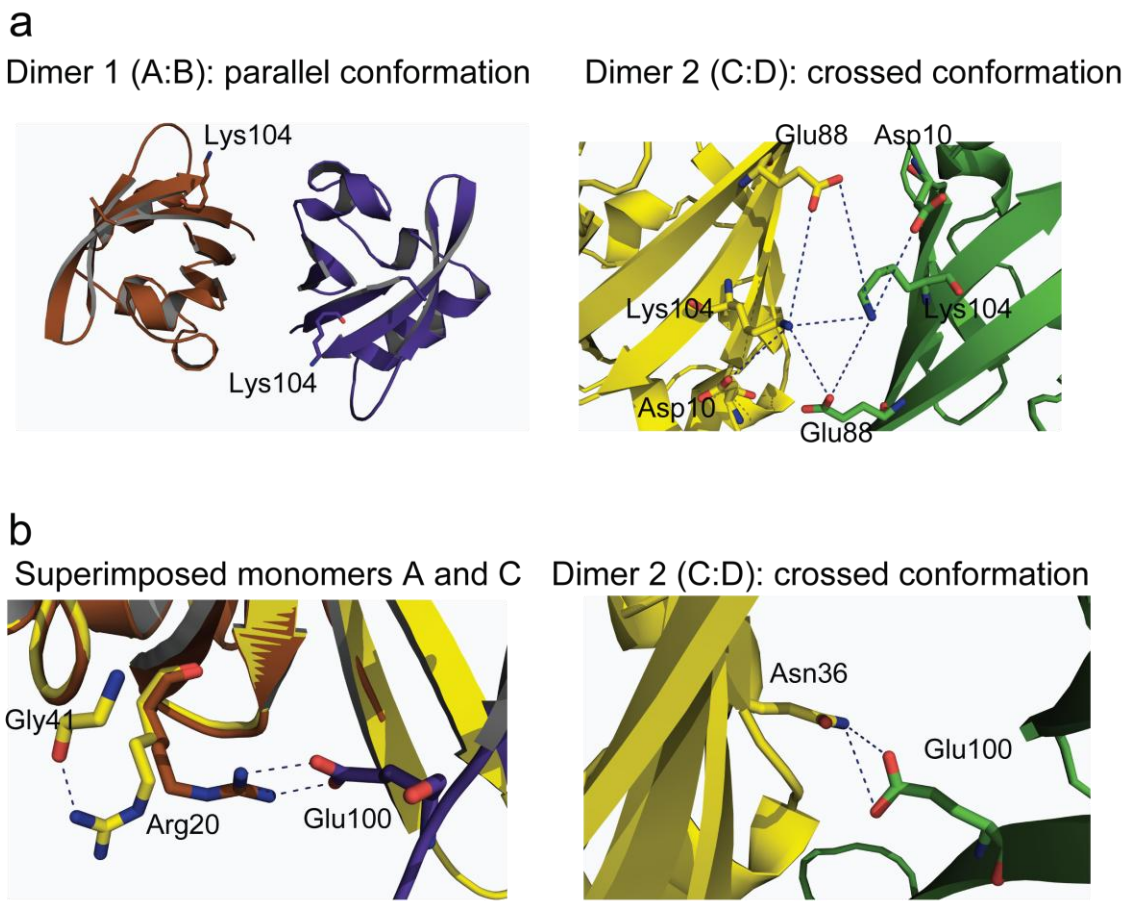


Figure 3-7. Electrostatics drive the switch from parallel to crossed conformation. Putatively important residues forming electrostatic interactions to stabilize the crossed conformation. Parallel and crossed conformations shown side-by-side. Blue dashed lines indicate putative hydrogen bonds. (a) Lys104 goes from being solvent exposed to forming a part of the dimerization interface in the cross conformation. (b) Superimposition of monomers A and C shows that Arg20 adopts different rotamers in Dimers 1 and 2. Glu100 interaction with Arg20 of Dimer 1 is substituted by Gln36 in Dimer2

\

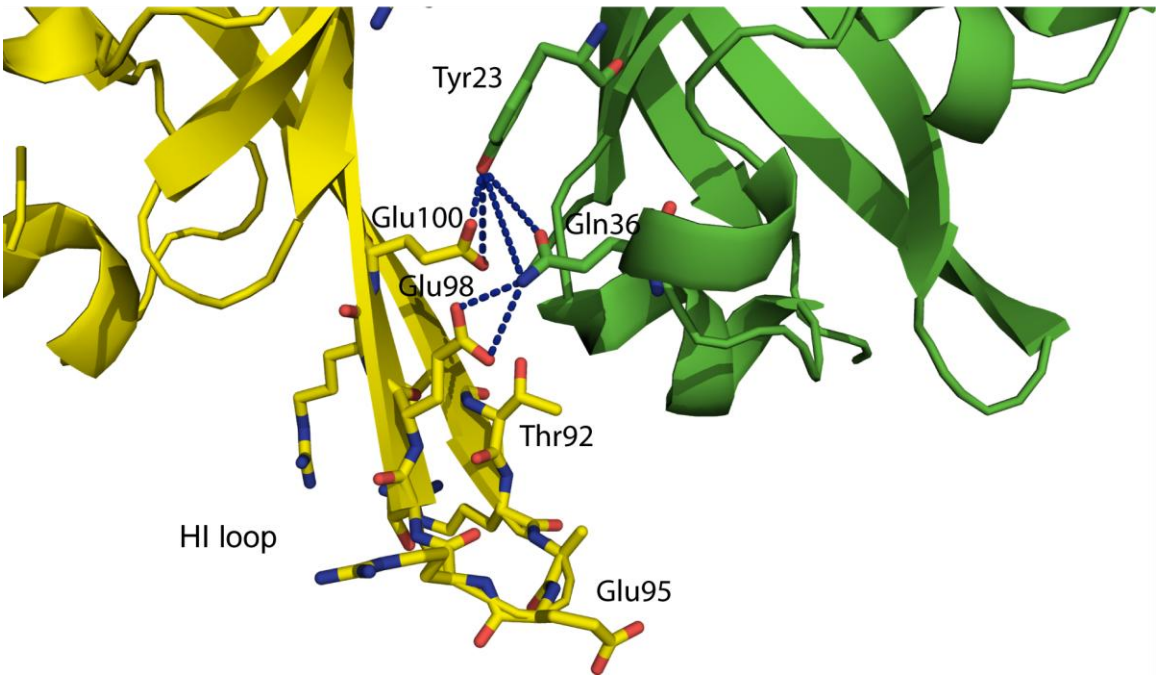


Figure 3-8. Putative interactions of the HI loop stabilizing its packing against the other monomer.

Chapter 4: NMR analysis of differential broadening of KinA PAS-A

By the end of the Chapter 2, the proper domain boundaries for KinA PAS-A were established. However, perhaps partially due both to the size of a dimeric protein and also to conformational exchange, weak signals in the ^{15}N-^{1}H HSQC remained a chief concern. Hence, guided by crystallographic structural information, we employed here in this chapter a number of NMR strategies to better identify the root cause of this broadening. First, we examined ^{15}N-^{1}H HSQC spectra of various mutants designed to disrupt crystallographically-defined interaction surfaces and examined how the broadening changed with these mutations. We subsequently employed triple resonance NMR approaches to assign the backbone chemical shifts of wild type and a Y23A mutant designed to disrupt a pivotal residue in the interaction interface.

MATERIALS AND METHODS

NMR sample preparation

For NMR samples, cells were grown in M9 minimal media containing 1g/L of $^{15}NH_4Cl$ for U-^{15}N samples and supplemented with 3g/L of [$^{13}C_6$]glucose for U-^{15}N/^{13}C labeled samples. The same protocol was applied in producing deuterated samples, except cells were grown in 99.9% D_2O M9 minimal media *(Conditions for ^{15}N-^{1}H HSQC spectra are detailed in the figures where they are presented).*

3D NMR spectroscopy of KinA PAS-A

Proteins for 3D NMR spectroscopy were produced as outlined in the above section. After purification, buffer conditions were 25 mM Tris (pH 7.5) and 100 mM NaCl. All proteins samples were spiked with 5 mM DTT. The final protein concentrations were 800 μM for wild type KinA PAS-A and 400 μM for the Y23A mutant. Triple resonance NMR data were recorded at 25°C with a Varian Inova 600 MHz spectrometer equipped with a cold probe, with nmrPipe (Delaglio, Grzesiek et al. 1995) for data processing and NMRview (Johnson 2004) for analysis. For wild-type KinA PAS-A, $^{15}N/^{13}C/^{2}H$ labeled protein was used to record 3D HNCA, HN(CO)CA, HNCACB, HN(COCA)CB, HNCO and ^{15}N-edited NOESY spectra. For the Y23A mutant, $^{15}N/^{13}C/^{1}H$ labeled protein was used to record 3D HNCACB, CBCA(CO)NH, HNCO, and simultaneous $^{15}N,^{13}C$-edited NOESY spectra. ^{15}N relaxation analyses were based on measurements of ^{15}N T_1 and T_2 recorded at 500 and 600 MHz by modified $^{15}N/^{1}H$ HSQC experiments. ^{15}N T_1 (T_2) values were determined from spectra with relaxation delays of 78,156, 256, 356, 478, 622, and 800 ms (17, 33, 50, 67, 83, 100, 117, 133 ms).

RESULTS

Comparison of ^{15}N-^{1}H HSQCs of wild type KinA PAS-A and mutants

In trying to identify the source of NMR broadening in ^{15}N-^{1}H HSQCs of KinA PAS-A, we looked to the two chief causes of this phenomenon: (1) As a dimer of 25 kDa molecular weight, rotational correlation time for the protein was slow, thus enhancing T_2 relaxation. (2) Interconversion at the dimeric interface—either monomer/dimer or among multiple dimeric forms as observed in the crystal could lead to intermediate exchange broadening at an interaction interface. To address these issues, mutations were designed to disrupt specific crystallographically-defined interfaces. As more thoroughly discussed in Chapter 3, a base network of hydrophobic interactions are provided by His12, Val 14, Ile22, Tyr23, Ala26, and Glu100 for both crystallographically defined conformations. Inspection of these residues for both structures revealed that Tyr23 seemed to be centrally positioned. First, we asked if the disruption of this seemingly key pivot for the dimerization interface could affect spectra? Thus, we generated the Y23A point mutant. Because Y23A caused a shift on gel-filtration elution volume consistent with a monomeric molecular weight (~16.1 kDa compared to wild type 31.1 kDa, **see Chapter 5, Fig. 5-1**), we hypothesized that this mutant would attenuate both of the chief causes of peak broadening mentioned above.

To see if this was indeed the case, we analyzed the peak intensities and line-widths for the Y23A ^{15}N-^{1}H HSQC spectra. In striking contrast to spectra of the wild type protein (**Fig. 4-1a**), the resonance peaks of the ^{15}N-^{1}H HSQC spectra of the Y23A mutant appear relatively narrow and uniformly intense, without any unusual differential broadening (**Fig. 4-1b**). These striking changes led us to conclude that we had indeed disrupted the dimerization interface.

As such, we believed that many residues would reside in new magnetic environments not influenced by a nearby dimerization partner. To investigate such a possibility, we compared ^{15}N-^{1}H HSQC spectra from both wild type and Y23A mutants collected in similar conditions. In comparison to wild type ^{15}N-^{1}H HSQC spectra, chemical shift changes were observed in a majority of residues. Moreover, a large number of new resonance peaks also appeared (**Fig. 4-1c**). Thus, we concluded that disrupting a key dimerization interaction—identified to be equally important to both crystallographically-defined conformations—could disrupt differential broadening.

Next we asked if other mutations, chosen to specifically disrupt one of the two crystallographically-defined dimers, cause large changes in NMR peak intensities and location. Thus, mutations were made to Ile89 and Gln36 to disrupt what appear to be interaction surfaces for the parallel and crossed conformation, respectively (**Chapter 3, Fig. 3-3**). An additional mutation was also made to Ile102, a residue providing hydrophobic interactions just outside the pivotal

Tyr23 in the common dimerization interface, albeit seemingly more emphasized in the crossed conformation. In contrast to the remarkable intensification of Y23A resonance peaks, the I89A/E, I102A, and Q36A mutants exhibited similar differential broadening to the wild type spectra (**Fig. 4-2**) suggesting that these mutants maintained dimeric status. Finally, comparing the ^{15}N-^{1}H HSQC spectra of the I89A/E, I102A, Q36A, and Y23A with that of wild type reveals only minor chemical shift changes within relatively similar chemical shift patterns. These data are indicative of no global changes in the overall fold for all the mutants presented.

Global chemical shifts cannot be correlated to dimerization affinity

Next we investigated whether global trends in chemical shift changes could be observed in the overlaid spectra, particularly in the context of dimerization. Based on Superdex 75 gel filtration results, we assumed that mutants would have the following dimerization affinities: Wild type>Q36A>I89E>I89A>I89E>I102A>Y23A (**see Chapter 5, Fig. 5-1**). To assess whether chemical shift changes could be correlated to monomerization, we overlaid all mutant spectra (**Fig. 4-3**). In general we could observe that gradual changes in particular subsets of chemical shift resonances from one mutant to another was occurring such that resonances for most mutants fit in a nearly linear pattern, suggesting that mutations altered a conformational equilibrium whereby

residues resided in a limited number of possible side chain orientations and/or environments. Furthermore, in a select number of peaks, chemical shift changes could be correlated to monomerization, whereby Y23A exhibited the most dramatic chemical shift changes compared to wild type (**Table 4-1**). However, at many peaks, the largest chemical shift changes occurred in I89E (ie. Gly19, Gly37 of **Fig. 4-3**), calculated by gel-filtration and analytical ultracentrifugation to be most dimeric of all the mutants (**Chapter 5**). Hence, the chemical shift changes cannot be strictly correlated to dimerization. It is then tempting to speculate that the various mutations of specific interfaces shift the conformational equilibrium of a specific set of residues in a differential manner according to the dimer interface that is disrupted.

Chemical shift changes partially correlate to conformation exchange

In this regard, we finally considered whether it was possible to at least partially deconvolute the effects of changes to the equilibrium of conformations from those based on dimerization affinity. Thus, we investigated whether mutations made to each conformation could shift chemical shift resonances in a specific direction such as to reflect that new magnetic environment. We looked at the residues Asn36 and Ile89, which appear to have crucial roles in the dimerization interface of only one of the two crystallographically-defined conformations. As detailed in Chapter 3, Asn36 provides a specific polar interaction with Glu100 only in the

crossed conformation while having no role in the parallel conformation. Likewise for the parallel conformation, Ile89 forms a critical interaction in a hydrophobic network also including Gly37 and Ile89. Inspection of I89A and Q36A spectra show that certain resonances of the wild type protein lie at an intermediate position between the corresponding resonances of I89A and Q36A (**Fig. 4-4)**. In this regard, when the chemical shifts of all mutants are compared (**Fig. 4-3**), only Q36A shifts certain resonances in this particular opposite direction. We attribute this effect either to local effects of mutation and/or global changes in dimer orientation. But to be fair, not every residue fits any specific pattern. We suggest that the disruption of one interface might be compensated for by other conformational changes at other interfaces. Furthermore, some resonances fits both the pattern whereby the largest chemical changes are exhibited by the most monomeric mutant and the pattern where WT has an intermediate chemical shift of Q36A and I89A. It is tempting to speculate here that these chemical shifts are reading both the conformational shift between dimeric conformations *and* that of the monomer-dimer equilibrium (**see Fig. 4-5).**

Table 4-1. Residues that have specific chemical patterns indicated. (*Only clearly distinguishable chemical shifts counted and all assignments are based on nearest assigned resonances).*

Monomerization: largest chemical shift changes occur with most monomeric Y23A and I102A.	Conformation shifting between dimeric states: WT has intermediate chemical shift of Q36A and I89A.	Ambiguous pattern: chemical shifts not quite in linear pattern, but WT generally has intermediate chemical shift of Q36A and I89A at least in one dimension.
Ala13	Leu15	Leu44
Gly19	Lys29	Ser74
Phe43	Gly33	Lys75
Gly37	Tyr34	
Met39	Lys45	
Trp82	Thr79	
	Glu88	

Nevertheless, mapping of residues whereby WT generally has an intermediate chemical shift of that of Q36A and I89A (listed in the second and third category in the table above) to monomer A of the parallel conformation shows that these residues generally cluster to Cα and Dα regions and the β-turn between Aβ and

Bβ strands (**Fig. 4-6**), suggesting that certain regions outside the dimerization interface are better readouts of a conformational equilibrium in solution.

These results combined with observations of crystallographically-defined structures tempt speculation that the wild type KinA PAS-A lies in equilibrium between not just monomer and dimer, but also between different dimeric conformations. The endpoints or transition points of this equilibrium may be represented by the dual crystal conformations. Hence, the wild type spectra represents an average of the chemical shifts for the equilibrium of conformations, whereas mutations designed to weaken specific endpoints/transitions shift the average of chemical shifts one direction or another, but in a differential manner for particular subsets of residues. New compensating interactions or strengthening of pre-existing interactions are reflected in a different set of chemical shift changes. These effects along with the inherent chemical shift changes of making mutations, makes it difficult to correlate chemical shift changes in a global manner. Thus, the differential broadening of individual resonances in the spectra may then represent the dynamics of individual residues at that interface as they rearrange from one orientation to another. In order to better understand the nature of this broadening, we decided to conduct NMR relaxation measurements as well as assign specific resonances for both wild type and Y23A to compare the effects of dimerization.

Basic tenets of NMR relaxation

Relaxation in NMR generally refers to processes that restore a non-equilibrium bulk magnetization vector precessing along the transverse x and y axes back to its equilibrium state on the +z axis with no net x,y magnetization. The mechanisms by which this occurs are T_1 and T_2 relaxation. Because T_2 relaxation is the chief concern for our protein, it merits a brief discussion. T_2 relaxation occurs due to phase incoherence of magnetization along the x and y planes caused by the effects of chemical shift anisotropy (CSA) and dipole–dipole (DD) coupling. T_2 relaxation is more significant for larger proteins because the slower a protein tumbles, the more efficient T_2 relaxation mechanisms become. Because the local fields created by chemical CSA and DD coupling also fluctuate with this tumbling, they cause unwanted relaxation.

A number of approaches can help attenuate T_2 relaxation. Probably the simplest and most robust solution is to deuterate the protein of interest (Gardner and Kay 1998). Less expensive means are to increase the tumbling rate of the protein. Various means are available to achieve this latter solution such as to increase temperature or lower the viscosity of the solution or reduce aggregation by changing solvent conditions—all of which were employed to reduce the effects of T_2 relaxation. However, these methods were unsuccessful in improving signal-to-noise ratio in triple resonance spectra for wild type protein. Thus, we were tempted to speculate that our broadening was not exclusively due to a slow

tumbling rate, but instead due to intermediate exchange at some interface. This phenomenon will be further addressed in the next section.

Basic Tenets of NMR: Relaxation due to conformational exchange

Dynamics inherent in proteins often lead to nuclei that interact with other nuclei in a manner whereby their local magnetic fields can influence one another. As such, dynamic motions specifically affect NMR chemical shift resonances such that this they are prime for investigation by NMR. The interactions that drive these dynamics are often grouped according to the rates at which they interconvert, k_{ex}. If the k_{ex} is smaller than the difference of the frequencies at which conformation resonates, $k_{ex}<\Delta\nu$, then this exchange process is termed slow exchange. Similarly, $k_{ex}>\Delta\nu$ and $k_{ex}\sim\Delta\nu$ are termed fast and intermediate exchange.

$k_{ex}<\Delta\nu$	$k_{ex}\sim\Delta\nu$	$k_{ex}>\Delta\nu$
Slow exchange	Intermediate exchange	Fast exchange

Of course, these rates at which nuclei exchange magnetic environments are relevant to the difference between their resonance frequencies because the faster this occurs, the less time (~1/k) the nuclei spends at that resonance and the higher the probability that an average resonance frequency is observed on the NMR timescale. If this rate is slower than the difference in resonance frequencies, then

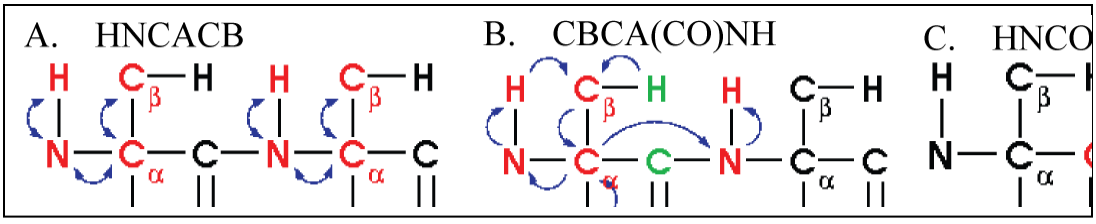
the nuclei of each respective conformation observes its own chemical shift. Intermediate rates lead to a blurring of this distinction such that resonance frequencies are broadened through the range between each magnetic environments' chemical shifts. The consequence of such broadening phenomenon is that T_2 relaxation rates for affected nuclei will be significantly greater than can be accounted for simply by rotational tumbling (see above discussion of this latter effect) (Ernst 1987; Cavanagh 1996).

As a result, two approaches were employed to obtain structural information about this domain: 1) Reacquire spectra on deuterated $^{2}H/^{13}C/^{15}N$ labeled samples and 2) crystallize the protein. The results of our crystallography were discussed in the crystallography section, while the NMR results from deuterated samples of PAS-A are discussed below.

Basic tenets of 3D NMR

$^{1}H/\ ^{13}C/\ ^{15}N$ labeled proteins were initially produced and the following 3D NMR spectra recorded: HNCACB, CBCA(CO)NH, HNCO (Grzesiek 1992; Wittekind 1993; Kay 1994; Muandiram 1994; Muhandiram 1994). These experiments have magnetization transfer pathways shown below.

(Transfer pathways adapted from the web site of Canadian National High Field NMR Centre (NANUC):



www.nanuc.ca/index_e.php).

The HNCACB spectra correlates HN(i)/N(i) with CA(i)/CB(i) and CA(i-1)/CB(i-1), while the CBCA(CO)NH spectra correlates HN(i)/N(i) with CA(i-1)/CB(i-1). As such, in the former spectra, by taking the ^{15}N and H^N cross-peak of one residue in one plane on the spectra and tracing the cross-peaks in the third dimension, one can identify the intra-residue $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ (peaks of opposite sign to those of $^{13}C^{\alpha}$) chemical shift as well as those same ^{13}C chemical shifts of the immediately preceding residue. Taking that residue's intra-residue ^{13}C chemical shifts, which can be distinguished from the inter-residue cross-peaks through the CBCA(CO)NH spectra, and identifying its cross-peaks in the ^{15}N and ^{1}H-N dimension completes one iterative cycle in the process of sequential assignment of the backbone of an entire polypeptide. In this manner, the CBCA(CO)NH spectra only has the inter-residue cross-peaks and thus, provides a means of distinguishing the intra-residue cross-peak in the HNCACB spectra. As a starting point for assignment, easily assignable amino acids are glycines for their lack of Cβ shift and upfield Cα shift near 45 ppm, alanines for their upfield Cβ shift near 15-20 ppm, and serines and threonines for their characteristic Cβ shifts being 65-75 ppm downfield of the Cα chemical shifts.

Basic tenets of deuteration of proteins

Deuteration is a method we employed to attenuate relaxation effects in NMR spectra. Deuteration has been shown to be very effective for improving NMR spectra such that it has been possible to assign the ^{1}H-^{15}N HSQCs for proteins as large as an 81.4 kDa monomer (Tugarinov, Muhandiram et al. 2002) and an 900 kDa multimer (Fiaux, Bertelsen et al. 2002). The replacement of the large population of protons with deuterons can significantly reduce the impact of the DD coupling component on the relaxation pathways for every nuclei in a protein. As already shown for deuterated samples of kinase domain (**Fig. 2-2**), major intensification and sharpening in ^{1}H-^{15}N HSQC spectra is observed such that previously overlapped regions are now well-resolved.

Deuteration of KinA PAS-A only partially effective in relieving broadening

Initially, three dimensional NMR experiments were collected on *non*-deuterated KinA PAS-A. Unfortunately only a very limited number of peaks in the spectra were detectable due to broadening by intermediate exchange. Our initial explanation for such a result was that the dimeric domain tumbled so slowly that signals relaxed too quickly by T_2 relaxation and/or that extensive intermediate exchange at some interface was broadening our signals beyond the point of detection. To solve these broadening problems, we made samples of deuterated KinA PAS-A.

Because of this prior knowledge of fast relaxation pathways exhibited in NMR spectra of non-deuterated KinA PAS-A, we were prompted to collect 3D NMR experiments on *deuterated* KinA PAS-A in a different way. Hence, 3D HNCA, HN(CO)CA, HNCACB and HN(COCA)CB were employed to separately obtain correlated resonances of the Cα and Cβ chemical shifts normally collected together in HNCACB and CBCA(CO)NH, allowing us to use experiments that are optimized for Cα or Cβ chemical shift detection, effectively increasing our signal-to-noise.

Spectra recorded on deuterated PAS-A in general exhibited much sharper linewidths. However, no new peaks and differential broadening still could be observed. It was evident that the assignments were possible only for certain long stretches of the sequence, but other long stretches had NMR signals too weak to assign (**Fig. 4-7a**). A total 158/324=48.8% of resonances could be assigned including those for ^{1}H-N (48.7%), C_a (52.8.1%), and C_b (49.3%) that produced generally correct TALOS predictions of secondary structure (Cornilescu, Delaglio et al. 1999). A likely explanation for these observations was an interface in KinA PAS-A was undergoing some intermediate exchange broadening that we initially speculated was due to either dimerization or the presence of a ligand.

Mapping of the assignments to our crystal structure reveals that overwhelmingly, the residues readily assigned through the experiments outlined above belonged outside the dynamic interface (**Fig. 4-8a)**. To investigate whether

this phenomenon was due to conformational flexibility as might be suggested by our dual dimeric crystal models or simply due to the exchange of a monomer-dimer equilibrium, we recorded ^{15}N-^{1}H HSQC spectra for various NMR samples at various dilutions.

Dilution series of KinA PAS-A NMR samples

^{15}N-^{1}H HSQC spectra were recorded for a number of samples at varying concentrations from 50 μM to 1 mM. In principle, if the intermediate exchange broadening was due to the monomer-dimer equilibrium, then dilution of NMR samples would shift the equilibrium to monomer subunits, thus allowing residues at the putative dimer interface to populate more of just one magnetic environment. However, no chemical shift changes could be observed in any spectra suggesting that the broadening was indeed due to intermediate exchange among different conformations of the dimeric PAS-A and that dissociation of the dimer was insignificant at these concentrations (**Fig. 4-9**). These data suggest a K_d for dimerization no greater than 50μM, consistent with equilibrium sedimentation determinations of dimerization affinities (**Chapter 5**).

3D NMR of Y23A mutant exhibits relief of broadening

In a similar fashion, we decided to see if the increased spectral linewidths seen in ^{15}N-^{1}H HSQC spectra of the Y23A mutant would lead to increased signal intensity in 3D NMR spectra. Resonance assignments for the ^{15}N-^{1}H HSQC

spectra for this mutant were attempted. Initially, 3D NMR experiments were collected for ~1mM protein but the signals from these experiments remained surprisingly low for a protein of this small size and quality ^{15}N-^{1}H HSQC spectra. Suspecting low affinity dimerization or association, we diluted the sample to 400 μM protein and saw a significant enhancement of signals such that we could assign a large majority of the protein (**Fig. 4-7)**. Thus, for Y23A, a total 267/324=82.4% of resonances could be assigned including those for ^{1}H-N (82.3%), Cα (85.1%), and Cβ (79.6%) chemical shifts that also generally produced correct TALOS predictions of secondary structure (Cornilescu, Delaglio et al. 1999). TALOS predicts secondary structure by comparing the chemical shifts of a given residue against a database of chemical shifts for residues with known phi and psi angles. Furthermore, in contrast to wild type PAS-A samples which had the additional benefit of deuteration, resonance assignments for the Y23A could be made even for residues at the dimer interface (**Fig. 4-8b).**

These results reveal the functional importance of Tyr23 to dimerization in the context of a well-concentrated NMR sample. Size exclusion chromatography reported the Y23A to be of monomeric molecular weight for concentrations from 100 μM to 1mM. However, this technique entails a dilution of the sample as it passes through resin. NMR, obviously, entails no such dilution, essentially reporting on the exact condition of the NMR sample at a defined concentration. Thus, our lack of 3D NMR signals at 1 mM concentrations compared to that at

400 μM suggests that the Y23A mutant retains some ability to dimerize or aggregate. It is then tempting to speculate that other residues that are crystallographically-defined to be important to dimerization allow Y23A residual multimerizing affinity.

Nevertheless, the fact remains that, at least at 400 μM concentration, signals are much more intense for Y23A than those for *deuterated* wild type KinA PAS-A at 0.8 mM concentration. This suggests that dimerization and the resulting increase in rotational tumbling is the chief cause of broadening for all peaks seen in wild type spectra, but that differential broadening is caused by the instability of dimer association that results in different dimer conformations exchanging in solution.

Basic tenets of measuring T_1/T_2 relaxation rates to determine rotational correlation times.

To further investigate the oligomerization of KinA PAS-A, we decided to measure ^{15}N T_1 and T_2 relaxation rates. T_1/T_2 relaxation measurements can be easily measured by adapting the ^{15}N-^{1}H HSQC pulse sequence. These experiments entail a series of HSQCs modified so that magnetization on nitrogen remains in various states of S_z for ^{15}N T_1 and S_x/S_y for ^{15}N T_2 for varying periods of time (Farrow, Muhandiram et al. 1994; Farrow, Zhang et al. 1995). The T_1 relaxation NMR experiment simply entails directing a 180° degree pulse on the x-

axis to bring z-magnetization completely to the negative z-axis. The delay time during which ^{15}N magnetization relaxes back to the positive z-axis is specified by a parameter in the pulse program, which will be varied among several spectra that are $^{15}N/^{1}H$ correlations. Of course, to record the remaining z-magnetization, a 90° pulse will be applied to bring magnetization to the x-axis so that it is recordable.

Similarly, T_2 relaxation measurements are made by adapting the HSQC so that a variable number of spin echoes are inserted after ^{15}N chemical shift is evolved and before magnetization transfer back to proton, thus increasing the delay time during which ^{15}N relaxes. The number of spin echoes inserted is specified by a parameter specific to the pulse sequence, which will be set differently for each determination in the series. The series of spectra for T_1 and T_2 relaxation measurements are subsequently processed and T_1 and T_2 extracted by fitting the decline in signal strength over time to a decreasing exponential. With these measures of T_1 and T_2 relaxation, we used the ratio of T_1/T_2 values to obtain rotational correlation times for the wild type and I89E and Y23A variants of KinA PAS-A (Kay, Torchia et al. 1989).

^{15}N T_1 and T_2 values were recorded for wild type and I89A and Y23A mutants and the ratios were found to fit a distribution (**Fig. 4-10).** After fitting using an algorithm implemented in PERL by Qiong Wu of our lab, we determined the following rotational correlation times for these proteins:

Table 4-2. Rotational correlation times for KinA PAS-A and Y23A and I89E mutants determined from ^{15}N T_1/T_2 relaxation measurements

	[1 mM]	[0.3 mM]
WT	13.4 ns	12.4 ns
I89E	13.1 ns	8.9 ns
Y23A	11.6 ns	8.2 ns

Wild type KinA PAS-A has a rotational correlation time of 13.4 ns and 12.4 ns for 1 mM protein and 0.3 mM protein concentrations, respectively (**Table 4-2**). This is consistent with those rotational correlation times of proteins on the order of 20-30 kDa (Wagner 1997). The rotational correlation time of the Y23A mutant was significantly lower at 8.2 ns for 0.3 mM protein, but at 1 mM concentration this number elevated to 11.6 ns, suggesting that the protein had residual dimeric affinity, supporting the idea that Y23A plays a significant, but not exclusive role in dimerization. A starker contrast between concentrations was seen with the I89E mutant that showed rotational correlation times of 13.1 ns and 8.9 ns at 1 mM and

0.3 mM protein, respectively. This could support the hypothesis that the parallel crystal and crossed conformations represent "loose" and "tight" dimer conformations; the I89A/E mutants interfere with the dimerization of the "loose" dimer, hence lowering dimerization affinity for the overall equilibrium of conformations (**Fig. 4-5**). At higher concentrations, this single residue becomes less important to the dimerization as other residues compensate. However, it remains unexplained why equilibrium sedimentation reports a stronger dimerization affinity of 1 μM for I89E, thus tighter than that of wild type at 8 μM (**Chapter 5**). Nevertheless, because of the lower rotational correlation time exhibited by the 0.3 mM I89E, we initially hypothesized that collecting ^{15}N-^{1}H HSQC spectra on this sample would attenuate line broadening in a similar manner as with the putatively monomeric Y23A mutant. However, differential broadening remained about the same at this concentration as with the higher concentration. A possible explanation for this result may be that an interface is dynamically switching between different dimeric states causing intermediate exchange broadening. Of course, the alternative explanation that the interface was suffering intermediate exchange due to a monomer-dimer equilibrium remains viable. Given that we are unable to affect spectra by diluting NMR samples, this latter explanation is less likely to be most prominent source of broadening. However, the fact that we can make mutations that target each of the crystallographically-defined dimers and achieve ^{15}N-^{1}H HSQC spectral changes further supports the

notion of this dynamic switching between different dimeric states. In any case, the affinity of dimerization cannot be so strong as to disallow flexibility at this interface. However, it must be sufficiently strong that diluting samples produces no changes in peak location or linewidth. As such, it can be speculated that the affinity should be in the low micromolar range.

Supporting this notion of a low micromolar affinity for the dimer complex are size-exclusion chromatography (SEC) and analytical ultracentrifugation (AUC) experiments to be detailed in chapter 5. Routinely, these techniques reports complexes with submicromolar affinities. The physiological context of such results will be discussed also in an upcoming chapter.

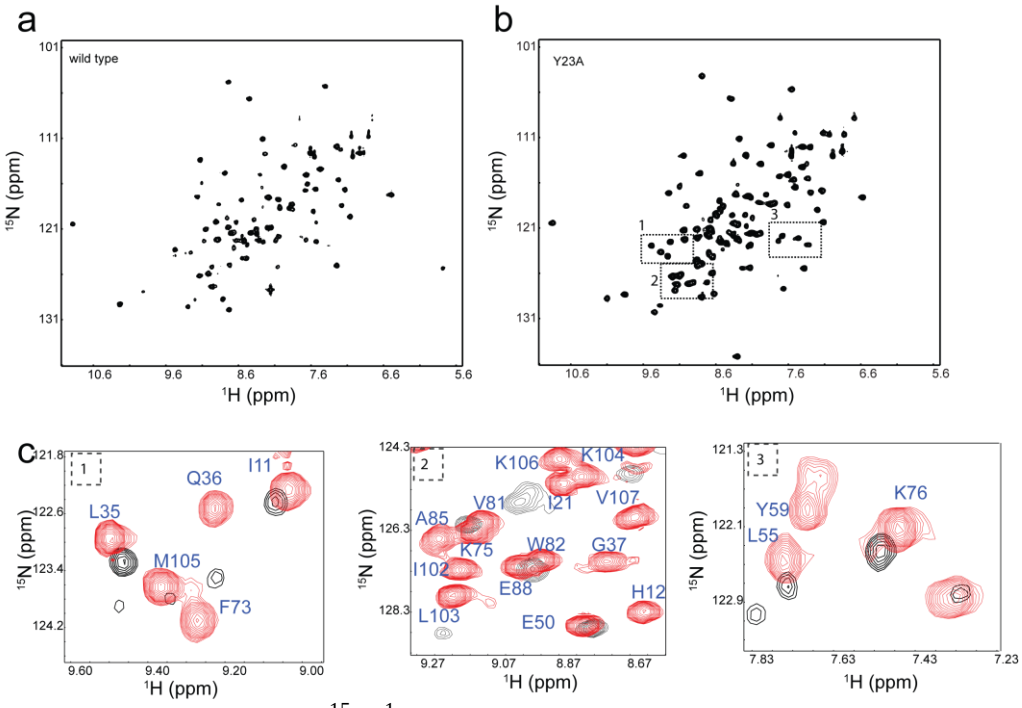


Figure 4-1. Comparison of ^{15}N-^{1}H HSQC spectra of 800 µM wild type versus 770 µM Y23A mutant. Both spectra recorded on 600 MHz spectrometer at 25°C and are plotted with equivalent contour levels. (a)^{15}N-^{1}H HSQC of wild type (nt=24). (b) ^{15}N-^{1}H HSQC of Y23A mutant (nt=32). (c) Zoom views of both ^{15}N-^{1}H HSQC of wild type (black) and Y23A (red) with its Y23A assignments as derived from 3D ^{15}N/13C/^{1}H triple resonance NMR experiments.

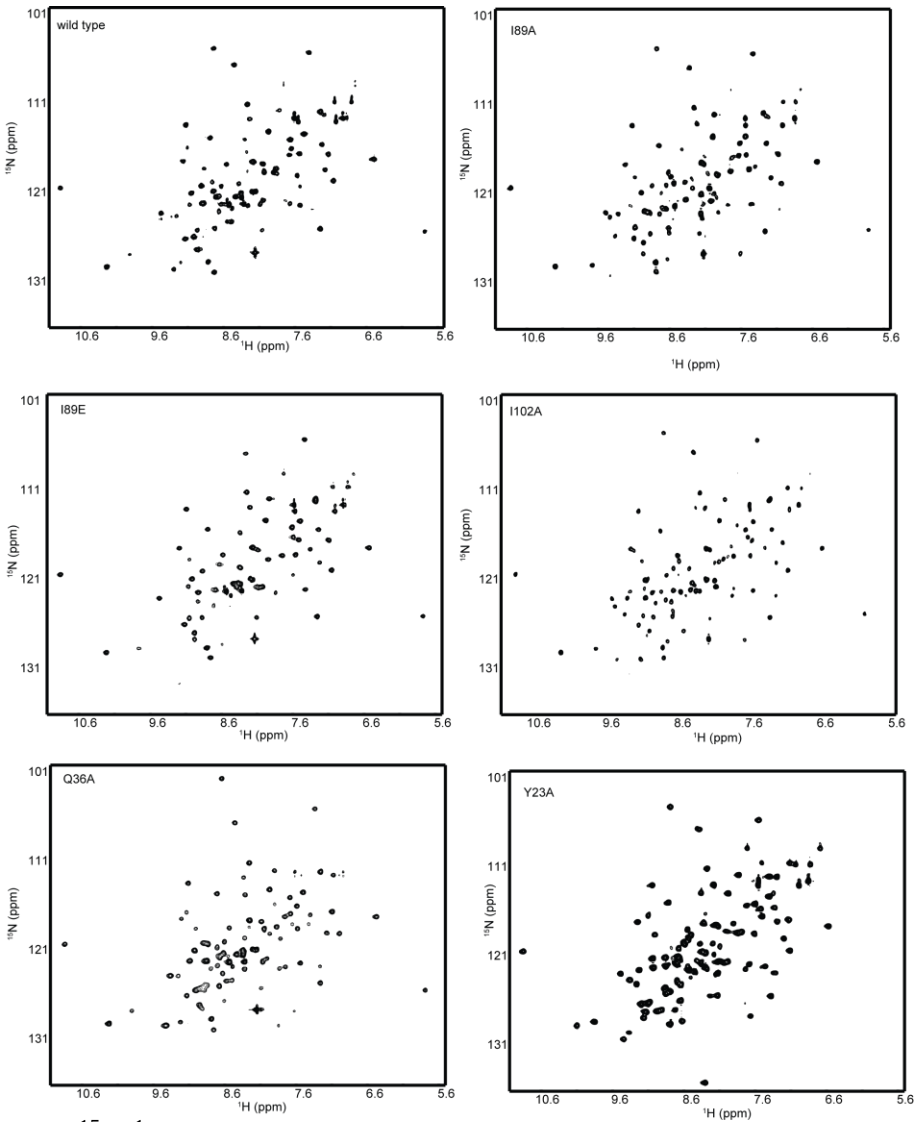


Figure 4-2. ^{15}N-^{1}H HSQC spectra of wild type and I89A, I89E, I102A, Q36A and Y23A mutants. All proteins are in buffer containing 25 mM Tris (pH 7.5) and 100 mM NaCl. All spectra recorded on 500 MHz spectrometer at 25°C. Proteins are all 366 μM in concentration except Y23A (420 μM) and Q36A (240 μM).

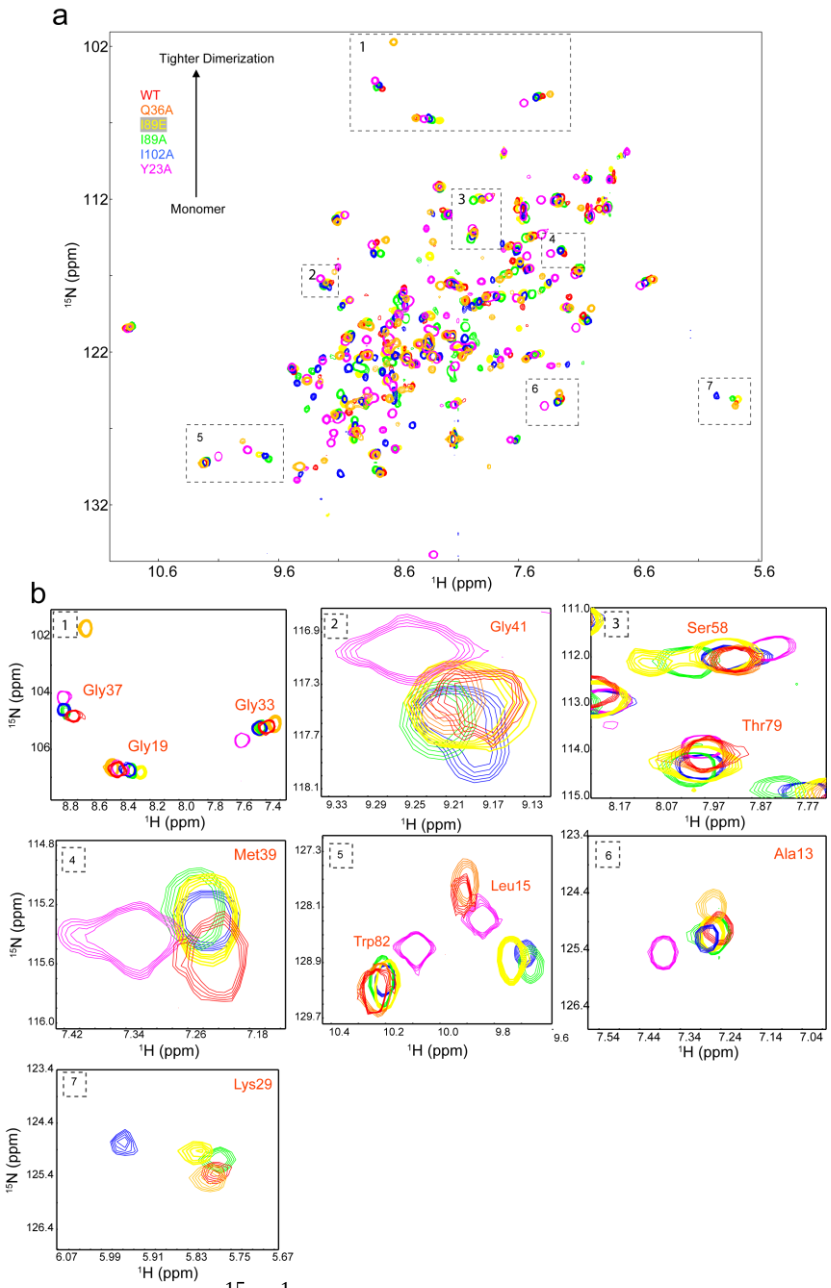


Figure 4-3. (a) Overlay of ^{15}N-^{1}H HSQC spectra of wild type versus several mutants. Conditions are as stated in Figure 4-2. (b) Boxed zones have corresponding zoom views. Assignments based on Y23A spectra whereby the nearest Y23A resonance was chosen to represent both wild type and mutant.

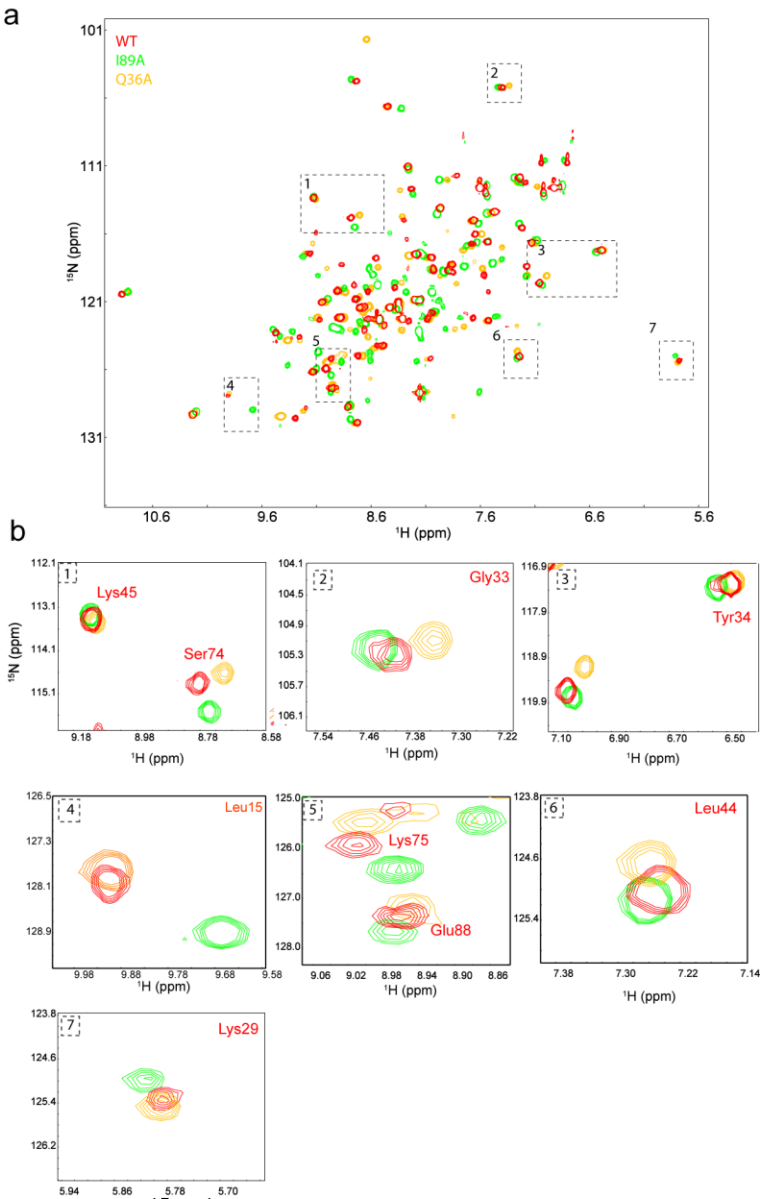


Figure 4-4. Overlay of ^{15}N-^{1}H HSQCs from wild type and Q36A and I89A mutants with corresponding excerpts of boxed regions. Conditions as stated in Fig. 4-2. (b) Boxed zones have corresponding zoom views. Assignments based on Y23A spectra whereby the nearest Y23A resonance was chosen to represent both wild type and mutant.

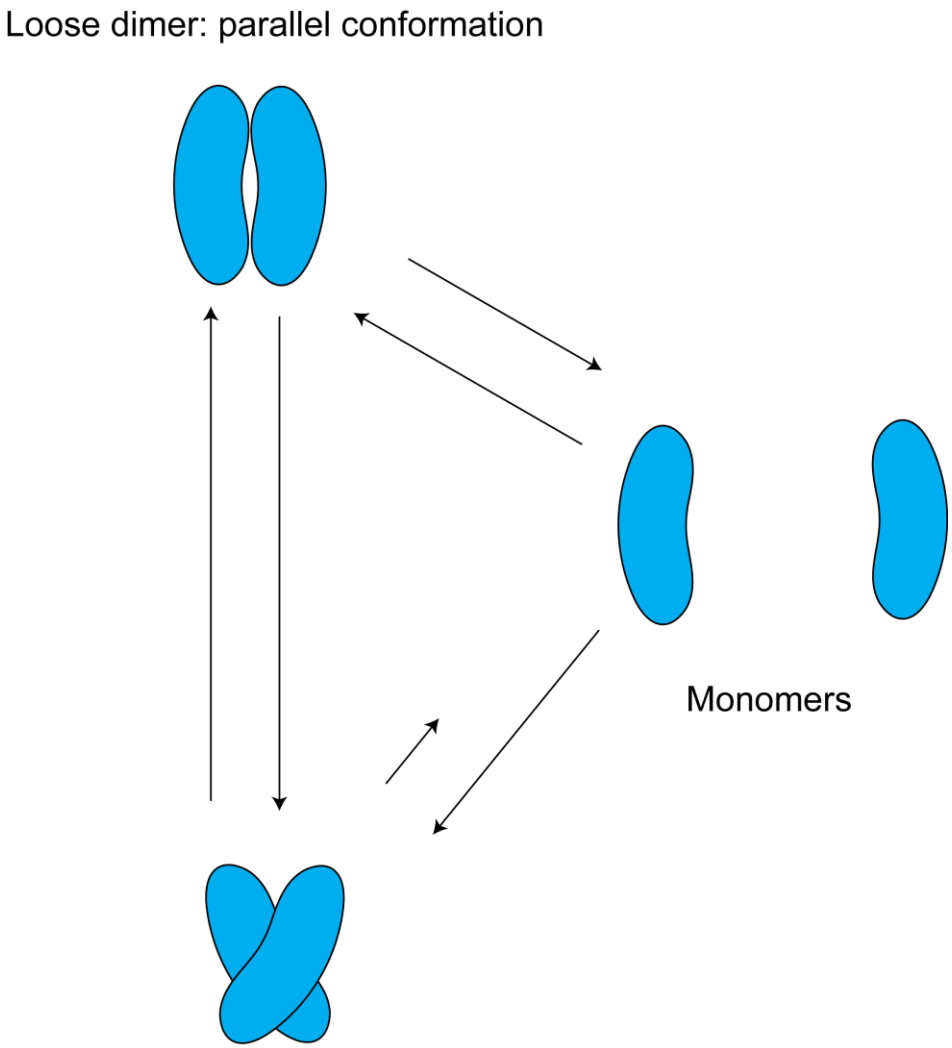


Figure 4-5. Model for putative conformation equilibria for KinA PAS-A. Unknown are the exact exchange rates between different conformations. We assume that the crystallographically-defined crossed and parallel conformations represent "tight" and "loose" dimers, respectively, based upon the differences in

buried surface area.

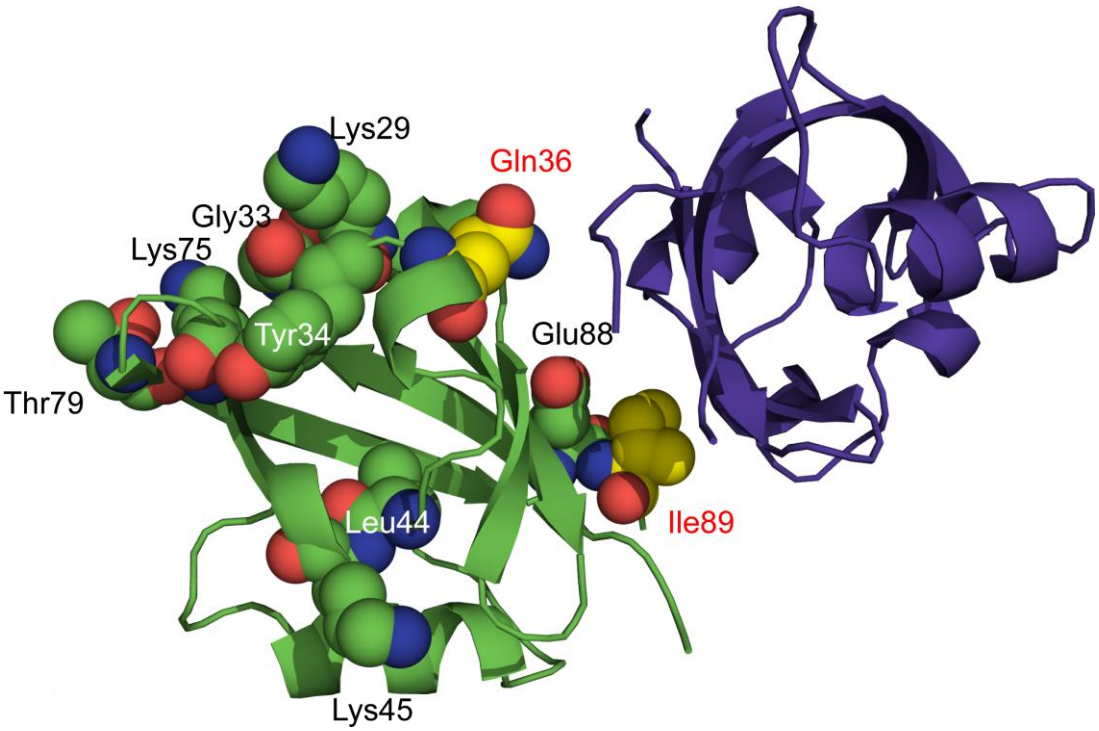


Figure 4-6. Clustering of residues with resonances that specifically shift with both I89A and Q36A mutations.

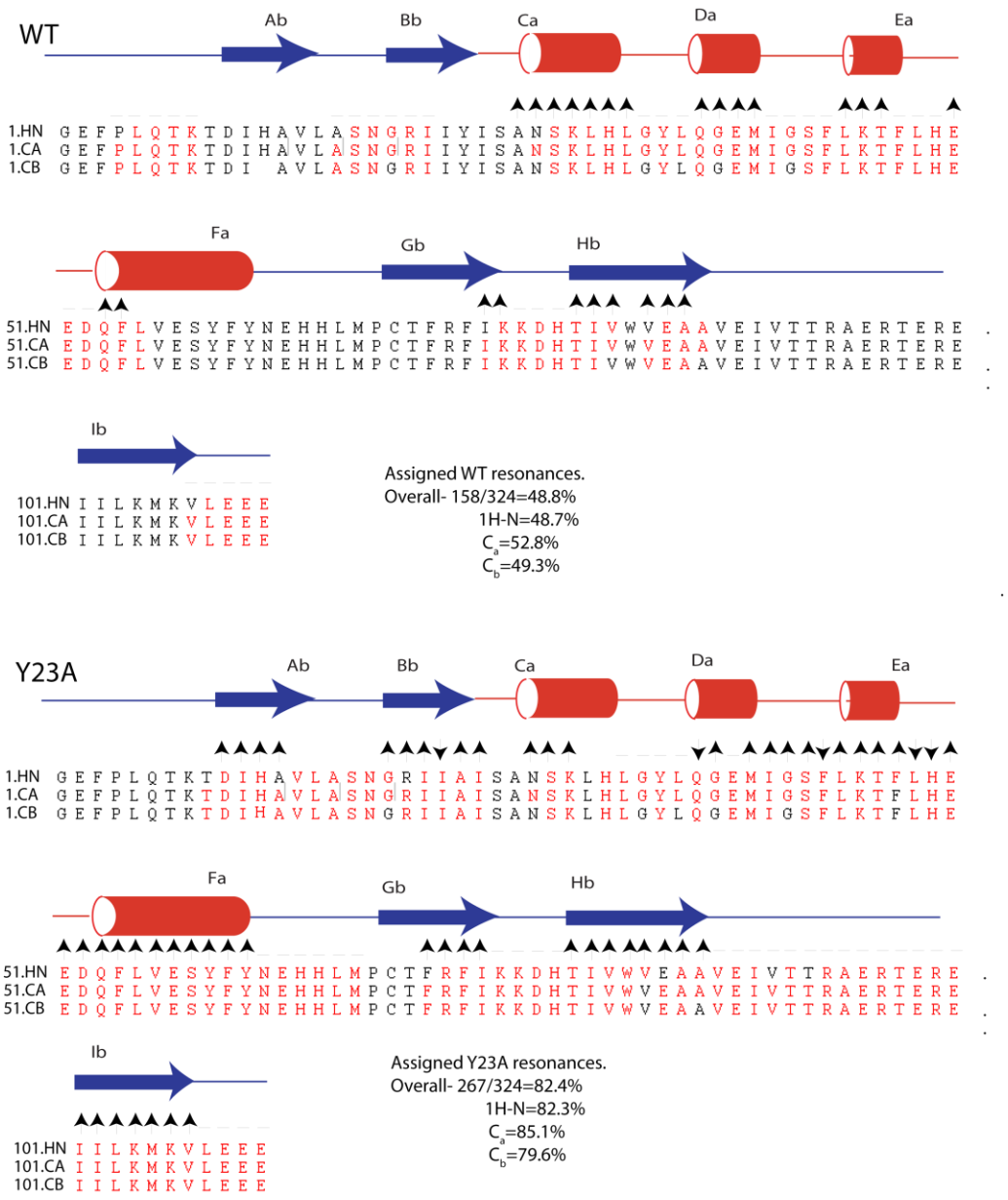


Figure 4-7. Resonances from wild type PAS-A and Y23A mutant that were either assigned (red) or unassigned (black). Arrows indicate locations correct (up) and incorrect (down) secondary structure predictions by TALOS as compared to the parallel conformation of the crystal structure presented in chapter 3.

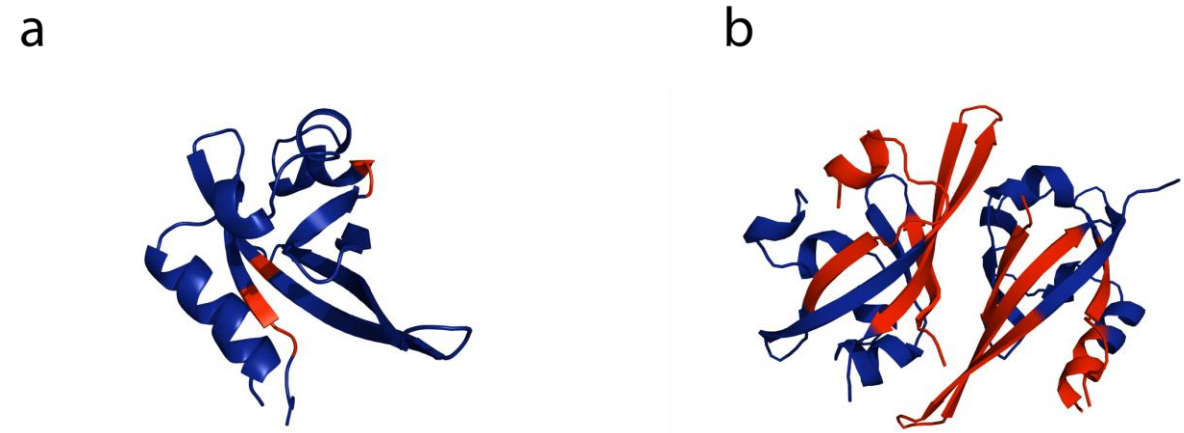


Figure 4-8. Mapping of residues that were assigned by triple resonance NMR experiments to the crossed conformation structure. Blue residues were assigned as judged by $^{15}N/^{1}H$ chemical shifts. Red residues were not assignable. (a) Assignable resonances in wild type KinA PAS-A spectra. (b) Assignable resonances in KinA PAS-A Y23A mutant spectra.

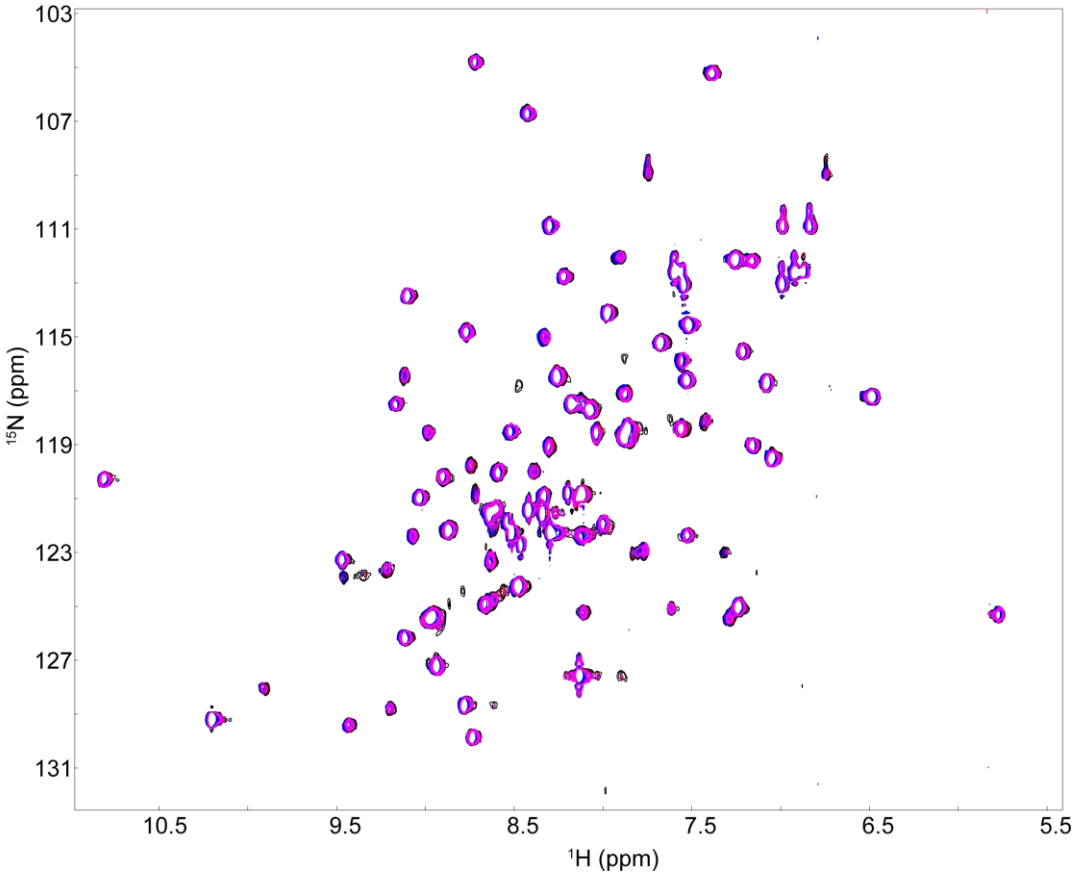


Figure 4-9. Overlaid ^{15}N-^{1}H HSQC spectra of NMR samples of KinA PAS-A at 50 μM (red, nt=128), 100 μM (magenta, nt=16 on 600 MHz spectrometer equipped with a cold probe), 450 μM (black, nt=32), and 1 mM (blue, nt=8) concentrations with the number of transients increased to compensate for signal-to-noise differences. All proteins are in 25 mM Tris (pH 7.5) and 100 mM NaCl. Spectra recorded on 800 MHz spectrometer unless otherwise indicated.

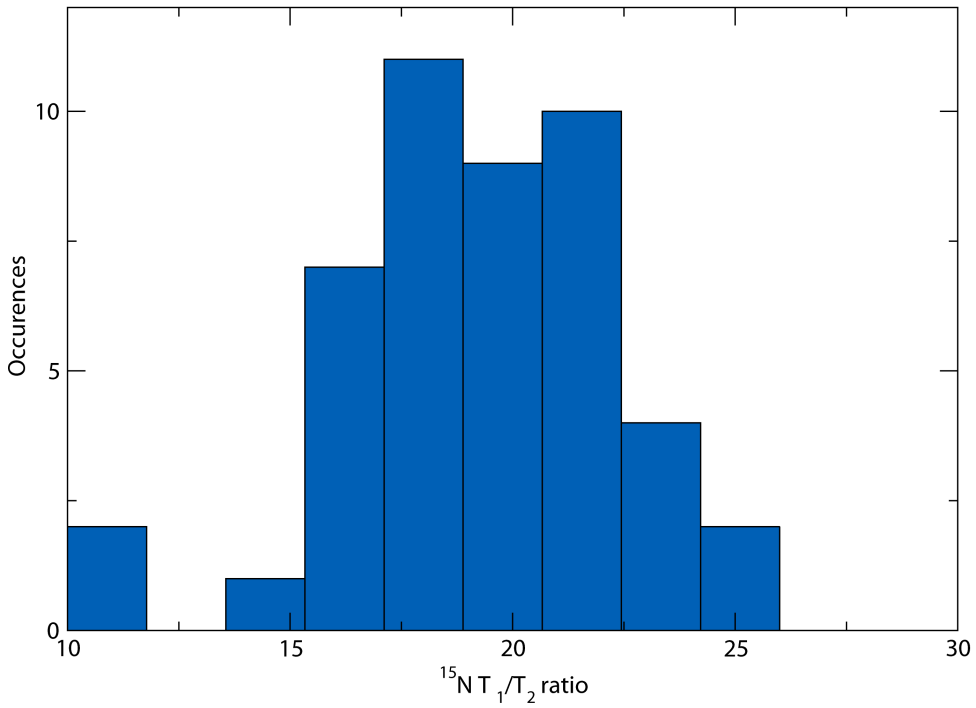


Figure 4-10. Histogram distribution of T_1/T_2 ratios of all peaks in wild type ^{15}N-^{1}H HSQC.

Chapter 5: Hydrodynamic radius, dimerization affinity, shape analysis of KinA PAS-A.

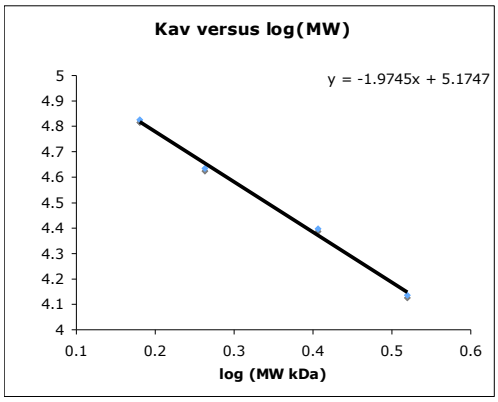
Our data thus far suggests that KinA PAS-A exists in a heterogeneous population because of dynamics in the dimerization interface. A number of techniques are available to obtain better resolution of this conformational equilibrium. In this regard, one general property of proteins that can be used to our advantage is that they have different shapes that are dictated not by the total number of amino acids, but by how they fold and how they multimerize. Size-exclusion chromatography and analytical ultracentrifugation are two techniques we employed to observe how mutations to KinA PAS-A change the hydrodynamic properties of the dimer.

MATERIALS AND METHODS

Size-exclusion Chromatography

Size-exclusion chromatography was performed using a Superdex 75 prep grade column from Pharmacia. Calibration of this column was performed using the standards albumin (67 kDa), ovalbumin (43 kDa), chymotrypsin (25 kDa), and Rnase (13.7 kDa) which had elution volumes of 59.4, 65.7, 76.7, and 85.3 mL, respectively. These elution volumes were subsequently fit to a equation of K_{av}

plotted against the logarithm of molecular weight, where $K_{av} = \frac{V_e - V_o}{V_t - V_o}$ and V_o is the void volume (45.6 mL) determined from injection of the high molecular weight compound dextran blue and V_t is the total volume of the column (122 mL).



Graph. 5-1

The equation for K_{av} and the logarithm of molecular weight was thus determined to be K_{av}=-1.97 log (MW) + 5.17. Solving for molecular weight then determines the equation correlating molecular weight in Da with elution volume (V_e) in mLs:

$MW = 10\hat{}(\frac{V_e}{38.5} + 6.42)$.

KinA PAS-A and mutant samples were purified as outlined in chapter 2. The buffer conditions after purification were 25 mM Tris (pH 7.5), 100 mM NaCl, and ~2 mM imidazole. Before concentration of samples for injection on the Superdex 75 prep grade column, samples were spiked with 5 mM DTT. After concentration of the samples to 250 μM concentration using an Amicon nitrogen pressured stir cell concentrator, samples were injected in a volume of 1 mL on the Superdex 75 prep grade column pre-equilibrated in 25 mM Tris (pH 7.5), 100

mM NaCl, and 5 mM DTT. The sample run was maintained at 0.5 mL/min flow rate for the duration of the run until the last peak at ~120-130 mL was observed.

Equilibrium Sedimentation Measurements

All KinA PAS-A proteins were purified as outlined in Chapter 2. The buffer conditions after purification were 25 mM Tris (7.5), 100 mM NaCl, and ~2 mM imidazole. Before concentration of samples for equilibrium sedimentation measurements, samples were spiked with 5 mM DTT. Concentrations were adjusted so that samples had absorbance at 280 nm of 0.2 and 0.7 except for I89A which only had an equilibrium sedimentation measurement only at the lower concentration due to lack of a tenth channel space. Thus the sample concentrations for KinA PAS-A and mutants were 17 μM and 61 μM except for Y23A which were 20 μM and 70 μM due to its predicted lower extinction coefficient (E_{280}) of 9970 $M^{-1}cm^{-1}$ compared to the E_{280} of 11,460 $M^{-1}cm^{-1}$ of the other constructs. Equilibrium analytical ultracentrifuge measurements were made with a Beckman XL-I analytical ultracentrifuge using a 4-position An60Ti rotor at 25°C. Each cell had a 6-channel carbon-epon centerpiece with two quartz windows giving an optical path length of 1.2 cm. Thus, nine [2(WT, I89E, I102A, Y23A)+ I89A] different 120 μL samples were measured together in one measurement with their respective buffers as blanks in the corresponding channel

below the sample channels such as to fill three rotor positions with a fourth rotor position serving as the balance. Each data set represents the average of 10 scans. Each cell was scanned stepwise (0.002 cm steps) at a wavelength of 280 nm and absorbance monitored relative to buffer. Experiments were carried allowing the samples to equilibrate at 24,000 and 35,000 rpm at 25°C. After equilibrium was reached (from 24 to 40 h), overspeed runs at 42,000 rpm were carried out to obtain baseline values of absorbance which were used in subsequent fits. The partial specific volumes of KinA PAS-A wild-type, I89E, I89A, I102A, and Y23A (0.7465, 0.7443, 0.7451, 0.7451, and 0.7469 $cm^3\ g^{-1}$, respectively) at 25°C were calculated from their amino acid compositions. The solvent density was calculated to be 1.002 $g{\cdot}mL^{-1}$ at 25°C. All data sets were collectively fitted to the appropriate equation using Beckman Optima XL-A/XL-1 data analysis software, version 4.0. Figures were replotted using Microsoft Excel except for the non-linear least squares fitting of data and residuals to this fit.

Velocity Sedimentation Measurements

All KinA PAS-A proteins were purified as outlined in Chapter 2. The buffer conditions after purification were 25 mM Tris (pH 7.5), 100 mM NaCl, and ~2 mM imidazole. Before concentration of samples for injection on velocity sedimentation measurements, samples were spiked with 5 mM DTT. Concentrations were adjusted so that samples had absorbance at 280 nm of 0.5.

Thus the sample concentrations for KinA PAS-A and mutants were 44 μM, except for Y23A which was 50 μM due to its lower extinction coefficient. Velocity sedimentation analytical ultracentrifugation measurements were made with a Beckman XL-I analytical ultracentrifuge equipped with a 12mm charcoal-filled double-sector centerpiece on sample volumes of 400 μL. Data were collected using interference optics at 55000 rpm at 25° C on all samples.

Basic tenets of Size-exclusion chromatography

Size-exclusion chromatography is based on the principle that porous particles can be used to retain molecules of different sizes (Coligan, Dunn et al. 2001). Molecules that are smaller than the pore size can enter the particles and thus, have a longer transit time on a column filled with those particles than larger molecules that cannot enter the particles. All molecules that are larger than the pore size experience total exclusion by the pore and elute together in what is termed the void volume. Vice versa, molecules much smaller than the pore size spend the longest residence time in the pore and elute together significantly later. For protein chromatography, small polypeptides and buffer molecules different from the equilibration buffer of the column elute in this fraction. Molecules smaller than the pore spend an average residence time there that depends on the molecule's size and shape, with the volume that they eventually elute at a volume (V_e) that follows a logarithmic function of the molecular weight. With this

principle, it is possible to obtain accurate calibrations for SEC columns by plotting the elution volumes for several standards of known molecular weights. For proteins of the size generally suitable for NMR, the Superdex 75 prep grade column from Pharmacia or columns with similar resin pore sizes can generally well resolve at least 5 kDa differences in proteins of the range from 5 kDa to 60 kDa. Although the most common application of SEC is to extrapolate a molecular weight, it must be noted that this is based on the assumption that the calibration standards and all other proteins have globular folds giving them a spherical shape. In such a case, molecular weight varies proportionally with the third power of the hydrodynamic radii of these proteins. Unfolded proteins or ones with highly elongated shapes will be excluded from pores more than globular proteins of equivalent molecular weight. As such, they elute at artificially low retention volumes, giving rise to high molecular weight estimates.

Comparison of SEC elution volumes for KinA PAS-A

Wild type KinA PAS-A and mutant derivatives were injected on a Superdex75 prep grade column from Pharmacia at 250μM protein concentration in a volume of 1mL for comparison (**Fig. 5-1**). Other concentrations were also injected, but generally insignificant differences were observed for these runs. The elution profile for KinA PAS-A calculates to a molecular weight of 31.1 kDa, consistent with a dimeric molecular weight. As more thoroughly discussed in Chapter 3, a

base network of hydrophobic interactions are provided by His12, Val 14, Ile22, Tyr23, Ala26, Glu88, and Glu100 for both crystallographically defined conformations. Inspection of these residues in both structures revealed that Tyr23 is positioned centrally in this network (**Chapter 3**). We thus, asked how disrupting this key pivot for the dimerization interface could affect the gel-filtration elution volume? The resulting elution profile for the Y23A mutant was shifted approximately 11.0 mL from the 74.2 mL elution volume wild type to 85.2 mL (**Table 5-1**), consistent with a monomeric molecular weight of 16.1 kDa. Just outside this pivotal residue, Ile102 seemed important to supporting hydrophobic moieties in the His12/Glu88 polar interaction. Perhaps also underscoring this residues importance to the dimerization interface, the mutation of Ile102 to alanine leads to a severe shift in gel-filtration elution volume of 7 mL from 74.2 mL to 81.2 mL (**Fig. 5-1**).

Table 5-1. Molecular weights as determined from Superdex75 prep grade column.

	Elution Volume, V_e (mL)	MW* calculated from SEC75 (kDa)
WT	74.2	31.1
Q36A	75.0	29.6
I89E	77.3	25.7
I89A	78.6	23.9
I102A	81.2	20.5
Y23A	85.2	16.1

*Refers to molecular weight as calculated from the equation, $MW = 10^{\wedge}(\frac{V_e}{38.5} + 6.42)$, as determined from calibration using the standards albumin (67 kDa), ovalbumin (43 kDa), chymotrypsin (25 kDa), and RNase (13.7 kDa) which had elution volumes of 59.43, 65.7, 76.7, and 85.3 mL, respectively. These elution volumes were subsequently fit to a equation of K_{av} versus Log (MW), where $K_{av} = \frac{V_e - V_o}{V_t - V_o}$ and V_o is the void volume (45.6 mL) determined from injection of the high molecular weight compound dextran blue and V_t is the total volume of the column (122 mL).

In contrast to these striking gel-filtration elution shifts, subtler size-exclusion chromatography changes are experienced by mutations aimed only at each individual crystallographic conformation. Ile89, which helps forms a hydrophobic network of contacts *flanking* the central β-sheet dimerization

interface for the parallel conformation, observes a 3.2 mL shift in elution volume when mutated to glutamic acid (**Fig. 5-1**). Similar small gel-filtration elution shifts are observed with the Q36A, which forms a polar interaction with Glu100 only for the crossed conformation. This mutant has a gel-filtration elution profile with a shift of ~0.81 mL from that of wild type. Of course, the gel-filtration elution volumes for these mutants would predict molecular weights intermediate of that between monomer and dimer (**Table 5-1**), an unlikely scenario given that ^{15}N-^{1}H HSQC spectra show that these mutations generally exhibit the same overall fold of wild type. However, because these spectral and gel-filtration changes are small, it is possible that even a single contact residue that is specific only to one conformation can significantly contribute to dimerization affinity.

It is tempting to explain our gel-filtration results through a model where the mutations at the interface were simply weakening dimerization affinity to different degrees. In this regard, we concluded in chapter 3 that chemical shift changes could not be strictly correlated to dimerization. To be fair, some resonances do observe the greatest chemical shift changes in the completely monomeric Y23A. However, in many other resonances, the greatest chemical shift changes occurred in I89E, calculated by gel-filtration to be most dimeric of all the mutants (**Chapter 4**). Moreover, only the Q36A mutant shifts certain resonances in a direction opposite from the pattern followed by resonances for wild type and I89A, suggesting an opposite conformational shift compared with

that of the other mutants. Based upon these gel-filtration data, we thus hypothesize that the mutants I89A/E, Q36A, and I102A do reduce dimerization affinity and this effect has characteristic chemical shift changes. However, these mutations also shift the equilibrium between different dimeric conformations, leading to another series of specific chemical shift changes. Moreover, the disruption of one interface leads to either new compensating interactions or the strengthening of pre-existing interactions that have also have specific chemical shift changes (**Chapter 4, Fig. 4-5)**. We thus conclude that the convolution of all these effects make correlation of chemical shift changes to a monomer-dimer model only possible for select subsets of resonances for all mutants.

Our results, thus far, suggest that a change to the equilibrium of dimeric conformations was at least partially responsible for the NMR changes. However, it is noteworthy that the I89E/A and I102A mutations shift elution volumes without flattening out the peak shapes for these gel-filtration profiles or giving rise to another slower eluting monomer peak, suggesting that shape changes are occurring in the equilibrium of PAS-A conformations. To gain a better understanding as to the structural mechanisms underlying changes in gel-filtration elution volume and also chemical shift resonances, we performed analytical ultracentrifugation on PAS-A and its mutant derivatives.

Basic tenets of equilibrium sedimentation

In order to describe the equilibrium sedimentation of protein molecules, one must first consider that in an ensemble, there are many individual particles. To describe all of these particles, one must theorize that the flow of the molecules, including both sedimentation *and* diffusion, is dependent on a gradient of an appropriate potential that includes the effects of both concentration and centrifugal field (Wills and Winzor 2001). This total potential can be defined as the following:

$$\tilde{\mu} = U_c + \mu_2 \quad \textbf{Eq. 5-1.}$$

where U_c is the centripetal potential and μ_2 is the chemical potential of solute. The flux of particles through a unit area (J) defined in the sample is calculated by multiplying the gradient of the total potential, $\left(\partial\tilde{\mu}/\partial r\right)$, by mobility, L. In computing the total potential gradient, the chemical potential is then treated as a function of temperature (T), pressure (p) and concentration (c). The actual solution density (ρ) provides the means of calculating the pressure term, which varies with rotor position. Thus, to calculate J, we take derivatives of both sides of Equation 5-1 to get Equation 5-2a and insert the formulas for centripetal potential (U_c) and chemical potential (u_2) individually to obtain Equation 5-2b:

$$J = L\left\{\partial\tilde{\mu}/\partial r\right\} = \left\{\partial U_c/\partial r + \partial\mu_2/\partial r\right\} \quad \textbf{Eq. 5-2a}.$$

$$J = L\left\{\omega^2 rM(1-\rho\tilde{v}_2) - \frac{RT}{c}\frac{dc}{dr}\right\} \quad \textbf{Eq. 5-2b.}$$

where the first term accounts for the centripetal force acting on the particles after correcting for buoyancy by the $(1-\rho\tilde{\nu}_2)$ term and the second term accounts for the flux due to the diffusion of a particle down a concentration gradient.

At equilibrium, *J* equals zero and equation 5-2 can be solved as follows:

$$\frac{1}{c}\frac{dc}{dr}=\frac{\omega^2 rM}{RT}(1-\rho\tilde{\nu}_2) \quad \textbf{Eq. 5-3.}$$

To determine a certain concentration at a specific radial position (r) from the meniscus position (a), we re-evaluate equation 5-3:

$$c(r)=c(a)\exp(\frac{\omega^2 M(1-\rho\tilde{\nu}_2)(r^2-a^2)}{2RT}) \quad \textbf{Eq. 5-4.}$$

where a is the position of the meniscus of the solution. Finally, to determine the molecular weight (M), one linearizes the equation:

$$\ln[c(r)]=\ln[c(a)]+\frac{\omega^2 M(1-\rho\tilde{\nu}_2)(r^2-a^2)}{2RT} \quad \textbf{Eq. 5-5.}$$

In such manner, if many species are present, then the concentration profile simply becomes the sum of exponentials seen in equation 5-5.

Dimerization affinities from equilibrium sedimentation for KinA PAS-A

With this basis for deriving an affinity of associating species, we collected successive absorbance (280nm) data until equilibrium of the system was reached both for speeds of the 25,000 and 35,000 rpm. Shown in **Fig 5-2a** are representative profiles from 25,000 rpm which was determined to be more

optimal for the dimerization affinities and molecular weights studied for KinA PAS-A (see below). To allow a qualitative comparison without computationally fitting this data, we have also chosen to plot absorbance (280nm) on a logarithmic scale (**Fig. 5-2b)**. Such a scale allows one to compare the relative slopes of our wild type and mutant profiles to make qualitative comparisons of affinity of dimerization as Eq. 5-5 shows that a greater slope of ln[C(r)] versus r equates to greater molecular weight values calculated. From this plot, one can observe that the I89E profile has a slightly steeper profile than wild type, resulting in a higher molecular weight determination and dimerization affinity. Y23A has a very shallow profile and thus indicates a significant weakening of dimerization affinity. However, the Beckman Origin-based software (version 3.01) can be used to more accurately determine the apparent molecular weight through non-linear least squares fitting. Fitting was performed for wild type, Y23A, I102A, I89A, and I89E, with residuals to the fit within appropriate deviation and spread (**Fig. 5-2c**). We fit these data to a monomer-dimer equilibrium model such that the a molecular mass for wild type was determined as 18.8±0.2 kDa (**Table 5-2)**. By assuming that the only species in equilibrium are monomer and dimer, then we can calculate the ratio of monomer and dimer that are needed to determine this molecular mass. This ratio can then be used to determine a K_d value for dimerization. Thus, in wild type, K_d for dimerization was determined to be 8 μM while Y23A fits a model whereby the only species present had a molecular weight

around 11.5±0.4 kDa, consistent with the complete lack of dimerization. Meanwhile, I89E had a monomer-dimer equilibrium model that predicts a molecular mass of 22.2±0.1kDa, thus producing an affinity tighter than wild type with K_d~1 μM. I102A and I89A fit a monomer-dimer equilibrium model predicting a molecular mass of 16.6±0.3kDa and 17.3±0.4kDa with resulting dimerization affinities of 32μM and 24uM, respectively. These values are generally consistent with the idea that mutant gel-filtration elution volumes intermediate between those of wild type and Y23A reflect intermediate dimerization affinities with the exception that I89E was measured as having a greater dimerization affinity than wild type. This suggests that this particular mutant shifts the equilibrium of dimer conformations to a tighter dimer conformation. To further assess such possibilities, we also performed velocity sedimentation.

Table 5-2. Molecular weights and dimerization affinities as determined from gel-filtration and equilibrium sedimentation.

	Elution Volume, V_e (mL)	MW* calculated from gel-filtration (kDa)	Molecular mass calculated from equilibrium sedimentation	K_d+ calculated from equilibrium sedimentation (μM)
WT	74.2	31.1	18.8±0.2 kDa	8
Q36A	75.0	29.6	NA	NA
I89E	77.3	25.7	22.2±0.1kDa	1
I89A	78.6	23.9	17.3±0.4kDa	24
I102A	81.2	20.5	16.6±0.3kDa	32
Y23A	85.2	16.1	11.5±0.4 kDa	monomer

*Refers to molecular weight as calculated from the equation, $MW = 10\text{^}(\frac{-V_e}{38.5} + 6.42)$, as determined from calibration using the standards albumin (67 kDa), ovalbumin (43 kDa), chymotrypsin (25 kDa), and Rnase (13.7 kDa) which had elution volumes of 59.4, 65.7, 76.7, and 85.3 mL, respectively. These elution volumes were subsequently fit to a equation of K_{av} versus Log (MW), where $K_{av} = \frac{V_e - V_o}{V_t - V_o}$ and V_o is the void volume (45.6 mL) determined from injection of the high molecular weight compound dextran blue and V_t is the total volume of the column (122 mL).

+K_D calculated from equation $K_d = M \exp[\frac{M_{mono}(1-v\rho)\omega^2(r-a)}{2RT} + K_a M^2 \exp[\frac{2M_{mono}(1-v\rho)}{2RT}]$ where M is the average molecular weight of the monomer/dimer combination and M_{mono} is the monomer molecular weight. A graph of the natural log of the concentration versus the radius squared will yield a curve that can be fitted to yield K_A using non-linear least squares fit.

Basic tenets of velocity sedimentation

The velocity sedimentation experiment is based on the AUC cell spinning with some angular velocity. The total force applied on particles in such a cell can be analyzed as the following:

$$F_{total} = \omega^2 rm - \omega^2 rm_o - fv \quad \textbf{Eq. 5-7.}$$

where the first term describes centripetal force with angular velocity ω, the second term describes buoyancy force for a displaced solvent mass m_o, and the third term describes the frictional force for the effective linear velocity v under the frictional coefficient f (Cantor and Schimmel 1980). Of course, if we assume that the displaced solvent mass at terminal velocity observes no net acceleration, then the sum of these forces should be zero. Furthermore, the displaced solvent mass can be taken simply as the partial specific volume $\tilde{v}_2$ multiplied by the solution density and mass, $m_o = m\tilde{v}_2\rho$.

Now the equation can be resolved into a value that can be defined as the Svedberg (S):

$$S = \frac{v}{\omega^2 r} = \frac{M(1-\rho\tilde{v}_2)}{N_a f} \quad \textbf{Eq. 5-8.}$$

To obtain S from AUC measurements, we define v as the displacement of the particles with a certain time t, and solve for terms that are measurable in the AUC experiment:

$$v = \frac{dr_b}{dt} = r_b \omega^2 S \quad \textbf{Eq. 5-9.}$$

where r_b is defined as the boundary of the moving particles under a certain angular velocity. After dividing both sides by r_b, the differential equation can integrated and solved:

$$\ln\left(\frac{r_b(t)}{r_b(0)}\right) = \omega^2 S(t - t_o) \quad \textbf{Eq. 5-10.}$$

When collecting velocity sedimentation data, one takes scans of the cell along the entire pathlength at certain time intervals.

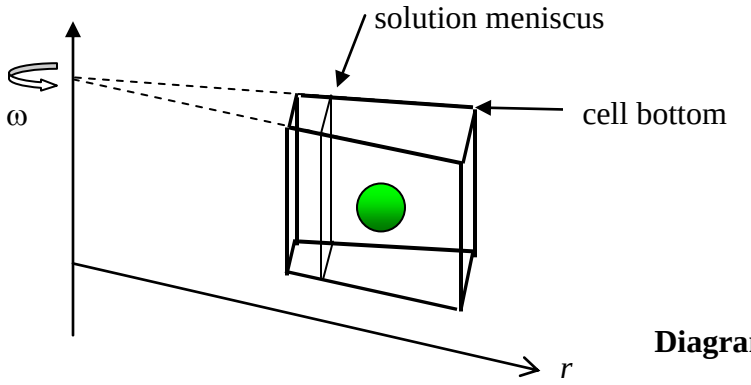


Diagram 5-1.

The readout for such an experimental setup is a series of curves that progress away from solution meniscus toward the cell bottom with increasing time:

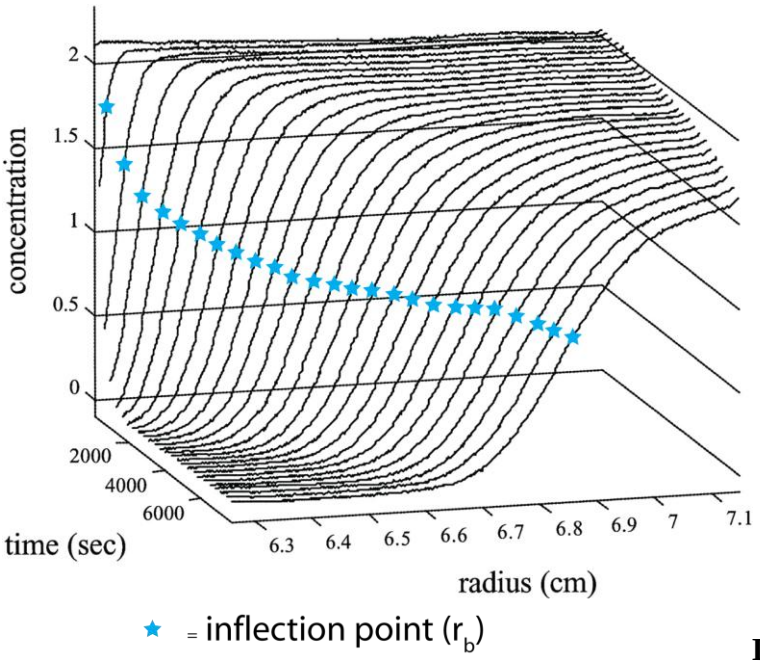


Diagram 5-2.

(Figure from Lebowitz et al., Protein Science 11: 2067-9)

In principle, plotting natural logarithm of the inflection point of each curve (r_b)—denoted by the blue star above—versus time (t), should produce a line with a slope that is directly proportional to *S*. Although this mathematical treatment provided the fundamental means of solving for *S*, newer computational approaches are now available that produces greater signal-to-noise and reduces time-independent systematic error. Such methods are employed by the g(s*) method Sedfit software (Schuck 2000), which we used to evaluate our data. In this approach, closely spaced data sets are subtracted to give time derivative profiles. Each boundary gives a separate Gaussian distribution that can be normalized such that these curves will superimpose. With a renormalized y-axis, one can average these curves to produce a gain in signal-to-noise ratio. The peak

of this averaged Gaussian gives the sedimentation coefficient while standard error is related to the diffusional coefficient (D).

Velocity sedimentation: PAS-A lies in a number of conformation states

To be certain if gel-filtration elution changes exhibited by KinA PAS-A mutants were reflecting either changes to dimerization affinity and/or changes to shape, sedimentation velocity profiles were measured for KinA PAS-A and various mutants. Shown in **Figure 5-3** are sample velocity sedimentation profiles for wild type and Y23A KinA PAS-A. It can be observed that the wild type profile has inflection points (r_b) that move slower than that of the Y23A, indicative of larger Svedberg species. Indeed in **Figure 5-4**, it can be observed that the wild type profile fit a model whereby two separate Svedberg constants of 1.89 and 2.42, both higher than that of Y23A (**Table 5-3**). In order to better understand these values, we recalibrated gel-filtration elutions using hydrodynamic radius instead of molecular weight. Given the hydrodynamic radius of 2.41 nm determined from this data, the calculated molecular weights would be 19.1 kDa and 24.5 kDa. The latter molecular weight is consistent with a dimeric domain, but the former value is intermediate of a monomeric and a dimeric mass. Meanwhile, the I89A and I89E mutants seemingly shift to the intermediate conformation seen in the wild type profile. Their major species calculate respective molecular weights of 18.4 kDa and 19.4 kDa when given hydrodynamic radii from gel-filtration. However,

new conformations also appear at 2.82 and 3.01 Svedberg constants, respectively. These conformations calculate molecular weights of 25.7 kDa and 28.9 kDa, consistent with dimers if again, it is assumed that hydrodynamic radii from gel-filtration are appropriate for this calculation. Here, we observe that even though hydrodynamic radius is reduced in I89A/E mutants, the minor but larger Svedberg species still calculates to be of dimeric mass. On the other hand, the retention of a species of intermediate Svedberg constant and molecular weight as compared to wild type profiles may suggest the existence of a common dissociating species and/or a static and stable species that is not altered by disruption of a single interface in an equilibrium of possible interfaces.

These compensatory changes of hydrodynamic radius and Svedberg constant tempts speculation that mutants aimed to specifically disrupt a particular dimerization interface observe new compensating interactions by other residues or the strengthening of pre-existing interactions. Perhaps, then the separate dimer subunits interact with each other through an interface that is not as centralized to the β-sheet core as shown for the crystal structure of the wild type, leading to "loose" dimer packing at an peripheral interface. This idea may then support the fact that chemical shift changes for these mutants intended to disrupt specific crystallographically-defined dimerization interfaces can only be correlated to dimer affinity for particular subsets of residues while other sets of chemical shift resonances follow more vague patterns (**Chapter 4**).

Table 5-3. Hydrodynamic radius, Svedberg constants, and molecular weights.

	Hydrodynamic Radius (nm)*	Svedberg constant	Molecular weight (kDa)$^{+}$
WT	2.41	2.42^{a} 1.89^{b}	24.5^{a} 19.1^{b}
I89E	2.23	2.07^{a} 3.01^{b}	19.4^{a} 27.9^{b}
I89A	2.17	2.02^{a} 2.82^{b}	18.4^{a} 25.7^{b}
I102A	2.04	1.67	14.3
Y23A	1.86	1.55	12.1

Superscripts a and b refers to separate peaks observed on velocity sedimentation with **a** as the major peak and **b** the minor peak.

Superscript (+) is the molecular weight calculated from hydrodynamic radius and Svedberg constant according to the formula: $S = \frac{M(1-\rho\tilde{v}_2)}{N_a f}$

Superscript* refers to hydrodynamic radius as calculated from the equation, $R = 10\text{^}(\frac{-V_e}{97.7} + 1.14)$, as determined from calibration using the standards albumin (3.55 nm), ovalbumin (2.73 nm), chymotrypsin (2.24 nm), and RNase (1.92 nm) which had elution volumes of 59.4, 65.7, 76.7, and 85.3 mL, respectively. These elution volumes were subsequently fit to a equation of K_{av} versus Log (R), where $K_{av} = \frac{V_e - V_o}{V_t - V_o}$ and V_o is the void volume (45.6 mL) determined from injection of the high molecular weight compound dextran blue and V_t is the total volume of the column (122 mL).

Less ambiguous are the velocity sedimentation profiles for the Y23A and I102A mutants. These respectively produce single sharp profiles with Svedberg constants of 1.55 and 1.67 *S* values that when coupled with hydrodynamic radius information from gel-filtration elution profiles, determine molecular weights of 12.1 and 14.3 kDa, both more consistent with a monomeric domain.

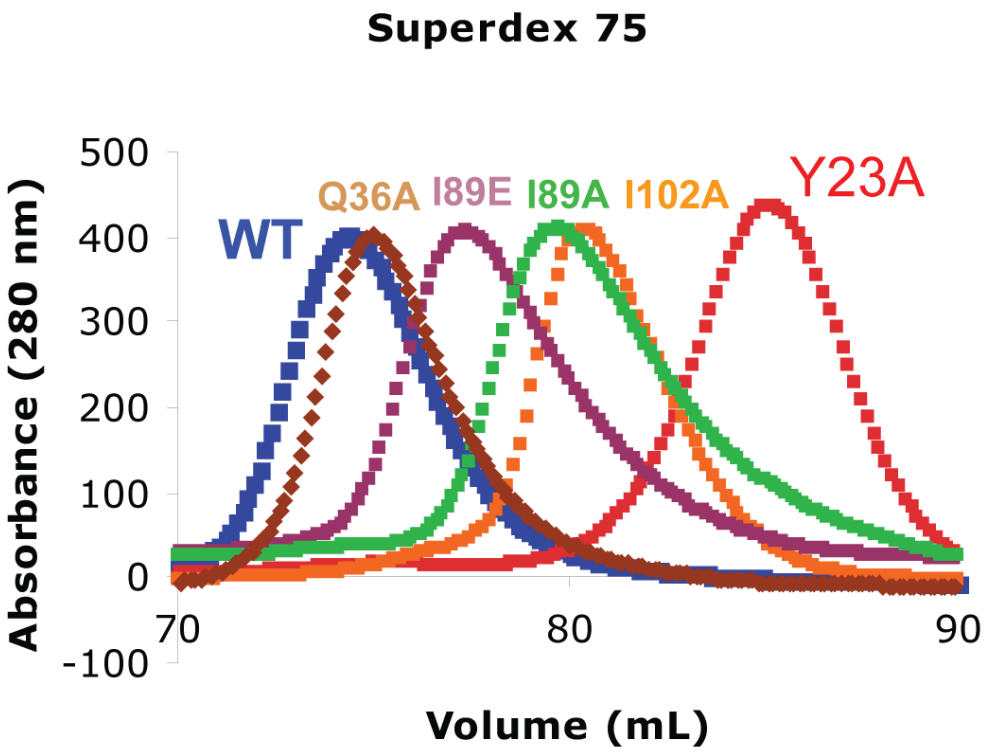


Figure 5-1. Size-exclusion chromatography profiles of wild type and various mutants from SEC75 column injected at ~250 μM (Pharmacia).

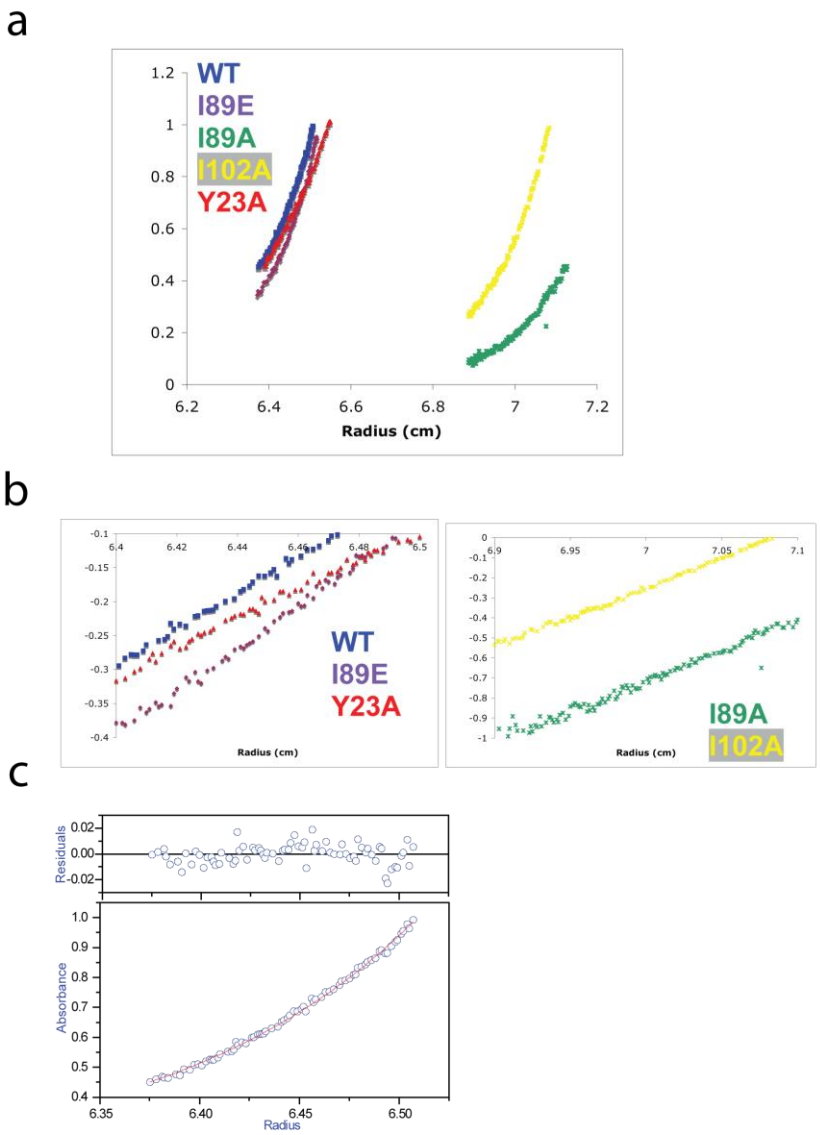


Figure 5-2. Equilibrium sedimentation of KinA PAS-A wild type (blue), I89E(purple), Y23A (red), and I102A (yellow, separate AUC cell) profiles taken for ~50μM protein at 24,000 rpm. (a) Profiles of equilibrium sedimentation obtained from analytical ultracentrifugation experiments. (b) Profiles of equilibrium sedimentation from (a) replotted on logarithmic scale to allow qualitative comparison of relative dimerization. (c) Sample residuals of fit to exponential function from wild type profile as determined by Origin to obtain the K_d of dimerization of 8 μM.

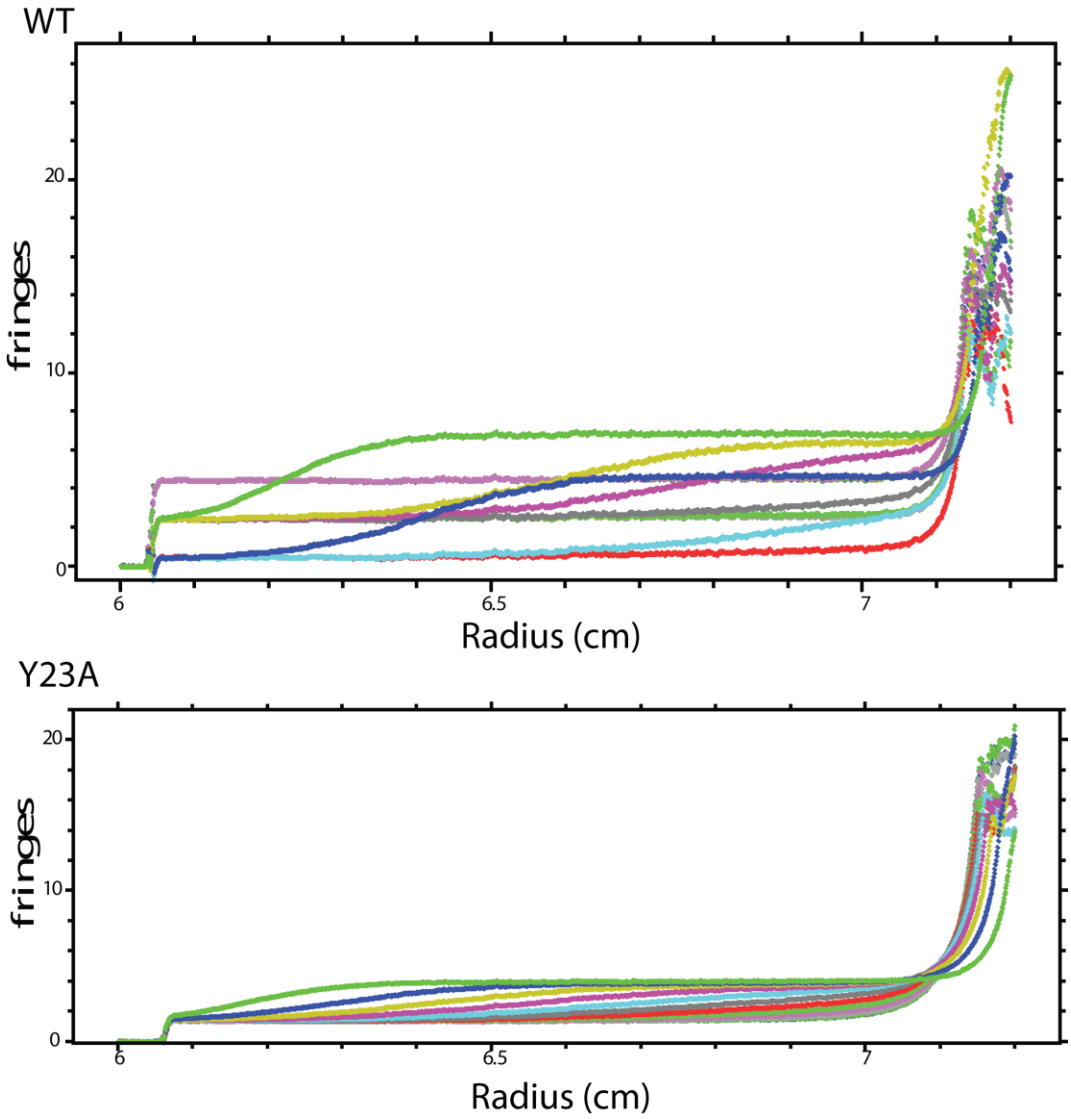
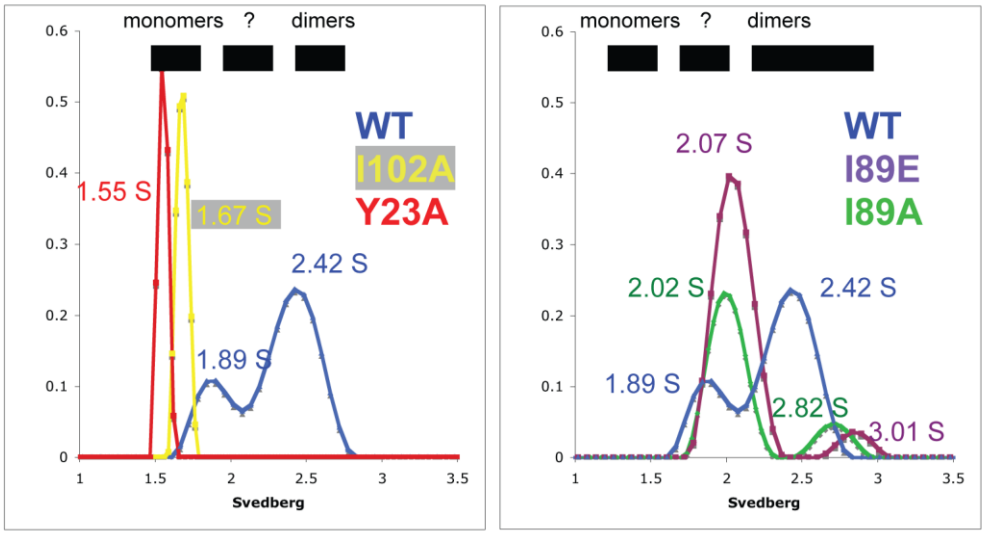


Figure 5-3. Raw velocity sedimentation profiles of WT and Y23A proceeding rightward with time in the AUC cell.



-

Figure 5-4. Sedimentation profiles of wild type and various mutants.

Chapter 6: Reconstitution of PAS-A mutants in full-length KinA.

Through the use of NMR, size-exclusion chromatography, and analytical ultracentrifugation, we have shown that mutations targeting crystallographically-defined interfaces change the equilibrium occurring in KinA PAS-A. It remains a question whether or not disrupting the PAS-A monomer/dimer equilibrium has an impact on KinA activity. In order to determine the direct impact of these mutants on the overall function of KinA, mutations were introduced in full-length protein and biochemically purified to assess the effects of these mutations on *in vitro* kinase activity. To test whether these mutations had physiological significance, we also expressed these mutations in live *Bacillus subtilis* cells and monitored activity from a sporulation-dependent promoter. From these data, we can begin to formulate a model for the PAS-A regulation of KinA activity.

MATERIALS AND METHODS

Expression and purification of KinA

An *E coli* system was used to facilitate protein expression using the established materials of our laboratory. Briefly, the DNA encoding *Bacillus subtilis* KinA was PCR-amplified from chromosomal DNA isolated from *B. subtilis* strain 1A40 (Bacillus Genetic Stock Center, Columbus, OH) and subcloned into the parallel vector pHis$_6$-parallel (Sheffield, Garrard et al. 1999) using EcoRI and XhoI

restrictions sites. We generated constructs encoding residues for the full-length KinA sequence (aa 1-606) and residues 151-606 for the ΔPAS-A protein. These constructs were overexpressed as a TEV-cleavable N-terminal His_6-tag fusion protein using *E coli* BL21 (DE3) by induction with 0.5 mM IPTG for 16 h at 20 °C. Cells harboring this overexpressed His_6-KinA protein were then resuspended in buffer containing 25mM Tris (pH 8.0), 100mM NaCl, 5mM EDTA, and 5% (v/v) glycerol. His_6-KinA was purified using Ni^{2+}-loaded NTA column, eluted using a linear 20mM-300mM imidazole gradient and subsequently exchanged into imidazole-free buffer. The His_6 tag was cleaved by incubating overnight with His_6-TEV protease. As with non-aggregated KinA PAS-A proteins, we found that we could bind this TEV-cleaved full-length construct to another Ni^{2+}-loaded NTA column, suggesting that our full-length KinA was being overexpressed in a non-aggregated form. In this manner, KinA could be separated from the cleaved His_6 tag as well as further purified by loading our protein sample to another Ni^{2+}-loaded NTA column. The cleaved KinA is capable of binding this column with low affinity (elutes with 20mM imidazole) (**Fig. 6-1**).

In vitro Kinase assay method

After purification with a Superdex 200 10/300GL column from Pharmacia, the final yields per liter of LB broth media were in the range of 7-12 mg/L of LB media. These purified products could be quantitated for activity in *in vitro* kinase

assays where KinA was incubated with γ-^{32}P labeled ATP and monitored with an autophosphorylation assay. Phosphorylation reactions were carried out as previously outlined by Wang et al (Wang, Fabret et al. 2001). In summary, reactions were carried out in a total volume of 40 µl in reaction buffer (25 mM Tris buffer (pH 7.5), 100 mM NaCl, 0.1 mM EDTA, 20 mM MgCl2, 5% [v/v] glycerol) containing 5 µCi of γ-^{32}P radiolabeled ATP (3,000 mCi/mmol; Du Pont-NEN) mixed with cold ATP to give a final concentration of 400 µM. Assay were initiated by ATP addition and the reaction was quenched by addition of 5X SDS-PAGE sample buffer after incubation at 25°C for specified lengths of time. Analyses of these samples were conducted by loading samples onto a 10% SDS-PAGE gel, followed by electrophoresis at a constant voltage (240 V) for 35 min or until the dye front had migrated at least 75% of the gel length. Background radiation during autoradiography was reduced by removing the lower portions of these gels containing the dye front along with unincorporated [γ-^{32}P]ATP. The gel was then dried at 80°C under vacuum and the gel exposed on a PhosphorImager screen for 16 h at room temperature and the bands of interest quantitated using with ImageQuant software (Molecular Dynamics).

In vivo KinA activation assay method

In order to confirm that the *in vitro* results above had significance physiologically, we were determined to perform *in vivo* measurements of kinase activity

experiments. Toward this end, we used the vector pdg-148 which has two different origins of replication to allow propagation not only in *B subtilis* but *E coli* as well, a feature which facilitates cloning (Joseph, Fantino et al. 2001). Furthermore, KinA was cloned into this vector such it is placed under control of an IPTG-inducible promoter. Previous reports have validated the approach of overexpressing KinA as a means of artificially inducing sporulation (Sheffield, Garrard et al. 1999; Fujita and Losick 2005; Veening, Kuipers et al. 2006)

The complementary DNA encoding *Bacillus subtilis* full-length KinA (1-606) was PCR-amplified from chromosomal DNA isolated from *B. subtilis* strain 1A40 (Bacillus Genetic Stock Center, Columbus, OH) and cloned into this pdg-148 vector using HindIII and SphI restrictions sites to build the pKinA plasmid. The plasmid pKinA, whereby KinA is placed under an IPTG inducible promoter, and its derivative mutant plasmids were propagated and isolated from *E. coli* MC1061. All future steps in *in vivo* experiments were next performed by Remco Kort, a post-doctoral fellow working in the laboratory of Klaas Hellingwerf at the University of Amsterdam. These plasmids were then transformed into *Bacillus subtilis* strain IIA-gfp (Veening, Kuipers et al. 2006). This strain has the genes for KinA and KinB deleted and GFP placed under the control of the spoIIA promoter. The use of this specific strain should allow fluorescence as a reporter of KinA activity as it has previously been shown that $KinA^-/KinB^-$ strains do not sporulate (Trach and Hoch 1993). In this manner, KinA activity is directly monitored

through a readout that comprehensively includes phosphotransfer throughout the entire sporulation phosphorelay as reconstituted in physiologically active cells (**Fig. 6-6 for schematic)**.

Method: Colony microscopy to qualitate changes induced by mutations

(Experiments performed by Remco Kort from the laboratory of Klaas Hellingwerf at the University of Amsterdam)

Colonies of *B subtilis* were observed and photographed using the stereoscopic 0.8 to 8x zoom microscope SMZ-1000 (Nikon Corporation, Tokyo, Japan), equipped with two 10x oculars and a 0.5x objective (working distance, 123.6 mm). The microscope contained an epifluorescence attachment and a DXM-1200 digital camera system (Nikon). The EclipseNet software package version 1.16.2 was used for control of the camera and image processing. Fluorescent *B. subtilis* reporter strains were monitored by the use of an HBO 103W/2 mercury short arc lamp (Osram Inc., Augsburg, Germany) and a long pass GFP filter (ex 460-500, DM505, BA510). All fluorescence pictures were taken with an exposure time of 1 s and default color balance settings (red gain, 30; green gain, 10; blue gain, 50) and processed identically.

***Method: Flow cytometry analysis (*Experiments performed by Remco Kort of the laboratory of Klaas Hellingwerf at the University of Amsterdam)**

To investigate whether differences seen by fluorescence microscopy could be quantitated, we next tried to perform flow cytometry analysis of these various constructs. Transformed *B. subtilis* cells harboring various constructs were first propagated in overnight liquid cultures in LB Km10 Cm5 and then re-diluted to OD_{600} = 0.1 in the same medium with 25 mM IPTG and grown under aerobic conditions for 4 hours. Subsequently, cells were diluted 40-fold in 0.2 µM-filtered Isoton II media (Beckman Coulter, Mijdrecht, The Netherlands) and directly measured on a Coulter Epics XL-MCL flow cytometer (Beckman Coulter) operating an argon laser (488 nm). For each sample, at least 10,000 cells were analyzed. Data containing the green fluorescent signals were collected by a fluorescein isothiocyanate filter, and the photomultiplier voltage was set to 800 V. Data were captured using Expo32 ADC cytometry list mode data acquisition and analysis software version 1.1.C. (Beckman Coulter) and further analyzed using Origin version 7 (OriginLab Corporation, Northampton, MA, USA).

RESULTS

Gel-filtration reveals that PAS-A keeps KinA in a more compact state

Size-exclusion chromatography using a Superdex 200 10/300GL column from Pharmacia reports a molecular weight of 163 kDa, roughly consistent with the

dimeric molecular weight previously established for this protein with calculated monomeric molecular weight of 69.1 kDa (Wang, Fabret et al. 2001). Site-directed mutant derivatives showed no changes in elution profile from that of wild type KinA at all concentrations. Therefore, the Y23A mutant can abolish dimerization of PAS-A alone, but this mutation is not sufficient to prevent dimerization in the full-length KinA based on size-exclusion chromatography (**Fig. 6-2)**. Thus, monomerization of PAS-A does not have impact on the overall dimerization. Given that the domains in KinA other than PAS-A each also have dimerization potential (Wang, Fabret et al. 2001), it is improbable that we would have been able to attenuate dimerization to these levels in the full-length construct such that we could affect a change on gel-filtration.

These results show that without PAS-A dimerization, KinA is still dimeric as measured by gel-filtration. Because dimerization is a prerequisite for trans-phosphorylation, we asked why PAS-A dimerization was required for efficient kinase activity (Wang, Grau et al. 1997)? Even if dimerization was attenuated, we remained skeptical that overall dimerization affinity could be attenuated to levels that would affect the local concentrations of individual monomer subunits of the dimeric histidine kinase domain with respect to one another. To be more explicit, we will assume that wild type KinA has nanomolar dimerization affinity given that PAS-A has dimerization affinity of 8 μM (**Chapter 5**) and also that the other domains in KinA have also been shown to have dimerization potential(Wang,

Fabret et al. 2001). Thus, ten or one-hundred fold attenuation of this affinity that we assume occurs from the Y23A mutation should not make a difference to the apparent *local* concentrations of already-dimeric histidine kinase domain subunits such as to affect an over 95% loss in the initial rate of activity. We thus hypothesized then that just the presence of the PAS-A domain keeps KinA in a conformation competent for activity, perhaps by sterically positioning the other domains such that they do not sterically interfere with the kinase domain.

To investigate such possibilities, we purified the PAS-A truncated protein and performed gel-filtration. Somewhat surprisingly, this construct actually appears as a higher molecular weight (~230kDa) species than the full-length KinA on Superdex200 10/300GL (Pharmacia), without eluting in the void volume that often indicates non-specific aggregation (**Fig. 6-2)**. This result may suggest that the presence of the PAS-A domain keeps KinA in a more compact conformational state that may be required for its activity.

Mutations have impact on KinA activity

To characterize the buildup of KinA autophosphorylation in a meaningful manner, we collected samples at 2, 4, 8, 16, 32, and 64 minutes for different concentrations of proteins. Most notably, truncation of PAS-A led to inhibition of KinA activity at all concentrations of protein such that it took long incubations (>8 minutes) with [γ-^{32}P] ATP to observe appreciable amounts of phosphorylation

(**Fig. 6-3**). At higher concentrations of protein, KinA and its mutant derivatives generally observed the similar kinase activities (**Fig. 6-4)**. However, at 4 μM protein, differences could be observed between wild type and mutant KinA constructs. By monitoring the progress of γ-^{32}P incorporation with proteins at this concentration, we could observe that certain mutations enhanced or inhibited autophosphorylation (**Fig. 6-5 and Table 6-1**). The rank order according to highest values of γ-^{32}P incorporation was Y23A>I102A>Wild type>I89A>I89E. To more rigorously quantitate these differences, we charted specific activities by calculating the build-up of γ-^{32}P incorporation over the first eight minutes. Y23A and I102A mutant KinA, with specific activities of 4.0±0.7 and 2.8±0.6 pmole ATP/nmole kinase-min, had increased activity as compared to wild type activity of 2.4±0.8 ρmole ATP/nmole kinase-min.

Table 6-1. Kinase activities of wild type KinA and several mutants.

	Specific Activity (pmole phosphate/nmole kinase-min)
I89E	0.8±0.2
I89A	1.3±0.6
WT	2.3±0.8
I102A	2.8±0.6
Y23A	4.0±0.6

Because the Y23A and I102A mutants cause monomerization of the PAS-A domain (**Chapter 5**), the results suggest that relieving dimerization in the PAS-A domain increases KinA activity, perhaps through the removal of some steric hindrance to fully active kinase function.

On the other hand, I89A and I89E mutant KinA have lower specific activity of 1.3±0.6 and 0.8±0.2 pmole ATP/nmole kinase-min compared to wild type activity (**Fig. 6-5 and Table 6-1**). Unlike for Y23A and I102A mutants, I89A and I89E do not have activities that correlate with elution volume on size-exclusion chromatography, as their elution volumes are intermediate of those for the putatively dimeric wild type and monomeric Y23A mutant. For I89E, a simple explanation for its inhibition of kinase activity is that equilibrium sedimentation reported a tighter dimerization affinity for this construct than that for wild type, consistent with the mechanism of steric hindrance by dimerization mentioned above. However, this explanation does not account for the inhibition observed by the I89A mutant, for which equilibrium sedimentation and gel-filtration report a lower affinity of dimerization. To account for kinase inhibition observed for both I89A and I89E, perhaps it is more relevant to refer to velocity sedimentation results where a shift to an intermediate Svedberg value was observed (**Chapter 5, Fig. 5-3**). This intermediate conformation may be more conducive to inhibition of the full-length KinA. In any case, we conclude from these results that

conformational differences or the monomerization of KinA could coordinate kinase activity in a differential manner.

Differences between mutants observed by colony microscopy (*Experiments performed by Remco Kort of the laboratory of Klaas Hellingwerf at the University of Amsterdam*)

We began this investigation by seeing if we could robustly observe fluorescence as activated specifically by the overexpression of KinA. To do this, we used the pKinA construct where KinA was placed under an IPTG-inducible promoter. These plasmids were then transformed into *Bacillus subtilis* strain IIA-gfp (Veening, Kuipers et al. 2006). This strain has the genes for KinA and KinB deleted and GFP placed under the control of the spoIIA promoter. The use of this specific strain should allow fluorescence as a reporter of KinA activity as it has previously been shown that $KinA^-/KinB^-$ strains do not sporulate (Trach and Hoch 1993). In this manner, KinA activity is directly monitored through a readout that comprehensively includes phosphotransfer throughout the entire sporulation phosphorelay as reconstituted in physiologically active cells (**Fig. 6-6 for schematic)**.

In this manner, we first proceeded to check if our construct observed the background fluorescence activation in the absence of IPTG either due to leaky expression of pKinA or spurious activation by other nascent phosphorylation

pathways. In such a situation, there would always be a basal level of fluorescence that would have to be subtracted from any results. From visual inspection, we concluded that no appreciable background fluorescence could be detected. In striking contrast to these results, the induction of KinA with 25 μM IPTG at the mid-exponential phase of growth (OD_{600}=0.4) in LB Km10 Cm5 medium led to the appearance of cells which produced bright green fluorescence throughout the main body of the colony with only the peripheral circle exhibiting diffuse and sparse fluorescence (**Fig. 6-7a**).

Next, we proceeded to test the *in vivo* effect of truncating PAS-A from KinA, a change that abolishes kinase activity *in vitro*. For purposes of comparison, we grew the appropriate vector-transformed *B. subtilis* cells of both pKinA and a similar construct whereby the PAS-A sequence is deleted from the coding sequence. Both types of cells were induced with 100μM IPTG at the mid-exponential phase of growth (OD_{600}=0.4) in LB Km10 Cm5 medium. From the outset, the PAS-A truncation observed very little fluorescence, while pKinA-transformed *B. subtilis* cells observed very bright fluorescence. After 42 hours of growth, the PAS-A truncation barely managed a faint fluorescent pattern along its peripheral circle (**Fig. 6-7a**).

Thus, the ability of our assay to discern a previously established large difference between wild type KinA and PAS-A truncated KinA *in vitro* kinase activity establishes that this difference also has ramifications in the context of the

entire phosphorelay pathway. We next hoped to test if both subtle differences in PAS-A dimeric conformations or the complete monomerization of PAS-A could also affect KinA activation of the phosphorelay.

To investigate this possibility, we transformed pKinA containing different, single amino acid mutations I89A, I89E, I102A, and Y23A. Cells were again induced with 100μM IPTG at the mid-exponential phase of growth (OD_{600}=0.4) in LB Km10 Cm5 medium. Although the results were not as striking as that of the PAS-A truncation, at earlier timepoints, certain differences could be observed. From visual inspection, the most robust differences could be observed only between 15 to 20 hours post-induction. From Figure 6-7b, we observed that Y23A mutants observed strong fluorescence with I102A a distant second. I89A had no appreciable differences from wild type or I89E, just from visual inspection (**Fig. 6-7b)**. To get a better idea as to the exact differences in fluorescence between different constructs, we quantitated these results using FACS-based analysis.

Comparisons of KinA activity by flow cytometry reveal differences

(*Experiments performed by Remco Kort of the laboratory of Klaas Hellingwerf at the University of Amsterdam*)

In this assay, the fluorescence intensity of single cells is measured using FACS. Profiles for each experiment can subsequently be fit to a Gaussian distribution. Fluorescence activation was inducible in a titratable manner up to 1000 μM IPTG

(**Fig. 6-8a**). In general, at higher IPTG concentrations, most of the Gaussian distributions for the various constructs centered around this same maximum level of fluorescence intensity (**Fig. 6-8a and Table 6-2**). However, PAS-A truncation of KinA notably led to the robust inhibition of KinA activation at all IPTG concentrations. This supports previous reports and our own *in vitro* kinase assays above that the presence of PAS-A is necessary for proper kinase function in KinA (Wang, Fabret et al. 2001).

Furthermore, at 25 μM IPTG, striking differences could be observed between wild type and mutant derivatives (**Fig. 6-8b**). By using the centers of the Gaussian distribution of fluorescence intensities (designated Mxc) as a measure for comparison between various constructs, we found notable differences between various constructs.

Table 6-2. Fluorescence activation values for wild type KinA and mutants at 25 μM IPTG induction.

	Mxc (maximum of Gaussian distribution of fluorescence intensities)
No IPTG	304
ΔPAS-A	376
I89E	450
WT	475
I89A	481
I102A	490
Y23A	570

We were satisfied to observe that without IPTG-induction, pKinA had the smallest Mxc of all at a value of 304, validating the expression of pKinA as necessary for robust phosphorelay activation as measured by this assay. Next, we wanted to determine exactly how important the presence of PAS-A was to KinA phosphorelay activation. In Figure 6-8b, we can see that the PAS-A truncation had a Mxc of 376, also below the Mxc for any construct containing PAS-A, wild type or mutant. However, with IPTG induction, pKinA exhibits a higher Mxc than the PAS-A truncation with a value of 475 (**Fig. 6-8b**). Thus, we were satisfied

that we could quantitate differences in fluorescence between a PAS-A truncation and the pKinA construct.

We next proceeded to quantitate the fluorescence activation of our various site-directed mutants. Y23A observed the highest Mxc value of 570, while I89E observed the lowest Mxc value at 450, in agreement with activity characterization for the above *in vitro* kinase assays. On the other hand, I89A and I102A had Mxcs intermediate of these values at 481 and 490, respectively. From these results, we concluded that conformational differences in PAS-A could coordinate phosphorelay activation in a differential manner.

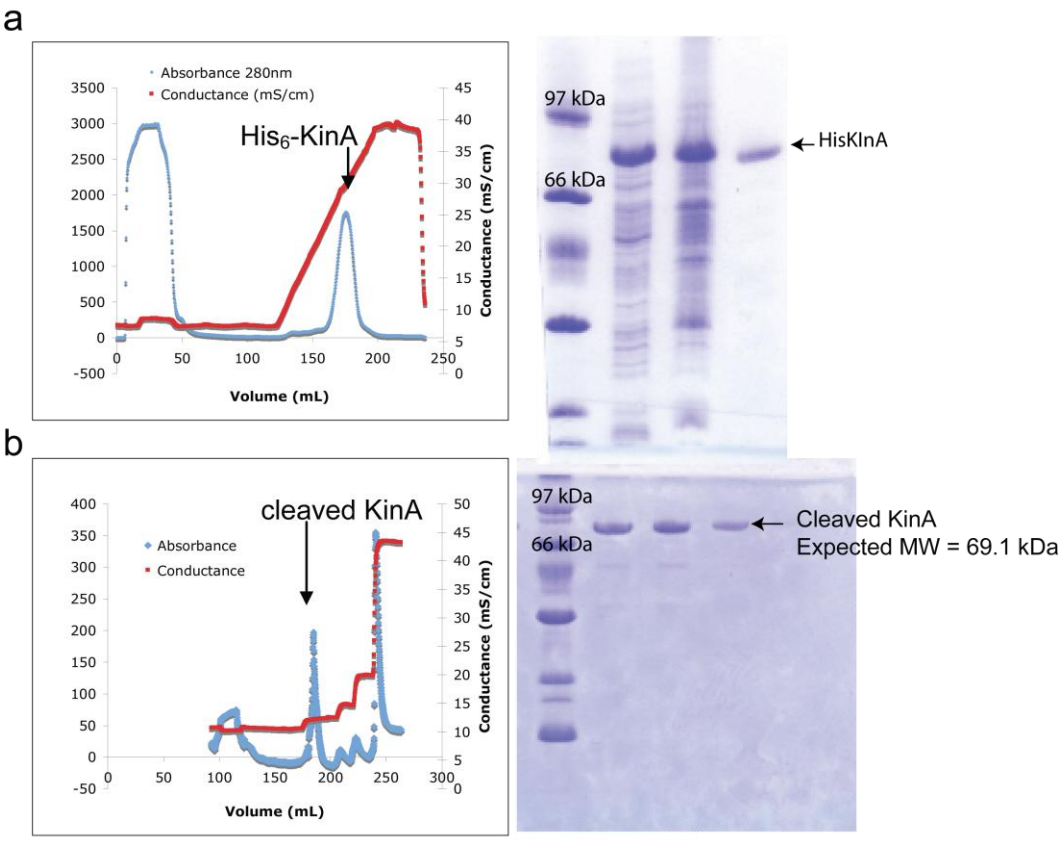


Figure 6-1. Purification of KinA (a) Elution of His_6-tagged KinA protein (aa 1-606) from Ni^{2+}-NTA resin with imidazole concentration as measured by salt conductance (mS/cm). As is common with this purification, the first peak at low imidazole concentrations exhibits flow-through of protein impurities commonly found in whole-cell lysates. The second peak (arrow) indicates where KinA elutes at ~160 mL. SDS-PAGE next to chromatogram shows comparison of insoluble (P) and soluble (S) fraction from whole cell lysate, and the major elution of His_6-KinA at ~160 mL (b) Passage of cleaved KinA through second Ni^{2+}-NTA resin sees KinA bind at low imidazole concentration (~20mM) at ~175 mL while the His_6-tag elutes at much higher imidazole concentration (250mM imidazole) at 250 mL. SDS-PAGE shows the purified product as it elutes from the second Ni^{2+}-NTA resin.

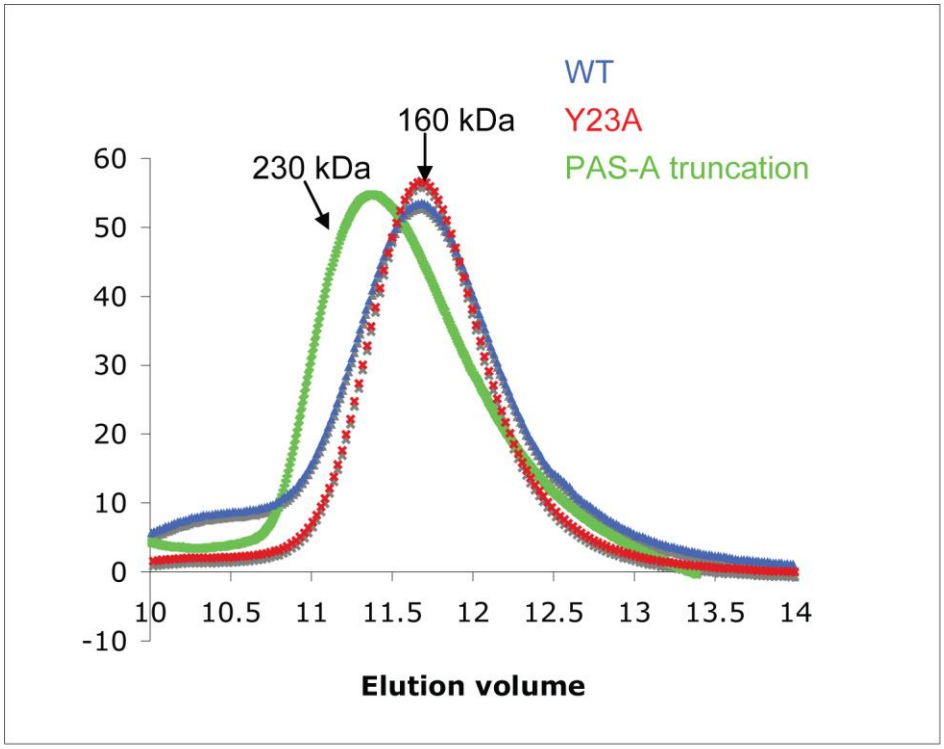


Figure 6-2. Size exclusion chromatography of full-length KinA versus Y23A mutant and PAS-A-truncated KinA. Column used was Superdex200 from Pharmacia. Proteins were injected at 100μM concentration. Void volume of this column is 8.12 mL.

\

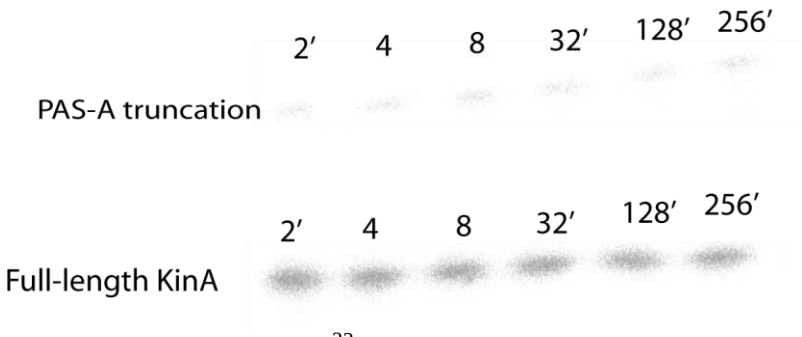


Figure 6-3. Progress of γ-^{32}P incorporation in 16μM PAS-A-truncated KinA versus full-length wild type KinA as monitored by phosphoimaging.

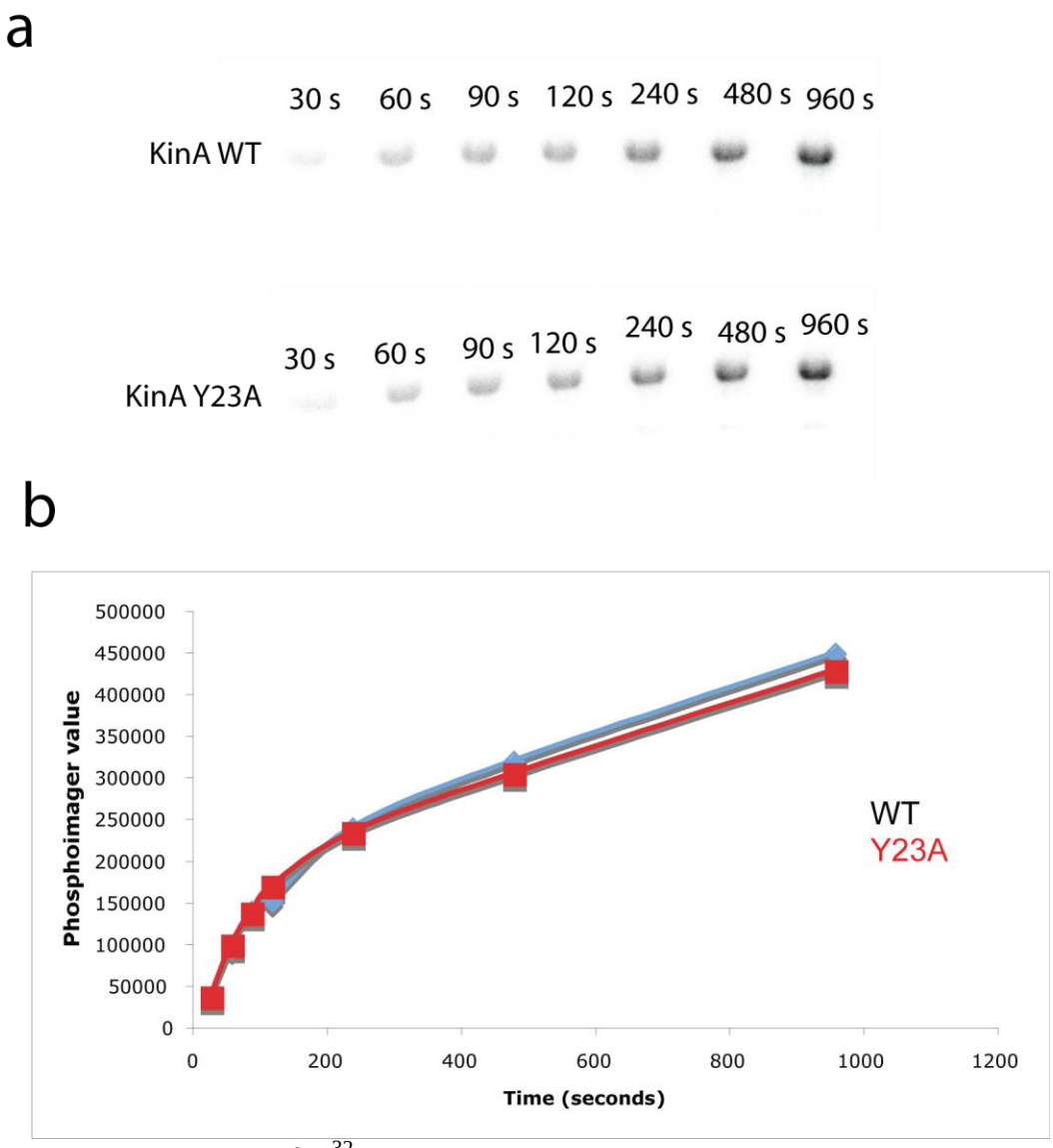


Figure 6-4. Progress of γ-^{32}P incorporation in 16 μM wild type KinA versus Y23A mutant as monitored by phosphoimaging. (a) Phosphoimaging of wild type KinA and Y23A mutant incubated with 400μM [γ-^{32}P]ATP for indicated timepoints followed by quenching by SDS-buffer. (b) Progress curve of quantification of bands from phosphoimaging in (a).

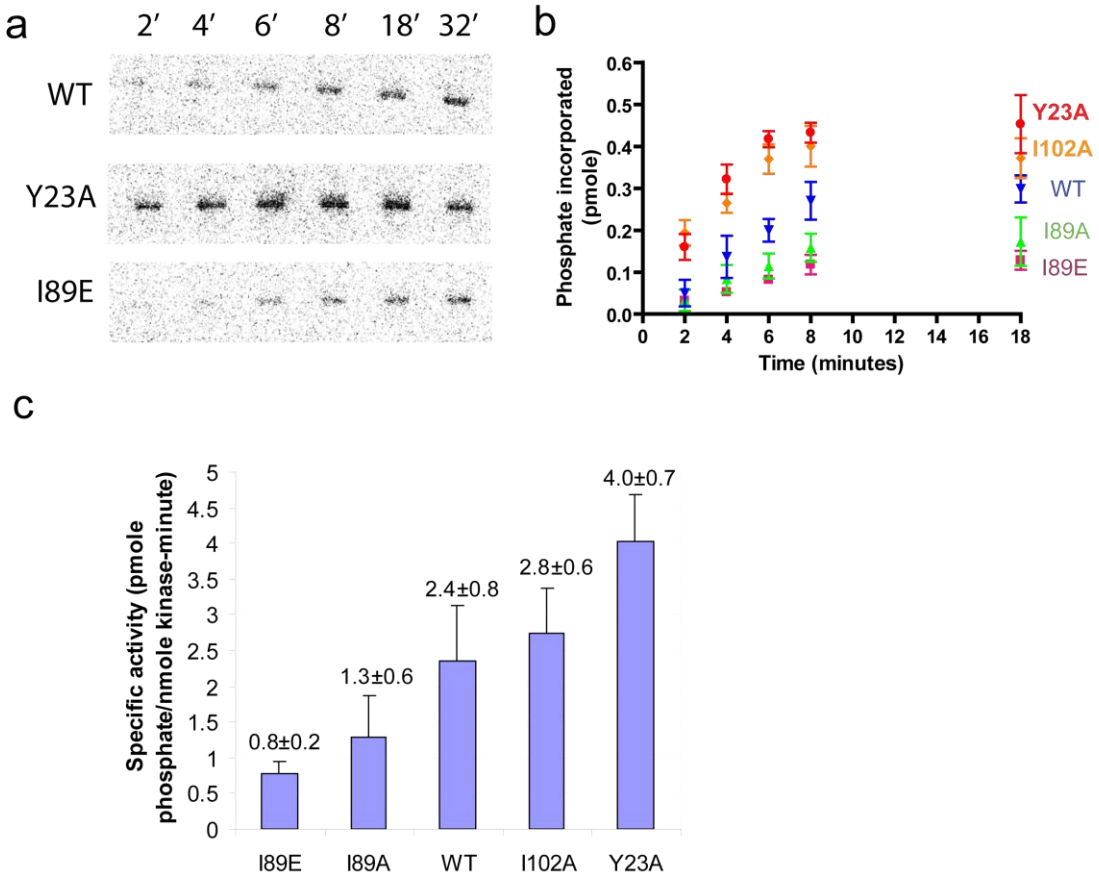


Figure 6-5. Kinase Assay results for KinA PAS-A wild type and various mutants. (a) Raw phosphoimager images. (b) Progress of ATP incorporation with time. (c) Specific activities.

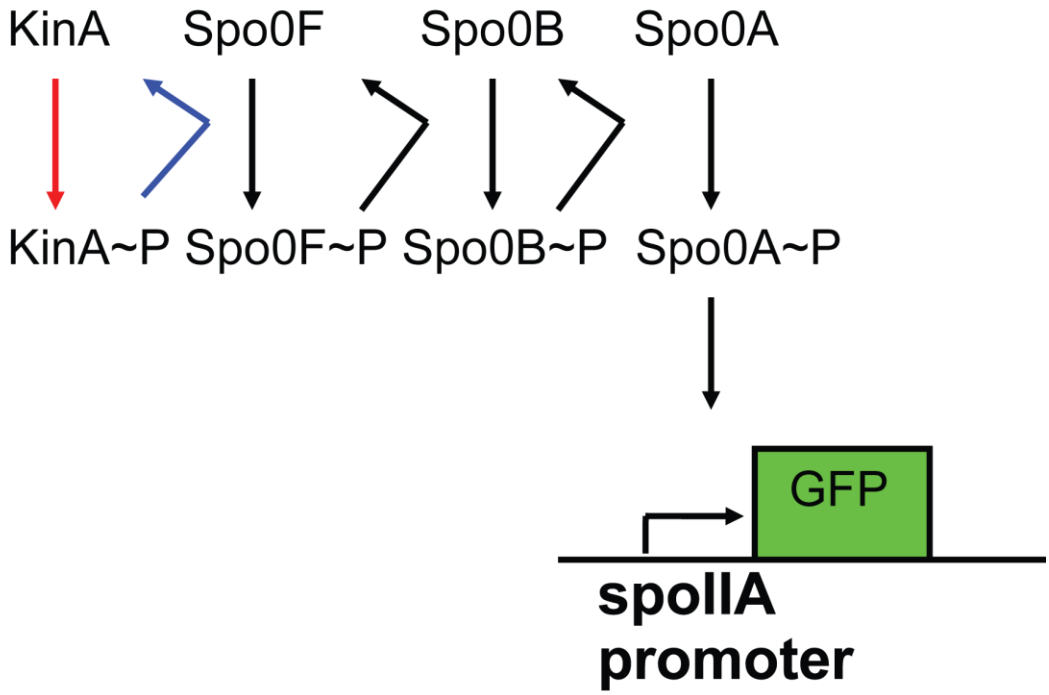


Figure 6-6. Logic of GFP reporter system for KinA activation.

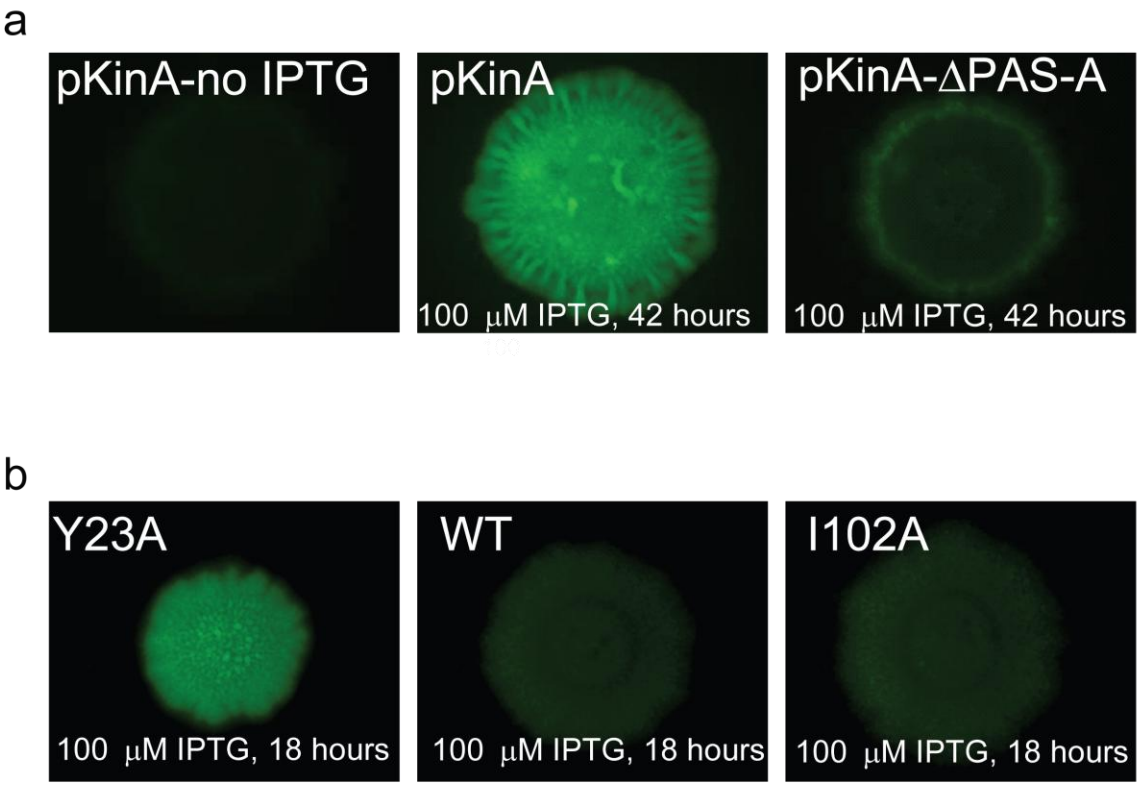
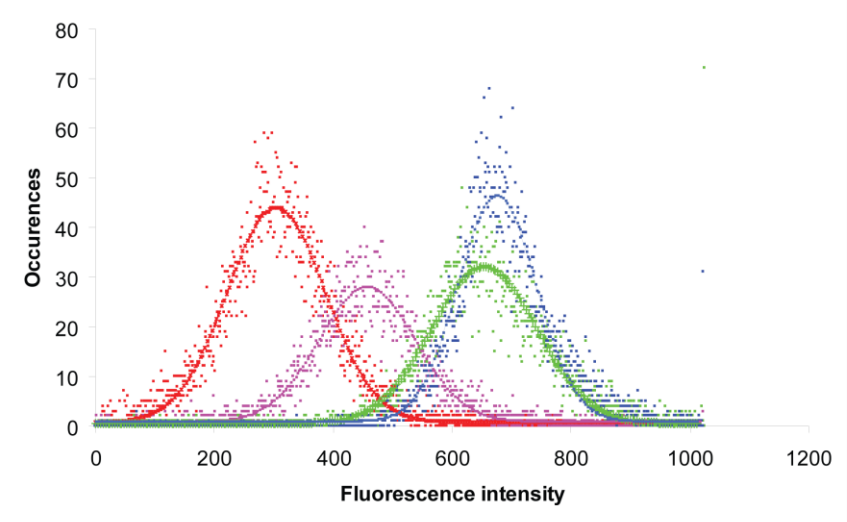


Figure 6-7. Colony microscopy. (a) PAS-A deletion analysis by colony microscopy. Colonies of Bacillus subtilis 1A700 harboring pKinA or pkinA-ΔPAS-A were photographed after 42 hours of incubation at 37 °C. (b) In vivo PAS-A site–directed mutant analysis. Colonies of Bacillus subtilis 1A700 harboring pDG148 containing KinA with different, single amino acid substitutions were photographed after 18 hours of incubation at 37 °C.

a



b

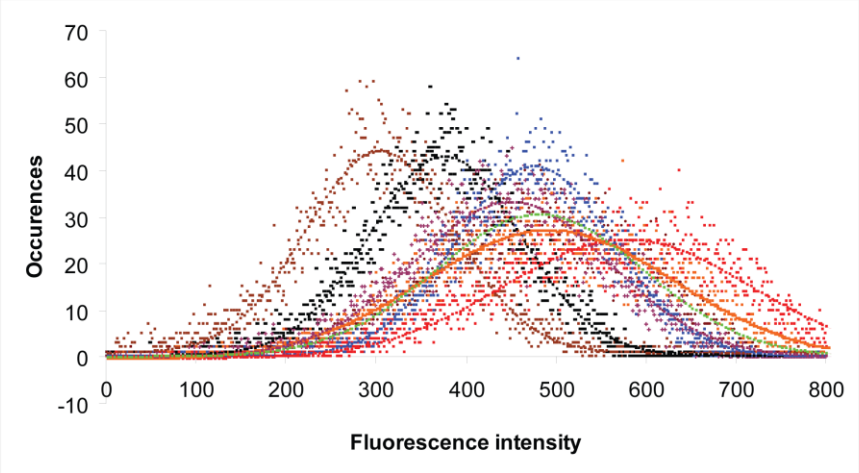


Figure 6-8. Flow cytometry measurements of IPTG-inducible KinA expression in *Bacillus subtilis* PspoIIA-gfp cells. (a) Overnight liquid cultures in LB Km10 Cm5 of *Bacillus subtilis* IIA-gfp harboring pKinA (pDG148-Stu with the full length KinA gene and N-term His_6-tag under control of Pspac) were diluted to OD660 = 0.1 in the same medium with 0, 25, 100 and 1000 μM IPTG and grown under aerobic conditions for 4 hours at 37°C. (b) Deletion and site-directed mutant analysis. Line plots indicate Gaussian fits.

Chapter 7: Conclusions and Future directions

Conclusions

The present work shows how a single PAS domain can modulate the activity of an interconnected effector domain without a direct interaction. Our model system is KinA, a soluble cytosolic protein. In **chapter 1**, we introduce the fact that PAS-A was important in maintaining KinA activity as Wang et al. showed that its truncation affected an over 95% loss in the initial rate of phosphorylation (Wang, Fabret et al. 2001). Thus we set out to establish the means by which we could investigate the molecular mechanism by which PAS-A is essential toward KinA activity. We began by purifying both the histidine kinase domain and PAS-A domain (**Chapter 2**). Each had mono-disperse peaks on gel-filtration columns and dispersed NMR spectra, evidencing that folded domains could be purified and isolated. We next asked if it was possible that PAS-A allosterically activated the kinase domain by directly binding it. The lack of chemical shift changes in titration experiments whereby ^{15}N-labeled PAS-A was titrated with unlabeled kinase domain and vice versa suggested that no such interaction existed, at least *in vitro*. We thus asked if PAS-A binds kinase domain only upon activation by ATP, established by Stephenson et al. as a ligand with micromolar affinity for this domain. To address such possibilities, we titrated ATP into PAS-A NMR

samples. We concluded from the lack of chemical shift changes in this experiment that ATP was not an effector for PAS-A mediated regulation of KinA. As such, we next turned our attention to the differential broadening seen in ^{15}N-^{1}H HSQC spectra for KinA PAS-A, suggestive of intermediate exchange behavior at certain interfaces. We hypothesized that such dynamics was relevant to the regulatory function of KinA PAS-A perhaps due to the binding of an unknown endogenous ligand or conformational exchange of between different signaling states.

To investigate this idea further, we attempted to obtain a high resolution structure for KinA PAS-A. However, the same differential broadening that suggested conformational changes were occurring for this domain also made NMR structure determination unfeasible. As such, in **chapter 3** we solved the x-ray crystal structure for KinA PAS-A to 1.98 Å resolution. In the structure, we observed a possible explanation for the dynamics observed in our NMR structure: two different dimer conformations could be observed in the crystallographic asymmetric unit. This dimeric oligomeric status would be confirmed by both gel-filtration and analytical ultracentrifugation studies. Furthermore, velocity sedimentation profiles report two shapes, one consistent with dimeric mass and the other consistent with mass intermediate between dimer and monomer (**Table 7-1)**. Whether these shapes reflect an equilibrium between monomer and dimer or two different dimeric conformations remains speculative. However, standard velocity sedimentation theoretical treatments precludes the idea that monomer-

dimer equilibria can produce profiles with two different shapes (Gilbert and Jenkins 1956; Gilbert and Gilbert 1973).

Table 7-1. Summary of wild type and mutant properties

	SEC75 Elution Volume	MW calculated from SEC75 (kDa)	R_H (nm) calculated from SEC75	Svedberg Constant	MW derived from Svedberg equation. (kDa)	Dimerization affinity from equilibrium sedimentation (μM)	Specific activity in *in vitro* kinase assay (ρmole ATP/ηmole kinase-Hour)
WT	74.2	31.1	2.41	2.42[a] 1.89[b]	24.0[a] 18.6[b]	8	2.3±0.8
Q36A	75.0	29.6	2.36	ND	ND	ND	ND
I89E	77.3	25.7	2.23	2.07[a] 3.01[b]	19.4[a] 25.2[b]	1	0.8±0.2
I89A	78.6	23.9	2.17	2.02[a] 2.82[b]	18.4[a] 25.6[b]	NA	1.3±0.6
I102A	81.2	20.5	2.04	1.67	12.5	32	2.8±0.6
Y23A	85.2	16.1	1.86	1.55	12.2	monomer	4.0±0.6

Subscripts a and b refers to separate peaks observed on velocity sedimentation with **a** as the major peak and **b** the minor peak.
Subscript (+) is the molecular weight calculated from hydrodynamic radius and Svedberg constant according to the formula: $S = \frac{M(1-\rho\tilde{\nu}_2)}{N_a f}$

We thus questioned the structural differences between the two dimer conformations. Both utilized the same β-sheet interface utilized by other PAS domains as an interaction surface to serve as the dimerization interface. However, the orientation that individual monomers from each dimer conformation adapted with respect to its partner subunit's β-strands was clearly different. Thus, for one dimer conformation, each subunit was oriented in a parallel manner with respect to the axis of the β-strand and hence, we termed this conformation the parallel conformation. The other conformation observed nearly an orthogonal

conformation with respect to this axis and hence, we termed this conformation the crossed conformation. Next we established that the foundation for both dimer conformations was supported by a common network of interactions located centrally to the β-sheet. These interactions are pivoted by the packing of respective Tyr23 residues in each monomer against each other. However, Ile102 also serves an important role in providing stabilizing van der Waal's contacts located just outside Tyr23 for both conformations. The interactions provided by these residues are maintained for both conformations by noticeable re-orientation of their side chains. Mutation of both these residues to alanine would affect striking changes on gel-filtration and analytical ultracentrifugation to suggest that both are monomeric (**Chapter 5**). Furthermore, compared to data for the Y23A mutant, Ile102 had a higher calculated hydrodynamic radius from gel-filtration elution profiles, but produced a reciprocal change in Svedberg constant (**see Table above**). We thus speculated that I102A had residual contacts that preserved some weak associations, producing a Svedberg species with "loose" dimer associations.

We next investigated the role of dimerization in NMR differential broadening and if monomerization of KinA PAS-A could relieve such broadening. Thus, we collected both ^{15}N-^{1}H HSQC and triple resonance NMR spectra for Y23A; the results of these data made possible the assignment of resonances belonging to residues at the dimer interface that were previously

unassignable in the wild type ^{15}N-^{1}H HSQC spectrum. From these data, we concluded that dimerization accounted for intermediate exchange broadening seen in the ^{15}N-^{1}H HSQC spectrum. However, the important question still remained as to which crystallographically-defined conformation was more relevant in solution.

Thus, we questioned the type of interactions that stabilized each crystallographically-defined conformation. Inspection of the change in surface area buried from one conformation to the other revealed that the type of interactions that were stabilizing the parallel conformation--which buries less surface area than the crossed conformation—were found to be hydrophobic. In the structure, we observed a key hydrophobic network in the regions above and below the central β-sheet dimer interface that consisted of Ile89 from one monomer providing van der Waals contacts with Gly37 and Ile40 from the other monomer. In contrast this stabilization by hydrophobic interactions, the crossed conformation appeared to be stabilized by electrostatic interactions. Our conclusion from the structure of the crossed conformation was that side chain reorientations by Glu88, Lys104, and Gln36 drove the formation of new electrostatic interactions between respective monomer subunits.

Thus, both crystallographically-defined conformations appeared to be stabilized by significant and distinctive interactions. However, we hypothesized that if one conformation was more relevant in solution, then mutations intended to disrupt the less important conformation would not lead to significant changes in

NMR spectroscopy, gel-filtration, and analytical ultracentrifugation. To test this idea, we generated both I89A/E and Q36A mutants. Ile89A/E produced gel-filtration elution volumes intermediate between those of putatively dimeric wild type and putatively monomeric Y23A and I102A mutants (**Table 7-1**). In principle then, if this mutant had weakened dimerization, then we would observe chemical shift changes in the same direction as that of Y23A, but at intermediate positions. However, although some resonances observed such patterns, others observed contradictory patterns—suggesting either that conformational changes other than monomerization were occurring and/or the conformational equilibrium had been shifted (**Chapter 4**). Perhaps then the maintenance of the mono-disperse shapes of the gel-filtration elution profiles reflects shape changes that support this idea. Indeed velocity sedimentation (**Chapter 5**) supports this idea; the resulting profiles observe a shift to the intermediate Svedberg species seen in wild type as well as the appearance of a new higher Svedberg species. Using hydrodynamic radii calculated from gel-filtration elution profiles, we calculated that the higher Svedberg species maintained dimeric mass, while the intermediate Svedberg species maintained the intermediate mass seen in wild type velocity sedimentation profiles. Thus, reciprocal changes to both Svedberg constant and hydrodynamic radius maintain the same calculated masses for I89A/E as that for wild type, similar to observations made between Y23A and I102A, suggesting that certain residues of I89A/E maintained dimerization contacts for this mutant.

In this regard, if the parallel conformation did indeed exist in solution, then we wondered if the crossed conformation—which is more thermodynamically favored from a buried surface area standpoint—was simply a crystallization artifact. However, the Q36A not only observed changes to gel-filtration elution volume (**see Table above**), but also to chemical shift resonances (**Chapter 4**). Furthermore, for many residues, the chemical shift resonance change was in a direction opposite from that elicited for the I89A/E mutant. From this, we conclude that both crystallographically-defined dimer conformations existed in solution. However, because the chemical shift changes are minor, we believe that no mutant is capable of shifting conformational equilibria toward just one conformation (**Fig. 4-5)**. Instead, we believe that the crystallographically-defined structures represent either transition states or end-points in an equilibrium of transiently stable structures.

With the establishment that we could shift the conformational equilibria occurring in PAS-A, we next proceeded to ask if these changes had impact on KinA activity and if so, would the changes be striking enough to affect the sporulation phosphorelay (**Chapter 6**). From both *in vitro* kinase assay and *in vivo* fluorescence reporter assays of Spo0IIA activation, we were able to show that truncating PAS-A essentially reduced the initial rate of phosphorylation to levels that are barely detectable, thus reconfirming the results of Wang et al. Mutations Y23A and I102A *increased* activity for both assays in a differential

manner whereby greater activity could be correlated with greater attenuation of dimerization, thus establishing that PAS-A's role in KinA is not simply to dimerize the protein for trans-phosphorylation. Rather, these results suggest that wild type PAS-A dimerization sterically inhibits the kinase domain such as to regulate its activity. In contrast to activation by the Y23A and I102A mutations, I89A and I89E inhibited KinA activity. Because these two mutations are implicated by biophysical experiments detailed above as being able to shift conformational equilibria observed in wild type, we thus suggest that certain conformations of PAS-A in the conformational equilibrium are more inhibitory of KinA activity than others. We also hypothesized that greater inhibition of KinA activity could be affected if we were truly able to only shift equilibrium of dimeric conformations without attenuating dimerization, which is established above in Y23A and I102A mutants as a way to allosterically activate KinA activity.

These results show that at physiological temperatures, PAS-A lies in a fleetingly steady state defined by an equilibrium of conformationally changing dimers. Mutations disrupting specific interfaces elicited specific Svedberg values (**Fig. 5-4**) that could be correlated to various activity levels for KinA activity (**Fig. 6-5c and Table 7-1**). Thus, we conclude that three general different activity states existed for KinA. In its regular wild type state, PAS-A samples several conformations, some of which lead varying levels of constrainment of the kinase

domain such that transphosphorylation is inhibited. The effective average conformation sampled then defines a "basal state" embodied by wild type KinA. However, with the shift to the 1.9 Svedberg species within the PAS-A resulting from mutation of Ile89, the kinase domain becomes populated mainly by an inhibited state. On the other hand, with the appearance of the 1.6-1.7 Svedberg species as seen in the Y23A and I102A mutants, the kinase domain becomes activated, thus defining the "active state" (**Fig. 7-1a**).

From a structural point of view, it would be interesting to consider the distinct modes of dimerization in the KinA PAS-A structure. We propose that activity states are modulated by varying conformations of dimeric PAS-A domains. The structural mechanism for this regulatory mode is probably based on the dynamic stacking of the Tyr23 residue on its partner monomer's corresponding Tyr23 residue. The interface uses this residue as its hinge with concerted side chain reorientations in other interfacial residues guiding the gradual change from the parallel to the crossed conformation. These conformations then modulate the exposure or efficiency of various other regions within KinA.

It is attractive to compare this mechanism of inhibition to that of the ArcB sensor kinase, whereby the weakening of dimerization by the reduction of inter-subunit disulfide bonds allosterically activates the interconnected histidine kinase (Malpica, Franco et al. 2004) (**Fig. 7-1b**). However, this model of linker region

assisted modulation of histidine kinase activity can now also be shown to also exist for a *cytosolic* protein precluding membrane-tethering as a prerequisite. Whether other PAS-containing systems will exhibit similar mechanistic details remains to be seen.

Nevertheless, in this thesis, we are able to present a possible structural mechanism for PAS-mediated allosteric inhibition of downstream kinase domains. This mechanism does not require direct binding of a regulatory domain to the kinase domain as is commonly observed for systems involving kinases. Instead we believe that we have evidence to show that a PAS domain can constrain a linker region in a differential manner that corresponds to the specific conformation adapted by the dimerization interface. Such mechanisms then affect downstream histidine kinase domains such that trans-phosphorylation is ineffective.

The studies described here should greatly assist in deciphering the structural roles PAS domains play in affecting allosteric changes to their cognate effector domains. In addition, a novel mechanism for conformational change in a PAS domain is presented here whereby a dimerization interface is not simply disrupted but rather shifted on the axis between the β-strands of the respective monomers of the dimer.

Exact role of PAS-A?

The exact role of PAS-A in the context of the full-length protein remains unknown. However, it seems likely that conformational changes exhibited by what seems like a very well evolved flexible interface are likely to play a role in the regulation of its activity. Through various point mutations, we were able to elicit several different conformations that were capable of modulating KinA activity in subtly different ways. It remains to be seen whether or not a ligand/nutrient signal can modulate the equilibrium of these various conformations (monomer/crossed/ parallel/ intermediate) to modulate kinase activity.

However, it remains possible that PAS-A does not detect any direct signal. In this regard, the other PAS domains in this protein could be the direct sensors of an environmental signal. Their binding of such a ligand could elicit dramatic conformational changes in the rest of the protein. In this regard, an inflexible dimeric PAS domain would hinder the ability to adopt such conformations. Classically, it has been thought that the existence of multiple PAS domains is for the purpose of signal integration; multiple opposing or synergistic signals can better modulate activity. However, simpler mechanisms remain. The ability of a dimerization domain to maintain its dimeric association even as other domains change the overall conformation of the protein could be necessary to preserve trans-phosphorylation in the histidine kinase.

Future directions

Future experiments should try to address whether KinA PAS-A is the direct signal sensor or simply an adaptor protein to allow flexible dimerization. In this regard, we should screen KinA PAS-A against our in-house developed small molecule ligand library using NMR chemical shift resonance changes as a measure of relative binding. In our library, such screens have produced ligands with micromolar affinity binding for hPASk PAS-A and HIF2α PAS-B (Amezcua, Harper et al. 2002; Erbel, Card et al. 2003). The finding of such compounds for KinA PAS-A should then lead to closer scrutiny of two structural facets. One is whether these ligands attenuate dimerization and the other is whether they shift conformational equilibria of dimeric conformations. As with the experiments detailed above, gel-filtration, NMR, and analytical ultracentrifugation should help address these issues. Upon the determination that these ligands do affect the conformation of KinA PAS-A, we will test whether they have additional implications on KinA activity and phosphotransfer through the sporulation phosphorelay that are measurable by *in vitro* kinase assays and *in vivo* fluorescence reporter assays as detailed in Chapter 6. The results of such experiments will allow us to speculate on whether physiological ligands control KinA activity through conformational changes in PAS-A which allosterically activate or inhibit the histidine kinase domain.

From a more structural standpoint, the conformational equilibria observed in KinA PAS-A suggests that perhaps other conformations exist in solution that could be defined by crystallography. In this regard, the mutants we determined in this thesis intended to disrupt specific dimerization interfaces and shift conformational equilibria may aid in the crystallization of these other structures. Of principal interest would be the determination of structures whereby the β-strands in the central β-sheet for each monomer of a dimer are oriented not parallel as in the parallel conformation or nearly orthogonal as in the crossed conformation, but in an intermediate conformation. The determination of such structures would support the idea that KinA PAS-A can shift through a range of concerted side chain reorientations via a stable base pivoted by Tyr23. Such findings would not only be interesting from a biophysical point-of-view, but would also produce more detailed mechanisms for how PAS domains physiologically regulate downstream histidine kinases.

However, the inability to crystallize such conformations may call for the generation of additional mutations intended to *strengthen* certain interactions, thus further shifting the equilibrium. In this regard, the complete or nearly complete shift to a specific conformation would allow us to investigate in *in vitro* and *in vivo assays* whether one of the crystallographically-defined conformations for PAS-A was indeed the more active or inactive signaling state for the full-length KinA.

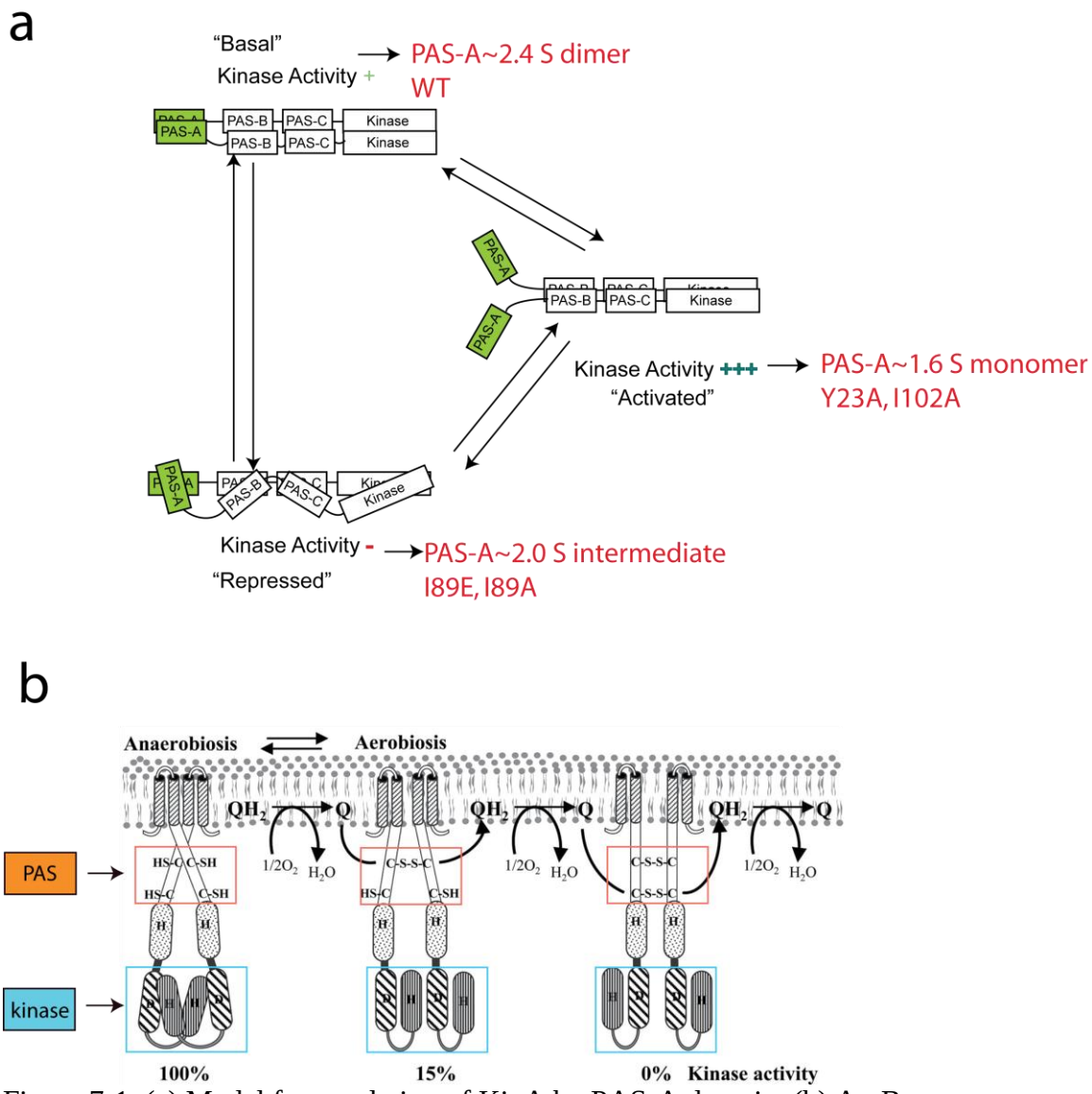


Figure 7-1. (a) Model for regulation of KinA by PAS-A domain. (b) ArcB sensor kinase forms covalent bonds in response to low redox potential. Ensuing conformational changes allosterically inhibits the histidine kinase. Figure adapted from Malpica et al.(Malpica, Franco et al. 2004).

Appendix I: Crystallography of KinA

MATERIALS AND METHODS

Crystallization and data collection

Crystals of KinA PAS-A (aa 10-117) were grown at 20°C using the hanging drop vapor diffusion method. The hanging drops were made by mixing 1 μl of ~7mg/ml protein solution with 1 μl well solution containing 13%–15% (w/v) PEG 10,000, 0.1 M ammonium acetate and 0.1 M bis-Tris (pH 5.5), originally determined from the sparse matrix screen Hampton Research Index Screen 2 (Hampton Research). Crystals of SeMet-labeled protein were obtained under similar conditions, except that the protein solution also contained 20mM DTT. Crystals grew over 4 d with dimensions of 80 X 80 X 100 μm. Before flash-cooling, crystals were transferred into cryo-protective solution composed of well solution and increasing glycerol, to a final glycerol concentration of 25% (v/v). Crystals were then mounted on a nylon loop and flash-frozen in liquid propane. The native and multiwavelength anomalous dispersion data were collected under cryogenic conditions (100 K) at the 19BM and 19ID beamlines, respectively at the Advanced Photon Source. Wavelengths of data collection are indicated in Table 1. Data were processed and reduced using HKL2000 (Otwinowski 1997).

The asymmetric unit of the native KinA PAS-A crystals contains four molecules of KinA PAS-A and has a solvent content of 49%.

Phase determination and structure refinement

For phase determination, the resolution range from 44.76 to 1.97 Å was used, and all 12 of the expected selenium sites were located using the program SHELXD (Schneider and Sheldrick 2002). After phase refinement by MLPHARE and density modification and four-fold averaging by DM *(1994)*, ARP/wARP (Morris, Zwart et al. 2004) was then used to automatically build 79% of the protein model into the experimental electron density. Remaining protein was then built manually using the program O (Jones, Zou et al. 1991). After a round of rigid body refinement using the native data, the resulting model was refined with CNS (Brunger, Adams et al. 1998), using rounds of simulated annealing, energy minimization and group, then later individual B-factor refinement, with interspersed manual revision and model building. The electron density for the protein main chain was clear for residues 9-90 and 98-108 of monomer A and 9-91 and 99-108 for monomer B, residues 9-110 of monomer C and residues 9-62 and 64-108 of monomer D. A Ramachandran plot analysis using the program PROCHECK (Laskowski 1993) showed that 99% of protein residues are in the most favored regions and 0.39% are in additionally allowed regions. Data collection and model refinement statistics are in Table 1. Except for line

drawings, all figures were generated using PyMOL with the APBS plugin (Baker, Sept et al. 2001) for surface electrostatics (http://pymol.sourceforge.net). All protein r.m.s. deviations are derived from Cα superpositions,

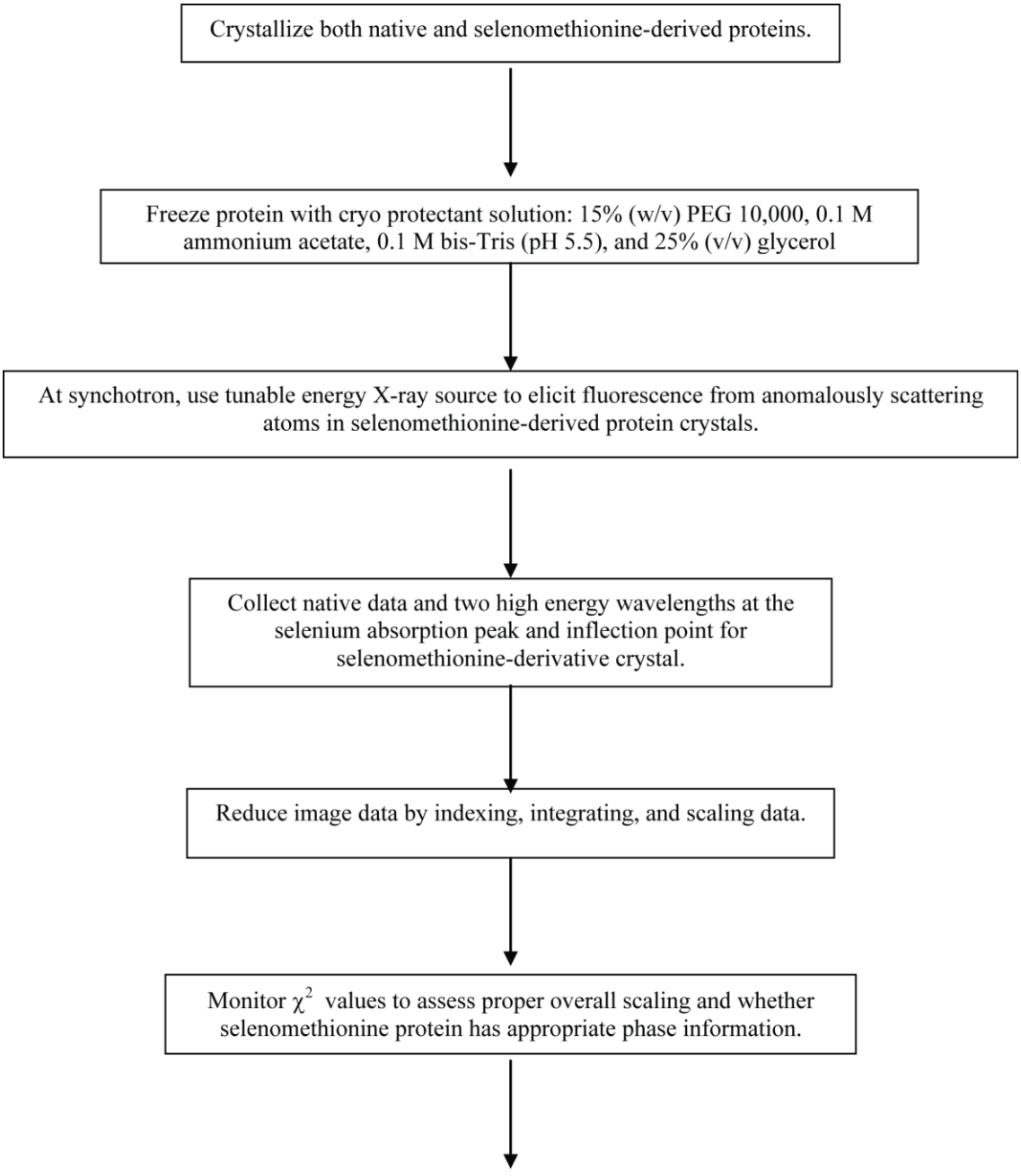
Crystallization as a means to solving protein structures

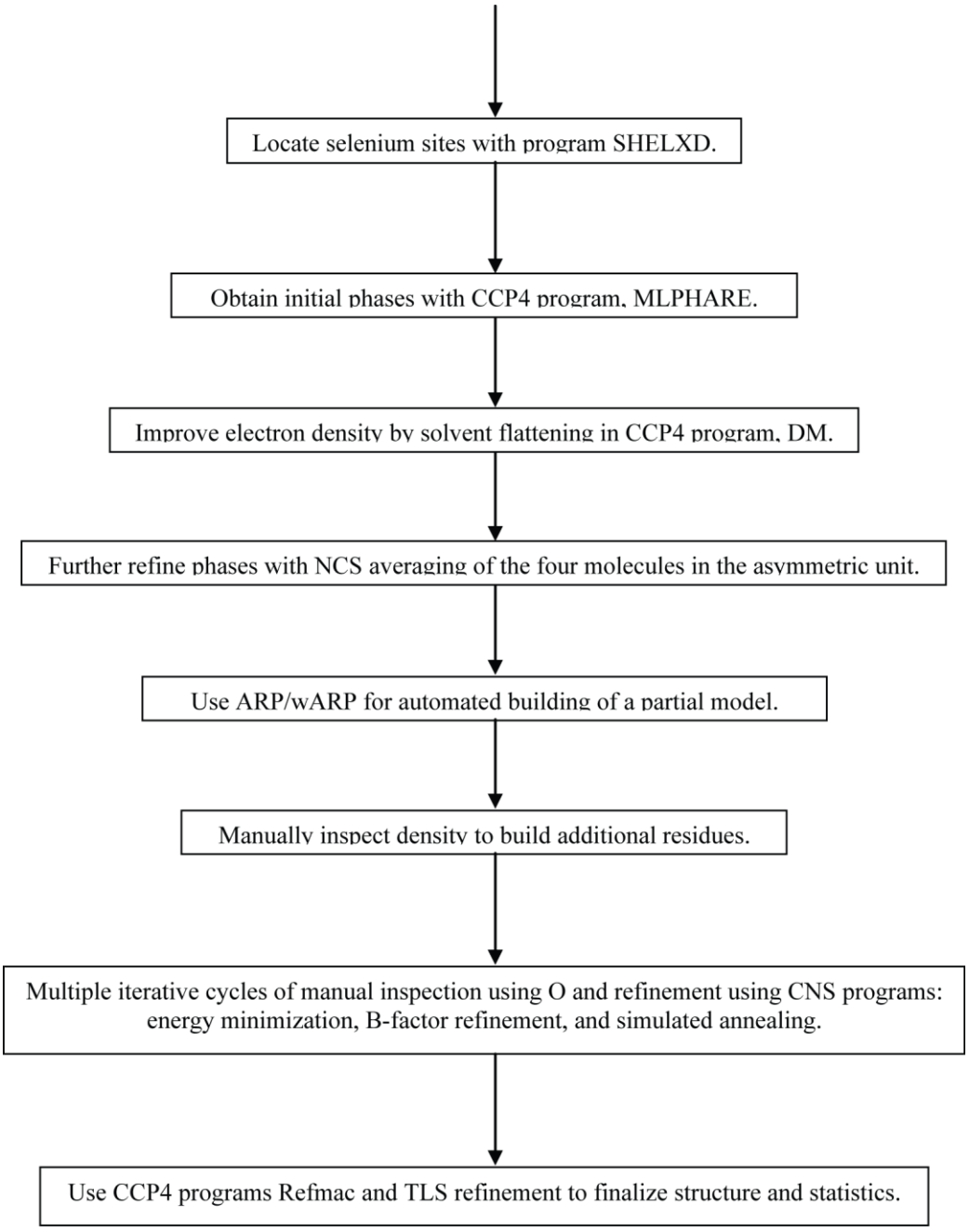
Although X-ray crystallography and NMR are very different in the physics and the types of samples they analyze, both require the use of Fourier transformations to analyze data. From the NMR section in this dissertation, we will see that, in a protein molecule, the Fourier transform is used to describe the complicated sinusoidal waves that are the convolution of free induction decays for specific nuclei. In similar fashion, a diffraction pattern is a Fourier transform that describes the wave function of regular repeating molecules of a crystal. Waves can be used to describe the physical appearance of an object because the addition of more high frequency functions leads to waves that have more intricate features such that even tight corners can be described. Thus, with the summation of enough high frequency functions, even a physical object like a protein structure can be described. Mathematically, this means that the more Fourier terms are included in the Fourier series, the higher the resolution of the electron density map generated.

The following describes the process of solving the crystal structure of KinA PAS-A. It is by no means a comprehensive analysis of crystallography, but

is a good reflection of my understanding of this intricate process. To simplify the process we undertook, I have included a flow chart, but individual steps in the flow chart are detailed afterward.

Diagram. A-1 Flow chart of process undertaken in crystallography




Locate selenium sites with program SHELXD.
Obtain initial phases with CCP4 program, MLPHARE.
Improve electron density by solvent flattening in CCP4 program, DM.
Further refine phases with NCS averaging of the four molecules in the asymmetric unit.
Use ARP/wARP for automated building of a partial model.
Manually inspect density to build additional residues.
Multiple iterative cycles of manual inspection using O and refinement using CNS programs: energy minimization, B-factor refinement, and simulated annealing.
Use CCP4 programs Refmac and TLS refinement to finalize structure and statistics.

Crystallization of KinA PAS-A

Crystals of KinA PAS-A were grown at 20°C using the hanging drop vapor diffusion method (**Fig. 3-1a)**. The hanging drops were made by mixing 1 μl of ~7mg/ml protein solution with 1 μl well solution containing 13%–15% (w/v) PEG 10,000, 0.1 M ammonium acetate and 0.1 M bis-Tris (pH 5.5), originally determined from the sparse matrix screen Hampton Research Index Screen 2 (Hampton Research). Crystals of SeMet-labeled protein were obtained under similar conditions, except that the protein solution also contained 20mM DTT. Crystals grew over 4 d with dimensions of 80 X 80 X 100 μm. Before flash cooling, crystals were transferred into cryoprotectant solution composed of well solution and increasing glycerol, to a final glycerol concentration of 25% (v/v). Crystals were then mounted on a nylon loop and flash-frozen in liquid propane.

Crystal Screening

Screening for well-diffracting crystals and effective cryoprotectants was performed using a Rigaku rotating copper anode X-ray generator. Data collection was done at a Cu-Kα wavelength (1.5418 Å) using an R-Axis IV image plate detector (Rigaku/MSC). Two 1º-oscillation images were collected at two

orientations of the crystal rotated by 90º in order to assess the quality of crystal diffraction, mosaicity, and resolution (**Fig. 3-1b**).

Synchotron data collection at the APS Structural Biology Center's Beamline 19-BM used a 3X3 mosaic CCD detector (SBC3) in the unbinned mode. The oscillation method was used for data collection and crystals were kept under a cold nitrogen gas stream to keep temperatures between –160 to –140ºC. The HKL2000 suite helped in judging the initial data indexing and quality (Otwinowski and Minor 1997). Our $P2_1$ space group necessitated collection of approximately 90º of data. Images were collected using 0.5º oscillations according to the Strategy function of the HKL2000 suite. A ten-fold attenuation of the beam was used to minimize overloading intensity observations on the detector. During data merging, any overloaded observations were disregarded.

The HKL2000 suite of Denzo and Scalepack indexed, integrated, and scaled the data, respectively. Reflections would be identified on a single image such that Denzo's autoindexing algorithm could predict the Bravais lattice of the crystal. Initial estimates predicted a P2 space group that we would later determine also had a screw axis. Using three consecutive images, the following procedures were implemented in order of first to last; refinement of the unit cell parameters, alignment of the rotation matrix of the crystal, adjustment of the distance to detector, positioning of the direct beam and determination of mosaicity. Next, the whole dataset was integrated and scaled with global refinement of parameters

such as unit cell and mosaicity. The scaling of multiple observations of reflections led to their eventual averaging, with measured errors used as weights during averaging. Scaling is necessary during data collection because intensities do not match from one diffraction image to another. Symmetry related reflections, however, must match and as such, these elements are used to normalize the data. To assess proper scaling and also if data is being collected properly in general, chi-squared (χ^2) values are monitored to see if they approximate unity such that the errors that are actually observed match what was predicted simply from scaling. χ^2 is essentially a measure of whether or not a reflection matches its symmetry related reflection. If not, weighting of measured errors in the averaging must be adjusted so that expected errors approximate observed errors. Often, many reflections are completely not weighted or thrown out to make sure χ^2 approach unity. If these adjustments do not correct χ^2, then either the crystal quality or data collection has to be reassessed.

The χ^2 value also gave us a means of assessing the quality of our selenomethionine-derived KinA PAS-A crystal. Preliminary fluorescence scans through the absorption edge of selenium produced weak signals, which is often indicative of poor molecule order in the crystal or low occupancy of the selenomethionine at methionine sites interpreted as being that these molecules were not well ordered in the crystal lattice. As we collected and scaled data on the selenomethione crystal, we monitored the χ^2 values. Because Friedel mates are

not expected to match for these crystals, it was expected that χ^2 values would be high if an anomalous signal was present and we simply treated the data scaling as if it were for native crystals. Based upon the non-unity of these "natively" scale χ^2 values, we were more assured that the MAD experiment would generate the necessary phases.

Data Processing

The native and multiwavelength anomalous dispersion data were collected under cryogenic conditions (100 K) at the 19BM and 19ID beamlines, respectively at the Advanced Photon Source. Wavelengths of data collection are indicated in the following Table 1. Data were processed and reduced using HKL2000 (Otwinowski 1997). The asymmetric unit of the native KinA PAS-A crystals contains four molecules of KinA PAS-A and has a solvent content of 49%.

Heavy atom localization and initial phase estimation

For phase determination, the resolution range from 44.76 to 1.97 Å was used, and all 12 of the expected selenium sites were located using the program SHELXD (Schneider and Sheldrick 2002). Two wavelength multiwavelength anomalous diffraction (MAD) data was used by this program to locate these heavy atoms (Crystallography Table). SHELXD initially looks for some heavy atom sites by using a Patterson method. Knowing the location of a few heavy atoms makes

direct methods searches for the rest of the heavy atoms much easier in what is termed Patterson seeding.

MAD Data Collection

Various methods are possible in phasing X-ray diffraction data. Phasing with multi-wavelength anomalous diffraction (MAD) uses the wavelength-dependence of the atomic structure factor for anomalously scattering atoms. At a certain specific wavelength in the absorption spectra for an atom, there is a precipitous increase in the imaginary component (f'') for scattering of photons.

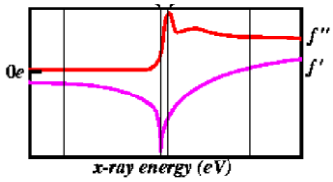


Diagram A-2

(Figure taken from University of Washington Biomolecular Structure web site: http://skuld.bmsc.washington.edu/scatter/)

Lower energy photons cannot excite the electronic transitions for atoms and thus, are re-scattered without a phase shift. However, high-energy photons excite these electron transitions, losing some of their momentum. The result is that fluorescence is emitted when these electronic transitions return to equilibrium with the incident photon becoming phase shifted with an imaginary component (f'') after it is scattered from the atom (Hendrickson 1991). The emitted fluorescence can be detected before crystal data collection as a means of checking for heavy atom incorporation. Along with an imaginary component (f'') to the

phase shift, there must also be a real component (f'). In theory, these two should be related by the Kramer's-Kronig relationship:

$$f'(\omega) = \frac{2}{\pi} \int_0^{\infty} \frac{\omega' f''(\omega') \partial \omega'}{\omega^2 - \omega'^2} \text{ .Eq. A-1}$$

However, in practice, they deviate from this equation because of the interaction of the scattering atom with its neighbors. As such, a typical MAD experiment entails collecting a peak wavelength λ_{peak} that maximizes f" and a second wavelength $\lambda_{inflection}$ that maximizes f', as also performed in our case.

The result of the phase shift induced through the scattering by an appropriate atom such as selenium—as with our experiment—is that Friedel's law no longer applies. That is, in a normal diffraction pattern, peaks that are centrosymmetric have the same intensity and phase. A brief proof using the standard Harker reconstruction for phase analysis is as follows: If one considers that a crystal is collected at a standard wavelength that has no anomalous scattering λ_{std} and another one where there is anomalous scattering λ_{anom}, then the equation representing the vector diagrams is:

$$F_{anom} = F_{std} + \Delta F_r + \Delta F_i. \textbf{ Eq. A-2}$$

Reflecting ΔF_i along the x-axis and reversing its sign as allowed under the rules for vector summation, then one obtains the imaginary structure factor contribution to the Friedel pair mate F_{anom}^-. As such, one can easily see that the F_{anom}^+ and F_{anom}^- are not equal in intensity or opposite in phase.

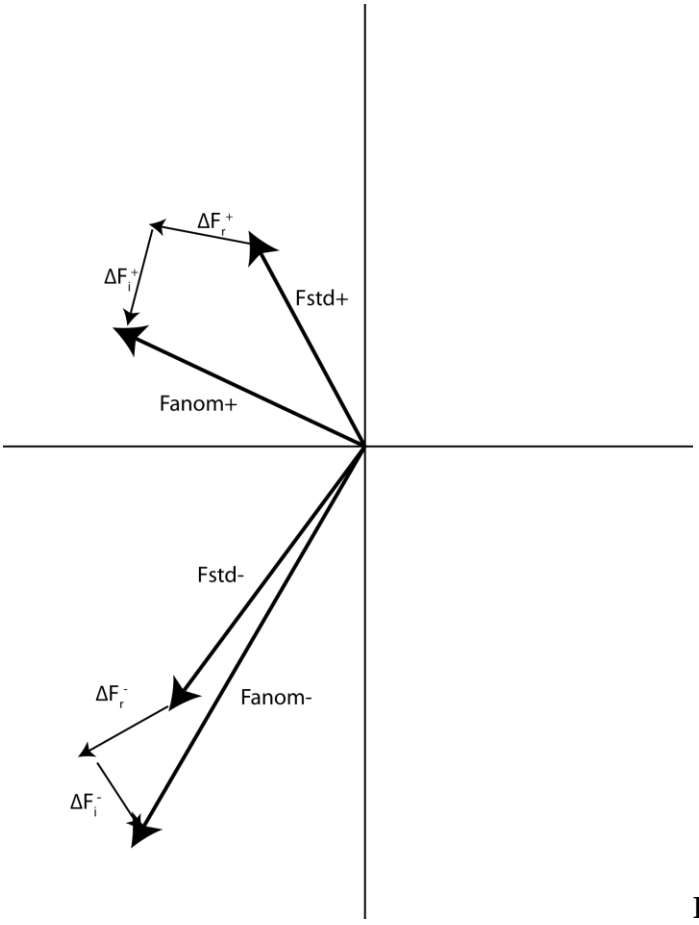


Diagram. A-3

In order to actually extract phase information, one now needs to solve for F_{std}: $F_{std}= F_{anom} - \Delta F_r^{+} - \Delta F_i+$. As such, to actually derive the vector solution, one has to draw the vector $- \Delta F_r$ with its tail at the origin and continue from that head the vector $- \Delta F_{r.}$ Finally, from this arrowhead, one draws a circle with radius F_{anom}. The solution to the phase problem here lies at the intersection of this circle and another circle with radius F_{std} originating from the origin. However, one derives two solutions that make phasing ambiguous.

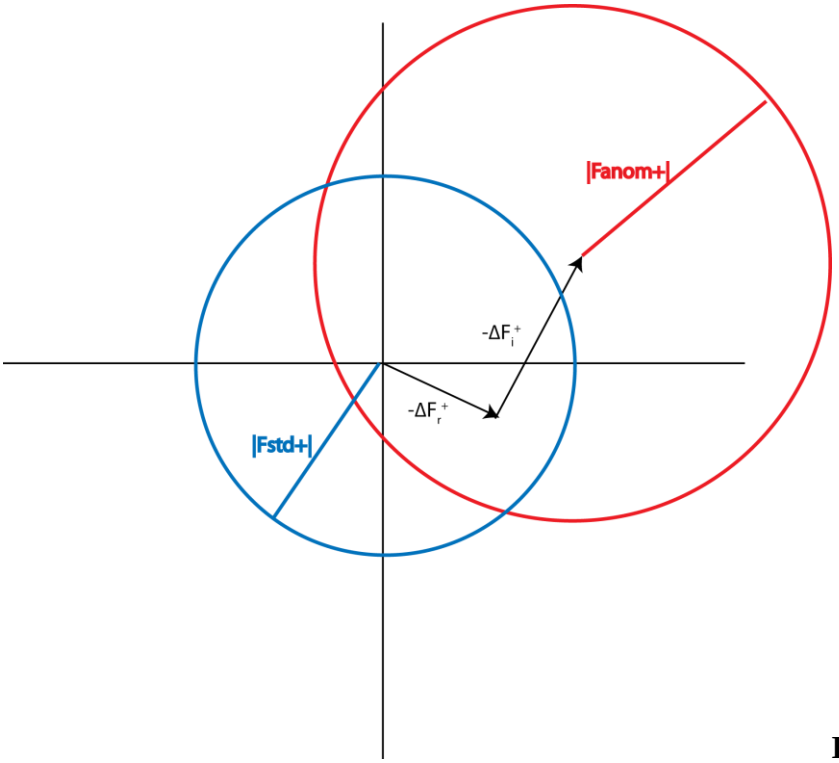


Diagram. A-4

To determine which solution is correct, one now needs to derive a second vector equation involving F_{std}. This is achievable by reflecting F_{anom}^- and all of its vector components across the real axis, which again is possible under the rules for solving vector equations.

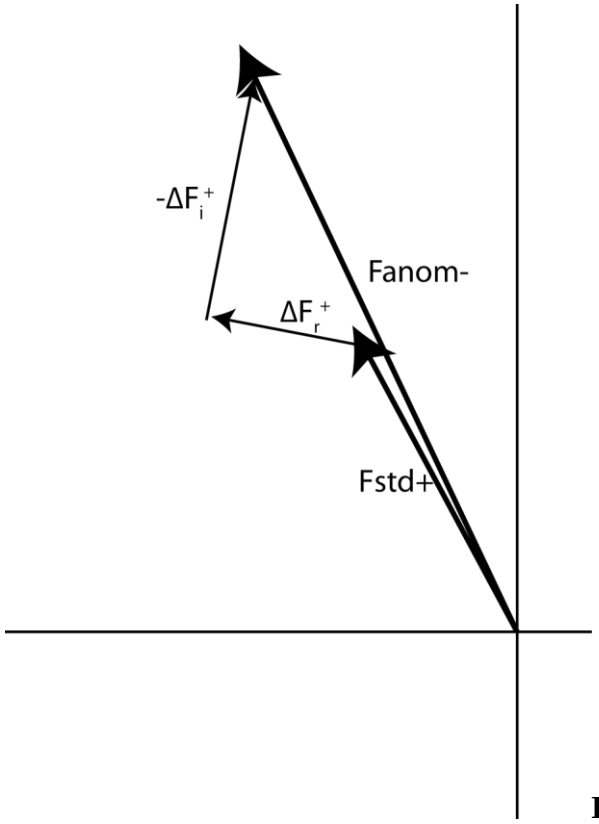


Diagram. A-5

From the diagram, one can see that $F_{std}^{-} = F_{std}^{+}$, $\Delta F_r^{-} = \Delta F_r^{+}$, and $\Delta F_i^{-} = -\Delta F_i^{+}$. Making these substitutions, one derives another equation: $F_{std} = F_{anom}^{-} - \Delta F_r - (-\Delta F_i)$. Again, deriving the vector diagram for this equation, one can derive a circle that also intersects the circle at the origin with radius F_{std}.

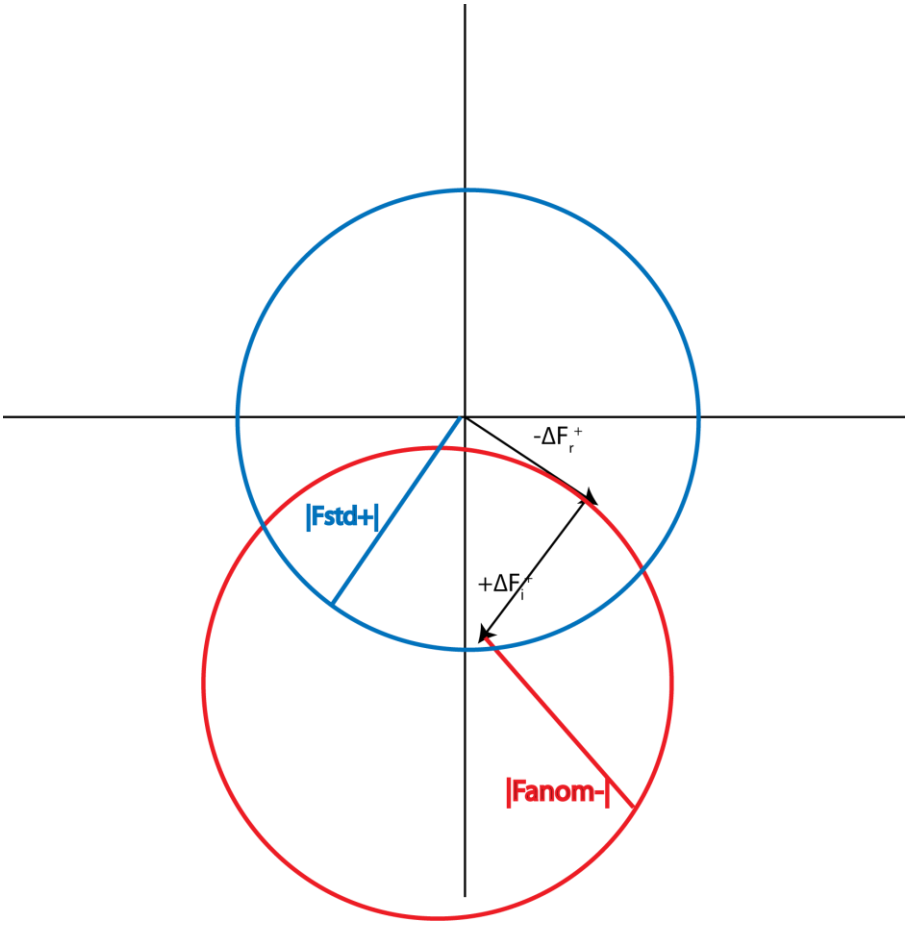


Diagram. A-6

Only one of the two solutions presented here will be the same as the solutions for the first equation involving F_{std} (Hendrickson, Horton et al. 1990; Hendrickson 1991; Rhodes 2000).

In summary, to take advantage of anomalous differences for phasing, one needs to collect data for a heavy atom derivative crystal at least for two wavelengths. First we collected data at an energy of 12661.23 eV that would not scatter photons anomalously. Next we collected data for the same crystal with the

same space group at a peak energy of 12659.57 that maximized anomalous differences. Since the crystal had not decayed by this point, we now collected at an additional energy of 12,657.79 eV that also scattered anomalously to obtain more phase information. This method for phasing heavy atom derivative data is distinct from isomorphous replacement (IR) methods where one collects data on multiple crystals (e.g. a native crystal and a heavy atom derivative crystal). Here one is relying on the additional contribution of a heavy atom to each reflection in the diffraction pattern to obtain a phase estimate (Rhodes 2000). The advantage here is that a tunable X-ray source is not necessary to phase the data for selenium derivative crystals. Nevertheless, even with these two robust methods of obtaining phases for heavy atom derivatives, one still needs to locate the heavy atom sites in the crystal lattice to phase the rest of the atoms. This is often done by Patterson methods.

Patterson Method to locate heavy atom sites

To locate heavy atom sites for small proteins and molecules, it is possible to use direct methods where one guesses the heavy atom positions are located and assesses if the derived phases estimate a reasonable electron density. However, Patterson methods are often a more powerful means for phasing heavy atom derivative data in most cases (Terwilliger and Eisenberg 1983; Weeks, Adams et al. 2003). Because the diffraction pattern represents a regular repeating set of

reflections, it can be represented as a Fourier series in what is termed the Patterson function. These functions are Fourier series without phases where the amplitude of each term is the square of the structure factor. Intuitively, if each reflection in the diffraction pattern represents a Fourier function of its path through many atom positions in the crystal, then the Fourier function of all reflection data represents information on the relative placement of each atom with respect to every other atom. The graphical representation of this is a map of the vectors between every atom. For small molecules, it would be easy to solve the original positions of these atoms based on the deconvolution of all the vectors. However, for large molecules, solving these atom positions becomes tremendously difficult, hence the use of heavy atom derivatives. By taking the Patterson function of the difference in amplitudes of structure factors between a heavy atom derivative crystal and a native crystal, we can derive all the inter-atomic vectors between just heavy atoms, a much simpler computational problem. As such, it is possible to locate the heavy sites and hence, their phases. From these phases, one can easily use Harker reconstructions to derive the phases for every other structure factor using the methods already mentioned above. These calculations are now automatically performed by software.

Phase refinement by Maximum likelihood statistical methods

Initial phases were estimated by the CCP4 program suite maximum likelihood heavy atom refinement and phase calculation (MLPHARE) (anonymous ftp 1994), using a maximum-likelihood mathematical treatment to generate phases (see below). To ensure that refinement is proceeding in a correct manner, R-factors are monitored to make sure they decrease during this process. The R-factor is a measure of the match between data and model as measured by the following terms:

$$R = \frac{\sum [\,|F_{obs}| - K\,|F_{calc}|\,]}{\sum |F_{obs}|} \quad \textbf{Eq. A-3}$$

where K is a scaling factor to ensure that the amplitudes of $|F_{calc}|$ match those of $|F_{obs}|$(Kleywegt and Brunger 1996; Adams, Pannu et al. 1997).

Improvement of protein electron density by solvent flattening

Density modification by solvent flattening was performed by the CCP4 program suite, DM (anonymous ftp 1994). This procedure attempts to add a constant density to the entire map. One begins by first setting negative electron density to zero. The inverse Fourier transform of this new electron density map generates a function. Now one needs to identify a boundary between protein density and solvent density and generate a cone with radius that approximates the protein features. The inverse Fourier transform of this cone generates a specific function. The coefficients of the function describing the new electron density map are

multiplied by this cone function and a new map generated from the adjusted function. The regions of this new map that appear solvent-like are zeroed, effectively producing an envelope that filters out the solvent region of the map.

Density modification by NCS averaging

Because the asymmetric unit of our crystal model contains four monomers, non-crystallographic symmetry (NCS) averaging could be applied. Since the individual subunits of each apparent dimer pair are related in space purely by rotation, a 2X2=4 fold averaging could be performed by DM (anonymous ftp 1994). NCS averaging is powerful in that one is taking advantage of the fact that individual subunits in an asymmetric unit are very similar and thus, make similar contributions to the structure factors and phases. Taking advantage of this fact can effectively increase signal-to-noise ratio. It must be clear that the symmetry must not be between crystallographic symmetry-related molecules and must be related almost purely by complete rotation. However, molecules related by just translation or a combination of rotation and translation can also contribute some additional signal-to-noise. To be more systematic, one can identify these rotation axes by generating Patterson maps. In Patterson maps, which again contains peaks corresponding to the vectors between atoms, crystallographic symmetry related molecules have very large peaks simply because of the number of molecules that pack together in a crystal that are rotated perfectly with respect to one another.

However, non-crystallographic symmetry rotated molecules will also produce self-vectors that are distinctly lower in intensity that those for crystallographic symmetry related molecules. From these distinctly weaker peaks, one can derive proper self-rotation functions that can be used in producing NCS averaging masks.

Maximum-likelihood refinement

The use of maximum-likelihood mathematical treatments by CNS and CCP4 warrants a brief description of this method. More traditional means of refinement are generally based on least-squares minimization of the difference between F_{obs} and F_{calc}. However, this type of minimization attempts to fit the deviations into a Gaussian distribution, thus making unnecessary assumptions about the quality of the data. To be more explicit, as with most scientific observations, phased structure factors should fit a Gaussian distribution. However, what we have in diffraction data are simply intensities without any phase information. As such, one needs to find the likelihood of having a certain phase given the data. With this in mind, the goal of Maximum-likelihood refinement is to obtain the best model that is the most consistent with the observations. The likelihood function is the probability of the data, given the model. The more formal nomenclature for this function is ρ(data; model). However, in empirical science, one really wants to know the probability of the model given the data, ρ(model; data). Fortunately,

Bayes' Theorem—which concerns the relationship between conditional probabilities—can relate the two functions:

$$p(\text{model};\text{data}) = \frac{p(\text{data};\text{model}) X\, p(\text{model})}{p(\text{data})} \quad \textbf{Eq. A-4}$$

The denominator ρ(data) is almost always simply a normalization factor that is not considered computationally. If there is no prior information about protein structures, then ρ(model) can be disregarded and one simply needs to consider ρ(data; model). However, there has been a great deal of knowledge collected regarding protein structures, and thus ρ(model) includes many known stereo-chemical restraints. Therefore, computationally, one takes logarithms of both sides of the equation and try to maximize the sum of log[ρ(model)] and log[ρ(data;model)] (Drenth 1999).

To be more accurate to crystallographic nomenclature, the more formal equation for the Bayesian equation of maximum likelihood phases is:

$$p(F_{calc};|\,F_{obs}\,|) = \frac{p(|\,F_{obs}\,|;F_{calc}) X p(F_{calc})}{p(|F_{obs}|)} \quad \textbf{Eq. A-5}$$

Here, F_{calc} contains known stereo-chemical constraints(McCoy 2002; McCoy 2004). The first advantage of this formalism is that error functions for F_{calc} and $|F_{obs}|$ can be integrated into the formula. These errors functions account for resolution-dependent coordinate error fractions and variance, as well as variance of observed structure factors. If these error models are completely accurate, then

unweighted maximization of the summation of logarithms of $p(|F_{obs}|;F_{calc})$ and F_{calc} should produce refined parameters that converge. However, inherent problems in maximum-likelihood algorithms often lead to the over-fitting of $p(|F_{obs}|;F_{calc})$. To handle this issue, this term is often less weighted in the initial stages of refinement. A CNS input script called "optimize_wa" was used on our part to find the optimal weighting term.

This idea that diffraction data can be overfit necessitated approaches to cross-validation. The R_{free} was created for this purpose. The R_{free} is calculated much like the R-factor except the $|F_{obs}|$ used are on a set of reflections set aside at the beginning of refinement to serve as an independent test, thus minimizing model bias (Bacchi, Lamzin et al. 1996; Kleywegt and Brunger 1996). In our dataset, 5% of the data was set-aside for this purpose. During the process of refinement, this parameter was closely monitored at each step to ensure the appropriate refinement of our model.

Structure refinement

ARP/wArp (Morris, Zwart et al. 2004) an automated model building program, was then used to automatically build 79% of the protein model into the experimental electron density. Remaining protein was then built manually using the program O (Jones, Zou et al. 1991). After a round of rigid body refinement using the native data, the resulting model was refined with CNS (Brunger, Adams

et al. 1998). CNS refinement entailed eight rounds of simulated annealing, energy minimization and group, then later individual B-factor refinement, with interspersed manual revision and model building. The electron density for the protein main chain was clear for residues 9-90 and 98-108 of chain A and 9-92 and 99-108 for polypeptide B, residues 9-110 of polypeptide C and residues 9-62 and 65-108 of polypeptide D. A Ramachandran plot analysis using the program PROCHECK (Laskowski 1993) showed that 99% of protein residues are in the most favored regions and 0.39% are in additionally allowed regions. Data collection and model refinement statistics are in Table 1. Except for line drawings, all figures were generated using PyMOL with the APBS plug-in (Baker, Sept et al. 2001)for surface electrostatics (http://pymol.sourceforge.net).

Simulated Annealing

Simulated annealing is a more widely used powerful algorithm that deserves discussion. As implemented in CNS, it is a maximum likelihood application attempting to lower free energies in various protein parameters such as torsion angles and bond lengths. As such, it is especially beneficial in the early stages of refinement (Brunger, Krukowski et al. 1990). The algorithm generates a maximum likelihood target (likelihood X prior) to produce a Boltzman distribution of energy potentials. From here, it allows molecules in the protein structure to freely move with the energy inherent at some high temperature (4000-

8000K). The slow and gradual decrease of this simulated temperature environment eventually allows molecules to find a global energy minima, rather than the false local minima that they existed prior to this refinement, thus minimizing model bias. We observed that this technique had tremendous impact in the first round of refinement, but its implementation in later rounds resulted in increases in R-factors.

Calculation of electron density maps

In order to visualize the fits between the current model and the data, we calculated σ_A-weighted electron density maps after every round of CNS refinement. These maps were read into the program *O* (Jones, Zou et al. 1991) so that we could compare the current model with the electron density. This is necessary because as we generated a more refined model, phases become more probable in correctness, thus leading to the ability to visualize previously unseen electron density. Two maps regularly created are σ_A-weighted $2m|F_{obs}|-D|F_{calc}|\ e^{i\alpha}$ and $m|F_{obs}|-D|F_{calc}|\ e^{i\alpha}$ maps (Read 1986). Simply put, σ_A, *m*, and D are resolution-dependent measures of coordinate error used to calculate phase probabilities, which again are crucial in maximum-likelihood phase refinement. These parameters are calculated with specific algorithms that use cross-validation data (also used to calculate R_{free}). Whereas the $2m|F_{obs}|-D|F_{calc}|\ e^{i\alpha}$ map gave us an idea of the model as it currently stands, the $m|F_{obs}|-D|F_{calc}|\ e^{i\alpha}$ map contoured at different weighted σ values gave

us an idea of the density that needed to be built or remodeled. Density that appeared at lower, but positive σ contours were very likely to be real density and not just noise. But vice versa, density that appeared at low negative σ contours were likely to be incorrectly built model that had biased electron density map generation such that it appears in $2m|F_{obs}|-D|F_{calc}|\ e^{i\alpha}$ maps.

Finally, CNS allowed us to generate simulated annealing omit maps to better refine certain areas in the model that did not seem to appeared incorrect compared to other standard X-ray crystal structures, but are defined clearly or unclearly by electron density. Again, the current model always biases electron density, but this CNS function allowed us to eliminate phase information from the target region in question, thus allowing simulated annealing using mainly phases from that target region. After nine iterative cycles of CNS refinement, we were able to lower R and R_{free} from 30.59% and 34.55% to 24.36% and 28.18%, respectively.

TLS refinement

As a final means of refining the structure of KinA PAS-A, we tried to account for the translation, libration, and screw-motion (TLS) of individual rigid bodies within the structure. The usual means of describing motion in atoms within a crystal structure is the B-factor, a measure of the mean isotropic atomic displacement. However, anisotropic motion has recently become more heavily

scrutinized as an obstacle to refinement of structures. In any structure, independent rigid-body motions occur by individual groups such as domains, secondary structure elements, or even groups within side chains (Painter and Merritt 2006). The diagram below describes this motion where an individual atom a distance (r) away from the origin can translate a distance (t) and librate (l) an angle λ, with their correlation described by screw-motion (s).

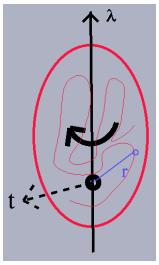


Diagram. A-7

(Diagram taken from Martyn Wynn's TLS tutorial as available for download on the CCP4 website: www.ccp4.ac.uk/courses/IUCr2005/PPT/mdw_tls_talk.ppt)

The TLSMD server at University of Washington (http:// Skuld.bmsc.washington.edu/~tlsmd) (Painter and Merritt 2005; Painter and Merritt 2006) suggested breaking each chain into 11 individual rigid-body groups. The large number of individual rigid-body groups here may underscore the dynamics we observed in both NMR and crystal studies. Finally, TLS refinement

by the CCP4 program Refmac5 affected a drop of R_{free} from 26.6% to 24.4% (**Table A-1**).

Table A-1: Data collection, structure determination and refinement

Data collection			
Crystal	Native	SeMet[a] *peak*	SeMet[a] *inflection*
Energy (eV)	12661.23	12659.57	12,657.79
Resolution range (Å)	50-1.71 (1.74-.71)	50.0-2.00 (2.03-2.00)	50.0-2.00 (2.03-2.00)
Unique reflections	42,220	29,676	29,140
Multiplicity	2.9	6.8	3.1
Data completeness (%)	88	99.3(100)	96.8 (80.7)
R_{merge} (%)[b]	4.1(41.8)	9.5 (26.5)	7.8 (30)
I/σ(I)	24.9 (1.92)	24.8 (7.0)	18.9 (2.8)
Wilson B-value (Å^2)	19.3	25.2	26.4

Phase determination	
Anomalous scatterer	selenium (12 possible sites)
Figure of merit (35-2.00 Å)	0.323

Refinement statistics	
Resolution range (Å)	44.76 -1.97 (2.02-1.97)
No. of reflections R_{work}/R_{free}	28,486/1029 (1806/69)
Atoms (non-H protein/solvent)	3213/383/8
R_{work} (%)	19.1 (20.8)
R_{free} (%)	24.4 (27.1)
R.m.s.d. bond length (Å)	0.018
R.m.s.d. bond angle (°)	1.724
Mean B-value (Å^2)	24.3
Missing residues	Chain A: 91-97 Chain C: None Chain B: 92-98 Chain D: 63-64

Data for the outermost shell are given in parentheses.

[a]Bijvoet-pairs were kept separate for data processing

[b]$R_{merge} = 100\ \Sigma_h\Sigma_i|I_{h,i} - \langle I_h\rangle|/\Sigma_h\Sigma_i I_{h,i}$, where the outer sum (h) is over the unique reflections and the inner sum (i) is over the set of independent observations of each unique reflection.

REFERENCES

Adams, P. D., N. S. Pannu, et al. (1997). "Cross-validated maximum likelihood enhances crystallographic simulated annealing refinement." Proc Natl Acad Sci U S A **94**(10): 5018-23.

Altschul, S. F., T. L. Madden, et al. (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res **25**(17): 3389-402.

Amezcua, C. A., S. M. Harper, et al. (2002). "Structure and interactions of PAS kinase N-terminal PAS domain: model for intramolecular kinase regulation." Structure **10**(10): 1349-61.

Anonymous ftp, D. L. (1994). "The CCP4 suite: programs for protein crystallography." Acta Crystallogr D Biol Crystallogr **50**(Pt 5): 760-3.

Appleman, J. A., L. L. Chen, et al. (2003). "Probing conservation of HAMP linker structure and signal transduction mechanism through analysis of hybrid sensor kinases." J Bacteriol **185**(16): 4872-82.

Aravind, L. and C. P. Ponting (1999). "The cytoplasmic helical linker domain of receptor histidine kinase and methyl-accepting proteins is common to many prokaryotic signalling proteins." FEMS Microbiol Lett **176**(1): 111-6.

Baca, M., G. E. Borgstahl, et al. (1994). "Complete chemical structure of photoactive yellow protein: novel thioester-linked 4-hydroxycinnamyl chromophore and photocycle chemistry." Biochemistry **33**(48): 14369-77.

Bacchi, A., V. S. Lamzin, et al. (1996). "A self-validation technique for protein structure refinement: the extended Hamilton test." Acta Crystallogr D Biol Crystallogr **52**(Pt 4): 641-6.

Baker, N. A., D. Sept, et al. (2001). "Electrostatics of nanosystems: application to microtubules and the ribosome." Proc Natl Acad Sci U S A **98**(18): 10037-41.

Barak, I., E. Ricca, et al. (2005). "From fundamental studies of sporulation to applied spore research." Mol Microbiol **55**(2): 330-8.

Bibikov, S. I., A. C. Miller, et al. (2004). "Methylation-independent aerotaxis mediated by the Escherichia coli Aer protein." J Bacteriol **186**(12): 3730-7.

Borgstahl, G. E., D. R. Williams, et al. (1995). "1.4 A structure of photoactive yellow protein, a cytosolic photoreceptor: unusual fold, active site, and chromophore." Biochemistry **34**(19): 6278-87.

Bracken, C. P., M. L. Whitelaw, et al. (2003). "The hypoxia-inducible factors: key transcriptional regulators of hypoxic responses." Cell Mol Life Sci **60**(7): 1376-93.

Briggs, W. R. and J. M. Christie (2002). "Phototropins 1 and 2: versatile plant blue-light receptors." Trends Plant Sci **7**(5): 204-10.
Bruick, R. K. (2000). "Expression of the gene encoding the proapoptotic Nip3 protein is induced by hypoxia." Proc Natl Acad Sci U S A **97**(16): 9082-7.
Brunger, A. T., P. D. Adams, et al. (1998). "Crystallography & NMR system: A new software suite for macromolecular structure determination." Acta Crystallogr D Biol Crystallogr **54**(Pt 5): 905-21.
Brunger, A. T., A. Krukowski, et al. (1990). "Slow-cooling protocols for crystallographic refinement by simulated annealing." Acta Crystallogr A **46 (Pt 7)**: 585-93.
Cantor, C. and P. Schimmel (1980). "Biophysical Chemistry Part II: Techniques for the study of biological structure and function." W. H. Freeman and Company.
Card, P. B., P. J. Erbel, et al. (2005). "Structural basis of ARNT PAS-B dimerization: use of a common beta-sheet interface for hetero- and homodimerization." J Mol Biol **353**(3): 664-77.
Card, P. B., P. J. Erbel, et al. (2006 (in review)). "Structural basis HIF/ARNT heterodimerization." J Magnetic Resonance (in review).
Card, P. B. and K. H. Gardner (2005). "Identification and optimization of protein domains for NMR studies." Methods Enzymol **394**: 3-16.
Carver, L. A., V. Jackiw, et al. (1994). "The 90-kDa heat shock protein is essential for Ah receptor signaling in a yeast expression system." J Biol Chem **269**(48): 30109-12.
Cavanagh, J. a. F. W. a. P. A. a. S. N. (1996). "Protein NMR Spectroscopy: Principles and Practice ".
Chen, J., A. Zou, et al. (1999). "Long QT syndrome-associated mutations in the Per-Arnt-Sim (PAS) domain of HERG potassium channels accelerate channel deactivation." J Biol Chem **274**(15): 10113-8.
Christie, J. M., T. E. Swartz, et al. (2002). "Phototropin LOV domains exhibit distinct roles in regulating photoreceptor function." Plant J **32**(2): 205-19.
Cohen, S. L., A. R. Ferre-D'Amare, et al. (1995). "Probing the solution structure of the DNA-binding protein Max by a combination of proteolysis and mass spectrometry." Protein Sci **4**(6): 1088-99.
Coligan, J., B. Dunn, et al. (2001). "Current protocols in protein science." John Wiley and Sons, Inc. **1**.
Cooney, P. H., P. F. Whiteman, et al. (1977). "Media dependence of commitment in Bacillus subtilis." J Bacteriol **129**(2): 901-7.
Cornilescu, G., F. Delaglio, et al. (1999). "Protein backbone angle restraints from searching a database for chemical shift and sequence homology." J Biomol NMR **13**(3): 289-302.

Crews, S. (1998). "Control of cell-lineage specific development and transcription by bHLH-PAS proteins." Genes Dev(9): 580-7.
Crews, S. T. and C. M. Fan (1999). "Remembrance of things PAS: regulation of development by bHLH-PAS proteins." Curr Opin Genet Dev **9**(5): 580-7.
Crosson, S. and K. Moffat (2001). "Structure of a flavin-binding plant photoreceptor domain: insights into light-mediated signal transduction." Proc Natl Acad Sci U S A **98**(6): 2995-3000.
Crosson, S., S. Rajagopal, et al. (2003). "The LOV domain family: photoresponsive signaling modules coupled to diverse output domains." Biochemistry **42**(1): 2-10.
Cusanovich, M. A. and T. E. Meyer (2003). "Photoactive yellow protein: a prototypic PAS domain sensory protein and development of a common signaling mechanism." Biochemistry **42**(17): 4759-70.
Delaglio, F., S. Grzesiek, et al. (1995). "NMRPipe: a multidimensional spectral processing system based on UNIX pipes." J Biomol NMR **6**(3): 277-93.
Denison, M. S., A. Pandini, et al. (2002). "Ligand binding and activation of the Ah receptor." Chem Biol Interact **141**(1-2): 3-24.
Drenth, J. (1999). Principles of protein X-ray crystallography **Springer, New York**
Dutta, R., L. Qin, et al. (1999). "Histidine kinases: diversity of domain organization." Mol Microbiol **34**(4): 633-40.
Dux, P., G. Rubinstenn, et al. (1998). "Solution structure and backbone dynamics of the photoactive yellow protein." Biochemistry **37**(37): 12689-99.
Erbel, P. J., P. B. Card, et al. (2003). "Structural basis for PAS domain heterodimerization in the basic helix--loop--helix-PAS transcription factor hypoxia-inducible factor." Proc Natl Acad Sci U S A **100**(26): 15504-9.
Ernst, R. a. B. G., Wokaun A (1987). "Principles of Nuclear Magnetic Resonance in one and two dimensions." Principles of Nuclear Magnetic Resonance in one and two dimensions: 1-610.
Farrow, N. A., R. Muhandiram, et al. (1994). "Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation." Biochemistry **33**(19): 5984-6003.
Farrow, N. A., O. Zhang, et al. (1995). "Spectral density function mapping using 15N relaxation data exclusively." J Biomol NMR **6**(2): 153-62.
Fiaux, J., E. B. Bertelsen, et al. (2002). "NMR analysis of a 900K GroEL GroES complex." Nature **418**(6894): 207-11.
Fujita, M. and R. Losick (2005). "Evidence that entry into sporulation in Bacillus subtilis is governed by a gradual increase in the level and activity of the master regulator Spo0A." Genes Dev **19**(18): 2236-44.

Gardner, K. H. and L. E. Kay (1998). "The use of 2H, 13C, 15N multidimensional NMR to study the structure and dynamics of proteins." Annu Rev Biophys Biomol Struct **27**: 357-406.
Genick, U. K., S. M. Soltis, et al. (1998). "Structure at 0.85 A resolution of an early protein photocycle intermediate." Nature **392**(6672): 206-9.
Gilbert, G. A. and R. C. Jenkins (1956). "Boundary problems in the sedimentation and electrophoresis of complex systems in rapid reversible equilibrium." Nature **177**(4514): 853-4.
Gilbert, L. M. and G. A. Gilbert (1973). "Sedimentation velocity measurement of protein association." Methods Enzymol **27**: 273-96.
Gong, W., B. Hao, et al. (2000). "New mechanistic insights from structural studies of the oxygen-sensing domain of Bradyrhizobium japonicum FixL." Biochemistry **39**(14): 3955-62.
Gong, W., B. Hao, et al. (1998). "Structure of a biological oxygen sensor: a new mechanism for heme-driven signal transduction." Proc Natl Acad Sci U S A **95**(26): 15177-82.
Gronenborn, A. M. and G. M. Clore (1996). "Rapid screening for structural integrity of expressed proteins by heteronuclear NMR spectroscopy." Protein Sci **5**(1): 174-7.
Grzesiek, S. (1992). J Am Chem Soc **114**: 6291-3.
Hamblen, M., W. A. Zehring, et al. (1986). "Germ-line transformation involving DNA from the period locus in Drosophila melanogaster: overlapping genomic fragments that restore circadian and ultradian rhythmicity to per0 and per- mutants." J Neurogenet **3**(5): 249-91.
Hammarstrom, M., N. Hellgren, et al. (2002). "Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli." Protein Sci **11**(2): 313-21.
Harper, S. M., J. M. Christie, et al. (2004). "Disruption of the LOV-Jalpha helix interaction activates phototropin kinase activity." Biochemistry **43**(51): 16184-92.
Harper, S. M., L. C. Neil, et al. (2003). "Structural basis of a phototropin light switch." Science **301**(5639): 1541-4.
Hecker, M., W. Schumann, et al. (1996). "Heat-shock and general stress response in Bacillus subtilis." Mol Microbiol **19**(3): 417-28.
Hefti, M. H., K. J. Francoijs, et al. (2004). "The PAS fold. A redefinition of the PAS domain based upon structural prediction." Eur J Biochem **271**(6): 1198-208.
Hendrickson, W. A. (1991). "Determination of macromolecular structures from anomalous diffraction of synchrotron radiation." Science **254**(5028): 51-8.
Hendrickson, W. A., J. R. Horton, et al. (1990). "Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD):

a vehicle for direct determination of three-dimensional structure." Embo J **9**(5): 1665-72.
Hoff, W. D., P. Dux, et al. (1994). "Thiol ester-linked p-coumaric acid as a new photoactive prosthetic group in a protein with rhodopsin-like photochemistry." Biochemistry **33**(47): 13959-62.
Hoff, W. D., A. Xie, et al. (1999). "Global conformational changes upon receptor stimulation in photoactive yellow protein." Biochemistry **38**(3): 1009-17.
Hoffman, E. C., H. Reyes, et al. (1991). "Cloning of a factor required for activity of the Ah (dioxin) receptor." Science **252**(5008): 954-8.
Huang, Z. J., I. Edery, et al. (1993). "PAS is a dimerization domain common to Drosophila period and several transcription factors." Nature **364**(6434): 259-62.
Huth, J. R., C. A. Bewley, et al. (1997). "Design of an expression system for detecting folded protein domains and mapping macromolecular interactions by NMR." Protein Sci **6**(11): 2359-64.
Jackson, F. R., T. A. Bargiello, et al. (1986). "Product of per locus of Drosophila shares homology with proteoglycans." Nature **320**(6058): 185-8.
Jiang, M., W. Shao, et al. (2000). "Multiple histidine kinases regulate entry into stationary phase and sporulation in Bacillus subtilis." Mol Microbiol **38**(3): 535-42.
Johnson, B. A. (2004). "Using NMRView to visualize and analyze the NMR spectra of macromolecules." Methods Mol Biol **278**: 313-52.
Jones, T. A., J. Y. Zou, et al. (1991). "Improved methods for building protein models in electron density maps and the location of errors in these models." Acta Crystallogr A **47 (Pt 2)**: 110-9.
Joseph, P., J. R. Fantino, et al. (2001). "Rapid orientated cloning in a shuttle vector allowing modulated gene expression in Bacillus subtilis." FEMS Microbiol Lett **205**(1): 91-7.
Kay, L. a. X., GY and Kay LE (1994). J. Magn Reson. Ser. A **109**.
Kay, L. E., D. A. Torchia, et al. (1989). "Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease." Biochemistry **28**(23): 8972-9.
Kleywegt, G. J. and A. T. Brunger (1996). "Checking your imagination: applications of the free R value." Structure **4**(8): 897-904.
Krogh, A., M. Brown, et al. (1994). "Hidden Markov models in computational biology. Applications to protein modeling." J Mol Biol **235**(5): 1501-31.
Laskowski, M., Moss, Thornton (1993). "PROCHECK: a program to check the stereochemical quality of protein structures." J Applied Crystallography **26**: 283-291.
Li, X., J. Xu, et al. (1997). "The human delta1261 mutation of the HERG potassium channel results in a truncated protein that contains a subunit

interaction domain and decreases the channel expression." J Biol Chem **272**(2): 705-8.
Liu, Y. (2003). "Molecular mechanisms of entrainment in the Neurospora circadian clock." J Biol Rhythms **18**(3): 195-205.
Malpica, R., B. Franco, et al. (2004). "Identification of a quinone-sensitive redox switch in the ArcB sensor kinase." Proc Natl Acad Sci U S A **101**(36): 13318-23.
McCoy, A. J. (2002). "New applications of maximum likelihood and Bayesian statistics in macromolecular crystallography." Curr Opin Struct Biol **12**(5): 670-3.
McCoy, A. J. (2004). "Liking likelihood." Acta Crystallogr D Biol Crystallogr **60**(Pt 12 Pt 1): 2169-83.
Meyer, B. K., M. G. Pray-Grant, et al. (1998). "Hepatitis B virus X-associated protein 2 is a subunit of the unliganded aryl hydrocarbon receptor core complex and exhibits transcriptional enhancer activity." Mol Cell Biol **18**(2): 978-88.
Miyatake, H., M. Mukai, et al. (2000). "Sensory mechanism of oxygen sensor FixL from Rhizobium meliloti: crystallographic, mutagenesis and resonance Raman spectroscopic studies." J Mol Biol **301**(2): 415-31.
Morris, R. J., P. H. Zwart, et al. (2004). "Breaking good resolutions with ARP/wARP." J Synchrotron Radiat **11**(Pt 1): 56-9.
Muhandiram, D. a. L. (1994). J Magn Reson Ser B **103**: 203-16.
Nambu, J. R., J. O. Lewis, et al. (1991). "The Drosophila single-minded gene encodes a helix-loop-helix protein that acts as a master regulator of CNS midline development." Cell **67**(6): 1157-67.
Neiditch, M. B., M. J. Federle, et al. (2005). "Regulation of LuxPQ receptor activity by the quorum-sensing signal autoinducer-2." Mol Cell **18**(5): 507-18.
Nishihara, T. (1979). "Growth and sporulation of auxotrophs of Bacillus subtilis in a medium with limited nutrients." Microbiol Immunol **23**(8): 727-34.
Otwinowski, M. (1997). "Processing of X-ray diffraction data collected in oscillation mode." Methods Enzymol **276**: 307-326.
Otwinowski, Z. and W. Minor (1997). "Processing of x-ray diffraction data collected in oscillation mode." Methods in Enzymology **276**: 307-326.
Painter, J. and E. A. Merritt (2005). "A molecular viewer for the analysis of TLS rigid-body motion in macromolecules." Acta Crystallogr D Biol Crystallogr **61**(Pt 4): 465-71.
Painter, J. and E. A. Merritt (2006). "Optimal description of a protein structure in terms of multiple groups undergoing TLS motion." Acta Crystallogr D Biol Crystallogr **62**(Pt 4): 439-50.

Park, H., S. K. Saha, et al. (1998). "Two-domain reconstitution of a functional protein histidine kinase." Proc Natl Acad Sci U S A **95**(12): 6728-32.
Parkinson, J. S. and E. C. Kofoid (1992). "Communication modules in bacterial signaling proteins." Annu Rev Genet **26**: 71-112.
Perego, M., S. P. Cole, et al. (1989). "Characterization of the gene for a protein kinase which phosphorylates the sporulation-regulatory proteins Spo0A and Spo0F of Bacillus subtilis." J Bacteriol **171**(11): 6187-96.
Piggot, P. J. and D. W. Hilbert (2004). "Sporulation of Bacillus subtilis." Curr Opin Microbiol **7**(6): 579-86.
Read, R. (1986). "Improved Fourier coefficients for maps using phases from partial structures with errors." Acta Crystallographica Section A **42**: 140-149.
Reinelt, S., E. Hofmann, et al. (2003). "The structure of the periplasmic ligand-binding domain of the sensor kinase CitA reveals the first extracellular PAS domain." J Biol Chem **278**(40): 39189-96.
Rhodes, G. (2000). "Crystallography made crystal clear (second edition)." Academic Press (Elsevier): 1-269.
Rossmann, M. G., D. Moras, et al. (1974). "Chemical and biological evolution of nucleotide-binding protein." Nature **250**(463): 194-9.
Rost, B. and C. Sander (1993). "Improved prediction of protein secondary structure by use of sequence profiles and neural networks." Proc Natl Acad Sci U S A **90**(16): 7558-62.
Rowland, S. L., W. F. Burkholder, et al. (2004). "Structure and mechanism of action of Sda, an inhibitor of the histidine kinases that regulate initiation of sporulation in Bacillus subtilis." Mol Cell **13**(5): 689-701.
Sadaie, Y. and T. Kada (1985). "Bacillus subtilis gene involved in cell division, sporulation, and exoenzyme secretion." J Bacteriol **163**(2): 648-53.
Sakamoto, K. and W. R. Briggs (2002). "Cellular and subcellular localization of phototropin 1." Plant Cell **14**(8): 1723-35.
Salomon, M., J. M. Christie, et al. (2000). "Photochemical and mutational analysis of the FMN-binding domains of the plant blue light receptor, phototropin." Biochemistry **39**(31): 9401-10.
Schneider, T. R. and G. M. Sheldrick (2002). "Substructure solution with SHELXD." Acta Crystallogr D Biol Crystallogr **58**(Pt 10 Pt 2): 1772-9.
Schuck, P. (2000). "Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling." Biophys J **78**(3): 1606-19.
Schultz, J., F. Milpetz, et al. (1998). "SMART, a simple modular architecture research tool: identification of signaling domains." Proc Natl Acad Sci U S A **95**(11): 5857-64.

Sheffield, P., S. Garrard, et al. (1999). "Overcoming expression and purification problems of RhoGDI using a family of "parallel" expression vectors." Protein Expr Purif **15**(1): 34-9.
Sonnhammer, E. L., S. R. Eddy, et al. (1998). "Pfam: multiple sequence alignments and HMM-profiles of protein domains." Nucleic Acids Res **26**(1): 320-2.
Stephenson, K. and J. A. Hoch (2001). "PAS-A domain of phosphorelay sensor kinase A: a catalytic ATP-binding domain involved in the initiation of development in Bacillus subtilis." Proc Natl Acad Sci U S A **98**(26): 15251-6.
Stephenson, K. and J. A. Hoch (2002). "Evolution of signalling in the sporulation phosphorelay." Mol Microbiol **46**(2): 297-304.
Stephenson, K. and R. J. Lewis (2005). "Molecular insights into the initiation of sporulation in Gram-positive bacteria: new technologies for an old phenomenon." FEMS Microbiol Rev **29**(2): 281-301.
Stock, J. B., A. J. Ninfa, et al. (1989). "Protein phosphorylation and regulation of adaptive responses in bacteria." Microbiol Rev **53**(4): 450-90.
Stragier, P. and R. Losick (1996). "Molecular genetics of sporulation in Bacillus subtilis." Annu. Rev. Genet. **30**: 297-41.
Swanson, H. I., K. Tullis, et al. (1993). "Binding of transformed Ah receptor complex to a dioxin responsive transcriptional enhancer: evidence for two distinct heteromeric DNA-binding forms." Biochemistry **32**(47): 12841-9.
Swartz, T. E., S. B. Corchnoy, et al. (2001). "The photocycle of a flavin-binding domain of the blue light photoreceptor phototropin." J Biol Chem **276**(39): 36493-500.
Taylor, B. L. and I. B. Zhulin (1999). "PAS domains: internal sensors of oxygen, redox potential, and light." Microbiol Mol Biol Rev **63**(2): 479-506.
Terwilliger, T. and D. Eisenberg (1983). "Unbiased three-dimensional refinement of heavy-atom parameters by correlation of origin-removed Patterson functions." Acta Crystallographica, Section A **39**: 813-7.
Trach, K. A. and J. A. Hoch (1993). "Multisensory activation of the phosphorelay initiating sporulation in Bacillus subtilis: identification and sequence of the protein kinase of the alternate pathway." Mol Microbiol **8**(1): 69-79.
Tuckerman, J. R., G. Gonzalez, et al. (2001). "Complexation precedes phosphorylation for two-component regulatory system FixL/FixJ of Sinorhizobium meliloti." J Mol Biol **308**(3): 449-55.
Tugarinov, V., R. Muhandiram, et al. (2002). "Four-dimensional NMR spectroscopy of a 723-residue protein: chemical shift assignments and secondary structure of malate synthase g." J Am Chem Soc **124**(34): 10025-35.

Utsumi, R. and M. Inoue (1991). "[Signal transduction via chimeric receptor (Tar-EnvZ, Taz)]." Tanpakushitsu Kakusan Koso **36**(2): 160-4.
Veening, J. W., O. P. Kuipers, et al. (2006). "Effects of phosphorelay perturbations on architecture, sporulation, and spore resistance in biofilms of Bacillus subtilis." J Bacteriol **188**(8): 3099-109.
Wagner, G. (1997). "An account of NMR in structural biology." Nature Structural Biology **4 supplement**: 841-4.
Wang, L., C. Fabret, et al. (2001). "Dissection of the functional and structural domains of phosphorelay histidine kinase A of Bacillus subtilis." J Bacteriol **183**(9): 2795-802.
Wang, L., R. Grau, et al. (1997). "A novel histidine kinase inhibitor regulating development in Bacillus subtilis." Genes Dev **11**(19): 2569-79.
Weeks, C. M., P. D. Adams, et al. (2003). "Automatic solution of heavy-atom substructures." Methods Enzymol **374**: 37-83.
Wills, P. and D. Winzor (2001). "
Exact theory of sedimentation equilbrium made useful." Progr. Colloid Polymer Sci **119**: 113-120.
Wittekind, M. a. L. (1993). J. Magn Reson Ser. B **101**: 201-05.
Xie, A., W. D. Hoff, et al. (1996). "Glu46 donates a proton to the 4-hydroxycinnamate anion chromophore during the photocycle of photoactive yellow protein." Biochemistry **35**(47): 14671-8.
Yang, Y. and M. Inouye (1993). "Requirement of both kinase and phosphatase activities of an Escherichia coli receptor (Taz1) for ligand-dependent signal transduction." J Mol Biol **231**(2): 335-42.
Yildiz, O., M. Doi, et al. (2005). "Crystal structure and interactions of the PAS repeat region of the Drosophila clock protein PERIOD." Mol Cell **17**(1): 69-82.
Yu, Q., A. C. Jacquier, et al. (1987). "Molecular mapping of point mutations in the period gene that stop or speed up biological clocks in Drosophila melanogaster." Proc Natl Acad Sci U S A **84**(3): 784-8.
Zhou, P., A. A. Lugovskoy, et al. (2001). "A solubility-enhancement tag (SET) for NMR studies of poorly behaving proteins." J Biomol NMR **20**(1): 11-4.
Zhulin, I. B., B. L. Taylor, et al. (1997). "PAS domain S-boxes in Archaea, Bacteria and sensors for oxygen and redox." Trends Biochem Sci **22**(9): 331-3.