2010

# Structural Studies on the Regulation of Sporulation in Bacillus and Antibiotic Sulfation

Matthew J. Bick

Recommended Citation

Bick, Matthew J., "Structural Studies on the Regulation of Sporulation in Bacillus and Antibiotic Sulfation" (2010). *Student Theses and Dissertations.* Paper 72.

# STRUCTURAL STUDIES ON THE REGULATION OF SPORULATION IN *BACILLUS* AND ANTIBIOTIC SULFATION

A Thesis Presented to the Faculty of
The Rockefeller University
in Partial Fulfillment of the Requirements for
the degree of Doctor of Philosophy

by

Matthew J. Bick

June 2010

# Structural Studies on the Regulation of Sporulation in *Bacillus* and Antibiotic Sulfation

Matthew J. Bick, PhD
The Rockefeller University 2010

The microbial world is rich with complexity. Bacteria have evolved several mechanisms to help them survive the harshest environments on earth. These range from complex developmental transformations to the synthesis of small molecule antibiotics that can fend off competing species. Both of these topics will be the focus of this thesis, specifically the regulation of endospore formation in *Bacillus* and the generation of a sulfated small molecule antibiotic from an uncultured soil bacterium. Using standard molecular biology methods and the techniques of X-ray crystallography, synthetic chemistry, and protein semisynthesis, we have:

(1) Solved the X-ray crystal structure of a *Geobacillus stearothermophilus* sporulation histidine kinase in complex with an inhibitor of the kinase reaction (Chapter 2).

(2) Developed an approach using the methods of protein semisynthesis and small molecule organic chemistry that will facilitate the study of active state conformations of a sporulation histidine kinase (Chapter 3).

(3) Developed a full-length sporulation histidine kinase expression and purification system so that we may gain insights into the signaling mechanism of these proteins through crystallography (Chapter 4).

(4) Developed an expression and purification system for an important stage II sporulation protein for future crystallographic studies (Chapter 5).

(5) Solved the X-ray crystal structures of two sulfotransferases from an uncultured soil bacterium in complex with the antibiotic teicoplanin (Chapters 6 and 7).

Dedicated to Elyse Blum, whose love and support during the arduous
writing process was invaluable to my success

# Acknowledgements

I thank Wuixan Shi of beamline X29A at the NSLS, Kanagalaghatta Rajashankar of beamline NE-CAT 24-ID at the APS, and Deena Oren of The Rockefeller University Structural Biology Resource Center. I thank Dr. Elizabeth Campbell of the Darst Laboratory, and Dr. Robert Landick of the University of Wisconsin - Madison for helpful discussions, as well as Drs. Alberto Marina and Wayne Hendrickson of Columbia University for providing the TM0853 encoding plasmid, as well as structural coordinates for the TM0853 autophosphorylation model. I thank Drs. Brian Chait and Tom Muir of The Rockefeller University for use of the mass spectrometry facilities.

I would also like to thank our funding sources. The *Geobacillus stearothermophilus* Genome Sequencing Project is funded by the NSF EPSCoR Program (Experimental Program to Stimulate Competitive Research Grant #EPS-9550478). The work described in this thesis is based in part on research conducted at the NE-CAT beam lines of the APS, supported by award RR-15301 from the NCRR at the NIH. Use of the APS is supported by the U.S. DOE, Office of Basic Energy Sciences, under contract No. W-31-109-ENG-38. X-ray data were also acquired at the X4A and X4C beamlines of the NSLS, BNL, which is operated by the New York Structural Biology Center. This work was supported by NIH GM081697 to S.A.D. The teicoplanin aglycone (Chapter 6) was generously provided by Supelco.

I would like to thank all of my former advisors over the years, Dr. Robert Rose and Chris Lane of the University of Rochester, Dr. Charles Rice of the Center for the Study of Hepatitis C at The Rockefeller University, and especially Dr. Peggy

MacDonald, a member of the Rice lab, whose patience and superb teaching were key factors in my decision to attend graduate school. Much of what I know about structural biology is the result of daily conversations with Dr. Valerie Lamour, a postdoc in the Darst lab. The assistance of many excellent undergraduate and graduate students, Oksana Sergeeva, Wenyan Jiang, Anne El-Shafei, and Katharina Häußermann greatly enhanced my efficiency at the bench.

Of course I would like to thank my thesis committee, Dr. C. Erec Stebbins, Dr. Tom Muir, Dr. Brian Chait, Dr. Wayne Hendrickson, and most importantly my advisor Dr. Seth Darst, who has provided me with sound guidance over the course of my thesis.

The research described in chapters 3, 6, and 7 could not have been possible without the enthusiastic dedication of my collaborators, Dr. Jean-Philippe Pellois of Texas A & M University, Jacob Banik and Dr. Sean Brady of the Laboratory of Genetically Encoded Small Molecules at The Rockefeller University, and especially Dr. Jung-Min Kee of the Muir Laboratory at Rockefeller, whose chemical wizardry helped bring a project back from the dead.

I would also like to thank my parents, Vincent and Judith Bick, for supporting my education, and my brother John Bick for friendship, wine, food, and squash. I would like to thank the developers at Bioware for offering diversions from the thesis so I could explode the heads of unsuspecting darkspawn. A large no thanks goes out to the utterly useless fools at Microsoft for thrusting Word upon the world. And of course this thesis would not have been possible without the constructive criticism of Mr. Krinkle, Petunia-Begonia, and Moofa, whose insightful suggestions included “7yuuuuuu7”, “adfb09aerh”, and “gggggggggggggjjjjjjjjjjjjjjjjjjj”.

# Table of Contents

# List of Figures

# List of Tables

# CHAPTER 1
## Introduction

### Bacterial Sporulation

---

Specialized differentiated cell types are used by many species of bacteria to help them cope with harsh environmental conditions. Endospores produced by a number of gram-positive bacteria from the Firmicutes, including the genera *Bacillus* and *Clostridium*, represent the hardiest class of these specialized cells. When bacteria sense that environmental conditions have become non-conducive for growth and reproduction, they will commit themselves to the process of sporulation. Bacteria sporulate in order to store their genetic material as an archived copy. As early as 1890 it was suggested that the onset of sporulation correlated with a decreasing concentration of available nutrients in the environment (Buchner, 1890). We now know that the primary signal for entry to sporulation is a lack carbon, nitrogen, and even sometimes phosphorus sources, or in other words, starvation (Piggot and Coote, 1976). It also appears that a high cell population density is a requirement for sporulation. In a resource deficient environment that is otherwise conducive to spore formation, bacteria will nevertheless fail to sporulate at low population densities (Grossman and Losick, 1988; Ireton et al., 1993). Secreted peptides accumulate in the extracellular medium as a bacterial population increases. When these peptides reach a certain threshold concentration, they interact with cell surface receptors and inform the bacteria that the environment can support sporulation (Waldburger et al., 1993).

Endospores have been studied since the earliest days of the field of microbiology. Robert Koch and Ferdinand Cohn simultaneously published the first morphological descriptions of *Bacillus anthracis* and *Bacillus subtilis* endospores in 1876 (Cohn, 1876; Koch, 1876). Koch's 1876 paper that described *B. anthracis* sporulation also led to the first confirmation of a specific organism as the cause of a specific disease, anthrax. Cohn's work focused more on the basic biology of sporulation, rather than its medical implications. Cohn had made the observation that almost all bacteria could effectively be killed at temperatures above 80 °C, but found that *B. subtilis* was an exception to this rule. His discovery of heat resistant endospores, which survived boiling and could resume bacterial growth when cooled, helped to disprove a prevailing theory at the time, spontaneous generation. In all of his experiments on heat sterilized organic materials, *B. subtilis* was the organism that survived sterilization. Cohn also noted that vegetatively growing bacteria, when moved to an environment inhospitable for growth, could switch from the vegetative to a spore-like state (Bulloch, 1938). To this day *B. subtilis* remains the most thoroughly studied of the endospore-forming bacteria, with *Clostridium* a distant second and other less well studied genera (*Thermoactinomyces*, *Sporolactobacillus*, *Sporosarcina*) filling out the remainder of spore research.

The ability to form endospores is thought to have arisen in the pre-oxygen era from a common ancestor of *Bacillus* and *Clostridium*, and became a characteristic of aerobic bacteria when *Bacillus* diverged from *Clostridium*. *Bacillus* and *Clostridium* have many sporulation regulatory proteins in common (Sauer et al., 1994). The fact that only one known truly gram-negative bacterial species, *Sporohalobacter*, has the ability to form endospores suggests that gram-positives acquired the ability to sporulate at some point after the divergence of gram-positive and most gram-negative species (Sonenshein, 2000b).

Many endospore-forming bacteria are important human pathogens. The *Clostridium* species *botulinum* and *tentani* both cause horrific human disease. *Clostridium perfringens* is the third leading cause of food poisoning in the United Kingdom and United States, and also the agent behind gas gangrene (Warrell, 2003). A weaponized, highly dispersible form of *B. anthracis* was used in the 2001 anthrax letter attacks in the United States. The ability of *B. anthracis* to produce tough endospores appears at least in part to play a role in its pathogenicity (Spencer, 2003).

Endospores are extremely tough. They can survive environments at the physical extremes of those found on earth, and even beyond. Because of their durability, spores have even been used to test the viability of life in space (Lorenz et al., 1968; Lorenz et al., 1969). Endospores are resistant to temperatures above 100 °C, many chemical solvents, ultraviolet and gamma radiation, hydrolytic enzymes, extreme desiccation including exposure to vacuum, and oxidizing agents (Nicholson et al., 2000). Physical structures of the spore such as the cortex and coat, which are not present in the vegetative cell, collectively contribute the spore's remarkable resistance (Figure 1.1). For example, during sporulation a unique group of proteins called $\alpha/\beta$-type small acid-soluble spore proteins (SASP) is synthesized in the forespore. SASP saturate the spore chromosome and act to protect the DNA from a variety of chemicals and mutagenic agents. When the spore germinates, it digests the SASP into free amino acids, which in turn are used for new protein synthesis (Setlow, 1988).

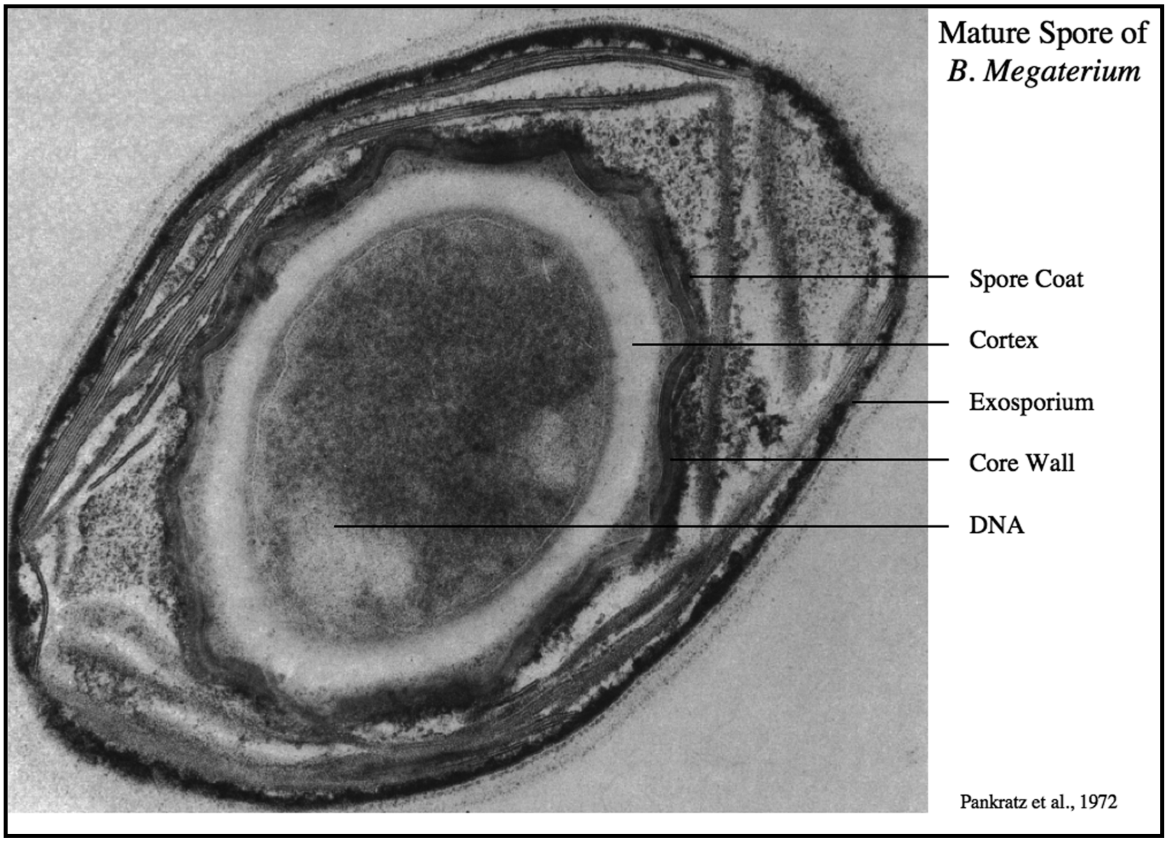


**Figure 1.1)** Electron micrograph cross-section showing the main structural features of a *Bacillus megaterium* endospore. All endospore forming species make a spore coat, core wall, and cortex, while the exosporium is specific to certain species. Adapted from (Beaman et al., 1972).

Sporulation may also serve as a strategy for relocating bacteria to an environment where nutrients are more abundant. Frequently, spores of a particular species are found in locations obviously removed from their point of origin. They can be dispersed by wind, flowing water, and living hosts to locations all over the earth, both above and below the surface (Priest, 1993). Spores exhibit incredible longevity. There have been reports of the revival of spores from 100,000 years ago (Gest and Mandelstam, 1987). In 1995, a somewhat controversial report detailed the recovery of *Bacillus sphaericus* spores from the gut of a bee fossilized in Dominican amber that were 25 - 40 million years old

(Cano and Borucki, 1995). The most incredible discovery of an ancient spore-forming species came in 2000, when Vreeland et al. reported the viability of a *Bacillus* species, designated 2-9-3, from a brine inclusion within a 250-million-year-old salt crystal from the Permian Salado Formation in New Mexico. Though the authors did not have direct evidence that the isolated *B. 2-9-3* were endospores, the fact that conditions of high salinity promote sporulation *in vitro* at least suggests that the bacteria were able to survive for such an extended period of time by existing in the endospore form (Vreeland et al., 2000).

However many advantages sporulation may impart on a bacterial species, there are disadvantages to the process as well. By the time the bacterium has reached stage II of sporulation (see below) it is fully committed to the process; there is no turning back (Dworkin and Losick, 2005; Freese, 1972; Parker et al., 1996). In addition to the time it takes to make a spore (~8 hr), germination of the spore and the return to the vegetative state is also a lengthy process. If, while the bacterium is in the midst of sporulating, the environmental conditions turn favorable again, competing bacterial species may come to dominate the local environment in terms of number and utilization of resources. This could explain why sporulation depends in part on a high population density, to ensure that there are enough of the same species around before bacteria cut themselves off from potential resources for such a long period of time (Sonenshein, 2000b).

Because the process is so energy intensive and potentially detrimental to the bacterium, sporulation is considered a survival mechanism of last resort (Piggot and Losick, 2002; Sonenshein, 2000a, b). In fact, *Bacillus* has developed a number of ways to put off sporulating for as long as possible. For example, *B. subtilis* that have entered the sporulation pathway produce an intercellular signaling protein that delays sporulation in sister cells. These same sporulating cells also export a killing factor, to which the spore formers are immune, that functions cooperatively with the delaying factor to lyse sister cells. The sporulating bacteria are then able to feed from the nutrients released by the lysed sister cells, and reverse their own sporulation (Gonzalez-Pastor et al., 2003). Other methods for staving off sporulation exist as well, including the production of degradative enzymes (proteases, nucleases, amylases) and antibiotics to kill competing bacteria in the local environment, the onset of cell motility to search for nutrients in alternate locations, and the uptake of exogeneous DNA through genetic competence (Grossman, 1995). Sporulation is also intimately tied to the cell cycle; cells can only make the transition from vegetative growth to sporulation after the initiation of DNA replication.

*Bacillus* passes through several distinct morphological stages on its journey to becoming an endospore (Figures 1.2 and 1.3). Numerous studies have documented and categorized the morphology of these stages (Kay and Warren, 1968; Ryter, 1965; Szulmajster, 1964; Young and Fitz-James, 1959a, b). The designation into seven distinct morphological stages of sporulation put forth by Ryter in 1965 has withstood the test of time (Ryter, 1965).

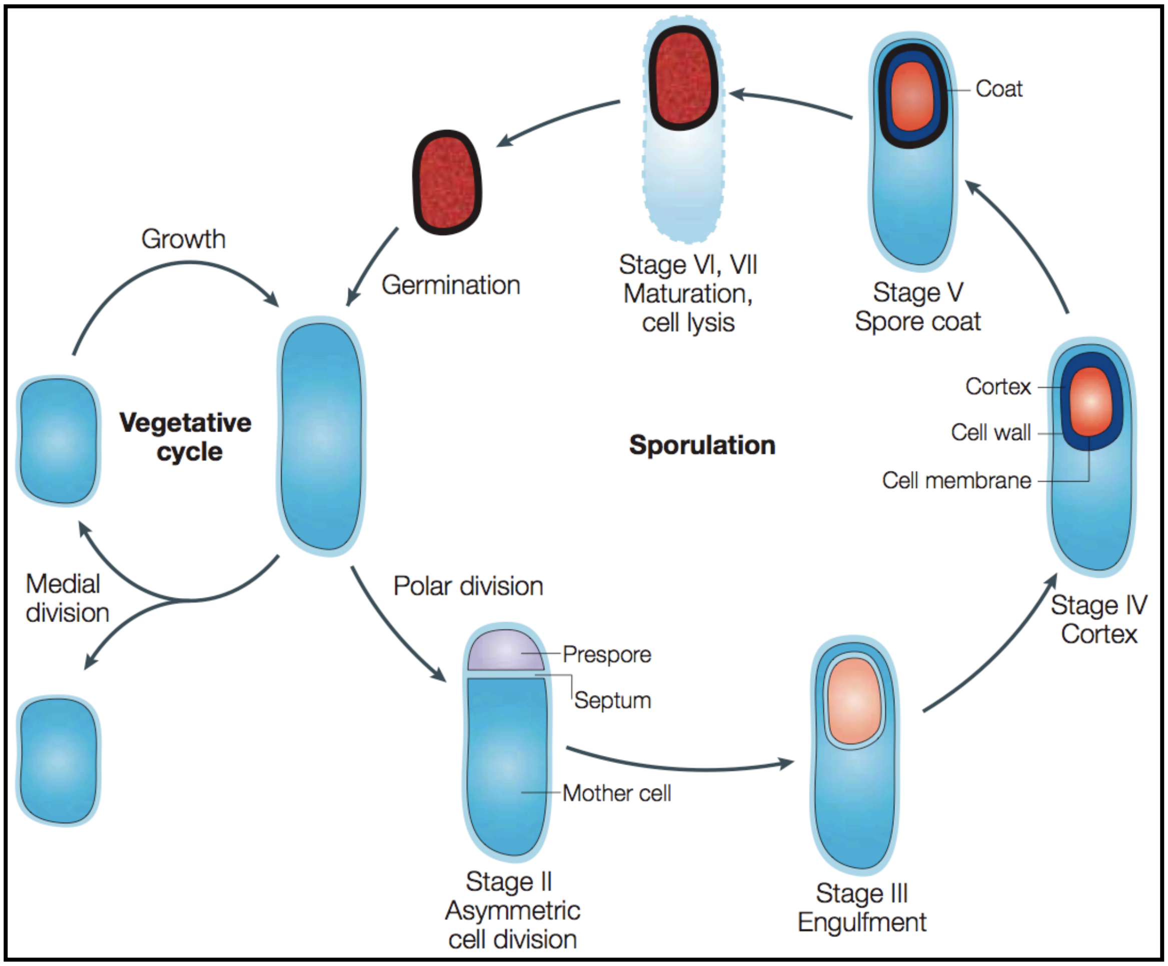


**Figure 1.2)** Key stages of *B. subtilis* sporulation. Stage 0 is defined as the growing cell. Stage I, which is not depicted in the figure, occurs when the two chromosomes after the final medial septation form an axial filament. This phenomenon is observed even in mutants blocked at stage 0, and may be the result of a change in ionic concentration in the cell when the cell encounters nutrient starvation. Stage II is defined by the completion of the asymmetrical division septum at one pole of the cell. Stage III occurs when the burgeoning prespore is fully engulfed by the mother cell. The deposition of a primordial germ cell wall and cortex between the membranes of the prespore defines stage IV. The two membranes of the prespore are derived from both the prespore itself and the mother cell. The formation of a thick protein coat around the cortex is Stage V. During stage VI the spore loses almost all of its water and acquires the characteristics that make it resistant to environmental conditions. At stage VII the mother cell breaks down and lyses, releasing the mature spore to the environment. From (Errington, 2003).

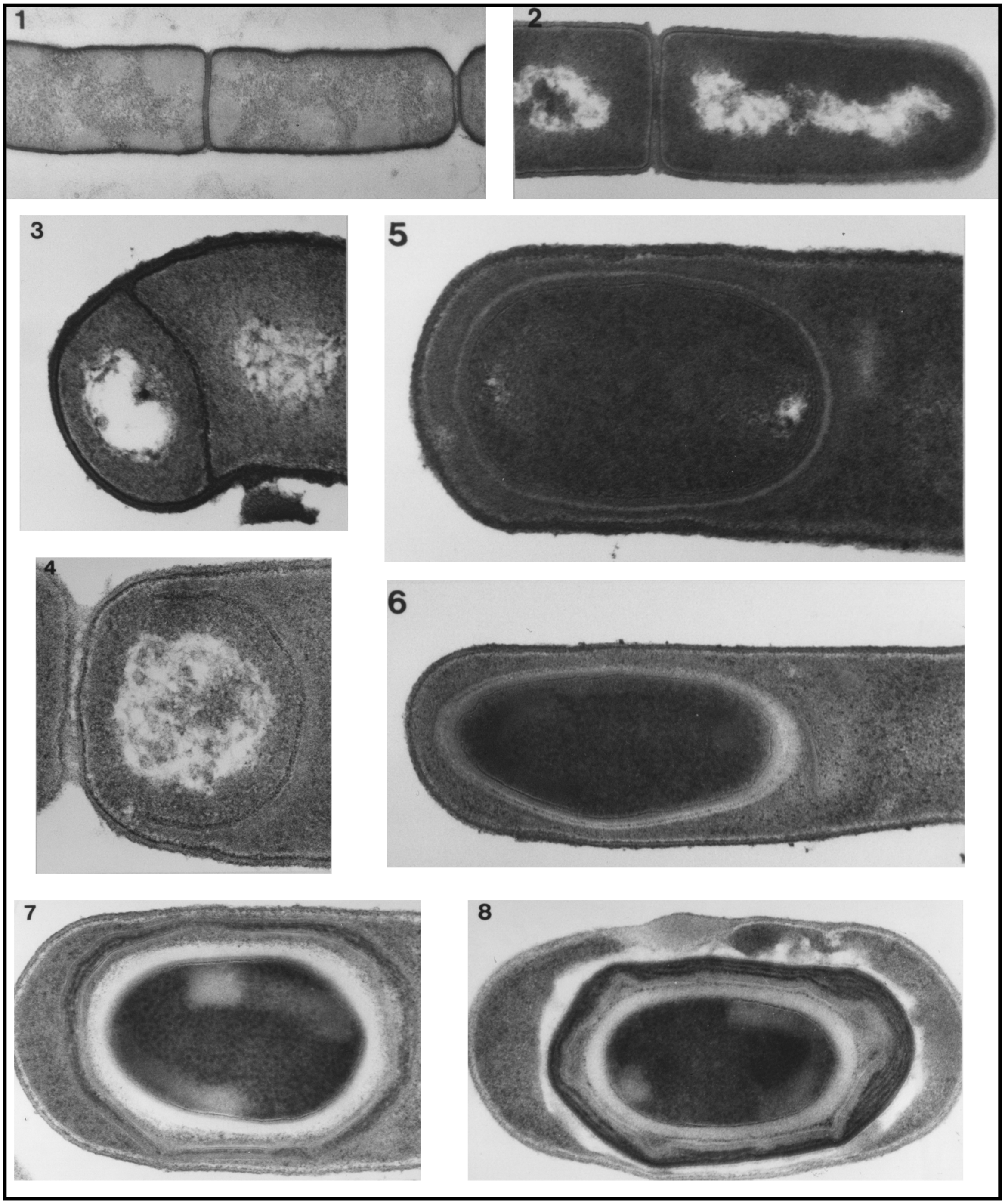


**Figure 1.3)** EM micrographs of *B. subtilis* captured during the various stages of sporulation. The photos mirror the Figure 2 schematic. Stage 0 - The vegetative cell and medial division (photo 1). Stage 1 and the formation of the chromosomal axial filament is shown (photo 2). Stages 2 through 7 are represented in photos 3 through 8, respectively. Adapted from (Jin, 1995).

The hallmark of sporulation is the presence of an asymmetrical division septum. Normally a bacterium will divide precisely in the middle to produce two identical daughter cells. Sporulating bacteria, however, divide asymmetrically into a small prespore at one pole of the cell, and a much larger mother cell. The asymmetric septum is thinner than the usual medial septum, but at least some peptidoglycan is generated in order to give the septum direction during synthesis. However, the peptidoglycan is digested by enzymes specific to the prespore peptidoglycan once the septum has completed and before the mother cell has engulfed the prespore (Guinand et al., 1979). Later on, the prespore becomes highly desiccated, mineralized, and acquires several protective protein coats that will help it survive exposed in the environment in a dormant state. The success of the prespore is entirely dependent upon the mother cell. Before realizing its own demise, the mother cell will dedicate all of its resources to the spore, for example by synthesizing spore coat proteins and assembling them around the forespore from its own cytoplasm.

In the past 40 years many of the genetic pathways governing the various stages of sporulation have been discovered. *B. subtilis* has been the organism of choice for these studies due to its genetic tractability, natural competence, and non-pathogenicity. Also, it is routine to make endospores of *B. subtilis* in the laboratory simply by starving the cells (Sterlini and Mandelstam, 1969). The activity of cell specific (predivisional, mother cell, or forespore) sigma factors and transcriptional regulators is crucial to the development of the spore (Figure 1.4). The stationary phase sigma factor, $\sigma^H$, is responsible for gene expression during the earliest stages of sporulation (Grossman, 1995). After the asymmetric division septum has formed, compartment specific sigma factors take over.

The first cell specific sigma to become active is $\sigma^F$ in the forespore. The appearance of $\sigma^F$ is followed shortly after by the mother cell specific $\sigma^E$, whose activation is dependent on $\sigma^F$. $\sigma^G$ replaces $\sigma^F$ in the forespore following engulfment (Stage III). Activation of the second mother cell sigma, $\sigma^K$ is dependent upon the activity $\sigma^G$, and thus also follows engulfment (Li and Piggot, 2001). The dependency of the two mother cell specific sigma factors on the activity of forespore sigma factors demonstrates that the two cells are in communication with one another, despite the presence of the division septum. The major vegetative sigma factor, $\sigma^A$, is active at all times during sporulation, in both the mother cell and forespore (Li and Piggot, 2001). Non-sigma regulatory factors also play a major role in sporulation. Chief among these is Spo0A. Every sporulation sigma factor depends either directly or indirectly on the activity of Spo0A in its phosphorylated form, Spo0A~P. Chapter two of this thesis describes the atomic resolution structure of a histidine kinase that is responsible for maintaining sufficient levels of Spo0A~P at the onset of sporulation. Because entry to sporulation is dependent upon the activity of histidine kinases, a brief introduction to them is given next.

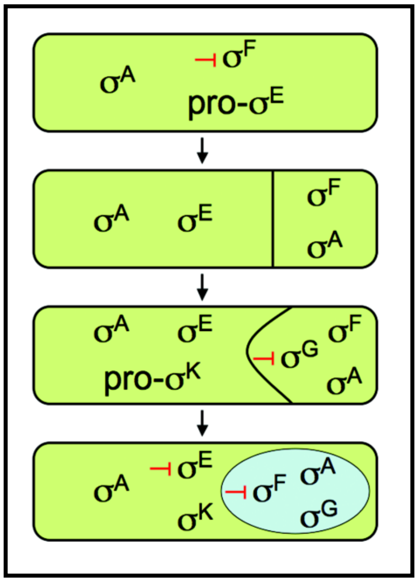


**Figure 1.4)** Schematic representation of compartment specific sigma factors. The vegetative sigma factor, $\sigma^A$, is always active during sporulation. $\sigma^F$ is the first compartment specific sigma to become active, followed closely by $\sigma^E$ in the mother cell. Later, $\sigma^F$ is inhibited in the prespore and $\sigma^G$ takes over, while in the mother cell $\sigma^K$ takes over for $\sigma^E$. Adapted from (Li and Piggot, 2001).

## Histidine Kinases and Two-Component Systems

---

### *Components of the System*

Two-component systems were first described in the 1980s in the context of bacterial chemotaxis, osmo-regulation, and nitrogen regulation. Protein pairs identified in these systems were found to be involved in sensing and responding to environmental stimuli (Bourret et al., 1989; Hess et al., 1988; Ninfa and Magasanik, 1986; Nixon et al., 1986). Since then, two-component systems have been recognized as the dominant signal transduction regulatory systems employed by bacteria and archaea. The number of two-component proteins varies between bacterial species. In *E. coli* there are 62 two-component proteins, while in *Helicobacter pylori* there are only 11 (Mizuno, 1997, 1998). They can often make up more than 2% of a bacterial genome, as is the case for species of *Synechocystis*, where there are at least 80 different two-component proteins (Mizuno et al., 1996). However, some bacteria, such as *Mycoplasma genitalium*, do not have any two-component proteins at all.

Two-component systems are composed of a histidine kinase (HK) - response regulator (RR) pair. The HK senses the environmental signal, while the RR responds to this signal to affect gene expression. Upon receiving the requisite stimulus, the HK will utilize the γ-phosphoryl group of ATP to undergo an autophosphorylation event on a conserved histidine residue. The phosphoryl group is subsequently transferred from the histidine of the HK to a conserved aspartate on the RR (Figure 1.5). Despite the high sequence and structural similarity among HKs and RRs, and the presence of sometimes hundreds of HK-RR pairs in a given species, bacteria maintain an exquisite level of signal transmission fidelity in two-component pathways (Laub and Goulian, 2007).

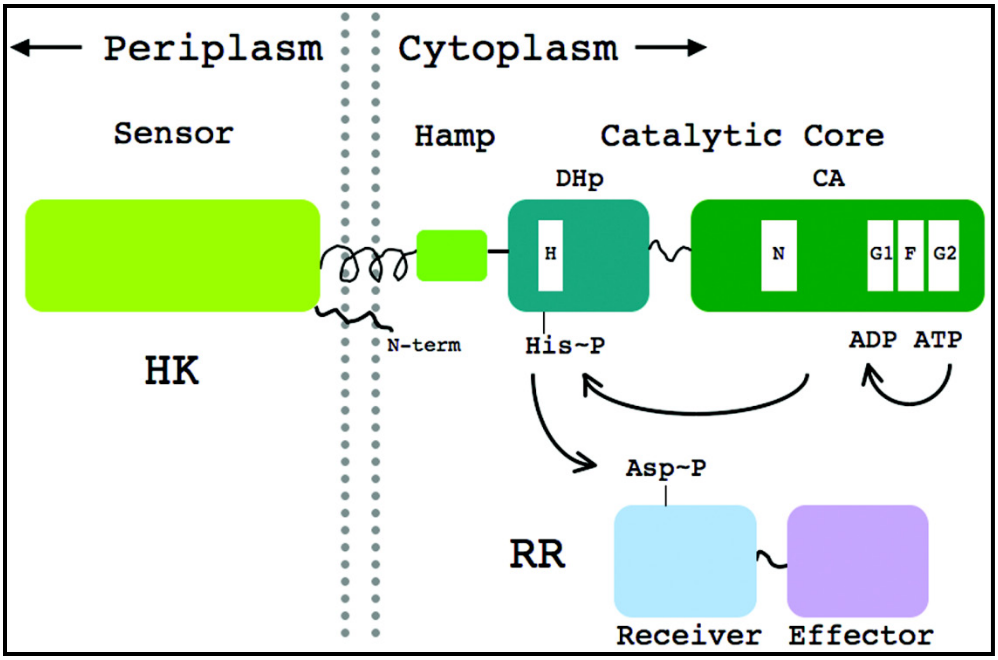


**Figure 1.5)** Two-component systems are composed of a histidine kinase - response regulator pair. The majority of histidine kinases from two-component systems are of the type I variety. They have an extracellular signal-sensing domain connected via transmembrane helices to a cytoplasmic kinase domain. An additional cytoplasmic signal-transducing region termed HAMP may separate the sensor and kinase domains. Histidine kinases autophosphorylate on a conserved histidine within a sequence motif called the H-box in the DHp domain. Other conserved sequence motifs (N, G1, F, G2) within the CA domain are responsible for binding ATP. The response regulator is a two-domain protein that receives phosphate from the kinase in its receiver domain. Phosphorylation induces structural changes in the effector domain, typically a DNA binding domain, which functions as a transcription factor to modulate gene expression.

HKs differ from eukaryotic serine/threonine/tyrosine kinases in several respects. First of all, HKs universally autophosphorylate, while some Ser/Thr/Tyr kinases donate phosphate directly from ATP to another protein substrate (some, such as receptor tyrosine kinases, also autophosphorylate). Often a given Ser/Thr/Tyr kinase will phosphorylate multiple targets, in contrast to the one-to-one specificity of an HK for its cognate RR. Secondly, HKs adopt a fold that is entirely different from the Ser/Thr/Tyr superfamily. Lastly, HKs catalyze the formation of a phosphoramidate bond, in contrast to the phosphoester formed by Ser/Thr/Tyr kinases.

The majority of histidine kinases are classified as type I HKs (Dutta et al., 1999; Fabret et al., 1999). They are typically membrane bound proteins comprised of an extracellular signal sensing domain and a cytoplasmic catalytic core (CC), or kinase domain. The sensor domains vary greatly among HKs in terms of sequence similarity, reflective of the different ligands to which they respond. In contrast, the CC is highly conserved. The sensor and CC domains are connected via a single transmembrane $\alpha$-helix. Many class I HKs (> 1/5th) have an additional cytoplasmic linker domain, termed HAMP (histidine kinase, adenylyl cyclase, methyl-accepting chemotaxis proteins and phosphatase) that connects the C-terminus of the transmembrane helix to the N-terminus of a conserved sequence motif termed the H-box within the CC. This linker domain can vary in length from about 40 to 180 residues (Figure 1.6). The transmembrane helix and HAMP domain have been implicated in the signal transduction reaction, and the HAMP domain offers an additional point of HK regulation (Aravind and Ponting, 1999; Fabret et al., 1999; Hulko et al., 2006; Williams and Stewart, 1999; Zhu and Inouye, 2004).

The CC can be further subdivided into a dimerization and histidine phosphotransfer (DHp) domain, and an ATP binding catalytic (CA) domain. The DHp domain mediates HK dimerization when two monomers interact to form a four-helix bundle (Figure 1.7). The DHp domain contains the invariant histidine residue, which is the site of phosphorylation, within the H-box. Other conserved motifs, the N, G1, F, and

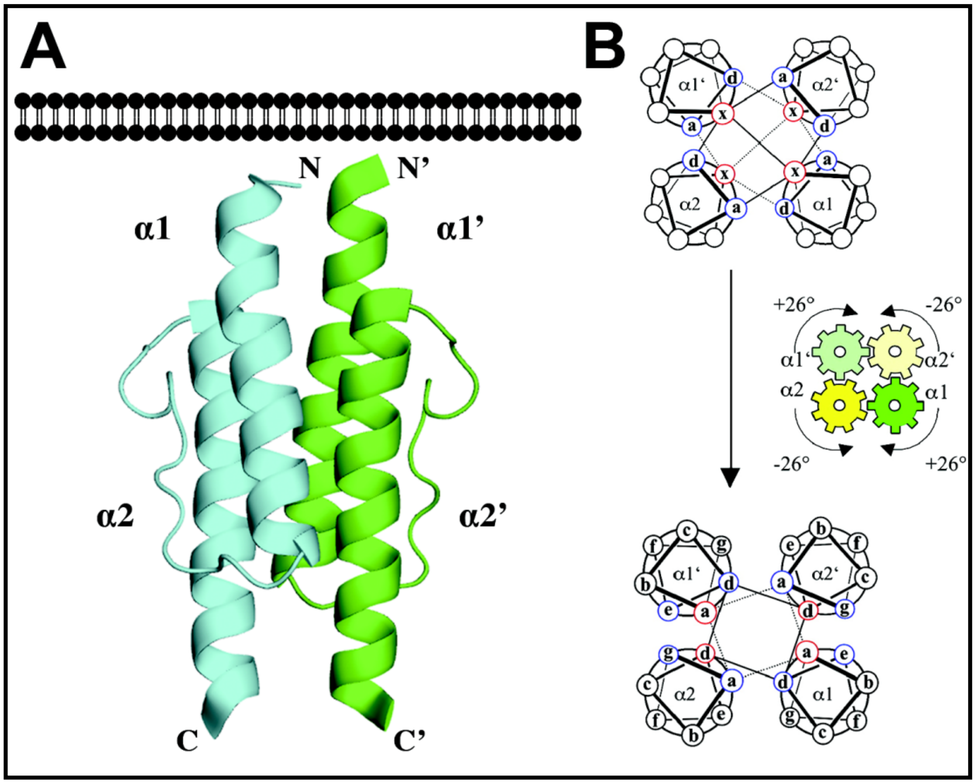


**Figure 1.6)** (A) The solution structure of an isolated HAMP domain from *Archaeoglobus fulgidus* (PDB code 2ASW). A parallel four-helix bundle makes up the dimer. The two helices of each monomer are equal in length and offset in relation to each other by a single helical turn. In the context of HKs, the HAMP domain is interspersed between the transmembrane helices and the DHp domain of the catalytic core. (B) Hulko et al. have proposed a model for HAMP mediated signal transduction that involves a rotary mechanism of the helices. The schematic illustrates the proposed rotational conversion between complementary knobs-to-knobs packing (top) and knobs-into-holes packing (bottom) (Hulko et al., 2006). In the knobs-to-knobs orientation, *x*-layer side chains point straight towards the central supercoil axis, while the *da*-layer side chains point sideways to form an interacting ring of residues enclosing the central cavity. Adapted from (Hulko et al., 2006).

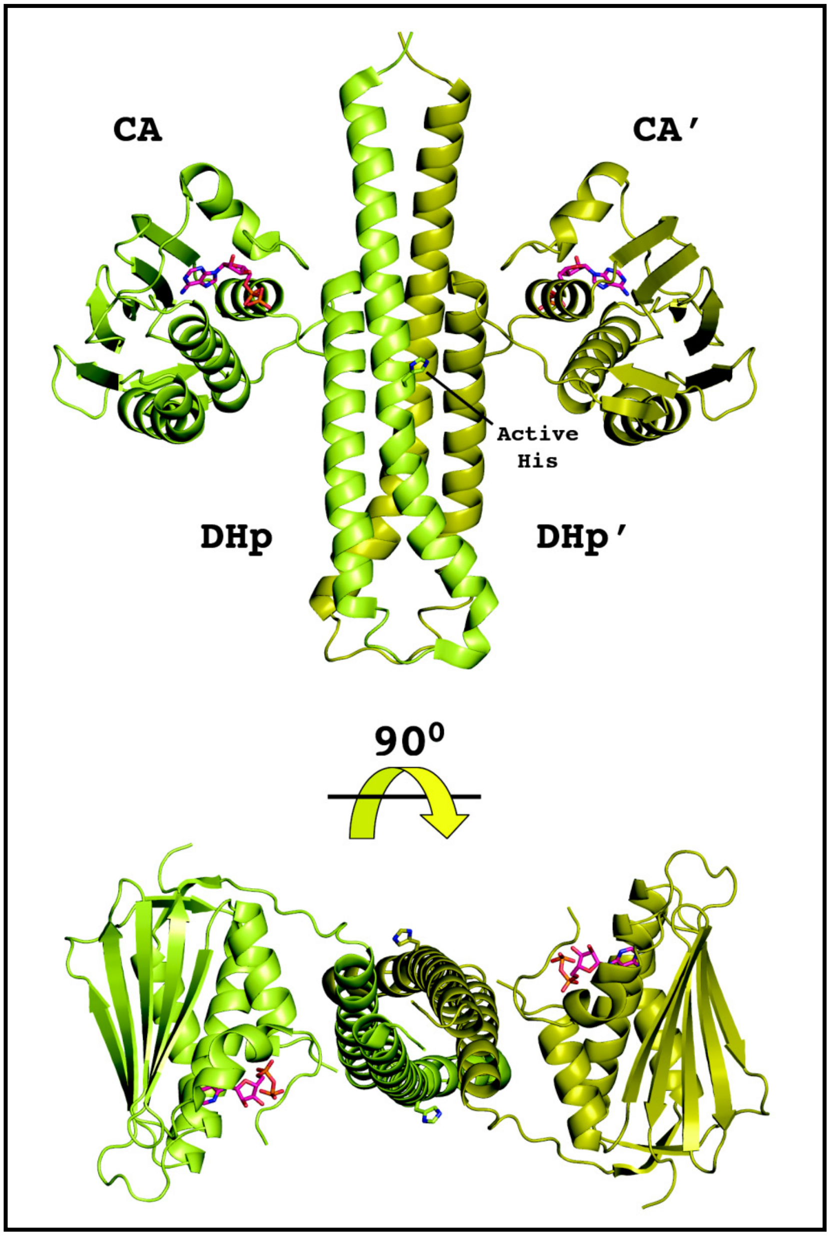


**Figure 1.7)** Two views of HK 0853 from *T. maritima*. Dimerization is mediated by the DHp domain. The DHp is connected to the CA domain via a flexible linker. The side chain of the active site histidine is shown as a stick representation. Also depicted are two molecules of ADPβN bound to each of the CA domains. Adapted from (Marina et al., 2005).

G2 boxes, are within the CA domain and define the boundary of the ATP binding pocket (Parkinson and Kofoid, 1992).

CA domains adopt the Bergerat ATP-binding fold seen in members of the GHKL superfamily of ATPases (Bergerat et al., 1997; Bilwes et al., 1999; Dutta and Inouye, 2000). Prototypical members of this family include the N-terminal domain of the topoisomerase GyrB, MutL-like DNA mismatch repair proteins, the N-terminal domain in the Hsp90 family of heat shock protein chaperones, and histidine kinases (Ban and Yang, 1998; Prodromou et al., 1997a; Prodromou et al., 1997b; Stebbins et al., 1997; Wigley et al., 1991). The Bergerat fold is made up of an $\alpha/\beta$-sandwich consisting of five antiparallel $\beta$ strands and three $\alpha$ helices. Several crystal structures are available for individual CA domains, and all show the same overall structure (Marina et al., 2001; Song et al., 2004). In addition, the crystal structures of two intact CC domains, one of which is described in detail in chapter 2 of this thesis, reveal the disposition of the CA domain relative to the DHp domain, and show how the two domains are connected through a flexible linker (Bick et al., 2009; Marina et al., 2005). Because of the CA domain's similarity to proteins from such diverse organisms, it is likely that the domain architecture of the CC is the result of a fusion event of two independently evolved domains, the DHp and CA domains.

Many of the classic studies that parsed the various enzymatic activities and structural features of the CC were conducted on the osmoregulator EnvZ from *E. coli*. EnvZ is a prototypical type I HK. Along with its cognate RR, OmpR, EnvZ regulates the expression of porin genes in response to changes in osmotic pressure (Hall and Silhavy, 1981; Mizuno et al., 1982). Results from biochemical experiments performed with EnvZ have since been shown to be applicable to type I HKs in general. The EnvZ DHp subdomain dimerizes on its own, forming an antiparallel four-helix bundle. The active site histidine is located towards the N-terminus of helix 1, where it faces the solvent and is accessible to both the CA domain for Phosphorylation and the RR for phosphotransfer (Tomomori et al., 1999). In contrast, the CA domain is monomeric when expressed separately. The kinase activity of the CC resides in the CA domain; in the presence of ATP and $Mg^{2+}$ it is possible to phosphorylate the DHp domain by a non-covalently linked CA domain. A phosphorylated DHp by itself is able to transfer the phosphoryl group to its cognate response regulator (Park et al., 1998).

In addition to their autophosphorylation and phosphotransfer activities, HKs also possess a phosphatase activity that helps to regulate the levels of phosphorylated RR (Egger and Inouye, 1997; Forst and Roberts, 1994). The RR is typically a two-domain transcription factor, whose N-terminal domain receives the signal from the HK and output domain binds DNA as either an activator or repressor of gene expression. Much of the phosphatase activity of HKs can be attributed to the DHp domain, as evidenced from the fact that the DHp domain on its own is competent to dephosphorylate its cognate RR, but the CA domain is not (Zhu et al., 2000). However, when the CA domain is covalently linked to the DHp domain, the phosphatase activity is enhanced relative to the DHp domain alone. Furthermore, ADP in the reaction mixture further enhances the HK phosphatase activity. This is true whether the DHp and CA domains are separate, or whether they are covalently linked, although if they are linked the activity is increased by at least 10-fold. ADP bound in the pocket of the CA domain may induce a conformational change that allows the CA domain to influence the DHp domain and stimulate its phosphatase activity (Zhu et al., 2000).

Dephosphorylation of the RR aspartate generates inorganic phosphate. The active site histidine of the HK in some instances is not crucial for this reaction, but in others, such as EnvZ, it does appear to play an essential role. Generation of phospho-His on the HK may be an obligatory intermediate for the phosphatase reaction (Carmany et al., 2003; Dutta and Inouye, 1996; Hsing and Silhavy, 1997; Kamberov et al., 1994; Zhu et al., 2000). In some cases, a phosphorylated RR is capable of donating the phosphoryl group back to its cognate HK to generate a phosphorylated HK (Wang et al., 2001). Despite many demonstrations of HK mediated RR dephosphorylation *in vitro*, the importance of the phosphatase reaction *in vivo* has been called into question in a recent review. The author points out that many of the experiments that established HK phosphatase function *in vitro* were performed with high concentrations of the HK relative to the RR, and that in the cell often it is the RR that is in excess (Kenney, 2010).

Histidine kinases function as homodimers. Historically, they were believed to carry out the autophosphorylation reaction exclusively *in trans*, meaning that the CA domain from one monomer phosphorylates the active histidine within the DHp domain of the opposite monomer. Again, the classic experiments that provided evidence for the trans mechanism were carried out on EnvZ. The first of these involved an EnvZ hybrid HK, Taz1. Taz1 is a fusion of the Tar (a chemotaxis HK which responds to aspartate) signal sensing and transmembrane domains with the EnvZ catalytic core. Taz1 is able to respond to aspartate, the Tar ligand, and stimulate the phosphorylation of OmpR. Experiments involving co-transformation of two mutant forms of Taz1, one with a valine in place of the active site histidine and the other with a missing CA domain, showed that the two mutants were able to complement each other for the autophosphorylation reaction. The authors of the study hypothesized that this result was due to subunit exchange between the Taz1 dimers and the formation of heterodimers in the membrane. Interestingly, complementation using the two Taz1 mutants did not restore the phosphatase activity for OmpR (Yang and Inouye, 1991). Generation of the Taz1 hybrid kinase also demonstrated for the first time the modular nature of the HK domain architecture, that the sensor domain from one HK could be fused with the cytoplasmic kinase domain of another to create a functional enzyme (Utsumi et al., 1989). This modularity has become a powerful tool in the field synthetic biology in recent years (Levskaya et al., 2005; Salis et al., 2009).

More evidence for trans phosphorylation came about a decade later when it was demonstrated that disulfide cross-linked EnvZ heterodimers, in which one monomer was defective for binding ATP and the other contained a mutation at the catalytic histidine, were still competent for the autophosphorylation reaction through complementation. For the first time, the authors also provided direct evidence that the cis mechanism of autophosphorylation, at least for EnvZ, does not occur, and that autophosphorylation does not result from dimer - dimer interactions (Cai and Inouye, 2003).

The dogma of the universality of trans autophosphorylation has recently been challenged, however (Casino et al., 2009; Pena-Sandoval and Georgellis, 2010). Using the classic approach of mixed dimers, Pena-Sandoval et al. showed that heterodimers of the HK ArcB failed to autophosphorylate, or in other words, there was no trans complementation between the two mutant monomers (Pena-Sandoval and Georgellis, 2010). Additional evidence for cis autophosphorylation has come from the recently published high-resolution crystal structure of a *Thermotoga maritima* HK0853-RR468 complex (Figure 1.8). In this structure the β-phosphorus of the bound ADPβN from one CA domain is less than half the distance to the active histidine εN (~11 Å) of the same monomer than it is to the histidine εN of the opposite monomer (~24 Å). The authors of the paper also demonstrate cis autophosphorylation experimentally, again by employing heterodimers. They confirm that HK0853 and a related type I HK, PhoR from *Staphylococcus aureus*, autophosphorylate *in cis* (Casino et al., 2009). It appears that at least some type I HKs do not obey the rule of trans autophosphorylation.

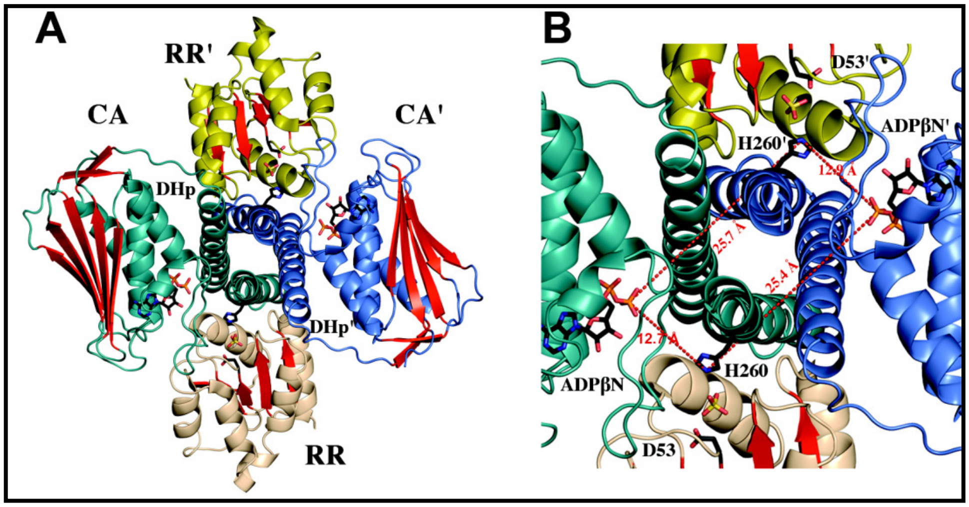


**Figure 1.8)** (A) A view of the HK0853-RR468 complex from *T. maritima*, looking down the four-helix bundle axis of the HK DHp domain. Shown as black sticks are the HK His 260 side chain, the RR Asp 53 side chain, the ADPβN bound to each CA domain, and two sulfates which mediate contacts between His 260 and Asp 53. (B) Distances determined in Pymol between the εN of His 260 and the β-phosphate of ADPβN from both monomers are labeled. The values calculated by Pymol using the deposited PDB file (PDB code 3DGE) differ slightly from the values reported in the publication by Casino et al. However, the proposed cis autophosphorylation mechanism is still valid based on the distance values reported here. Adapted from (Casino et al., 2009).

## *Chemistry of Two-Component Systems*

The chemistry carried out by the HK and RR of two-component systems can be summarized by the following equations:

1. Autophosphorylation: HK-His + ATP $\leftrightarrow$ HK-His~P + ADP
2. Phosphotransfer: HK-His~P + RR-Asp $\leftrightarrow$ HK-His +RR-Asp~P
3. Dephosphorylation: RR-Asp~P + $H_2O$ $\leftrightarrow$ RR-Asp + Pi

All three reactions require a divalent cation, with $Mg^{2+}$ presumably filling that role *in vivo* (Stock et al., 2000). Phosphorylation of histidine occurs in all branches of life, and is not restricted to bacterial two-component systems (Steeg et al., 2003). Both the $N_3$ and $N_1$ positions of the histidine imidazole ring can be phosphorylated (Hultquist et al., 1966), and both forms have been found in proteins (Figure 1.9). However, phosphorylation of $N_3$ is the most commonly observed site for modification (Surette et al., 1996). $N_3$-phosphohistidine is thermodynamically more stable than $N_1$-phosphohistidine, as measured by their respective rates of hydrolysis (Hultquist et al., 1966). $N_1$-phosphohistidine at pH 2.4 (46 ºC) has a half-life of 5 min, while $N_3$-phosphohistidine has a 25 min half-life under the same conditions (both isoforms are fairly stable under basic conditions). The only known example of a protein containing $N_1$ phosphohistidine comes from the bacterial phosphoenolpyruvate:sugar phosphotransfer system (Waygood et al., 1985).

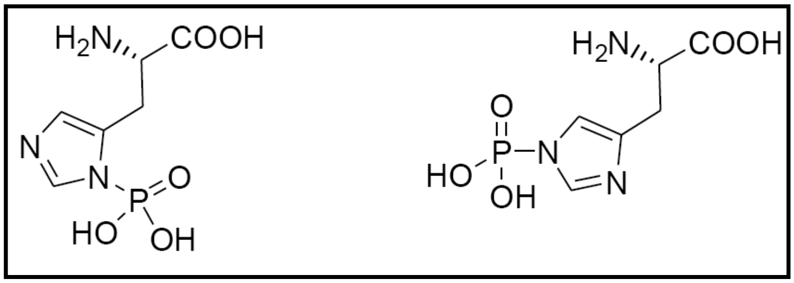


**Figure 1.9)** Chemical structures of 1 (left) and 3 (right) phosphohistidine. 3-phospho-His is almost always observed in proteins.

Phosphohistidine and phosphoaspartate are both considered ‘high-energy’ molecules as free amino acids. The difference between the two arises when they are viewed in the context of a protein. In a protein, the phosphoramidate bond has a free energy of hydrolysis between -12 and -14 kcal/mol, making phospho-His much more unstable than phospho-Ser, phospho-Thr, and phospho-Tyr and the corresponding phosphoester bond (Stock et al., 1990) (Table 1.1). This negative free energy is not much different for the phosphoramidate of a free amino acid (-8 to -10 kcal/mol). Likewise, the free energy for phosphoaspartate as the free amino acid is -10 to -13 kcal/mol. But unlike phospho-His, phospho-Asp has a free energy of +2 in the context of a protein (Table 1.1). These values make sense when considering the role of phospho-His primarily as a short-lived catalytic donor, with its free energy capable of driving phosphorylation of a substrate (e. g. a response regulator). Phosphorylation of Asp on the other hand often promotes large, long-range conformational changes in a protein, as in the case of response regulators. The difference observed for the free energies of phospho-Asp implies that in the context of a protein, the free energy of phospho-Asp is heavily dependent on its interactions with the protein of which it is a part (Cho et al., 2001). This leads to an interesting observation for two-component systems, specifically the back reaction from the RR to the HK in two-component systems, and in particular, the sporulation phosphorelay in *B. subtilis*, for in that system phosphoryl transfer occurs from Asp to His.

Table 1.1

| Phosphorylated Residue | $\Delta G^{o}_{\text{hydrolysis}}$ (kcal/mol) Intrinsic | $\Delta G^{o}_{\text{hydrolysis}}$ (kcal/mol) Protein |
|---|---|---|
| Phosphoserine | -3 | -6.5 |
| Phosphotyrosine | N/A | -9.5 |
| Phosphohistidine | -8 to -10 | -12 to -14 |
| Phosphoaspartate | -10 to -13 | +2 |

**Table 1.1)** The free energies of hydrolysis for the most common phospho amino acids. ΔG values are given for the free amino acid, and for the same in the context of a protein. Based on (Fukami and Lipmann, 1983; Hubler et al., 1989; Jencks, 1968; Pickart and Jencks, 1984; Shizuta et al., 1975; Weigel et al., 1982; Wylie et al., 1988).

### *HK Sensor Domains and Signaling*

Now that much is understood about how the catalytic core functions, attention has turned to HK sensor domains. Unlike the CC domain of HKs, sensor domains vary greatly both in sequence and structure, a consequence of the different signals to which sensor domains respond. The signaling molecules that bind to sensor domains also show tremendous diversity. Typically they are small molecule ligands (Cheung and Hendrickson, 2008, 2009; Emami et al., 2009; Sevvana et al., 2008), but can also be divalent cations (Cho et al., 2006), peptides (Geisinger et al., 2008), oxygen (Gilles-Gonzalez et al., 2008), and even light (Purcell et al., 2007; Swartz et al., 2007) and temperature (Aguilar et al., 2001). Whereas the HK kinase domain is always cytoplasmic, sensor domains can be extracellular, cytoplasmic, or membrane embedded.

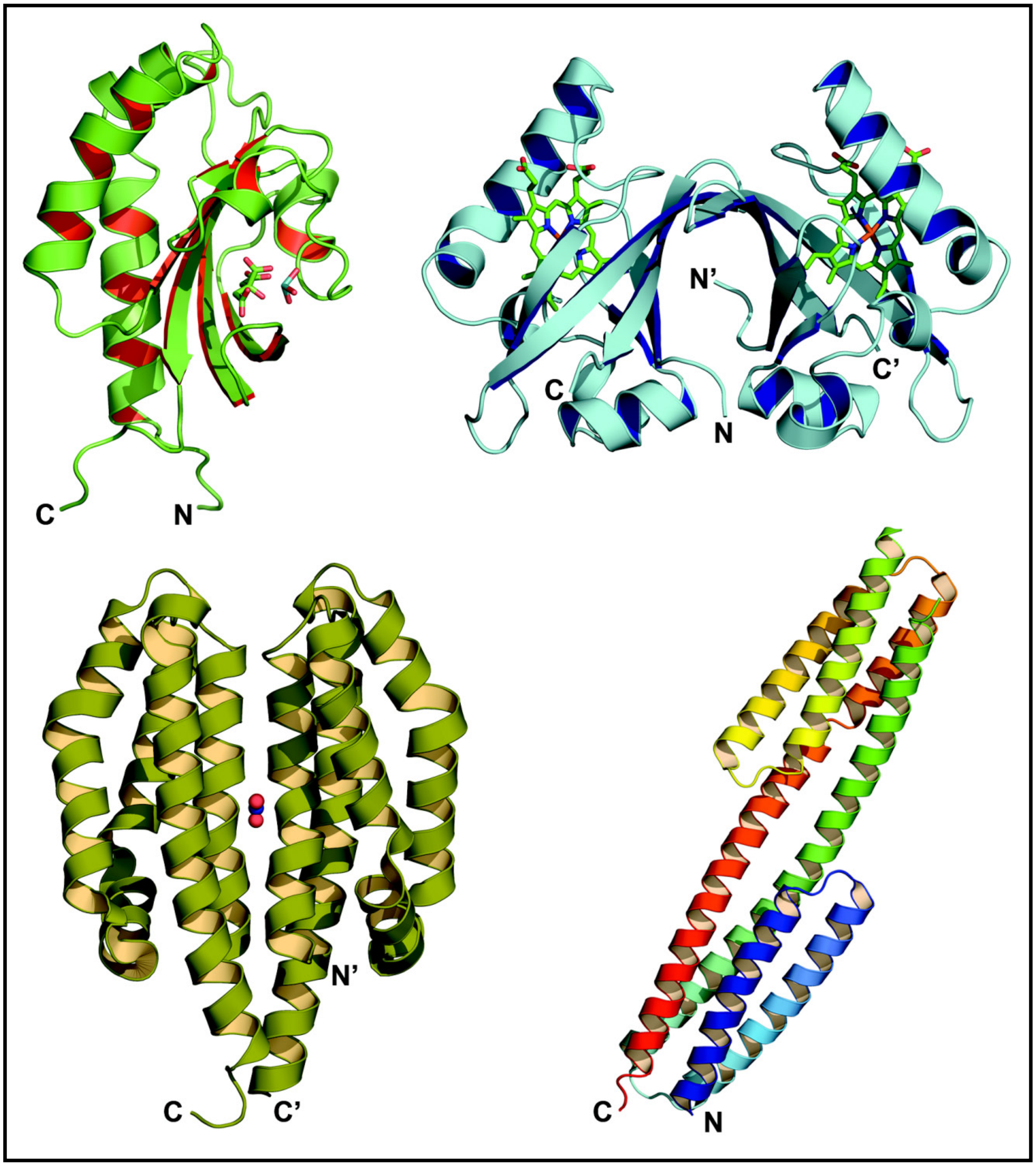
Extracellular sensor domains can be classified according to the three dimensional structure they adopt. Currently three classes exist: those that adopt the PDC (PhoQ-DcuS-CitA) fold, a mixed $\alpha/\beta$ structural motif with topological similarities to the PAS Per/ARNT/Sim) domains found in many signal transduction pathways (Cheung and Hendrickson, 2008; Moglich et al., 2009; Reinelt et al., 2003), those that are entirely $\alpha$-helical (Moore and Hendrickson, 2009), and those that have a similar fold to periplasmic binding proteins (Cheung et al., 2009) (Figure 1.10).

There are no currently available structures of membrane embedded sensor domains, though biochemical studies on a number of HKs have shown that signals are indeed sensed within the transmembrane segments (Albanesi et al., 2009; Jensen et al., 2008). It is possible that KinB from the sporulation phosphorelay detects signals via its transmembrane segments. HKs with several membrane spanning regions and no obvious extracellular domain are the most likely candidates to utilize this mechanism of signal sensing. Based on secondary structural predictions (Rost et al., 2004), KinB fits this description.

Cytoplasmic sensors often contain PAS domains (Ayers and Moffat, 2008; Gong et al., 1998). These sensors can be found in the context of entirely cytoplasmic HKs, at the very N-terminus of membrane bound HKs, or in between the last transmembrane segment and the CC. The cytoplasmic HKs KinA and KinE of the sporulation phosphorelay are predicted to encode three and four PAS domains in their sensor regions, respectively. Other structural motifs found in cytoplasmic sensor domains are the GAF (Ho et al., 2000) and PHY (phytochrome) folds (Yang et al., 2008). Regardless of the domain structure, sensor domains, and HKs overall, form homodimers. It is presumed that the signaling step requires the dimer state. However, it has been shown for some sensors that binding of a single ligand to only one protomer of the dimer is sufficient to induce signaling (Yeh et al., 1996). Because there is no available structure of a full-length HK, either cytoplasmic or membrane associated, we know comparatively little regarding the mechanism of HK signal transduction. How signaling occurs can only be inferred based on the structures of isolated CCs and sensor domains, and from what is known about other transmembrane signaling proteins, such as chemoreceptors. In chemoreceptors, very small (1-2 Å) conformational changes induced by ligand binding are transmitted from a sensor domain membrane-spanning helix to the cytoplasmic catalytic domain by a relative piston-like displacement mechanism between helices (Falke and Hazelbauer, 2001).

**Figure 1.10** (adjacent page) Four representative HK sensor domain crystal structures. (Top Left) Citrato-molybdate bound structure of the extracellular sensor domain of CitA from *Klebsiella pneumoniae* (PDB code 1P0Z) in its monomeric form. CitA is a member of the PDC group of HK sensor domains. (Top Right) The PAS-containing sensor domain of FixL (PDB code 2VV6) from *Bradyrhizobium japonicum* in its dimeric ferric form. The FixL sensor domain is cytoplasmic, and is located directly N-terminal to the kinase domain (C-terminal to the transmembrane helices). (Bottom left) The nitrate periplasmic sensor domain of NarX (PDB code 3EZH) from *Escherichia coli*. The NarX dimer contains a single ligand-binding site at the dimer interface. (Bottom right) The TorS (PDB code 3I9Y) extracellular sensor domain from *Vibrio parahaemolyticus*. TorS is entirely helical. Adapted from (Ayers and Moffat, 2008; Cheung and Hendrickson, 2009; Moore and Hendrickson, 2009; Reinelt et al., 2003).

Figure 1.10



Two recent structures of HK sensor domains with and without ligand bound suggest that HKs also transmit their signals via a piston displacement movement of the transmembrane helices (Cheung and Hendrickson, 2009; Moore and Hendrickson, 2009) (Figure 1.11). However, structures of full-length HKs are needed to confirm this hypothesis. Chapter 4 of this thesis lays the foundation for obtaining the structure of a full-length cytoplasmic HK. It is possible that the signaling mechanism is the same for all HKs, both cytoplasmic and membrane-bound, because for both the signaling and kinase domains are connected by a long helix.

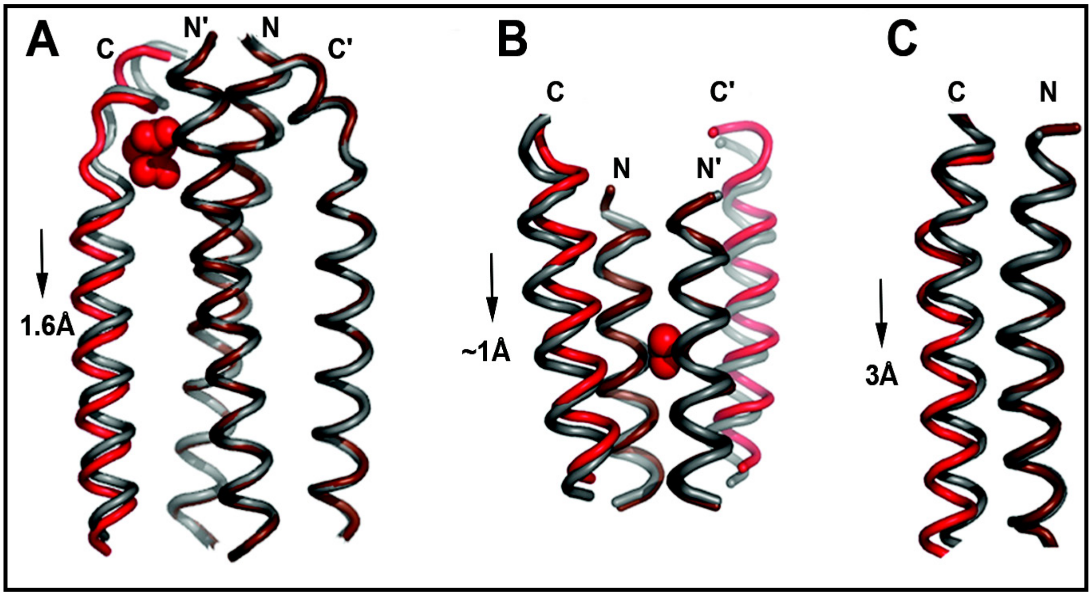


**Figure 1.11)** (A) Depiction of the N- and C-terminal helices of the periplasmic sensor domain from the aspartate receptor Tar, in the unliganded (grey) and liganded (red) form. A piston-like displacement of the C-terminal helix is induced upon binding aspartate to one of the monomers. (B) Nitrate induced piston displacement of the C-terminal signaling helix in the periplasmic sensor domain of NarX. Nitrate bound NarX is shown in red, unliganded in grey. (C) Crystal structures of the apo form of the N and C-terminal helices of the TorS sensor domain from *V. parahaemolyticus* (grey) and *E. coli* (red). A piston-like displacement is observed in the C-terminal helix between the two structures.

## Histidine Kinases in the Sporulation Phosphorelay

In the 1990s, phosphotransfer pathways resembling two-component systems were discovered in lower eukaryotes, including a system in *Saccharomyces cerevisiae* that regulates an osmosensing MAP kinase cascade (Maeda et al., 1994). The presence of additional intermediate response regulators and phosphotransferases in the eukaryotic systems led to the term phosphorelay in order to describe them. Phosphorelays allow for greater levels of fine-tuning the flow of phosphate than do simple two-component systems because of the presence of additional domains; these additional domains offer more targets for positive and negative regulation. HKs in eukaryotic phosphorelays appear to be exclusively of a hybrid type; the C-terminus of the HK catalytic core is fused directly to the RR. Regardless of domain organization or number of proteins in the signaling system, the sequence of phosphoryl transfer is always the same between prokaryotes and eukaryotes. In prokaryotes, phosphoryl transfer proceeds from His to Asp, whereas in eukaryotes this transfer is merely duplicated to His-Asp-His-Asp. Phosphorelays are not restricted to unicellular eukaryotes. *Arabidopsis thaliana* has an ethylene inducible signal transduction system that is mediated by a phosphotransfer event involving a hybrid HK, etr1 (Chang et al., 1993). To date, however, there have not been any reports of two-component or phosphorelay systems in animal species. In addition, the frequency of phosphorelay systems for any given eukaryote is much lower than two-component systems in bacteria. Sometimes only a single system is present, as in the case of *S. cerevisiae*.

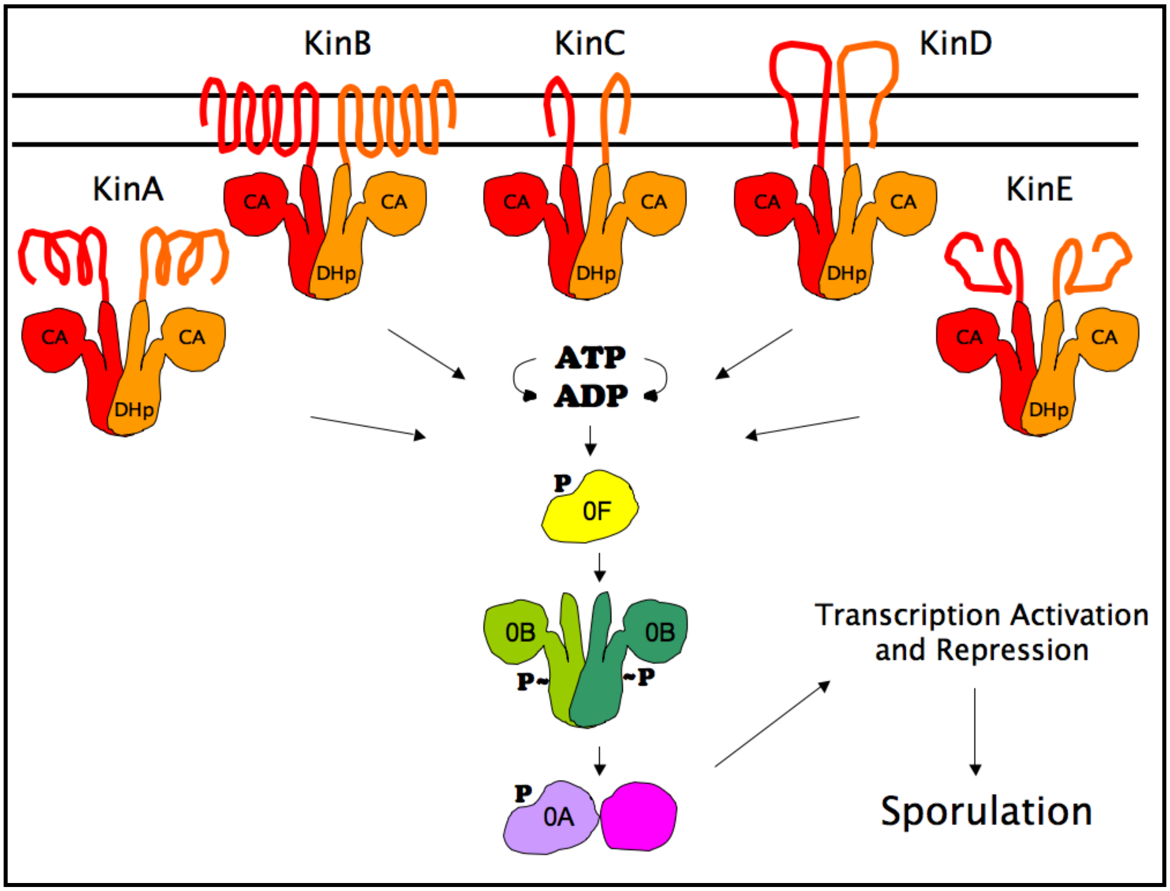


**Figure 1.12)** The *Bacillus* sporulation phosphorelay. Signals are sensed by one of five histidine kinases, KinA through KinE, resulting in an autophosphorylation event in the DHp domain. The phosphoryl group is passed to Spo0A via the response regulator Spo0F and the phosphotransferase Spo0B. Spo0A is the master regulator of sporulation gene expression.

Though prokaryotic phosphorelays occur with less frequency, they do exist (Zhang and Shi, 2005). In fact, the first described phosphorelay was not eukaryotic, but rather the phosphorelay controlling entry to sporulation in *B. subtilis* (Burbulys et al., 1991) (Figure 1.12). The gatekeepers of the sporulation phosphorelay are a group of five type I HKs, KinA through KinE (KinA-KinE are not hybrid HKs as in eukaryotic phosphorelays) (Jiang et al., 2000b). KinB, KinC, and KinD are membrane bound HKs, while KinA and KinE are cytoplasmic.

As type I HKs, the sporulation kinases bind and utilize ATP to autophosphorylate on a histidine residue in their DHp domains after sensing certain environmental cues. The phosphoryl group is then transferred to an Asp residue of the single domain RR Spo0F (Spo0F lacks the DNA binding domain present in many RRs), then to a His residue of the phosphotransferase Spo0B. Spo0B resembles the catalytic core of type I HKs; it shares a dimeric structure with a central four-helix bundle, containing the active His residue, flanked by two CA-like domains that are unable to bind ATP (Varughese et al., 1998). Finally, Spo0B donates the phosphoryl group to an Asp residue of the RR Spo0A, the master transcription factor that regulates expression of sporulation genes in the predivisional cell (Figure 1.12).

Phosphorylated Spo0A (Spo0A~P) binds to "0A boxes" on the chromosome (Hoch, 1993), and acts either as an activator or repressor of gene expression. Spo0A~P has 121 genes under its direct control, and at least 500 under indirect control (Fawcett et al., 2000; Molle et al., 2003). One notable member of the proteins whose expression is regulated by Spo0A~P is the forespore specific sigma factor, $\sigma^F$. Genes under the control of Spo0A~P can be roughly divided into two categories, those that require a low-threshold level of Spo0A~P, and those that require a high-threshold level. Interestingly, the low-threshold genes appear to be those that would act to avert sporulation, such as genes for cannibalism or fruiting body formation (Branda et al., 2001; Fujita et al., 2005; Gonzalez-Pastor et al., 2003). In order for the bacterium to sporulate, levels of Spo0A~P must accumulate gradually, and this accumulation must be mediated by the phosphorelay; overexpression of KinA, KinB, or KinC is sufficient to induce sporulation, but overexpression of a constitutively activate form of Spo0A, Spo0A-Sad67, is not (Fujita and Losick, 2005; Ireton et al., 1993). It is believed the reason for this is the fact that in normally sporulating cells, low-threshold genes are expressed before high-threshold genes, and that immediate induction or repression of both low- and high-threshold genes via the activity of Spo0A-Sad67 inhibits sporulation (Fujita and Losick, 2005).

Accumulation of Spo0A~P not only depends upon the activity of KinA-KinE, but also on an increased rate of synthesis of Spo0A, and the neutralization of various antagonists of the phosphorelay and repressors of Spo0A-dependent genes (Burbulys et al., 1991). Antagonists exist for each member of the phosphorelay. The small protein Sda has been shown to inhibit the autophosphorylation of KinA and KinB, and most likely KinC-KinE (discussed at length in Chapter 2) (Bick et al., 2009; Burkholder et al., 2001; Rowland et al., 2004). Three of the eleven *B. subtilis* Rap (response regulator aspartate phosphatase) family members, RapA, RapB, and RapE, promote dephosphorylation of phosphorylated Spo0F (Spo0F~P), possibly by stimulating the Spo0F autophosphatase activity (Jiang et al., 2000a; Perego et al., 1996; Perego et al., 1994). The sporulation kinases themselves have phosphatase activity against Spo0F~P. The well-studied Spo0E phosphatase, along with two of its homologs, YisI and YnzD, dephosphorylate Spo0A~P; overexpression of Spo0E abolishes the ability to sporulate (Ohlsen et al., 1994; Perego, 2001).

The signals to which KinA-KinE respond are currently unknown, although KinA, KinC, and KinE of *B. subtilis* possess PAS domains in their signal sensing regions (Stephenson and Hoch, 2002). PAS domains are known to sense the overall energy level of a cell, redox potential, and oxygen levels, in addition to serving other functions (Taylor and Zhulin, 1999). The amino-terminal most PAS domain of KinA is a catalytic

ATP-binding domain with nucleoside diphosphate kinase-like activity and is essential to the activity of KinA (Stephenson and Hoch, 2001; Wang et al., 2001). PAS domains B and C are involved in dimerization of the KinA signal-sensing domain (Wang et al., 2001).

Crystal structures of phosphorelay members have provided great insight regarding the propagation of phosphate in the system. Spo0F has been crystallized in both its metal free and metal bound states ($Mn^{2+}$, and $Ca^{2+}$) (Madhusudan et al., 1997; Madhusudan et al., 1996; Mukhopadhyay et al., 2004; Varughese et al., 2006; Zapf et al., 2000) (Figure 1.13). Upon binding metal, the active site of Spo0F is rearranged to create a divalent cation binding site. The phosphotransfer reaction from HKs to RRs, and in the phosphorelay from Spo0B to Spo0A, is divalent cation dependent (Lukat et al., 1990). Spo0F and Spo0A provide the metal for the transfer. $Mn^{2+}$ is believed to be the preferred metal for the HK to Spo0F phosphotransfer reaction because of its higher affinity for Spo0F compared with $Mg^{2+}$ ($K_d$ of 2.5 mM vs. 20 mM), and because of the presence of several nitrogen-containing side chains in the Spo0F active site, for which $Mn^{2+}$ ions have a greater affinity than $Mg^{2+}$ (Bock et al., 1999; Mukhopadhyay et al., 2004).

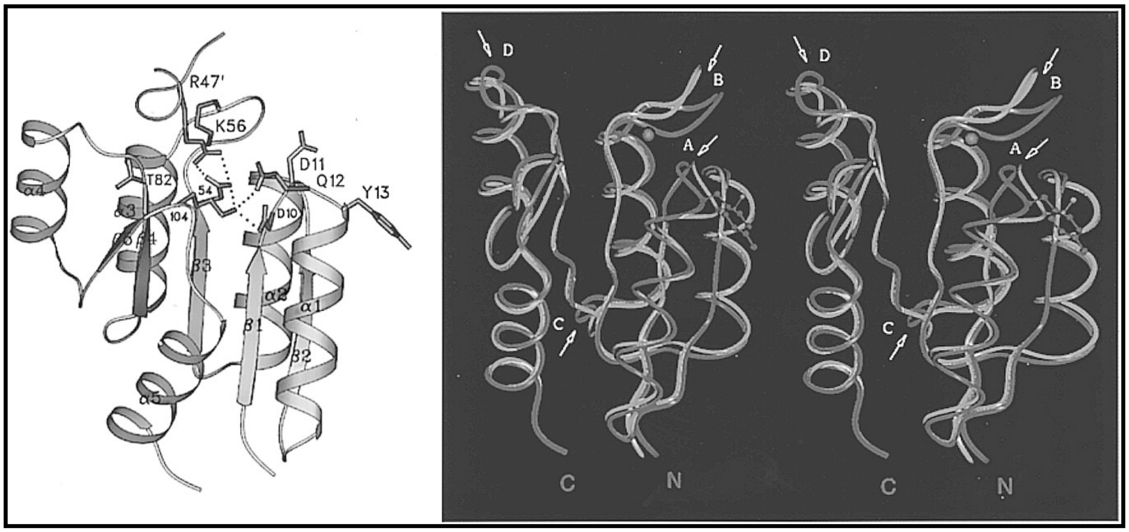
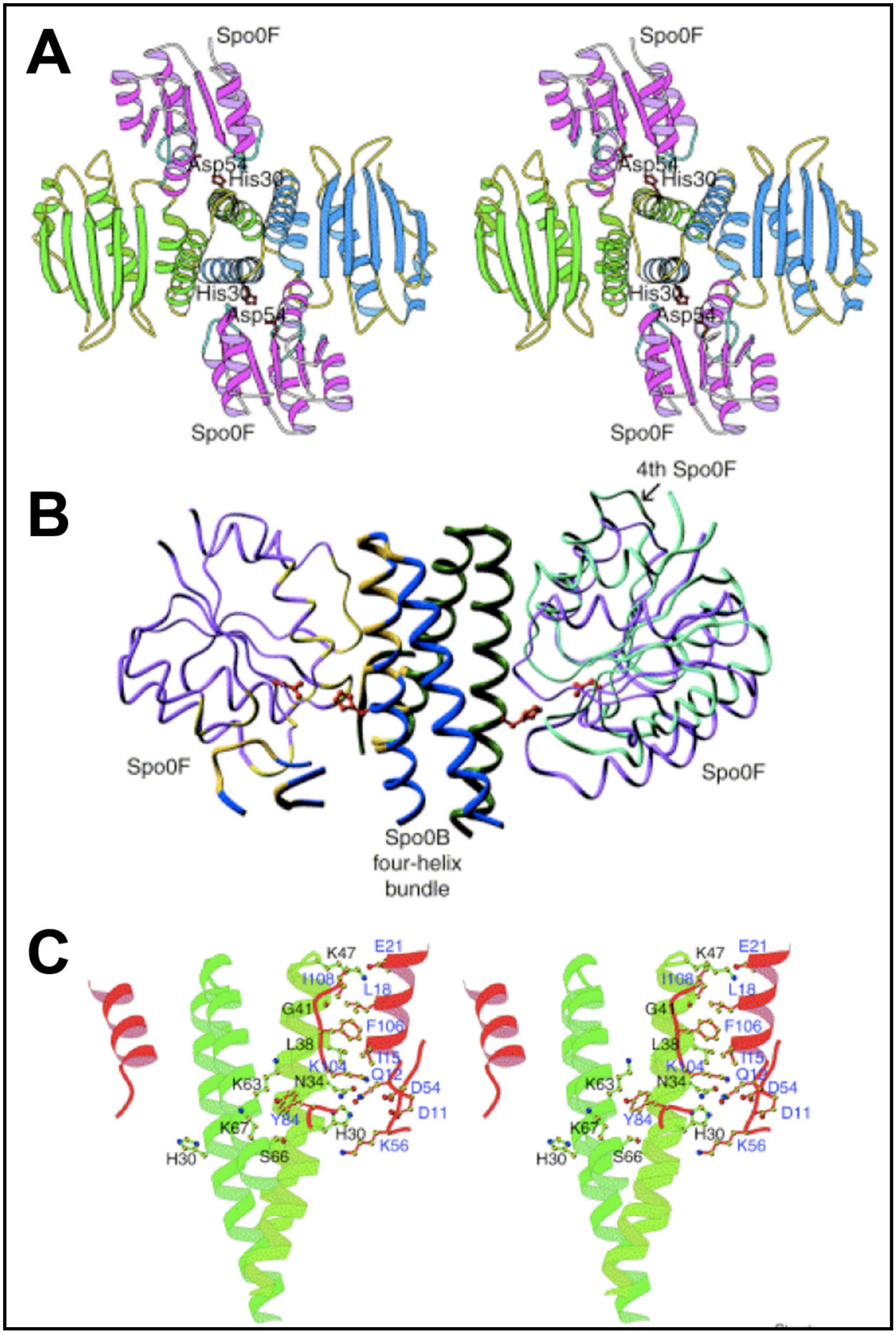


**Figure 1.13)** Comparison of the metal-free and metal-bound forms of Spo0F. On the left is the structure of the wild-type Spo0F molecule. The active site side chains are shown along with Y13 and R47 from the neighboring molecule. D54 is the site of phosphorylation from KinA. On the right is a superposition of the non-metal and the $Ca^{2+}$ bound forms of Spo0F. Note that the metal-bound Spo0F in this figure has a mutation from Y13S. The regions where the two structures differ significantly, around the β1-α1 loop, the β3-α3 loop, the β3-α4 loop, and the β4-α4 loop, are designated A-D, respectively. Taken from (Madhusudan et al., 1997).

The structure of Spo0F, with the active site Asp either unphosphorylated or as a berylloflouride ($BeF_3^-$) phospho-Asp mimic, in a complex with the phosphotransferase Spo0B has also been solved (Figure 1.14) (Varughese et al., 2006; Zapf et al., 2000). Until the recently published structure of an HK-RR pair (Casino et al., 2009), the Spo0F-Spo0B structures had been used as a model of HK-RR interactions. Spo0F shows significant deviations in two loop regions around its active site when in a complex with Spo0B compared with the Spo0F structure on its own. The primary contact surface between Spo0F and Spo0B is composed of hydrophobic residues from each protein, which in Spo0F are highly conserved among response regulators. These contact points in

**Figure 1.14** (adjacent page) Interaction between Spo0F and Spo0B. (A) Steareoview of the Spo0B-Spo0F interaction looking down the axis of the four-helix bundle. The Spo0B dimer is shown in blue and green, while the two molecules of Spo0F are shown in magenta. The sites of phosphorylation are colored red. (B) View of the four-helix bundle along with the two Spo0F molecules (magenta). The ribbon diagram in cyan on the right-hand side of the molecule is a fourth Spo0F molecule from the asymmetric unit of the crystal (the AU had two Spo0Bs and 4 Spo0Fs), shown displaced by ~9° with respect to the other molecules. This fourth Spo0F had a lower occupancy and was oriented differently compared with the other three molecules of Spo0F. Interacting surfaces of Spo0F and Spo0B are shown in gold on the left. (C) Steareoview of the interaction of helix $\alpha1$ of Spo0F with the four-helix bundle of Spo0B. Shown are the side chains that associate through hydrophobic interactions, as well as the active-site residues His30 from Spo0B and Asp54 from Spo0F. Taken from (Zapf et al., 2000).

Figure 1.14


A
Spo0F
Asp54
His30
His30
Asp54
Spo0F
B
4th Spo0F
Spo0F
Spo0F
Spo0B
four-helix
bundle
C
K47
E21
I108
L18
G41
F106
L38
I15
K104
Q12
K63
N34
D54
D11
K67
Y84
H30
K56
H30
S66

Spo0F occur around the active site Asp 54 residue. Recognition specificity is then gained through interaction outside of the hydrophobic patch among non-conserved residues in the β4-α4 loop.

The structures of the various members of the phosphorelay have been known for some time now, but the histidine kinases from this system have eluded structural determination. HKs are notoriously difficult to crystallize, primarily because they are often membrane bound proteins, but also because of the flexibility between the CA and DHp domains of the catalytic core. In the next chapter of this thesis, we describe the 2.0 Å crystal structure of the cytoplasmic catalytic core of KinB from the moderate thermophile, *Geobacillus stearothermophilus*, in a complex with the autophosphorylation inhibitor Sda. This structure takes us one step further towards understanding an important bacterial signaling pathway at atomic resolution detail.

# References

**Aguilar, P.S., Hernandez-Arriaga, A.M., Cybulski, L.E., Erazo, A.C., and de Mendoza, D.** (2001). Molecular basis of thermosensing: a two-component signal transduction thermometer in Bacillus subtilis. EMBO J *20*, 1681-1691.

**Albanesi, D., Martin, M., Trajtenberg, F., Mansilla, M.C., Haouz, A., Alzari, P.M., de Mendoza, D., and Buschiazzo, A.** (2009). Structural plasticity and catalysis regulation of a thermosensor histidine kinase. Proceedings of the National Academy of Sciences of the United States of America *106*, 16185-16190.

**Aravind, L., and Ponting, C.P.** (1999). The cytoplasmic helical linker domain of receptor histidine kinase and methyl-accepting proteins is common to many prokaryotic signaling proteins. FEMS microbiology letters *176*, 111-116.

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

**Ban, C., and Yang, W.** (1998). Crystal structure and ATPase activity of MutL: implications for DNA repair and mutagenesis. Cell *95*, 541-552.

**Beaman, T.C., Pankratz, H.S., and Gerhardt, P.** (1972). Ultrastructure of the exosporium and underlying inclusions in spores of Bacillus megaterium strains. Journal of bacteriology *109*, 1198-1209.

**Bergerat, A., de Massy, B., Gadelle, D., Varoutas, P.C., Nicolas, A., and Forterre, P.** (1997). An atypical topoisomerase II from Archaea with implications for meiotic recombination. Nature *386*, 414-417.

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

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

**Bock, C.W., Katz, A.K., Markham, G.D., and Glusker, J.P.** (1999). Manganese as a replacement for magnesium and zinc: Functional comparison of the divalent ions. Journal of the American Chemical Society *121*, 7360-7372.

**Bourret, R.B., Hess, J.F., Borkovich, K.A., Pakula, A.A., and Simon, M.I.** (1989). Protein phosphorylation in chemotaxis and two-component regulatory systems of bacteria. The Journal of biological chemistry *264*, 7085-7088.

**Branda, S.S., Gonzalez-Pastor, J.E., Ben-Yehuda, S., Losick, R., and Kolter, R.** (2001). Fruiting body formation by Bacillus subtilis. Proceedings of the National Academy of Sciences of the United States of America *98*, 11621-11626.

**Buchner, H. (1890).** Ueber die Ursache der Sporenbildung beim Milzbrandbacillus. Zentr Bakt Parasitenk *8*, 1-6.

**Bulloch, W.** (1938). The History of Bacteriology (London, New York [etc.], Oxford University Press).

**Burbulys, D., Trach, K.A., and Hoch, J.A.** (1991). Initiation of sporulation in B. subtilis is controlled by a multicomponent phosphorelay. Cell *64*, 545-552.

**Burkholder, W.F., Kurtser, I., and Grossman, A.D.** (2001). Replication initiation proteins regulate a developmental checkpoint in Bacillus subtilis. Cell *104*, 269-279.

**Cai, S.J., and Inouye, M.** (2003). Spontaneous subunit exchange and biochemical evidence for trans-autophosphorylation in a dimer of Escherichia coli histidine kinase (EnvZ). Journal of molecular biology *329*, 495-503.

**Cano, R.J., and Borucki, M.K.** (1995). Revival and identification of bacterial spores in 25- to 40-million-year-old Dominican amber. Science (New York, NY *268*, 1060-1064.

**Carmany, D.O., Hollingsworth, K., and McCleary, W.R.** (2003). Genetic and biochemical studies of phosphatase activity of PhoR. Journal of bacteriology *185*, 1112-1115.

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

**Chang, C., Kwok, S.F., Bleecker, A.B., and Meyerowitz, E.M.** (1993). Arabidopsis ethylene-response gene ETR1: similarity of product to two-component regulators. Science (New York, NY *262*, 539-544.

**Cheung, J., and Hendrickson, W.A.** (2008). Crystal structures of C4-dicarboxylate ligand complexes with sensor domains of histidine kinases DcuS and DctB. The Journal of biological chemistry *283*, 30256-30265.

**Cheung, J., and Hendrickson, W.A.** (2009). Structural analysis of ligand stimulation of the histidine kinase NarX. Structure *17*, 190-201.

**Cheung, J., Le-Khac, M., and Hendrickson, W.A.** (2009). Crystal structure of a histidine kinase sensor domain with similarity to periplasmic binding proteins. Proteins *77*, 235-241.

**Cho, H.S., Pelton, J.G., Yan, D., Kustu, S., and Wemmer, D.E.** (2001). Phosphoaspartates in bacterial signal transduction. Curr Opin Struct Biol *11*, 679-684.

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

**Cohn, F.** (1876). Untersuchengen über Bakterien. IV. Beiträge zur Biologie der Bacillen. Beiträge zur Biologie der Pflanzen *2*, 249-276.

**Dutta, R., and Inouye, M.** (1996). Reverse phosphotransfer from OmpR to EnvZ in a kinase-/phosphatase+ mutant of EnvZ (EnvZ.N347D), a bifunctional signal transducer of Escherichia coli. The Journal of biological chemistry *271*, 1424-1429.

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

**Dutta, R., Qin, L., and Inouye, M.** (1999). Histidine kinases: diversity of domain organization. Molecular microbiology *34*, 633-640.

**Dworkin, J., and Losick, R.** (2005). Developmental commitment in a bacterium. Cell *121*, 401-409.

**Egger, L.A., and Inouye, M.** (1997). Purification and characterization of the periplasmic domain of EnvZ osmosensor in Escherichia coli. Biochemical and biophysical research communications *231*, 68-72.

**Emami, K., Topakas, E., Nagy, T., Henshaw, J., Jackson, K.A., Nelson, K.E., Mongodin, E.F., Murray, J.W., Lewis, R.J., and Gilbert, H.J.** (2009). Regulation of the xylan-degrading apparatus of Cellvibrio japonicus by a novel two-component system. The Journal of biological chemistry *284*, 1086-1096.

**Errington, J.** (2003). Regulation of endospore formation in Bacillus subtilis. Nature reviews *1*, 117-126.

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

**Falke, J.J., and Hazelbauer, G.L.** (2001). Transmembrane signaling in bacterial chemoreceptors. Trends in biochemical sciences *26*, 257-265.

**Fawcett, P., Eichenberger, P., Losick, R., and Youngman, P.** (2000). The transcriptional profile of early to middle sporulation in Bacillus subtilis. Proceedings of the National Academy of Sciences USA *97*, 8063-8068.

**Forst, S.A., and Roberts, D.L.** (1994). Signal transduction by the EnvZ-OmpR phosphotransfer system in bacteria. Research in microbiology *145*, 363-373.

**Freese, E.** (1972). Sporulation of bacilli, a model of cellular differentiation. Current topics in developmental biology *7*, 85-124.

**Fujita, M., Gonzalez-Pastor, J.E., and Losick, R.** (2005). High- and low-threshold genes in the Spo0A regulon of Bacillus subtilis. Journal of bacteriology *187*, 1357-1368.

**Fujita, M., and Losick, R.** (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 & development *19*, 2236-2244.

**Fukami, Y., and Lipmann, F.** (1983). Reversal of Rous sarcoma-specific immunoglobulin phosphorylation on tyrosine (ADP as phosphate acceptor) catalyzed by the src gene kinase. Proceedings of the National Academy of Sciences of the United States of America *80*, 1872-1876.

**Geisinger, E., George, E.A., Muir, T.W., and Novick, R.P.** (2008). Identification of ligand specificity determinants in AgrC, the Staphylococcus aureus quorum-sensing receptor. The Journal of biological chemistry *283*, 8930-8938.

**Gest, H., and Mandelstam, J.** (1987). Longevity of microorganisms in natural environments. Microbiological sciences *4*, 69-71.

**Gilles-Gonzalez, M.A., Gonzalez, G., Sousa, E.H., and Tuckerman, J.** (2008). Oxygen-sensing histidine-protein kinases: assays of ligand binding and turnover of response-regulator substrates. Methods Enzymol *437*, 173-189.

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

**Gonzalez-Pastor, J.E., Hobbs, E.C., and Losick, R.** (2003). Cannibalism by sporulating bacteria. Science (New York, NY *301*, 510-513.

**Grossman, A.D.** (1995). Genetic networks controlling the initiation of sporulation and the development of genetic competence in Bacillus subtilis. Annual review of genetics *29*, 477-508.

**Grossman, A.D., and Losick, R.** (1988). Extracellular control of spore formation in Bacillus subtilis. Proceedings of the National Academy of Sciences of the United States of America *85*, 4369-4373.

**Guinand, M., Vacheron, M.J., Michel, G., and Tipper, D.J.** (1979). Location of peptidoglycan lytic enzymes in Bacillus sphaericus. Journal of bacteriology *138*, 126-132.

**Hall, M.N., and Silhavy, T.J.** (1981). Genetic analysis of the ompB locus in Escherichia coli K-12. Journal of molecular biology *151*, 1-15.

**Hess, J.F., Oosawa, K., Kaplan, N., and Simon, M.I.** (1988). Phosphorylation of three proteins in the signaling pathway of bacterial chemotaxis. Cell *53*, 79-87.

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

**Hoch, J.A.** (1993). Regulation of the phosphorelay and the initiation of sporulation in Bacillus subtilis. Annual review of microbiology *47*, 441-465.

**Hsing, W., and Silhavy, T.J.** (1997). Function of conserved histidine-243 in phosphatase activity of EnvZ, the sensor for porin osmoregulation in Escherichia coli. Journal of bacteriology *179*, 3729-3735.

**Hubler, L., Gill, G.N., and Bertics, P.J.** (1989). Reversibility of the epidermal growth factor receptor self-phosphorylation reaction. Evidence for formation of a high energy phosphotyrosine bond. The Journal of biological chemistry *264*, 1558-1564.

**Hulko, M., Berndt, F., Gruber, M., Linder, J.U., Truffault, V., Schultz, A., Martin, J., Schultz, J.E., Lupas, A.N., and Coles, M.** (2006). The HAMP domain structure implies helix rotation in transmembrane signaling. Cell *126*, 929-940.

**Hultquist, D.E., Moyer, R.W., and Boyer, P.D.** (1966). The preparation and characterization of 1-phosphohistidine and 3-phosphohistidine. Biochemistry *5*, 322-331.

**Ireton, K., Rudner, D.Z., Siranosian, K.J., and Grossman, A.D.** (1993). Integration of multiple developmental signals in Bacillus subtilis through the Spo0A transcription factor. Genes & development *7*, 283-294.

**Jencks, W.P.** (1968). In *Handbook of Biochemistry*, H.A. Sober, ed. (Cleveland, The Chem. Rubber Co.), pp. J-144-J-149.

**Jensen, R.O., Winzer, K., Clarke, S.R., Chan, W.C., and Williams, P.** (2008). Differential recognition of Staphylococcus aureus quorum-sensing signals depends on both extracellular loops 1 and 2 of the transmembrane sensor AgrC. Journal of molecular biology *381*, 300-309.

**Jiang, M., Grau, R., and Perego, M.** (2000a). Differential processing of propeptide inhibitors of Rap phosphatases in Bacillus subtilis. Journal of bacteriology *182*, 303-310.

**Jiang, M., Shao, W., Perego, M., and Hoch, J.A.** (2000b). Multiple histidine kinases regulate entry into stationary phase and sporulation in Bacillus subtilis. Molecular microbiology *38*, 535-542.

**Jin, S.** (1995). PhD Thesis. (Boston, Tufts University School of Medicine).

**Kamberov, E.S., Atkinson, M.R., Chandran, P., and Ninfa, A.J.** (1994). Effect of mutations in Escherichia coli glnL (ntrB), encoding nitrogen regulator II (NRII or NtrB), on the phosphatase activity involved in bacterial nitrogen regulation. The Journal of biological chemistry *269*, 28294-28299.

**Kay, D., and Warren, S.C.** (1968). Sporulation in Bacillus subtilis. Morphological changes. The Biochemical journal *109*, 819-824.

**Kenney, L.J.** (2010). How important is the phosphatase activity of sensor kinases? Curr Opin Microbiol *13*, 168-176.

**Koch, R.** (1876). Untersuchungen über Bakterien. V. Die Aetiologie der Milzbrand Krankheit, begrandet auf Entwicklurgs geschichte des *Bacillus anthracis*. Beiträge zur Biologie der Pflanzen *2*, 277-308.

**Laub, M.T., and Goulian, M.** (2007). Specificity in two-component signal transduction pathways. Annual review of genetics *41*, 121-145.

**Levskaya, A., Chevalier, A.A., Tabor, J.J., Simpson, Z.B., Lavery, L.A., Levy, M., Davidson, E.A., Scouras, A., Ellington, A.D., Marcotte, E.M., *et al.*** (2005). Synthetic biology: engineering Escherichia coli to see light. Nature *438*, 441-442.

**Li, Z., and Piggot, P.J.** (2001). Development of a two-part transcription probe to determine the completeness of temporal and spatial compartmentalization of gene expression during bacterial development. Proceedings of the National Academy of Sciences of the United States of America *98*, 12538-12543.

**Lorenz, P.R., Hemenway, C.L., and Hotchin, J.** (1968). The biological effectiveness of solar electromagnetic radiation in space. Life Sci Space Res *6*.

**Lorenz, P.R., Orlob, G.B., and Hemenway, C.L.** (1969). Survival of microorganisms in space. Space Life Sci *1*, 491-500.

**Lukat, G.S., Stock, A.M., and Stock, J.B.** (1990). Divalent metal ion binding to the CheY protein and its significance to phosphotransfer in bacterial chemotaxis. Biochemistry *29*, 5436-5442.

**Madhusudan, Zapf, J., Hoch, J.A., Whiteley, J.M., Xuong, N.H., and Varughese, K.I.** (1997). A response regulatory protein with the site of phosphorylation blocked by an arginine interaction: Crystal structure of Spo0F from Bacillus subtilis. Biochemistry *36*, 12739-12745.

**Madhusudan, Zapf, J., Whiteley, J.M., Hoch, J.A., Xuong, N.H., and Varughese, K.I.** (1996). Crystal structure of a phosphatase-resistant mutant of sporulation response regulator Spo0F from Bacillus subtilis. Structure *4*, 679-690.

**Maeda, T., Wurgler-Murphy, S.M., and Saito, H.** (1994). A two-component system that regulates an osmosensing MAP kinase cascade in yeast. Nature *369*, 242-245.

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

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

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

**Mizuno, T.** (1998). His-Asp phosphotransfer signal transduction. Journal of biochemistry *123*, 555-563.

**Mizuno, T., Kaneko, T., and Tabata, S.** (1996). Compilation of all genes encoding bacterial two-component signal transducers in the genome of the cyanobacterium, Synechocystis sp. strain PCC 6803. DNA Res *3*, 407-414.

**Mizuno, T., Wurtzel, E.T., and Inouye, M.** (1982). Cloning of the regulatory genes (ompR and envZ) for the matrix proteins of the Escherichia coli outer membrane. Journal of bacteriology *150*, 1462-1466.

**Moglich, A., Ayers, R.A., and Moffat, K.** (2009). Structure and signaling mechanism of Per-ARNT-Sim domains. Structure *17*, 1282-1294.

**Molle, V., Fujita, M., Jensen, S.T., Eichenberger, P., Gonzalez-Pastor, J.E., Liu, J.S., and Losick, R.** (2003). The Spo0A regulon of Bacillus subtilis. Molecular microbiology *50*, 1683-1701.

**Moore, J.O., and Hendrickson, W.A.** (2009). Structural analysis of sensor domains from the TMAO-responsive histidine kinase receptor TorS. Structure *17*, 1195-1204.

**Mukhopadhyay, D., Sen, U., Zapf, J., and Varughese, K.I.** (2004). Metals in the sporulation phosphorelay: manganese binding by the response regulator Spo0F. Acta Crystallogr D *60*, 638-645.

**Nicholson, W.L., Munakata, N., Horneck, G., Melosh, H.J., and Setlow, P.** (2000). Resistance of Bacillus endospores to extreme terrestrial and extraterrestrial environments. Microbiol Mol Biol Rev *64*, 548-572.

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

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

**Ohlsen, K.L., Grimsley, J.K., and Hoch, J.A.** (1994). Deactivation of the sporulation transcription factor Spo0A by the Spo0E protein phosphatase. Proceedings of the National Academy of Sciences of the United States of America *91*, 1756-1760.

**Park, H., Saha, S.K., and Inouye, M.** (1998). Two-domain reconstitution of a functional protein histidine kinase. Proceedings of the National Academy of Sciences of the United States of America *95*, 6728-6732.

**Parker, G.F., Daniel, R.A., and Errington, J.** (1996). Timing and genetic regulation of commitment to sporulation in Bacillus subtilis. Microbiology (Reading, England) *142 (Pt 12)*, 3445-3452.

**Parkinson, J.S., and Kofoid, E.C.** (1992). Communication modules in bacterial signaling proteins. Annual review of genetics *26*, 71-112.

**Pena-Sandoval, G.R., and Georgellis, D.** (2010). The ArcB sensor kinase of Escherichia coli autophosphorylates by an intramolecular reaction. Journal of bacteriology *192*, 1735-1739.

**Perego, M.** (2001). A new family of aspartyl phosphate phosphatases targeting the sporulation transcription factor Spo0A of Bacillus subtilis. Molecular microbiology *42*, 133-143.

**Perego, M., Glaser, P., and Hoch, J.A.** (1996). Aspartyl-phosphate phosphatases deactivate the response regulator components of the sporulation signal transduction system in Bacillus subtilis. Molecular microbiology *19*, 1151-1157.

**Perego, M., Hanstein, C., Welsh, K.M., Djavakhishvili, T., Glaser, P., and Hoch, J.A.** (1994). Multiple protein-aspartate phosphatases provide a mechanism for the integration of diverse signals in the control of development in B. subtilis. Cell *79*, 1047-1055.

**Pickart, C.M., and Jencks, W.P.** (1984). Energetics of the calcium-transporting ATPase. The Journal of biological chemistry *259*, 1629-1643.

**Piggot, P.J., and Coote, J.G.** (1976). Genetic aspects of bacterial endospore formation. Bacteriological reviews *40*, 908-962.

**Piggot, P.J., and Losick, R.** (2002). Sporulation Genes and Intercompartmental Regulation. In *Bacillus subtilis* and Its Closest Relatives, A.L. Sonenshein, J.A. Hoch, and R. Losick, eds. (Washington, D. C., ASM Press), pp. 483-517.

**Priest, F.G.** (1993). Systematics and Ecology of *Bacillus*. In *Bacillus subtilis* and Other Gram-Positive Bacteria, A.L. Sonenshein, J.A. Hoch, and R. Losick, eds. (Washington, D. C., ASM Press), pp. 3-16.

**Prodromou, C., Roe, S.M., O'Brien, R., Ladbury, J.E., Piper, P.W., and Pearl, L.H.** (1997a). Identification and structural characterization of the ATP/ADP-binding site in the Hsp90 molecular chaperone. Cell *90*, 65-75.

**Prodromou, C., Roe, S.M., Piper, P.W., and Pearl, L.H.** (1997b). A molecular clamp in the crystal structure of the N-terminal domain of the yeast Hsp90 chaperone. Nature structural biology *4*, 477-482.

**Purcell, E.B., Siegal-Gaskins, D., Rawling, D.C., Fiebig, A., and Crosson, S.** (2007). A photosensory two-component system regulates bacterial cell attachment. Proceedings of the National Academy of Sciences of the United States of America *104*, 18241-18246.

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

**Rost, B., Yachdav, G., and Liu, J.** (2004). The PredictProtein server. Nucleic Acids Res *32*, W321-326.

**Rowland, S.L., Burkholder, W.F., Cunningham, K.A., Maciejewski, M.W., Grossman, A.D., and King, G.F.** (2004). Structure and mechanism of action of Sda, an inhibitor of the histidine kinases that regulate initiation of sporulation in Bacillus subtilis. Molecular cell *13*, 689-701.

**Ryter, A.** (1965). [Morphologic Study of the Sporulation of Bacillus Subtilis.]. Annales de l'Institut Pasteur *108*, 40-60.

**Salis, H., Tamsir, A., and Voigt, C.** (2009). Engineering bacterial signals and sensors. Contrib Microbiol *16*, 194-225.

**Sauer, U., Treuner, A., Buchholz, M., Santangelo, J.D., and Durre, P.** (1994). Sporulation and primary sigma factor homologous genes in Clostridium acetobutylicum. Journal of bacteriology *176*, 6572-6582.

**Setlow, P.** (1988). Small, acid-soluble spore proteins of Bacillus species: structure, synthesis, genetics, function, and degradation. Annual review of microbiology *42*, 319-338.

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

**Shizuta, Y., Beavo, J.A., Bechtel, P.J., Hofmann, F., and Krebs, E.G.** (1975). Reversibility of adenosine 3':5'-monophosphate-dependent protein kinase reactions. The Journal of biological chemistry *250*, 6891-6896.

**Sonenshein, A.L.** (2000a). Bacterial Sporulation: a Response to Environmental Signals. In Bacterial Stress Responses, G. Storz, and R. Hengge-Aronis, eds. (Washington, D. C., ASM Press), pp. 199-215.

**Sonenshein, A.L.** (2000b). Endospore-Forming Bacteria. In Prokaryotic Development, Y.V. Brun, and L.J. Shimkets, eds. (Washington, D. C., ASM Press), pp. 133-150.

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

**Spencer, R.C.** (2003). Bacillus anthracis. Journal of clinical pathology *56*, 182-187.

**Stebbins, C.E., Russo, A.A., Schneider, C., Rosen, N., Hartl, F.U., and Pavletich, N.P.** (1997). Crystal structure of an Hsp90-geldanamycin complex: targeting of a protein chaperone by an antitumor agent. Cell *89*, 239-250.

**Steeg, P.S., Palmieri, D., Ouatas, T., and Salerno, M.** (2003). Histidine kinases and histidine phosphorylated proteins in mammalian cell biology, signal transduction and cancer. Cancer Lett *190*, 1-12.

**Stephenson, K., and Hoch, J.A.** (2001). PAS-A domain of phosphorelay sensor kinase A: a catalytic ATP-binding domain involved in the initiation of development in Bacillus subtilis. Proceedings of the National Academy of Sciences of the United States of America *98*, 15251-15256.

**Stephenson, K., and Hoch, J.A.** (2002). Evolution of signalling in the sporulation phosphorelay. Molecular microbiology *46*, 297-304.

**Sterlini, J.M., and Mandelstam, J.** (1969). Commitment to sporulation in Bacillus subtilis and its relationship to development of actinomycin resistance. The Biochemical journal *113*, 29-37.

**Stock, A.M., Robinson, V.L., and Goudreau, P.N.** (2000). Two-component signal transduction. Annual review of biochemistry *69*, 183-215.

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

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

**Swartz, T.E., Tseng, T.S., Frederickson, M.A., Paris, G., Comerci, D.J., Rajashekara, G., Kim, J.G., Mudgett, M.B., Splitter, G.A., Ugalde, R.A., *et al.*** (2007). Blue-light-activated histidine kinases: two-component sensors in bacteria. Science (New York, NY *317*, 1090-1093.

**Szulmajster, J.** (1964). [Biochemistry of Sporogenesis in B. Subtilis.]. Bulletin de la Societe de chimie biologique *46*, 443-481.

**Taylor, B.L., and Zhulin, I.B.** (1999). PAS domains: internal sensors of oxygen, redox potential, and light. Microbiol Mol Biol Rev *63*, 479-506.

**Tomomori, C., Tanaka, T., Dutta, R., Park, H., Saha, S.K., Zhu, Y., Ishima, R., Liu, D., Tong, K.I., Kurokawa, H., *et al.*** (1999). Solution structure of the homodimeric core domain of Escherichia coli histidine kinase EnvZ. Nature structural biology *6*, 729-734.

**Utsumi, R., Brissette, R.E., Rampersaud, A., Forst, S.A., Oosawa, K., and Inouye, M.** (1989). Activation of bacterial porin gene expression by a chimeric signal transducer in response to aspartate. Science (New York, NY *245*, 1246-1249.

**Varughese, K.I., Madhusudan, Zhou, X.Z., Whiteley, J.M., and Hoch, J.A.** (1998). Formation of a novel four-helix bundle and molecular recognition sites by dimerization of a response regulator phosphotransferase. Molecular cell *2*, 485-493.

**Varughese, K.I., Tsigelny, I., and Zhao, H.** (2006). The crystal structure of beryllofluoride Spo0F in complex with the phosphotransferase Spo0B represents a phosphotransfer pretransition state. Journal of bacteriology *188*, 4970-4977.

**Vreeland, R.H., Rosenzweig, W.D., and Powers, D.W.** (2000). Isolation of a 250 million-year-old halotolerant bacterium from a primary salt crystal. Nature *407*, 897-900.

**Waldburger, C., Gonzalez, D., and Chambliss, G.H.** (1993). Characterization of a new sporulation factor in Bacillus subtilis. Journal of bacteriology *175*, 6321-6327.

**Wang, L., Fabret, C., Kanamaru, K., Stephenson, K., Dartois, V., Perego, M., and Hoch, J.A.** (2001). Dissection of the functional and structural domains of phosphorelay histidine kinase A of Bacillus subtilis. Journal of bacteriology *183*, 2795-2802.

**Warrell, D.A.** (2003). Oxford textbook of medicine, 4th edn (Oxford ; New York, Oxford University Press).

**Waygood, E.B., Erickson, E., el Kabbani, O.A., and Delbaere, L.T.** (1985). Characterization of phosphorylated histidine-containing protein (HPr) of the bacterial phosphoenolpyruvate:sugar phosphotransferase system. Biochemistry *24*, 6938-6945.

**Weigel, N., Kukuruzinska, M.A., Nakazawa, A., Waygood, E.B., and Roseman, S.** (1982). Sugar transport by the bacterial phosphotransferase system. Phosphoryl transfer reactions catalyzed by enzyme I of Salmonella typhimurium. The Journal of biological chemistry *257*, 14477-14491.

**Wigley, D.B., Davies, G.J., Dodson, E.J., Maxwell, A., and Dodson, G.** (1991). Crystal structure of an N-terminal fragment of the DNA gyrase B protein. Nature *351*, 624-629.

**Williams, S.B., and Stewart, V.** (1999). Functional similarities among two-component sensors and methyl-accepting chemotaxis proteins suggest a role for linker region amphipathic helices in transmembrane signal transduction. Molecular microbiology *33*, 1093-1102.

**Wylie, D., Stock, A., Wong, C.Y., and Stock, J.** (1988). Sensory transduction in bacterial chemotaxis involves phosphotransfer between Che proteins. Biochemical and biophysical research communications *151*, 891-896.

**Yang, X., Kuk, J., and Moffat, K.** (2008). Crystal structure of Pseudomonas aeruginosa bacteriophytochrome: photoconversion and signal transduction. Proceedings of the National Academy of Sciences of the United States of America *105*, 14715-14720.

**Yang, Y., and Inouye, M.** (1991). Intermolecular complementation between two defective mutant signal-transducing receptors of Escherichia coli. Proceedings of the National Academy of Sciences of the United States of America *88*, 11057-11061.

**Yeh, J.I., Biemann, H.P., Prive, G.G., Pandit, J., Koshland, D.E., Jr., and Kim, S.H.** (1996). High-resolution structures of the ligand binding domain of the wild-type bacterial aspartate receptor. Journal of molecular biology *262*, 186-201.

**Young, I.E., and Fitz-James, P.C.** (1959a). Chemical and Morphological Studies of Bacterial Spore Formation : I. The Formation of Spores in Bacillus cereus. The Journal of biophysical and biochemical cytology *6*, 467-482.

**Young, I.E., and Fitz-James, P.C.** (1959b). Chemical and morphological studies of bacterial spore formation. II. Spore and parasporal protein formation in Bacillus cereus var. alesti. The Journal of biophysical and biochemical cytology *6*, 483-498.

**Zapf, J., Sen, U., Madhusudan, Hoch, J.A., and Varughese, K.I.** (2000). A transient interaction between two phosphorelay proteins trapped in a crystal lattice reveals the mechanism of molecular recognition and phosphotransfer in signal transduction. Structure *8*, 851-862.

**Zhang, W., and Shi, L.** (2005). Distribution and evolution of multiple-step phosphorelay in prokaryotes: lateral domain recruitment involved in the formation of hybrid-type histidine kinases. Microbiology (Reading, England) *151*, 2159-2173.

**Zhu, Y., and Inouye, M.** (2004). The HAMP linker in histidine kinase dimeric receptors is critical for symmetric transmembrane signal transduction. The Journal of biological chemistry *279*, 48152-48158.

**Zhu, Y., Qin, L., Yoshida, T., and Inouye, M.** (2000). Phosphatase activity of histidine kinase EnvZ without kinase catalytic domain. Proceedings of the National Academy of Sciences of the United States of America *97*, 7808-7813.

# CHAPTER 2

## How to Switch Off a Histidine Kinase: Crystal Structure of *Geobacillus stearothermophilus* KinB with the Inhibitor Sda

### Introduction

*Bacillus* will go to great lengths to avoid sporulation, including even killing its own kind to scavenge their nutrients (Gonzalez-Pastor et al., 2003). However, if the environment becomes too inhospitable for growth, sporulation will ensue. If a bacterium is forced to sporulate, then it must also ensure that the spore is endowed with a high fidelity copy of the chromosome. Nothing could be worse for a spore than having to use a defective chromosome when it returns to vegetative growth. Therefore, *Bacillus* will confirm the integrity of its genome before sporulating (Ireton and Grossman, 1994; Lemon et al., 2000; Mandelstam et al., 1971). If there are defects, the cell will prevent sporulation, regardless of the environmental nutritional availability. One way cells do this is through the action of a protein called Sda (suppressor of *dnaA*).

Sda was first identified in mutants of *B. subtilis* that were defective for DNA replication but nevertheless could still form endospores (Burkholder et al., 2001). Through a series of elegant genetic experiments a previously unannotated intergenic region of the *B. subtilis* chromosome was shown to encode a 46 amino acid protein, Sda, which was overexpressed in *dnaA*, *dnaB*, and *dnaD* mutants. DnaA is essential for replication initiation; it binds to specific sites within the chromosomal origin and promotes assembly of the replisome (Moriya et al., 1999; Yoshikawa, 1993). The roles of DnaB and DnaD in replication initiation are not well understood. The presence of consensus DnaA binding motifs directly upstream of the *sda* open reading frame led to the hypothesis that DnaA directly regulates *sda* expression (Burkholder et al., 2001).

Burkholder et al. determined that Sda prevents the accumulation of Spo0A~P by directly inhibiting the autophosphorylation reaction of KinA, and likely also KinB (Burkholder et al., 2001; Rowland et al., 2004), providing a developmental checkpoint to ensure that the forespore receives an intact copy of the chromosome. The molecular mechanism of KinA inhibition by Sda has been studied using biochemical and biophysical approaches (Rowland et al., 2004; Whitten et al., 2007). Rowland et al. (Rowland et al., 2004) used NMR to determine that Sda adopts a helical hairpin fold in solution. A cluster of conserved, primarily hydrophobic residues exposed on one face of the hairpin was shown to mediate the interaction with KinA, which occurs via the KinA DHp domain with a stoichiometry of $Sda_2/KinA_2$. Analysis of the effects of Sda on KinA functions led to the conclusion that Sda acts as a molecular barrier to block KinA reactions that require communication between the CA and DHp domains (autophosphorylation, dephosphorylation), but that Sda does not interfere with phosphoryl transfer from KinA~P to Spo0F (Rowland et al., 2004).

A subsequent structural study (Whitten et al., 2007) used a combination of small-angle X-ray scattering (SAXS) and neutron scattering to derive a structural model of the Sda/KinA-CC complex in which Sda binds to the KinA DHp domain at the end distal to the DHp-CA linker. Contrary to Rowland et al. (Rowland et al., 2004), Whitten et al. (Whitten et al., 2007) concluded that Sda inhibition of KinA involves a conformational change within the DHp domain that allosterically affects the DHp-CA 'hinge', rather than a simple molecular barricade.

In this chapter we investigate the structural basis for Sda inhibition of a sporulation HK, KinB from *Geobacillus stearothermophilus* (*Gst*). We determined a 2.0 Å resolution X-ray crystal structure of the intact cytoplasmic CC of KinB complexed with Sda. Structural and biochemical analyses reveal that Sda inhibits autophosphorylation, phosphotransfer, and phosphatase activities of the HK by acting as a simple molecular barricade to sterically block access to the phosphoacceptor His-residue on the DHp domain.

## Results

### *Crystallization and Structure Determination of Sda/KinB-CC*

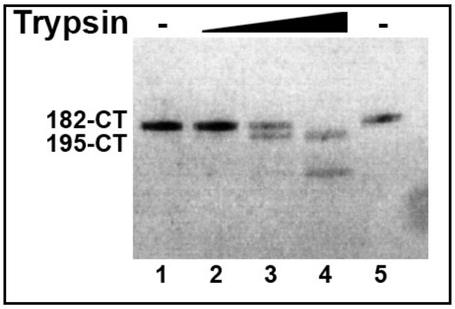


**Figure 2.1**) KinB Trypsinolysis.

KinB, Sda, and Spo0F were identified by sequence-based searches of the incomplete *Geobacillus stearothermophilus* strain 10 genome database (http://www.genome.ou.edu/bstearo.html), and the coding sequences were PCR sub-cloned from *Gst* strain 10 genomic DNA. A combination of secondary structure predictions (Rost, 1996) and the results of limited proteolysis and mass-spectrometry led to the identification of a stable structural core of KinB comprising KinB residues 195-434(C-term), (KinB-CC for Catalytic Core). The proteolysis results were the same in the presence or absence of Sda (Figure 2.1), indicating that Sda is unlikely to interact with the membrane-proximal end of the DHp domain.

Genetic results in *Bsu* strongly suggest that Sda effectively inhibits KinB as well as KinA (Rowland et al., 2004; Trach and Hoch, 1993), but KinB inhibition was not shown directly *in vitro*. In the presence of $\gamma$-[$^{32}$P]ATP, *Gst* KinB-CC underwent spontaneous autophosphorylation (Figure 2.4B, lane 1), and this reaction was inhibited by addition of purified *Gst* Sda (Figure 2.4B, lanes 2-4). The effectiveness of *Gst* Sda inhibition of *Gst* KinB autophosphorylation was similar to *Bsu* Sda inhibition of *Bsu* KinA (Rowland et al., 2004). The coexpressed (Campbell and Darst, 2000) and purified Sda/KinB-CC complex, comprising full-length *Gst* Sda (residues 1-46) and residues 195-434 of *Gst* KinB plus four N-terminal vector-derived residues (GPHM), crystallized in space group $P3_221$. The crystals diffracted to better than 2.0 Å resolution (Figure 2.2, Table 2.1). Crystals only formed in the presence of ADP-$Mg^{2+}$, but not with ATP or the non-hydrolyzable ATP analog, AMPPNP, and not in the absence of nucleotide. The structure was solved using MIRAS phases obtained from a methyl-mercury derivative (Table 2.1), combined with subsequent phased molecular replacement solutions that placed the individual DHp and CA domains from HK TM0853 (Marina et al., 2005) (Figure 2.3). An atomic model of Sda/KinB-CC was built and refined to an $R/R_{\text{free}}$ of 20.0/21.5 at 2.0 Å resolution (Figure 2.4C). Electron density for the 18 C-terminal residues of each KinB monomer (residues 417-434), internal loop residues 198-206 (monomer A) or 201-205 (monomer B), and the 4 C-terminal residues of Sda (residues 43-46), was weak or absent and these regions were presumed to be disordered.

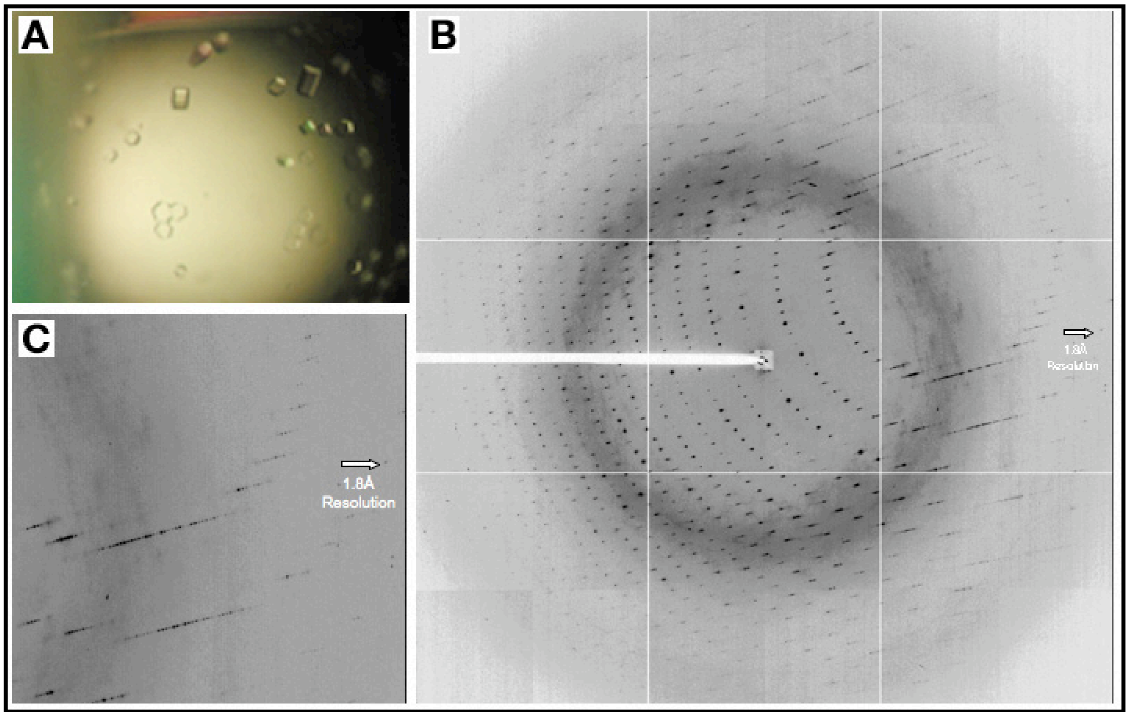


**Figure 2.2)** (A) KinB 195-CT/Sda crystals, grown in a 5 Molar excess of ADP relative to the protein, and 0.2M LiCl and 40% MPD. (B) and (C) KinB/Sda diffraction images (beamline NE-CAT 24-ID-E, APS).

Table 2.1) Crystallographic Analysis

**Diffraction data**

| **Data set**[a] | **Wavelength (Å)** | **Resolution (Å)** | **No. of reflections** (Tot./Unique) | **Completeness** | **$I/\sigma(I)$** | **$R_{sym}$**[b] **(%)** | No. of sites |
|---|---|---|---|---|---|---|---|
| Native | 0.97949 | 50-2.03<br>(2.10-2.03) | 361,127/52,393 | 98.0<br>(92.2) | 31.6<br>(4.5) | 5.3<br>(43.0) | |
| Methyl-HgCl | 1.00800 | 50-2.69<br>(2.79-2.69) | 152,032/38,514 | 88.7<br>(55.3) | 14.2<br>(1.5) | 13.1<br>(53.4) | 2 |

Crystal space group: $P3_221$
Unit cell: a = b = 66.11 Å, c = 315.57 Å
Figure of Merit[c] (45-2.7 Å); 0.18

**Refinement** (against Native data set)

Resolution: 42.41-2.03 Å
$R_{cryst}/R_{free}$[d] 0.200/0.215
rmsd bond lengths: 0.008 Å
rmsd bond angles: 1.096°
rmsd B factor (all atoms): 0.832
Number of water molecules: 373
Residues in favored/allowed/disallowed regions of the Ramachandran plot: 463 (98.5%)/6 (1.3%)/1 (0.2%)

[a] Both data sets were collected at the NE-CAT beamline 24ID-C, APS, Argonne National Laboratory, Argonne, IL; [b] $R_{sym} = \Sigma|I - \langle I\rangle|/\Sigma I$, where $I$ is observed intensity and $\langle I\rangle$ is average intensity obtained from multiple observations of symmetry-related reflections; [c] Figure of merit as calculated by SHARP; [d] $R_{cryst} = \Sigma||F_{observed}| - |F_{calculated}||/\Sigma|F_{observed}|$, $R_{free} = R_{cryst}$ calculated using 5% random data omitted from the refinement.

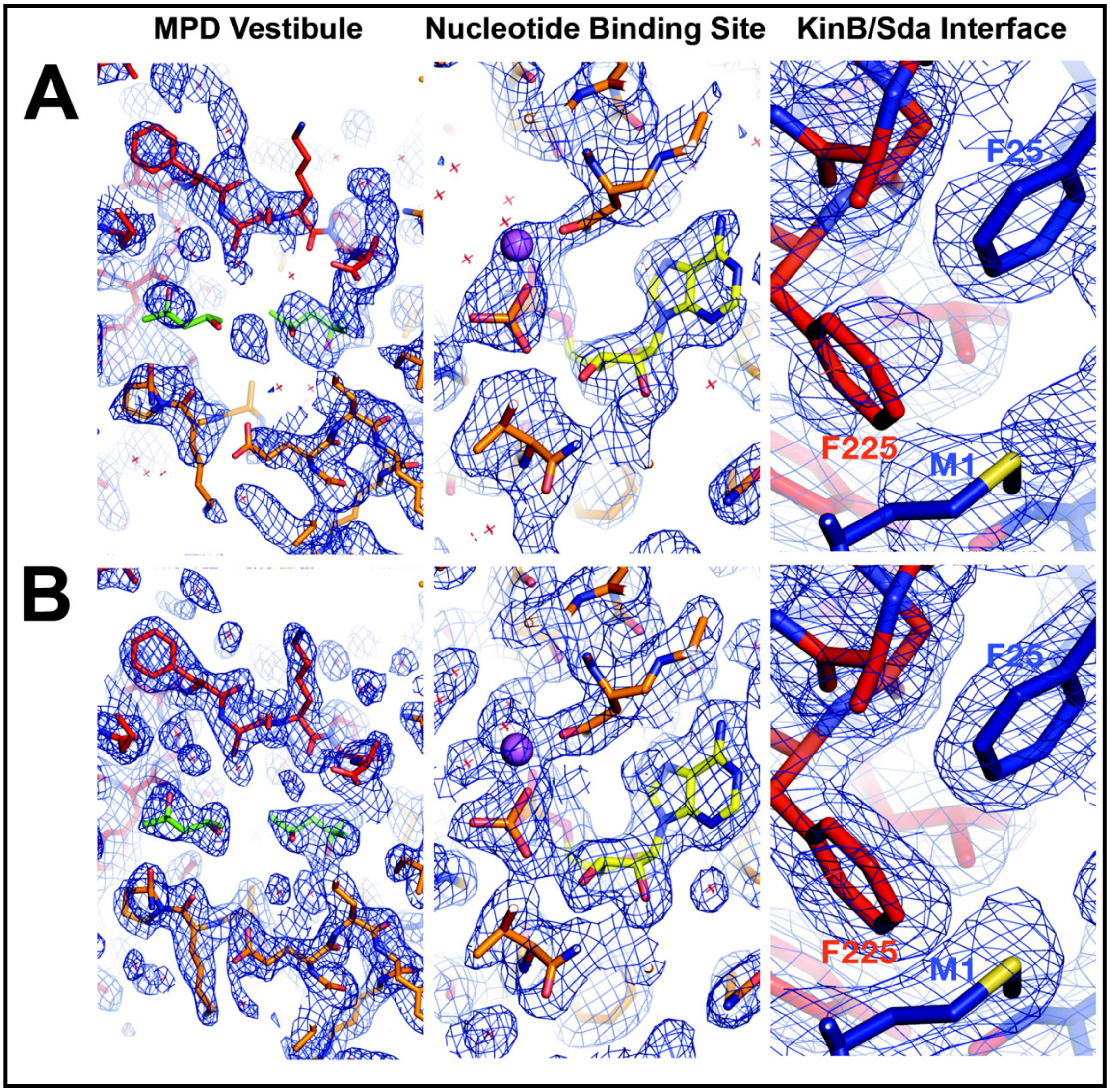


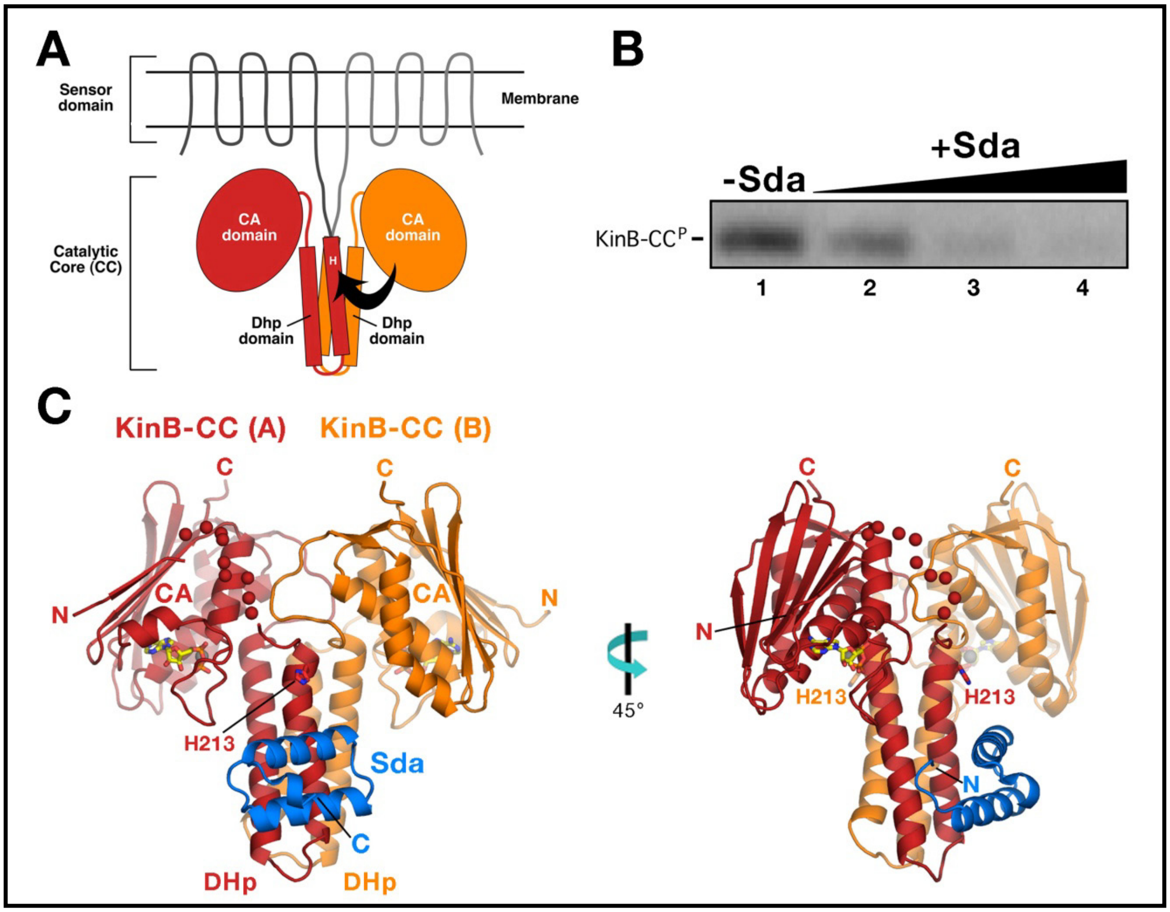
**Figure 2.3)** Electron density maps. Views of experimental (A) and final 2.0 Å resolution $2|F_o| - |F_c|$ (B) maps, contoured at 1σ and sectioned at key locations in the Sda/KinB-CC structure. The model is shown as sticks and colored according to Figure 2.4. Nitrogen and oxygen atoms are colored blue and red, respectively. Water molecules are represented as red stars. Left Panel - The MPD vestibule at the top of the DHp domain. Shown is a view looking from the N-terminus down the four-helix-bundle axis. MPD molecules are colored green. Middle Panel - Nucleotide binding site in the CA domain. ADP carbons are colored yellow. The $Mg^{2+}$ ion coordinating the α- and β-phosphates of the ADP is represented as a magenta sphere. Right Panel - Face-to-side Phe-Phe interaction between KinB Phe225 and Sda Phe25. Also shown is Met1 of Sda. The sulfur atom of Met1 is colored yellow.

### *Overall Structure*

The Sda/KinB-CC structure reveals the expected homodimer of KinB-CC (Marina et al., 2005), but bound to only one Sda molecule (Figure 2.4C). This stoichiometry of $Sda_1/KinB\text{-}CC_2$ was unexpected, as the results of previous biophysical studies (Rowland et al., 2004; Whitten et al., 2007) were consistent with an $Sda_2/KinA\text{-}CC_2$ stoichiometry in solution. We used nanoflow electrospray ionization-mass spectrometry (ESI-MS) (Sharon and Robinson, 2007) to precisely measure the mass of the purified Sda/KinB complex in solution. We found that the complex consisted predominantly (>90%) of an $Sda_2/KinB\text{-}CC_2$ complex (Table 2.2, Figure 2.5). Instead of binding a molecule of Sda, the KinB-CC chain B DHp domain (Figure 2.4C) makes a crystal packing interaction with a crystallographic symmetry-related copy of itself, sterically occluding the Sda binding site (Figure 2.14), likely explaining the discrepancy. Although the affinity of the Sda/sporulation-HK interaction has not been measured, it is expected to be quite high (Rowland, 2004). This is confirmed by our ability to purify the *in vivo* assembled Sda/KinB-CC complex through four chromatographic steps (see Methods) without the addition of extra Sda. Since the stability of the Sda/KinB-CC interaction appears to be due primarily to hydrophobic interactions (see Rowland, 2004, as well as below), it may be that it is destabilized by the high concentration of methyl-pentanediol (MPD), present at approximately 3 M in the crystallization solution.

**Figure 2.4** (adjacent page) The *Gst* Sda/KinB-CC complex. (A) Schematic diagram illustrating structural features of the *Bacillus* sporulation HK KinB. The colored domains (DHp and CA domains), comprising the cytoplasmic CC, have been crystallized in this study. The HK functions as a homodimer (one monomer is colored dark red, the other orange). Autophosphorylation is believed to occur *in trans*, with the orange CA domain phosphorylating the phosphoacceptor-His residue (labeled 'H') of the red DHp domain, though this has not been confirmed for KinB. (B) Sda inhibits spontaneous autophosphorylation of KinB-CC. Incubation of KinB-CC with $\gamma$-[$^{32}$P]ATP results in autophosphorylation, as monitored by SDS-PAGE and phosphorimagery (lane 1). Addition of increasing concentrations of Sda (1, 5, or 25-fold molar excess over KinB-CC, lanes 2-4, respectively) inhibits the autophosphorylation reaction. (C) Structure of the Sda/KinB-CC complex. Ribbon diagrams, with each protein chain color-coded: KinB-CC monomer A, dark red; KinB-CC monomer B, orange; Sda, blue. The KinB DHp and CA domains are labeled. ADP molecules bound to the CA domains are shown in stick format, with carbon atoms colored yellow. The associated $Mg^{2+}$ ions are shown as grey spheres. The side chain of the phosphoacceptor His213 is also shown. Disordered segments near the N-terminus of the KinB monomers are shown as spheres.

Figure 2.4


A
Sensor domain
Membrane
Catalytic Core (CC)
CA domain
CA domain
H
Dhp domain
Dhp domain
B
-Sda
+Sda
KinB-CC$^P$
1
2
3
4
C
KinB-CC (A)
KinB-CC (B)
C
C
N
CA
CA
N
H213
Sda
C
DHp
DHp
45°
C
C
N
H213
H213
N

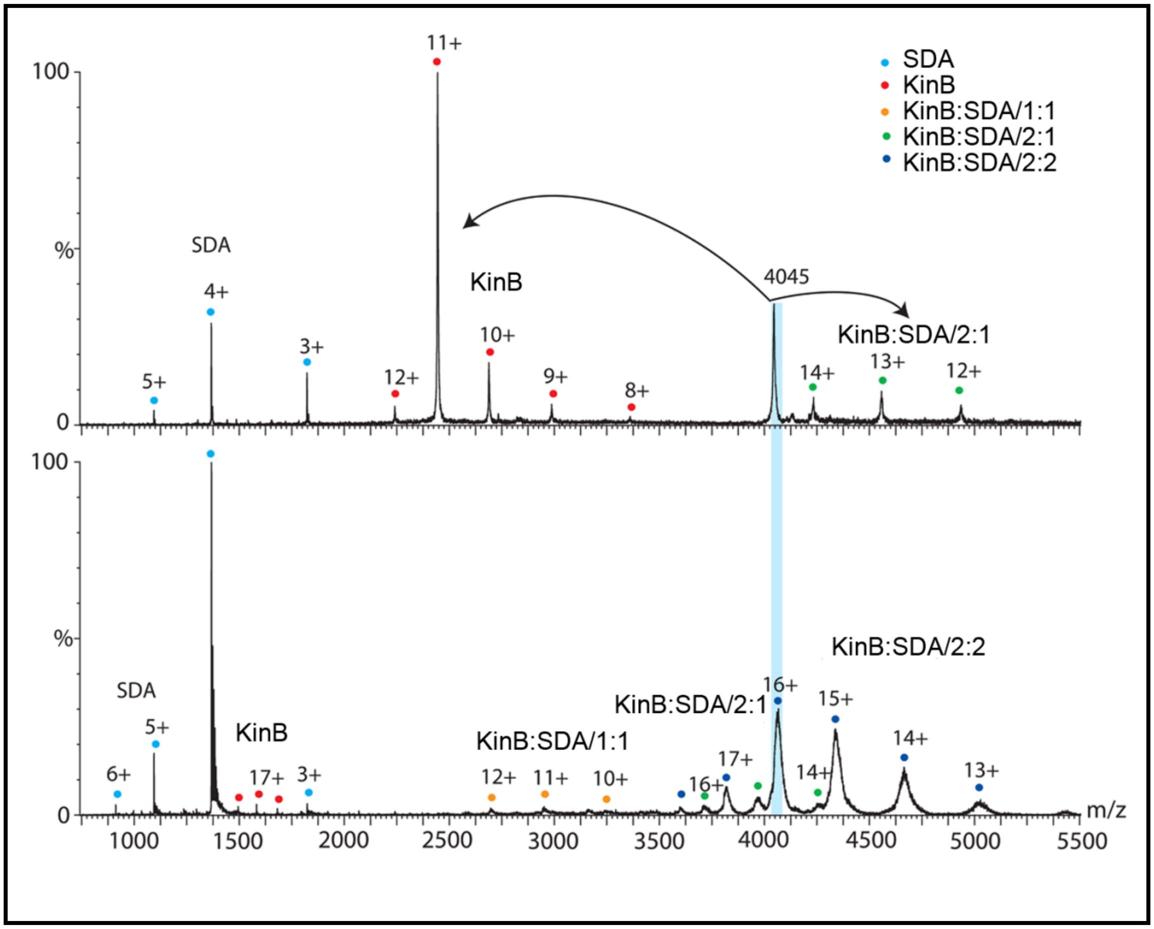


**Figure 2.5)** ESI-MS of Sda/KinB-CC. Mass spectrum of Sda/KinB-CC complex (lower panel) showing that most of the complex (~ 90 %) is present in 2:2 stoichiometry (dark blue circles), approximately 10% is present in 1:2 stoichiometry (green circles), and a small fraction of the complex comprises 1:1 stoichiometry (orange circles). In the region of 1000-2000 m/z there are peaks present for free Sda molecules (light blue circles) and very low intensity peaks for free KinB-CC (red circles). The stoichiometry of the major $Sda_2/KinB\text{-}CC_2$ complex was confirmed by tandem MS (upper panel). Shaded in blue is the isolated 16+ charge state of the $Sda_2/KinB\text{-}CC_2$ complex, which gave rise to dissociated Sda and KinB-CC as well as "stripped" complexes ($Sda_1/KinB\text{-}CC_2$) at higher m/z.

Table 2.2

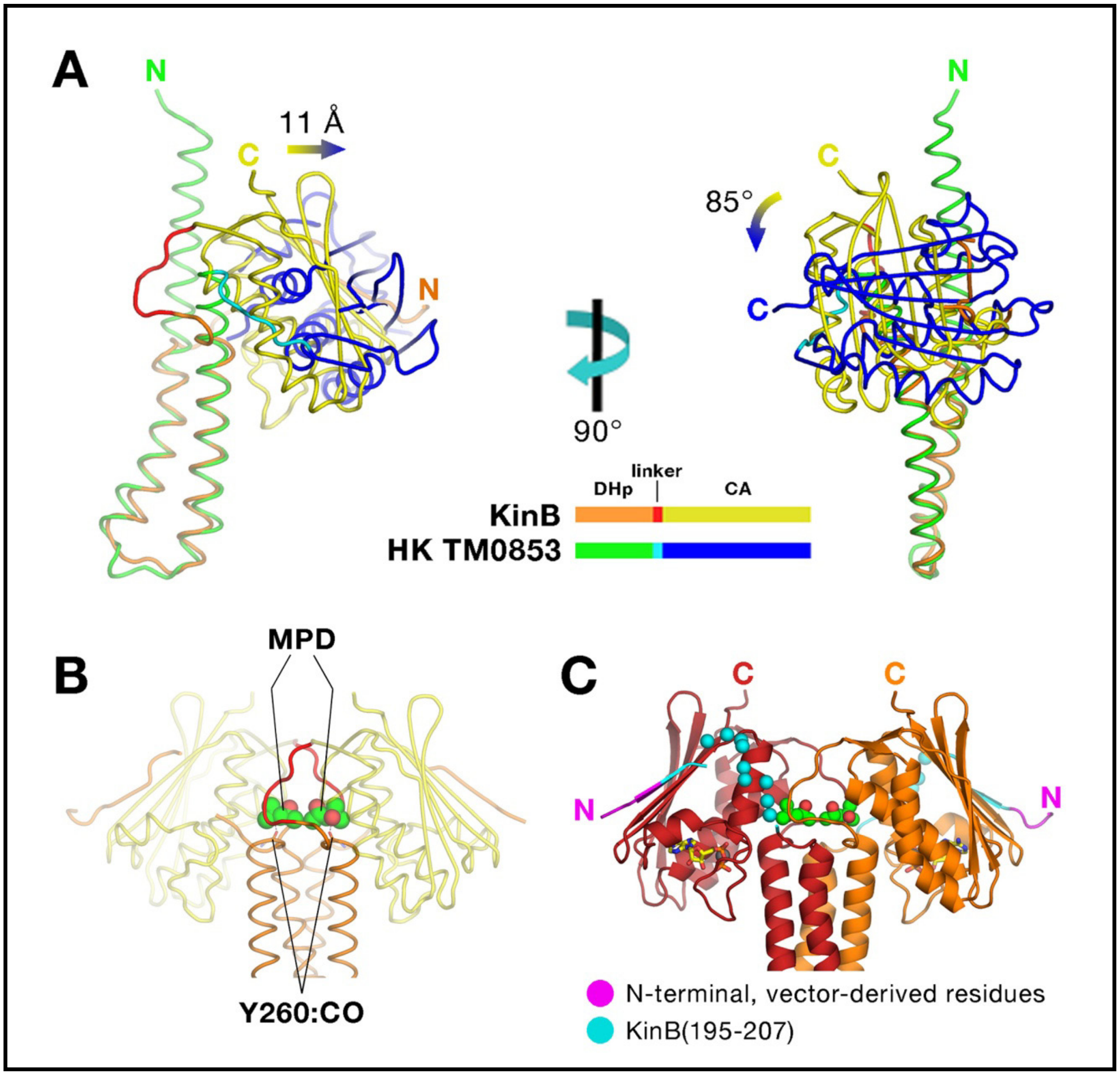
| Protein/complex | Theoretical mass, Da | Experimental mass, Da |
|---|---|---|
| Sda | 5,486 | 5,468 ± 0 |
| KinB-CC | 26,977 | 26,891 ± 2 |
| Sda:KinB-CC/1:1 | 32,463 | 32,460 ± 44 |
| Sda:KinB-CC/2:1 | 59,440 | 59,407 ± 13 |
| Sda:KinB-CC/2:2 | 64,926 | 65,018 ± 66 |

**Table 2.2)** Nanoflow ESI-MS of Sda, KinB-CC, and Sda/KinB-CC.

The KinB-CC homodimer shares similarities with the HK TM0853 structure (Marina et al., 2005), but significant differences are also noted (Figure 2.6A). The overall structure of the KinB CA domain is nearly identical to the HK TM0853 CA domain (1.01 Å root-mean-square-deviation over 116 α-carbon positions). Most of the KinB DHp domain superimposes well with that of HK TM0853, revealing that the disposition of the CA domains with respect to the DHp domains is very different (Figure 2.6A). Alignment of the KinB CA domains with the HK TM0853 CA domains requires a rotation of 85° about an axis roughly perpendicular to the axis of the DHp domain helices, and a translation of the CA domain outward away from the KinB homodimer 2-fold axis by 11 Å (Figure 2.6A). This 'pinched-in' disposition of the CA domains of KinB may be stabilized by the binding of two solvent MPD molecules at the top of the DHp domain, in between the two CA domains (Figure 2.6B). Both secondary structure prediction (Rost, 1996) and structural alignment with HK TM8053 (Figure 2.6A) indicates that the second DHp

**Figure 2.6** (adjacent page) Unusual structural features of KinB-CC. (A) Comparison of the CC's of KinB (orange/red/yellow) and TM0853 (green/cyan/blue). The two structures, shown as backbone worms, were superimposed on overlapping helices of the DHp domains, which superimpose closely (1.12 Å rmsd over 43 α-carbon positions corresponding to KinB chain B residues 211-231 and 239-260). The superposition reveals the different paths of the DHp-CA linkers (KinB-CC, red; TM0853, cyan) and the misalignment of the CA domains (KinB-CC, yellow; TM0853, blue). (B) Solvent MPD molecules (shown in CPK format with carbon atoms in green) bound in a vestibule at the top of the DHp helices (orange), and surrounded by the DHp-CA linkers (red) and the CA domains (transparent yellow). The MPD molecules hydrogen bond to the carbonyl oxygens of Tyr260, disrupting the second DHp α-helix. (C) Disordered N-terminal arm of KinB-CC. Rather than forming an α-helix at the N-terminus (as in TM0853, see Figure 6.2A), the N-terminal 17 residues of KinB-CC form a sixth β-strand at the end of the CA domain β-sheet, which is connected to the beginning of the first DHp α-helix by a disordered loop (residues 195-207, cyan spheres).

Figure 2.6



α-helix should continue to at least residue 263, but the MPD molecules hydrogen bond to the carbonyl oxygens of Tyr260 (MPD-O2:Tyr260-O distance, 2.7 Å), apparently prematurely disrupting the second DHp α-helix (Figure 2.6B).

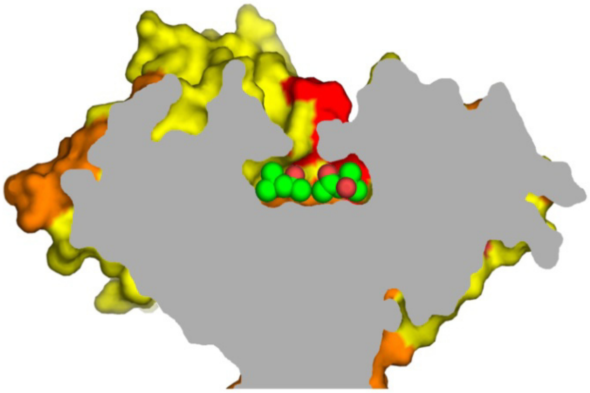


Figure 2.7) Surface view of MPD vestibule. KinB-CC is shown as a molecular surface (DHp domain, orange; DHp-CA linker, red; CA domain, yellow) and cross-sectioned to show the MPD vestibule. MPD molecules are represented in CPK format, with carbon and oxygen atoms colored green and red, respectively.

The MPD molecules bind in an otherwise hydrophobic vestibule formed by the top of the DHp α-helices, the DHp-CA linker, and the CA domains, and are completely buried inside the protein (Figure 2.7). Moreover, secondary structure prediction (Rost, 1996) and structural alignment with HK TM0853 (Figure 2.6A) also indicates that the first DHp α-helix should extend through to the N-terminus of the KinB-CC construct, corresponding to KinB residue 195, but the extension of the α-helix N-terminal to KinB residue 211 is incompatible with the presence of the MPD molecules. Instead, the N-terminal segments of the KinB-CC construct (four vector-derived residues, plus KinB residues 195-210) form a partially disordered loop that allows the N-terminal arms to add a sixth β-strand onto the conserved five-stranded β-sheet of the CA domains (Figure 2.6C).

A large-scale conformational change is required to reposition the CA domain with respect to the DHp domain for autophosphorylation to occur (Marina et al., 2005). We entertained the idea that the structural differences described above were somehow a consequence of Sda binding, and that the extra β-strand, allowed by the unstructured N-terminal tail, played a role in Sda-mediated inhibition of autophosphorylation, perhaps by tethering the CA domain to the DHp domain and restricting its motion. Two structural observations argue against this hypothesis. First, N-terminal residues participating in forming the sixth β-strand include vector-derived residues (magenta in Figure 2.6C), and second, the sequence register of the sixth β-strand with respect to the CA domains is different for the two KinB-CC monomers, indicating that this interaction is not specific. Nevertheless, we tested this hypothesis by comparing Sda inhibition of KinB-CC and an N-terminal truncation of KinB-CC beginning at residue 208, termed ΔN-KinB-CC, that lacks the N-terminal arm and therefore would not be able to form the putative tethering sixth β-strand. There was no obvious difference in the effectiveness of Sda inhibition of KinB-CC or ΔN-KinB-CC, indicating that the KinB-CC N-terminal arm does not play a role in Sda-mediated inhibition of autophosphorylation activity (data not shown).

### *Sda and the Sda/KinB Interface*

Sda adopts the antiparallel helical hairpin structure as reported for the NMR structure of *Bsu* Sda (Rowland et al., 2004). Sda α-helix 1 (residues 6-19) connects to α-helix 2 (residues 23-36) by a short interhelical loop (residues 20-22). The N-terminal loop (residues 1-5) curves back towards KinB, where it is wedged between Sda and KinB and participates in interactions with the KinB DHp domain. The C-terminal loop (residues 37-46) folds back on the opposite face of the hairpin, away from KinB. Electron density for the C-terminal four residues (43-46) is missing, and this region is presumed to be disordered. This C-terminal tail is required for degradation of Sda by the ClpXP protease, which in turn is required for efficient induction of sporulation after transient replicative stress (Ruvolo et al., 2006).

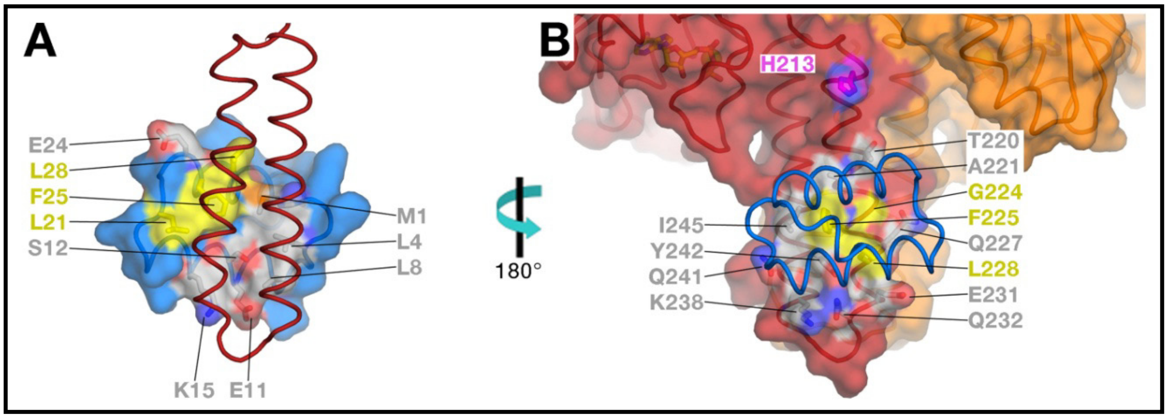
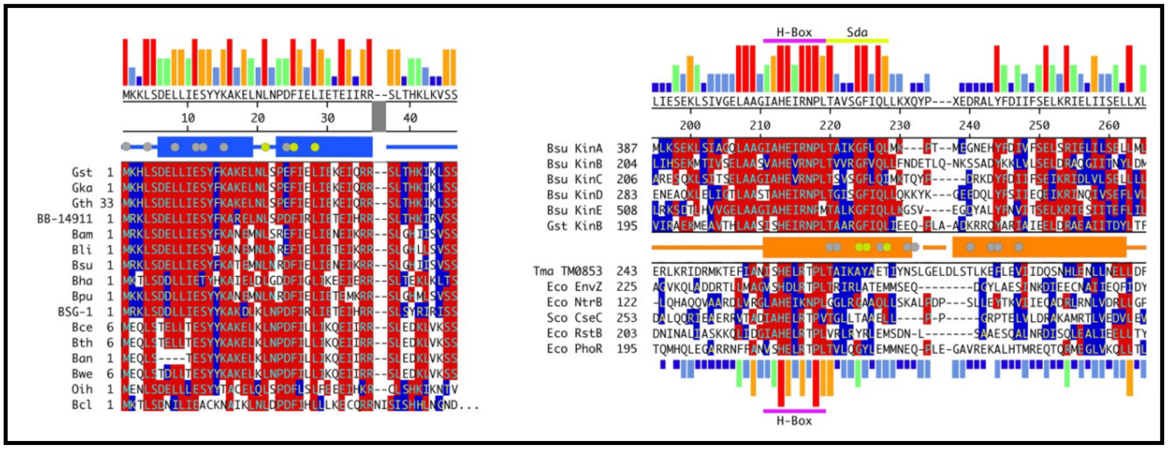


**Figure 2.8)** Structural and functional features of the Sda/KinB interface. (A) The KinB binding surface of Sda, viewed from the KinB perspective. The KinB chain A DHp helices are shown as a dark red backbone worm. Sda is shown as a blue backbone worm, along with a transparent molecular surface. Side chains that contact KinB (≤4 Å) are shown and color-coded (carbon, light grey; nitrogen, blue; oxygen, red; sulfur, orange), except the carbon atoms of three residues (Leu21, Phe25, and Leu28) shown to be important for *Bsu* KinA binding (Rowland et al., 2004) are colored yellow. (B) The Sda binding surface of KinB, viewed from the Sda perspective. Sda is shown as a blue backbone worm. KinB is shown as a backbone worm (chain A, dark red; chain B, orange), along with a transparent molecular surface. Side chains that contact Sda (≤4 Å) are shown and color-coded (carbon, light grey; nitrogen, blue; oxygen, red), except the carbon atoms of three residues (Gly224, Phe225, and Leu228) shown to be important for Sda binding (Figure 2.11A) are colored yellow.

**Figure 2.9** (adjacent page) (left) Sequence conservation among 16 Sda orthologs. The sequence at the top shows the consensus sequence, while the histogram above it denotes the level of sequence conservation at each position (red bar, 100% conserved; dark blue bar, less than 20%). The sequences are shown in one-letter amino acid code and identified by species at the left (annotation of the species codes shown is given at the end of the Discussion). The numbers at the beginning of each line indicate amino acid positions relative to the start of each protein sequence. The numbers at the top (underneath the consensus) indicate the amino acid position in *Gst* Sda. Positions in the alignment that share >50% identity with the consensus are indicated by red shading, while positions that share >50% homology are indicated by blue shading. The α-helices in the *Gst* Sda structure are indicated above the Sda sequences as blue rectangles, connecting loops are indicated by a continuous blue line. *Gst* Sda positions that contact KinB (≤4 Å) are denoted by grey dots, with the three residues shown to be important for Sda binding to *Bsu* KinA (Rowland et al., 2004) denoted by yellow dots. (right) Sequence conservation within the DHp domains of *Bacillus* sporulation HKs (top set of sequences) and other two-component, homodimeric sensor HKs (bottom set of sequences). The consensus sequence is shown on top, while the histogram above it denotes the level of conservation at each position for just the sporulation HKs (red bar, 100% conserved; dark blue bar, less than 20%). The histogram at the bottom denotes sequence conservation among all the HKs. The sequences are shown in one-letter amino acid code and identified by species at the left (Sco, *Streptomyces coelicolor*). The numbers at the beginning of each line indicate amino acid positions relative to the start of each protein sequence. The numbers at the top (underneath the consensus) indicate the amino acid position in *Gst* KinB. Positions in the alignment that share >50% identity with the consensus are indicated by red shading, while positions that share >50% homology are indicated by blue shading. The α-helices in the *Gst* KinB structure are indicated below the *Gst* KinB sequence as orange rectangles, connecting loops are indicated by a continuous orange line. *Gst* KinB positions that contact Sda (≤4 Å) are denoted by grey dots, with the three residues shown to be important for KinB binding to Sda (Figure 2.11A) denoted by yellow dots. The H-box, a sequence motif conserved among all the HKs, is denoted by the horizontal purple line above and below the top and bottom histograms, respectively. The Sda binding motif conserved only among the sporulation HKs is denoted by the horizontal yellow line above the sporulation HK histogram.

Figure 2.9



The Sda/KinB interface buries 726 Å$^2$ of predominantly hydrophobic surface (total for both proteins), and the most highly conserved and functionally important residues involved in the interface are non-polar (Figure 2.8). Rowland et al. (Rowland, 2004) investigated a number of amino acid substitutions in *Bsu* Sda at positions revealed by the *Gst* Sda/KinB-CC structure to be involved in the Sda/KinB interface (Figure 2.8A), including substitutions at Sda positions 8, 11, 12, 15, 21, 24, 25, and 28 (*Bsu* and *Gst* Sda numbering is identical; Figure 2.9). Leu21 and Phe25 were found to be essential for Sda binding to KinA and inhibition of KinA autophosphorylation, while substitution of Leu28 resulted in partial loss of function (Rowland et al., 2004). These three residues (Leu21, Phe25, Leu28; yellow in Figure 2.8A) form a prominent, contiguous hydrophobic patch on one face of Sda, and are universally conserved among Sda homologs (Figure 2.9). Also contributing to the hydrophobic patch is the side-chain of Met1, which is wedged between KinB and the rest of Sda and makes Van der Waals' interactions with Phe225 of KinB. *Bsu* Sda-Met1 interactions with KinA were noted by Rowland et al. (Rowland, 2004) from NMR chemical-shift experiments. Rowland et al.

also noted that substitutions of *Bsu* Sda Met19 resulted in partial loss of function. A shorter Leu side-chain is found at this position in *Gst* Sda, and it does not make Van der Waal's contact (using a cutoff of ≤4 Å) with KinB, suggesting that the role of the hydrophobic residue found at position 19 in Sda function is species specific.

Sda binds at the bottom of the KinB DHp domain, distal to the membrane, near the turn connecting the two KinB DHp helices, with the axes of the Sda helices roughly perpendicular to the KinB DHp helices (Figure 2.4C). Sda interacts with residues from both DHp α-helices, as well as from the turn connecting them. Sda does not interact with the KinB phosphoacceptor His213 (the closest approach of any Sda atom to His213 is

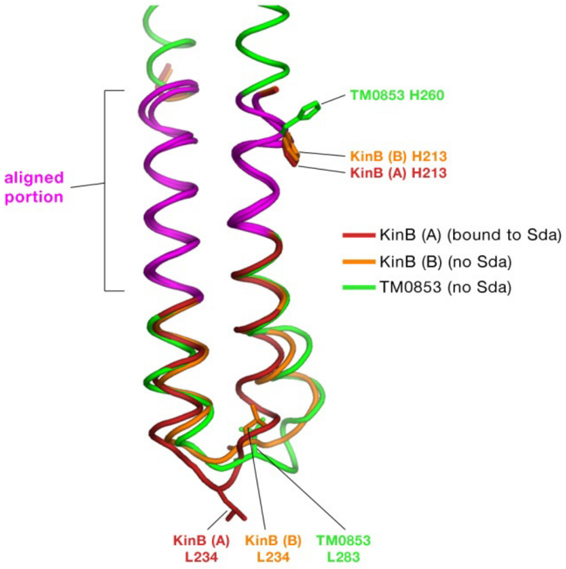


**Figure 2.10)** Comparison of DHp helices with and without Sda. Superimposition of KinB-CC DHp domains from KinB-CC chain A (bound to Sda), KinB-CC chain B (no Sda), and TM0853 (no Sda). The portion of the three DHp domains used in the alignment is colored magenta. Chain B aligns quite well with TM0853. The bottom of the first DHp α-helix of the DHp domains from TM0853 and KinB-CC chain B (both without Sda) splay outward, while the same region of KinB-CC chain A (bound to Sda) is parallel with the second DHp domain α-helix. In the absence of Sda, a Leu residue (KinB-CC Leu234, TM0853 Leu283) in the loop connecting the DHp α-helices point towards the hydrophobic core of the four-helix bundle. In the presence of Sda, the first DHp α-helix adopts a more parallel structure, and Leu234 flips out towards the solvent.

about 7 Å (Figure 2.8B)), does not interact with any residues in the conserved H-box, and also does not interact with any CA-domain residues.

The KinB chain B (not bound to Sda) DHp α-helices superimpose very well onto TM0853 (Figure 2.6A), but superimposition of KinB chain A (bound to Sda) reveals a small conformational change in the first DHp α-helix (Figure 2.10). In the two proteins in the absence of Sda, the bottom two turns of the first DHp α-helix splay outward away from the second DHp α-helix, while the same portion of KinB chain A (bound to Sda) is straight and parallel with the second DHp α-helix (Figure 2.10). In order to accommodate the inward bend of the first α-helix, KinB chain A Leu234, in the middle of the loop connecting the two DHp α-helices, flips out, while in KinB chain B and TM0853 the corresponding Leu points in between the two α-helices to occupy the space created by the outward bend of the first DHp α-helix (Figure 2.10). We believe this conformational change induced by Sda binding is highly localized and does not propagate to other parts of the structure, since α-carbons of the KinB DHp α-helices, excluding residues that interact with Sda, superimpose extremely closely with HK TM0853 (rmsd of 0.18 Å over a total of 21 α-carbons; residues 211-219 of the first DHp α-helix, including the phosphoacceptor His213, and residues 246-260 of the second α-helix; Figure 2.10).

As expected from the hydrophobic nature of the KinB binding determinants on Sda (Figure 2.8A), the KinB surface buried by the interaction with Sda is also predominantly hydrophobic (Figure 2.8B). Among the residues of KinB that interact with Sda (Figure 2.8B), sequence analysis reveals a motif uniquely conserved among the sporulation HKs, immediately following the H-box, which is conserved among all HKs

(Figure 2.9). KinB residues Gly224, Phe225, and Leu228 make the majority of contacts with Sda. Substitutions at these positions confer strong resistance to Sda inhibition (Figure 2.11A, 2.12), and the mutant KinB proteins do not bind Sda *in vitro*, as judged by native gel electrophoresis mobility shift analysis (not shown), while substitutions at other positions (Thr220, Gln227, Glu231) are relatively benign (Figure 2.11A, 2.12). All of these mutants, with the exception

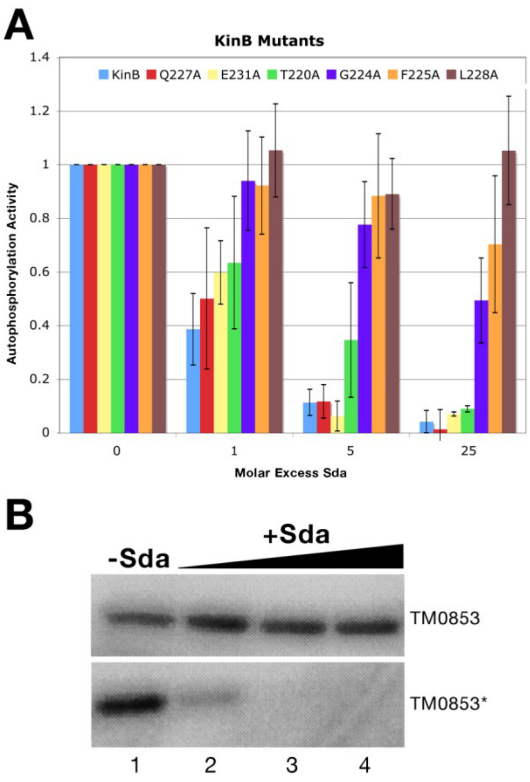


**Figure 2.11)** An Sda binding motif on the HK. (A) Effect of Ala substitutions in KinB on Sda inhibition. Single Ala substitutions were introduced at residues of KinB making extensive contacts with Sda. Autophosphorylation of wild-type KinB-CC (blue) and mutants was monitored in reactions containing increasing concentrations of Sda (0, 1, 5, and 25-fold molar excess over KinB-CC). The results are expressed as normalized histograms (autophosphorylation activity in the absence of Sda for each HK was used as reference), with error bars denoting the standard error from a minimum of three experiments for each histogram. KinBs harboring Q227A (red) or E231A (yellow) substitutions were essentially as sensitive to Sda inhibition as wild-type KinB, while G224A (purple), F225A (orange), and L228A (brown) were highly resistant (T220A was intermediate). (B) Three-residue Sda binding motif. Autophosphorylation by the CC of the homodimeric HK TM0853 is insensitive to Sda (1, 5, and 25-fold molar excess Sda, lanes 1, 2, and 3, respectively), while the triple-mutant TM0853* (A271G/Y272F/T275L, introducing side-chains corresponding to KinB G224, F225, and L228) is inhibited very effectively.

of Gly224Ala, expressed to wild-type levels, were soluble, and exhibited normal autophosphorylation activity. The KinB-CC-Gly224Ala substitution expressed to lower levels and was less soluble than wild-type KinB-CC suggesting a mild folding defect, but was otherwise normal in its autophosphorylation activity. KinB-Phe225 participates in a face-to-side Phe-Phe interaction with Phe25 of Sda (Hunter et al., 1991) (Figure 2.3C), while the Gly at position 224 of KinB allows the close approach of Sda-Phe25. Any side-chain at this position of KinB would clash with Sda-Phe25. Surprisingly, a KinB-Gln227Ala substitution did not significantly affect binding to, or inhibition by, Sda (Figure 2.11A, 2.12). Gln 227 is highly conserved among sporulation HKs (Figure 2.9) and makes several contacts with the crucial Leu21 of Sda. Less surprisingly, substitution of KinB-Glu231 also resulted in wild-type levels of Sda binding

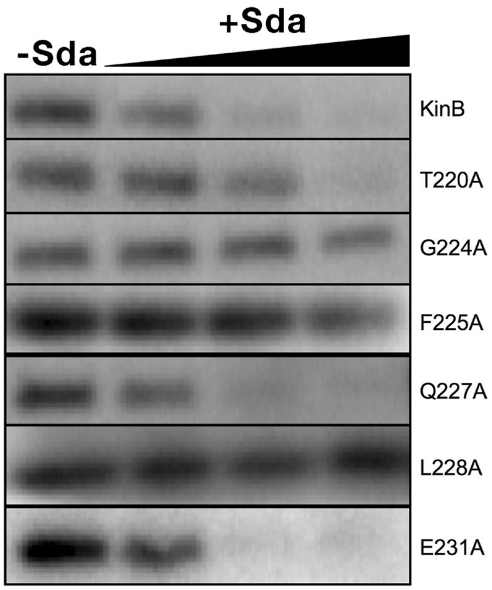


**Figure 2.12)** Effect of Ala substitutions in KinB on Sda inhibition. Ala substitutions in KinB were generated based on contacts with Sda observed in the crystal structure. Autophosphorylation reactions were carried out for 1 hr at 25°C. Reactions were quenched and analyzed by SDS-PAGE and phosphorimagery. The leftmost column shows the autophosphorylation level of each mutant in the absence of Sda. Columns 2, 3, and 4 contained a 1, 5, and 25-molar excess of Sda, respectively (with respect to KinB-CC). KinB-CC mutants G224A, F225A, and L228A were strongly resistant to Sda inhibition.

and sensitivity. This residue makes contacts primarily with Sda-Lys15, a residue shown not to be important for binding KinA (Rowland et al., 2004).

To underscore the importance of KinB residues Gly224, Phe225, and Leu228 for Sda function, and to test whether these three side chains are sufficient for Sda binding and inhibition of autophosphorylation activity, we constructed a triple-mutant of *Tma* HK TM0853 (Ala271Gly, Tyr272Phe, Thr275Leu, hereafter termed HK TM0853*), placing the GFXXL motif into the context of an HK not normally subject to Sda inhibition. While autophosphorylation activity of wild-type HK TM0853 was completely unaffected by Sda, HK TM0853* was inhibited at least as effectively as KinB (Figure 11B). Thus, the sporulation kinase GFXXL motif (Figure 2.11B) appears to be sufficient for Sda binding and inhibition of autophosphorylation.

## *Sda/KinB Stoichiometry*

The unexpected 1:2 stoichiometry of Sda:KinB-CC observed in the crystal structure led us to investigate the functional stoichiometry of autophosphorylation inhibition by Sda. One molecule of Sda may only inhibit one monomer of the KinB homodimer, so that two Sda molecules bound to the KinB homodimer would be required for complete inhibition of autophosphorylation. Alternatively, one molecule of Sda may be sufficient to inhibit autophosphorylation by both KinB monomers within the homodimer. Furthermore, since HK autophosphorylation occurs *in trans* (i.e. the chain A-CA domain phosphorylates His213 of chain B, and *vice versa*; (Ninfa et al., 1993; Qin et al., 2000; Yang and Inouye, 1991)), a molecule of Sda bound to chain A could inhibit chain A-CA domain function, inhibiting phosphorylation of chain B-His213 by the chain

A-CA domain, or it could inhibit chain A-DHp domain function by inhibiting phosphorylation of chain A-His213 by the chain B-CA domain (research published subsequent to the completion of the work described in this chapter has called into question the universality of HK trans phosphorylation (Casino et al., 2009; Pena-Sandoval and Georgellis), and we have not shown whether KinB phosphorylates *in cis* or *in trans*, see discussion).

To address these questions, we generated mixed dimers of KinB-CC, in which one monomer was resistant to Sda binding and inhibition (Sda$^R$), while the other was wild-type. The KinB-Phe225Ala mutation was used to generate Sda$^R$ KinB-CC (Figure 2.11A, 2.12). KinB-CC proteins (wild-type and Sda$^R$), either with or without an N-terminal His$_6$-tag, were co-expressed in *Eco*. Co-expression was expected to produce three dimeric KinB-CC species, homodimers of His$_6$-KinB-CC or KinB-CC, as well as the desired heterodimer (His$_6$-KinB-CC/KinB-CC). Purification of the heterodimers was achieved using Ni$^{2+}$-affinity chromatography and elution with a shallow imidazole gradient. Untagged KinB-CC homodimers flowed through the column, while the heterodimer eluted earlier in the gradient than the tagged homodimers. The purified heterodimers were tested in autophosphorylation assays. Autophosphorylation of the heterodimers was not fully inhibited, even in the presence of a 25-fold molar excess of Sda (Figure 2.13). Instead, only phosphorylation of the monomer that was competent to bind Sda was inhibited, while phosphorylation of the Sda$^R$ monomer was unaffected (Figure 2.13). Heterodimers in which both of the monomers were Sda-sensitive were inhibited to wild-type levels (data not shown). We conclude that Sda directly inhibits phosphoacceptor function of the DHp domain to which it is bound. In other words, Sda bound to the chain A-DHp domain inhibits phosphorylation of chain A-His213 by the chain B-CA domain without significantly affecting phosphorylation of chain B-His213 by the chain A-CA domain.

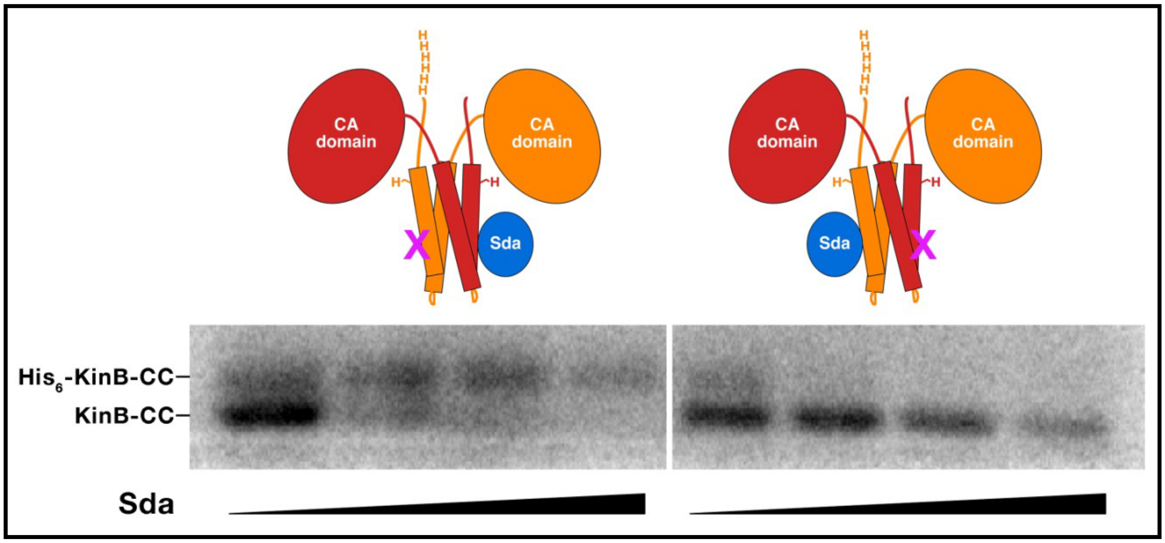


**Figure 2.13)** Sda inhibits DHp function only of the DHp domain to which it is bound. Mixed dimers of $His_6$-KinB-CC and KinB-CC, containing either wild-type (Sda-sensitive) or $Sda^R$ monomers, were generated and purified. The $His_6$-KinB-CC had a slightly lower mobility by SDS-PAGE, allowing the autophosphorylation of each monomer to be monitored as a function of Sda concentration (0, 1, 5, and 25-fold molar excess, left to right). In the left panel, the $His_6$-KinB-CC (orange monomer in the schematic) is $Sda^R$ due to an F225A substitution (illustrated in the schematic by the magenta 'X'), while the KinB-CC (dark red) is competent to bind Sda. Phosphorylation of KinB-CC is inhibited by Sda, while phosphorylation of $His_6$-KinB-CC is resistant. In the right panel, the $His_6$-KinB-CC is competent to bind Sda, while the KinB-CC harbors the $Sda^R$ mutation. In this case, phosphorylation of $His_6$-KinB-CC is inhibited by Sda, while phosphorylation of KinB-CC is resistant.

We also conclude that the Sda/KinB-CC complex is 2:2 in solution, and presumably the Sda/KinB complex is 2:2 in the normal biological context at the plasma membrane *in vivo*, since an Sda molecule bound to each KinB monomer is required for the complete inhibition of autophosphorylation activity that is observed. The 1:2 stoichiometry observed in the crystal structure is likely due to a crystal packing artifact (Figure 2.14).

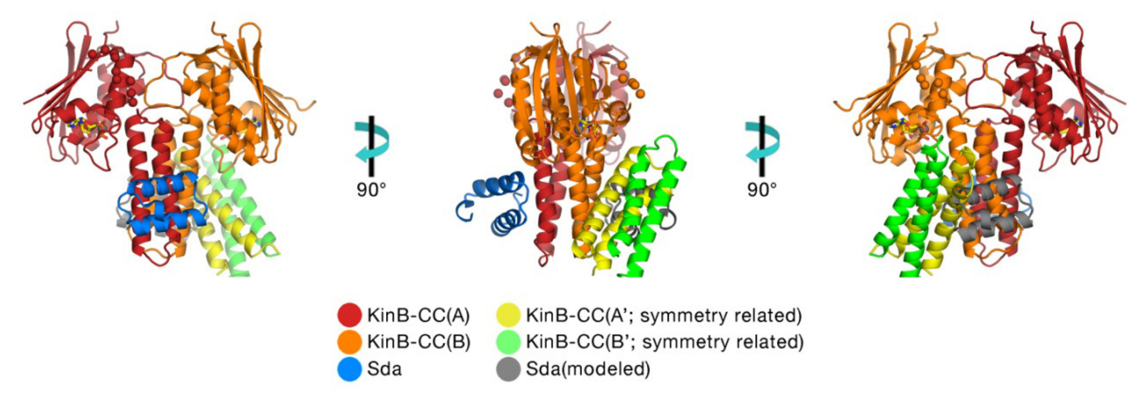


**Figure 2.14)** Symmetry-related KinB-CC occludes the Sda binding site on one KinB-CC monomer. Three orthogonal views are presented, showing the location of a symmetry-related KinB-CC DHp domain (colored green, chain A; yellow, chain B). The KinB-CC DHp domain chain B makes a crystal packing interaction with a crystallographic symmetry-related copy of itself. When Sda is modeled on to the KinB-CC DHp domain chain B (KinB-CC chain B, orange; modeled Sda, grey), clear steric clashes between the modeled Sda and the symmetry-related KinB-CC DHp domain are revealed. This feature of the crystal packing likely explains the Sda$_1$/KinB-CC$_2$ stoichiometry observed in the crystal structure.

### *Mechanism of Autophosphorylation Inhibition by Sda*

In both available structures of intact HK CC's, this structure and HK TM0853 (Marina et al., 2005), the disposition of the catalytic CA domain with respect to the phosphoacceptor DHp domain is inappropriate for the autophosphorylation reaction; the side chain of the phosphoacceptor His and the β-phosphate of the nucleotide bound to the CA domain are about 25 Å apart and incorrectly oriented for the phosphorylation reaction. Marina et al. (Marina, 2005) constructed a model for the autophosphorylation reaction by docking the CA domain onto the DHp domain in a manner consistent with phosphotransfer from ATP to the His phosphoacceptor. Superimposition of the KinB CA and DHp domains onto the HK TM0853 autophosphorylation model reveals steric clashes between the positioned CA domain and Sda (Figure 2.15A). These clashes would be expected to inhibit the autophosphorylation reaction.

### *Sda Inhibits Phosphotransfer to Spo0F*

Rowland et al. (Rowland et al., 2004) reported that Sda inhibited HK autophosphorylation by blocking communication between the CA and DHp domains, a conclusion consistent with our structural analysis (Figure 2.15A). Rowland et al. (Rowland et al., 2004) also reported that Sda did not interfere with phosphotransfer from the HK phospho-His residue to the Asp residue of Spo0F, the next protein in the sporulation phosphorelay. This finding is inconsistent with structural analysis based on a SAXS/neutron scattering model of the Sda/KinA-CC complex and a model of the Spo0F/KinA-CC complex, which indicated that binding of Sda or Spo0F to the KinA-DHp domain were mutually exclusive due to steric clashes (Whitten et al., 2007). We constructed a structural model of the Spo0F/KinB-CC complex (with the Spo0F-Asp approaching KinB-His213) based on the crystal structure of the Spo0F/Spo0B complex (PDB ID 1F51) (Zapf et al., 2000). Such a model is expected to be fairly representative since Spo0F binds to a His-containing four-helix bundle on Spo0B that is structurally similar to the HK-DHp domain. The precise details of the Spo0F/KinB-CC model are

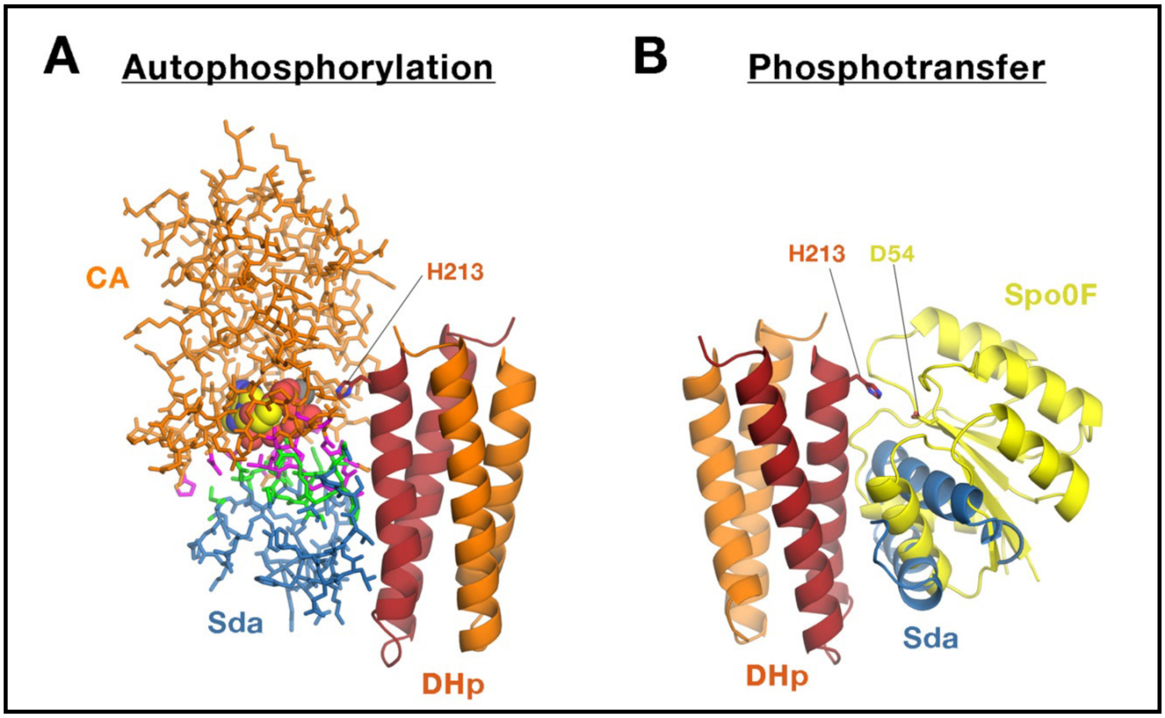


**Figure 2.15)** Mechanism of Sda inhibition of autophosphorylation and phosphotransfer. (A) Mechanism of Sda inhibition of autophosphorylation. The KinB DHp domain is shown as a ribbon diagram, with the side chain of the phosphoacceptor His213. Sda is shown in stick format. Superimposed on the Sda/KinB-DHp structure is a KinB CA domain (shown in stick format but with the ADP and $Mg^{2+}$ shown as CPK spheres) poised for the autophosphorylation reaction (modeled according to Marina et al. (Marina et al., 2005)). Sda is colored blue, except atoms within 4 Å of the modeled CA domain are colored green. The KinB CA domain is colored orange, except atoms within 4 Å of Sda are colored magenta. The green and magenta atoms on Sda and KinB-CA, respectively, illustrate the extent of steric clash between the bound Sda and the modeled CA domain, indicating that phosphorylation of His213 by the CA domain would be sterically blocked by the bound Sda. (B) Mechanism of Sda inhibition of phosphotransfer to Spo0F. Sda (blue) and the KinB DHp domain (dark red and orange) are shown as a ribbon diagram, with the side chain of His213. Superimposed on the Sda/KinB-DHp structure is Spo0F (yellow ribbon, with the side chain of the phosphoacceptor Asp54) poised for phosphotransfer (modeled based on the Spo0F/Spo0B crystal structure (Zapf et al., 2000)). Extensive steric overlap between Sda and Spo0F indicates that the phosphotransfer reaction would be sterically blocked by the bound Sda.

]

not critical, however, as the superposition of the modeled position of Spo0F and the crystallographically observed position of Sda on the KinB DHp domain result in severe steric overlap (Figure 2.15B). Thus, our structural modeling predicts that Sda binding to the KinB-DHp domain should interfere with phosphotransfer from KinB-His213 to Spo0F, consistent with the modeling of Whitten et al., but inconsistent with the conclusions of Rowland et al (Rowland et al., 2004).

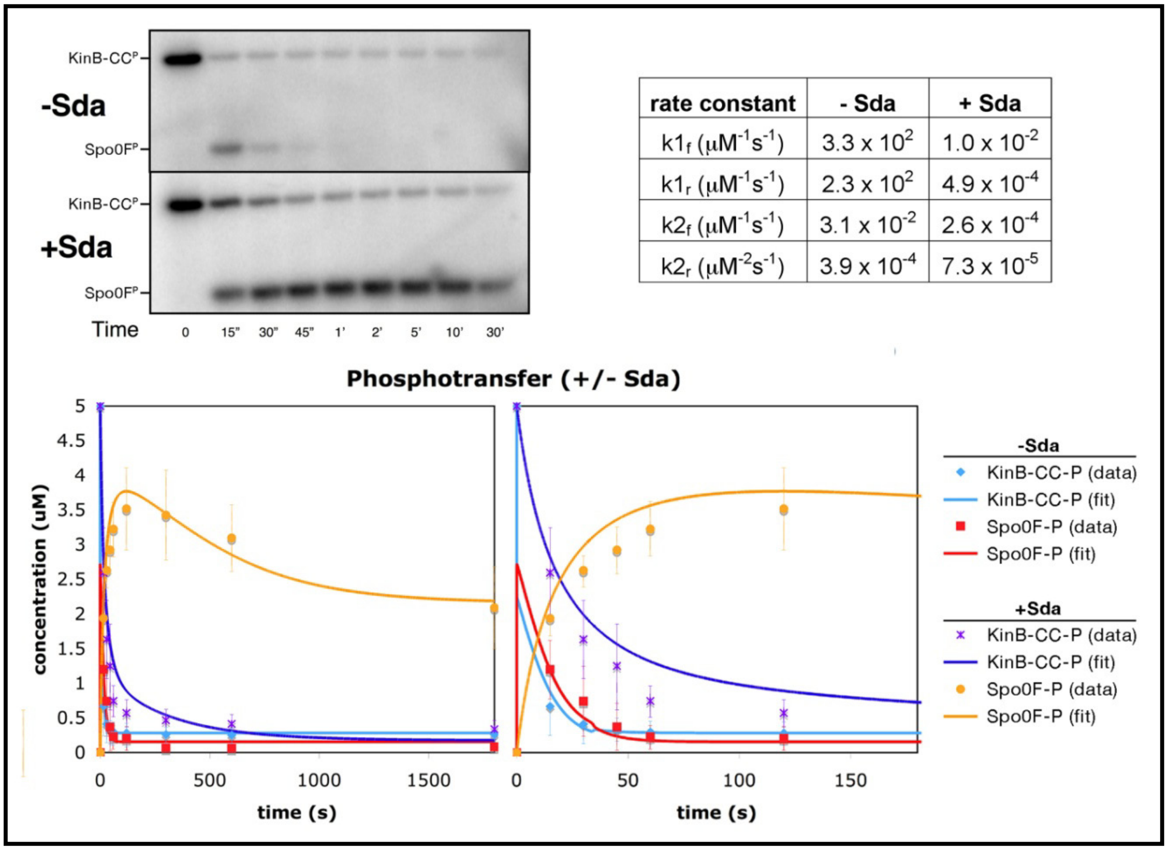
To address these inconsistencies, we measured the effect of Sda on the time course of phosphotransfer from KinB-CC$^{P}$ to Spo0F in an *in vitro* assay. In this assay, KinB-CC was allowed to autophosphorylate by incubation with γ-[$^{32}$P]ATP, excess [$^{32}$P]ATP was removed by rapid gel filtration over a desalting column, then Spo0F, and Sda, when present, was added. The time course of transfer of [$^{32}$P] from KinB-CC~P to Spo0F was monitored by SDS-PAGE and autoradiography (Figure 2.16). The results indicate the occurrence of two expected reactions, phosphotransfer from KinB-CC~P to Spo0F:

$$1)\qquad \text{KinB-CC}{\sim}\text{P} + \text{Spo0F} \underset{k1_r}{\overset{k1_f}{\leftrightarrow}} \text{KinB-CC} + \text{Spo0F}{\sim}\text{P}$$

with forward and reverse rate constants $k1_f$ and $k1_r$, respectively, and dephosphorylation of Spo0F~P by the phosphatase activity of KinB-CC (Hsing and Silhavy, 1997; Russo and Silhavy, 1993; Stock et al., 2000; Zhu et al., 2000):

$$2)\qquad \text{Spo0F}{\sim}\text{P} + \text{KinB-CC} \underset{k2_r}{\overset{k2_f}{\leftrightarrow}} \text{Spo0F} + \text{KinB-CC} + \text{P}_i$$

The phosphotransfer reaction (reaction 1) causes the rapid disappearance of KinB-CC~P and the appearance of Spo0F~P, while the phosphatase reaction (reaction 2) causes the disappearance of Spo0F~P over a longer period of time. This is particularly evident in the absence of Sda (Figure 2.16).



| rate constant | - Sda | + Sda |
|---|---|---|
| $k1_f$ ($\mu M^{-1}s^{-1}$) | $3.3 \times 10^{2}$ | $1.0 \times 10^{-2}$ |
| $k1_r$ ($\mu M^{-1}s^{-1}$) | $2.3 \times 10^{2}$ | $4.9 \times 10^{-4}$ |
| $k2_f$ ($\mu M^{-1}s^{-1}$) | $3.1 \times 10^{-2}$ | $2.6 \times 10^{-4}$ |
| $k2_r$ ($\mu M^{-2}s^{-1}$) | $3.9 \times 10^{-4}$ | $7.3 \times 10^{-5}$ |

**Figure 2.16)** Sda inhibits the rate of phosphotransfer from KinB-CC$^{P}$ to Spo0F, as well as KinB phosphatase activity towards Spo0F~P. (upper left panel) Autoradiographs showing the time course of phosphotransfer from KinB-CC~P to Spo0F in the absence (top) and presence (bottom) of a 25-fold molar excess of Sda. These experiments were repeated and quantitated in triplicate, and the results plotted in the lower panel (data points). The plot on the left shows the full time course (up to 0.5 hr), while the plot on the right shows the same data but with an expanded time scale. The lines show the results of the best-fit rate parameters (shown in the table) assuming the reaction scheme described in the text.

In the presence of saturating Sda, the kinetics were slow enough to capture with the 15 s time points. The rate constant $k1_f$ was estimated from the rate of the initial disappearance of KinB-CC~P to be approximately 0.015 $mM^{-1}s^{-1}$. This matches well with best-fit kinetic parameters (Figure 2.16). In the absence of Sda, phosphotransfer to Spo0F, as judged by the disappearance of KinB-CC~P, was essentially complete by the first time point of 15 seconds. It was not possible to collect shorter time points reliably using our protocol. A lower bound for $k1_f$ was estimated to be about 0.1 $mM^{-1}s^{-1}$, but this value is likely to be significantly in error. Indeed, best fit kinetic parameters yield a value for $k1_f$ of 330 $mM^{-1}s^{-1}$, more than three orders of magnitude higher than our lower bound estimate, and more than four orders of magnitude higher than the rate in the presence of Sda (Figure 2.16). However, the fitted parameters are not reliable due to the absence of time points shorter than 15 seconds. Nevertheless, it is clear that the presence of Sda inhibits the rate of phosphotransfer from KinB-CC~P to Spo0F by a minimum of one order of magnitude and possibly up to four orders of magnitude. Note also that the presence of Sda inhibits the forward rate of the phosphatase reaction ($k2_f$) as well, by about two orders of magnitude, which is not surprising since this reaction likely involves the interaction of Spo0F~P with the KinB-CC-DHp domain in a manner similar to the Spo0F/KinB-CC~P interaction required for the phosphotransfer reaction.

## Discussion

Sda is a potent inhibitor of KinB autophosphorylation (Figure 2.4B), a potent inhibitor of phosphotransfer from KinB~P to Spo0F, as well as a potent inhibitor of KinB phosphatase activity toward Spo0F~P (Figure 2.16). Our structural and biochemical analyses reveal that Sda binds at the end of the KinB-DHp domain, distal to the membrane, and prevents molecular transactions with the DHp domain to which it is bound by acting as a simple molecular barricade to sterically block communication between the KinB CA and DHp domains required for autophosphorylation, as well as to sterically block communication between Spo0F and the KinB DHp domain required for the phosphotransfer and phosphatase activities (Figure 2.16). The Sda/KinB molecular interface observed in the crystal structure (Figure 2.8) explains the effects of amino acid substitutions previously studied in Sda (Rowland et al., 2004) as well as substitutions introduced into KinB (Figure 2.11A).

In agreement with our conclusions, Rowland et al. (Rowland et al., 2004) concluded that Sda inhibits KinA autophosphorylation activity by acting as a molecular barricade to block the required communication between the HK DHp and CA domains. Contrary to our findings, however, Rowland et al. (Rowland et al., 2004) concluded, based on measurement of a single time point and not rate measurements, that Sda does not interfere with phosphotransfer to Spo0F, and proposed that Sda binds to the top of the DHp domain (proximal to the membrane) and interacts with the DHp-CA linker region, a binding site that would not interfere with Spo0F binding but could hinder the conformational mobility of the CA domain required for autophosphorylation. Our structural analysis indicates that binding of Sda and Spo0F to the DHp domain must be mutually exclusive (Figure 2.15B). Consistent with this, we show that Sda strongly retards the rate of phosphotransfer from KinB-CC~P to Spo0F (Figure 2.16).

Whitten et al. (Whitten et al., 2007) used a combination of SAXS and neutron scattering data to generate a structural model of the Sda/KinA-CC complex. Analysis of their model in the context of a structural model of DHp/CA domain communication during autophosphorylation led Whitten et al. to conclude that Sda did not present significant steric interference to DHp/CA domain communication, but rather that Sda inhibition activity must arise from allosteric effects on the HK structure. While the gross features of the Whitten et al. model, in which Sda binds to the KinA DHp domain at the end distal to the DHp-CA linker, with the Sda α-helices roughly perpendicular to the DHp α-helices, agree with the Sda/KinB-CC crystal

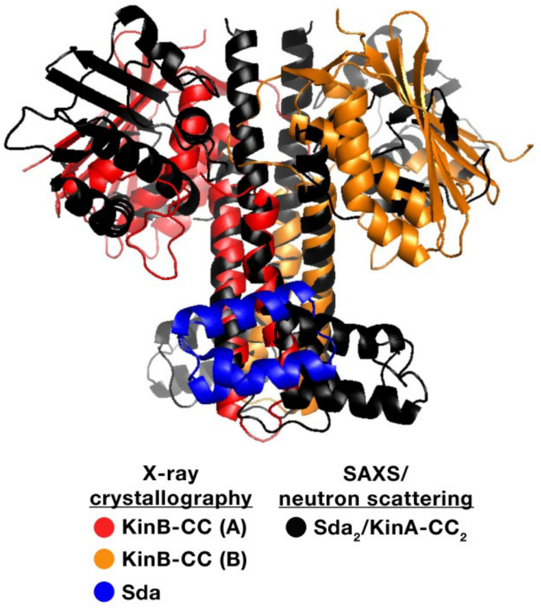


**Figure 2.17)** Superimposition of the Sda/KinB-CC crystal structure (KinB-CC chain A, red; KinB-CC chain B, orange; Sda, blue) with the SAXS/neutron scattering model of Sda/KinA (Whitten et al., 2007). Superimposition was by the HK DHp domain only. In the SAXS/neutron scattering model, Sda is bound in a different position on the DHp domain, and the orientation of Sda is rotated by approximately 180° (about an axis perpendicular to the DHp domain helices.

structure, the precise positioning of Sda, as well as the orientation of Sda with respect to the DHp domain, significantly differ due to the low resolution of the SAXS/neutron scattering method (Figure 2.17). The positioning and orientation of Sda observed in our crystal structure presents significant steric clashes with the CA domain modeled in position for the autophosphorylation reaction (Figure 2.15A), consistent with Sda acting as a simple steric block of CA domain communication with the DHp domain.

Also consistent with this conclusion are two additional observations from our study. First, conformational changes in the DHp domain induced by Sda binding appear to be restricted to residues that directly interact with Sda and are not propagated to other parts of the structure (Figure 2.6A), and second that a small set (three) of amino acid substitutions directly within the HK binding surface for Sda confers Sda sensitivity to an otherwise insensitive HK (Figure 11B).

In their SAXS/neutron scattering analysis, Whitten et al. (Whitten et al., 2007) observed that binding of Sda to KinA-CC, an 11% increase in mass, resulted in no significant change in the radius of gyration but caused a significant decrease in the maximum linear dimension of the complex. In the structural modeling, this was realized through a 'compaction' of the KinA CA domains in towards the homodimer two-fold axis compared with KinA alone (Whitten et al., 2007). We observe a qualitatively similar 'compaction' of the CA domains in our crystal structure of the Sda/KinB-CC complex (Figure 2.6A). It is tempting to propose that this compaction is a consequence of Sda binding, and that the compaction plays a role in the inhibition of autophosphorylation activity. We do not believe this to be the case for the following reasons. Sda binding to KinB in our crystal structure is asymmetric (i.e. an Sda molecule is bound to only one monomer of the KinB dimer; Figure 2.4C), and we have shown that Sda inhibition is also asymmetric (i.e. Sda inhibits DHp function of only the DHp domain to which it is bound, not both DHp domains of the KinB dimer (Figure 2.13)). The compaction of the CA domains in our crystal structure, however, is symmetric (Figure 2.4C); even though Sda is bound to only one KinB-CC monomer in the crystal, both CA domains are compacted in an essentially symmetric KinB-CC dimer. If the CA domain compaction seen in our crystal structure was a consequence of asymmetric Sda binding and played a role in asymmetric Sda inhibition of autophosphorylation, the symmetry of the compaction would be expected to match the symmetry of the inhibition, which is not the case. We believe the crystallization conditions stabilize the compacted CA domain architecture seen in our crystal structure (Figure 2.6B).

If the Sda functional mechanism is to bind the HK DHp domain and sterically block CA domain and Spo0F communication with the phosphoacceptor His, it might seem that Sda could be a more effective and direct inhibitor if it bound directly at the His residue, rather than peripheral to it (Figure 2.4C). However, the phosphoacceptor His is located within a sequence motif that is highly conserved among all two-component HKs, the H-box (Figure 2.9), so a small protein like Sda would not be able to achieve specificity for the sporulation HKs if its binding determinant included the H-box. In fact, Sda does not interact with any residues within the H-box but instead interacts with a region adjacent to the H-box at the base of the DHp domain that is conserved among *Bacillus* sporulation HKs, but not among all HKs (Figure 2.9). Specificity of Sda for the sporulation HKs is derived from at most three residues within this region (the GFXXL motif), since substitution of three residues within this region into a HK not normally inhibited by Sda gives rise to robust Sda inhibition (Figure 2.11B). Sda has been shown to inhibit the sporulation kinases KinA (Burkholder et al., 2001; Rowland et al., 2004) and KinB (Figure 2.4B), but the high conservation of the GFXXL motif among the sporulation HKs (Figure 2.9) suggests that Sda inhibits all of the sporulation HKs. Response regulators (such as Spo0F) interact directly with the phosphoacceptor His but still achieve exquisite specificity for their cognate HKs, but these proteins cover a much larger surface of the HK-DHp domain that includes the H-box region as well as the base of the DHp (where Sda binds, Figure 2.15B), explaining how specificity is achieved (Skerker et al., 2008).

In the KinB structure the disposition of the CA domain with respect to the DHp domain is inappropriate for the autophosphorylation reaction. The structure gives little support to either a cis or trans mechanism of autophosphorylation for KinB. However, the ε nitrogen of the side chain of the phosphoacceptor His is approximately 5.5 Å closer to the β-phosphate of the nucleotide bound to the CA domain of the same monomer than it is to the same atom of the opposite monomer (19.3 Å compared with 24.8 Å, respectively). In addition, the length of the DHp - CA linker would appear to allow either cis or trans autophosphorylation.

## Accession Numbers

Structure coordinates and structure factors have been deposited in the Protein Data Bank under ID code 3D36.

## Annotation of Bacterial Species Names in Figure 3A

Bacterial species names corresponding to the abbreviations in the left column are annotated below, along with the NCBI GenInfo (gi) number for the protein sequence: Gst (*Geobacillus stearothermophilus*); Gka (*Geobacillus kaustophilus HTA426*, gi[56421062]); Gth (*Geobacillus thermodenitrificans NG80-2*, gi[138896100]); BB-14911 (*Bacillus sp. NRLL B-14911* gi[89100817]); Bam (*Bacillus amylolequefaciens FZB42*, gi[154686829]); Bli, (*Bacillus licheniformis ATCC 14580*, gi[52081119]); Bsu, *Bacillus subtilis subsp. Subtilis str. 168*, gi[50812270]); Bha, (*Bacillus halodurans C-125*, gi[15613884]); Bpu, (*Bacillus pumilus SAFR-032*, gi[157693071]); BSG-1, (*Bacillus sp. SG-1*, gi[149183160]); Bce (*Bacillus cereus G9241*, gi[47569333]); Bth (*Bacillus thuringiensis serovar israelensis ATCC 35646*, gi[75761262]; Ban (*Bacillus anthracis str. Ames*, gi[30264409]); Bwe (*Bacillus weihenstephanensis KBAB4*, gi[8920636]); Oih (*Oceanobacillus iheyensis THE831*, gi[22777674]); Bcl (*Bacillus clausii KSM-K16*, gi[56962934]).

# Materials and Methods

## *Cloning and Mutagenesis*

KinB derivatives, Spo0F, and Sda were PCR amplified from *Gst* strain 10 genomic DNA. KinB derivatives were cloned into the NdeI and BamHI sites of both pET21a (Novagen) and a pET28a-based vector containing an N-terminal PreScission™ Protease (GE healthcare) cleavable $His_6$-tag. Spo0F was cloned into the NdeI and BamHI sites of a pET28-based vector. Sda was cloned into pET21a. KinB(208-CT) was subcloned from KinB-CC into the NdeI and BamHI sites of the pET28a-based vector. For coexpression, KinB-CC and Sda were subcloned into the pET28a-based vector such that the two proteins were coexpressed from a single promoter (Campbell and Darst, 2000). In the resulting vector, KinB-CC contained an N-terminal, PreScission™ Protease cleavable $His_6$-tag and Sda was untagged. After cleavage with PreScission™ Protease, KinB-CC harbors four vector derived residues (GPHM) at its N-terminus. KinB-CC mutants (Figure 2.11A) were generated using the megaprimer method (Perrin and Gilliland, 1990).

Plasmid pET24c-TM0853 was kindly provided by Dr. Alberto Marina (Marina et al., 2005). A single megaprimer containing three amino acid substitution sites was used to generate TM0853* (Figure 2.11B).

### *Expression and Purification*

KinB derivatives, Spo0F, Sda, and Sda/KinB complexes were all expressed in *Eco* Rosetta2 DE3 (Novagen). Cultures were grown at 37°C to an $OD_{600}$ of 0.6. Cells were then transferred to 16 °C, IPTG was added to 0.1 mM, and the cells were incubated overnight. Cells were harvested by centrifugation, resuspended in lysis buffer (20 mM Tris-HCl, pH 7.0, 150 mM NaCl, 10 mM $MgCl_2$, 5 mM imidazole, 5% glycerol) and stored at -80 °C.

For purification, phenylmethylsulfonyl fluoride and β-mercaptoethanol were added (each to 1 mM), and the cells were lysed using a continuous flow French press (EmulsiFlex-C5™, Avestin) and clarified by centrifugation. For Sda purification, the supernatant was heat-treated at 65 °C for 15 min, and aggregated material was removed by centrifugation. This yielded an Sda preparation that was more than 95% pure as judged by SDS-PAGE. Heat-treatment of KinB derivatives and Spo0F resulted in aggregation and was not used. Instead, KinB derivatives and Spo0F were applied to a $Ni^{2+}$-charged HiTrap metal chelating column (GE Healthcare), the column was washed extensively with loading buffer, then $His_6$-tagged proteins were eluted with lysis buffer containing 200 mM imidazole. To cleave the $His_6$-tag, PreScission™ Protease was added in a 1:100 mass:mass ratio and the protein was dialyzed overnight at 4°C against 2 L of 20 mM Tris-HCl, pH, 7.0, 50 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT. Protein for crystallization was further purified to homogeneity by passing the sample over a GST-affinity column (to remove the GST-tagged PreScission protease), followed by anion exchange chromatography (HiTrap Q, GE Healthcare) and size exclusion chromatography. TM0853 (and TM0853*) was expressed and purified as previously described (Marina et al., 2005), with the exception that aliquots were made following the heat treatment step.

### *Crystallization and Structure Determination*

Purified Sda/KinB-CC complex was concentrated to 10 mg/ml by centrifugal filtration (Vivascience), and a 5-fold molar excess of ADP (Sigma) was added. Crystals of the Sda/KinB-CC complex were grown at 4 °C in sitting drops using vapor diffusion by mixing equal volumes of protein solution (1 ml at 10 mg/ml in 20 mM Tris-HCl, pH 7.0, 50 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT) and crystallization solution (200 mM LiCl, 40% MPD). Crystals (small hexagonal plates) appeared overnight and grew to a maximum size of about 50 X 50 X 20 µm in one week. Microseeding techniques were employed to grow larger crystals (about 150 X 150 X 50 µm). Methyl-HgCl derivatized crystals (Table 2.1) were prepared by soaking crystals for 10 min. in crystallization solution supplemented with 2.5 mM methyl-HgCl (Sun and Radaev, 2002). Both native and methyl-HgCl crystals were directly flash frozen in liquid ethane and stored in liquid nitrogen prior to data collection. Diffraction data was processed using HKL2000 (Otwinowski and Minor, 1997). The structure was phased with data from the methyl-HgCl derivative (Table 2.1). Two Hg sites were located within the asymmetric unit using SHELX (Sheldrick, 2008). Heavy atom refinement, phasing, and density modification calculations were performed using SHARP (de La Fortelle et al., 1997). SAD phases alone, calculated from the single-wavelength methyl-HgCl data, yielded an uninterpretable electron density map, but gave a SIRAS map with clear elements of secondary structure when combined with the isomorphous signal from the native dataset.

The SIRAS phases were used as the basis for phased molecular replacement searches (http://zonker.bioc.aecom.yu.edu/cgi-bin/inhouse/bruteptf/bruteptf.cgi) using BRUTEPTF (Strokopytov et al., 2005), which placed the individual DHp and CA domains from HK TM0853 (Marina et al., 2005). An excellent MIRAS experimental electron density map (Figure 2.3) was obtained by combining the SIRAS phases with the phases calculated from the molecular replacement solution, along with subsequent density modification using RESOLVE (Terwilliger, 2000).

A partial model was built into the 2.0 Å resolution MIRAS map using RESOLVE (Terwilliger, 2003). Additional model building was carried out using O 10.0.3 (Jones et al., 1991). Iterative cycles of refinement and model building were carried out using CNS (Adams et al., 1997), with final rounds of refinement using RefMac (Murshudov et al., 1997). The final model was refined to an $R/R_{\text{free}}$ of 20.0/21.5. Structure validation using MolProbity (Davis et al., 2007) gave an overall clash score of 5.74 (98th percentile for a structure at 2.0 Å resolution), revealed 98.5% of residues in the most favored regions of the Ramachandran plot, with 1 residue (0.15%) in a disallowed region (Leu207 of chain B, adjacent to a disordered loop).

### *ESI-MS*

The Sda/KinB-CC complex was analyzed using a modified Q-ToF 2 mass spectrometer (Waters, Manchester, UK). For the analysis the complex was buffer exchanged into 200 mM ammonium acetate (pH 7.5), and aliquots of ~2 ml were introduced into the mass spectrometer via nanoflow capillaries. The following experimental conditions were used: capillary voltage, 1.6 kV; RF Lens, 65 V; cone voltage, 65 V; accelerating voltage, 8 V. Pressure at the ion transfer stage was maintained at $1.6 \times 10^{-1}$ mbar. The pressure at the collision cell was maintained at $3.6 \times 10^{-4}$mbar.

### *Mixed-Dimer Preparation*

Plasmid pET21a-KinB-CC (no affinity tag, ampicillin resistance) and pET28a-KinB-CC (N-terminal $His_6$-tag, kanamycin resistance) were co-transformed into *Eco* Rosetta2 DE3 and expressed as described for the Sda/KinB-CC complex. Following lysis, the supernatant was passed over a $Ni^{2+}$-charged HiTrap metal chelating column (GE Healthcare), allowing any untagged homodimer KinB-CC to flow through. The column was then eluted with a linear gradient from 5 mM to 200 mM imidazole, allowing the separation of $His_6$-KinB-CC/KinB-CC heterodimers from $His_6$-KinB-CC homodimers. Glycerol was added to 15% and aliquots were snap frozen in liquid nitrogen and stored at -80 °C.

### *Autophosphorylation Assays*

For the KinB-CC mutant and mixed-dimer assays, 50 pmol of KinB-CC was incubated in kinase buffer (50 mM 3- 4-(2-Hydroxyethyl)-1-piperazinyl propanesulfonic acid [HEPPS], pH 8.5, 20 mM KCl, 5 mM $MgCl_2$, 5% glycerol) at 25 °C with either 0, 1, 5, or 25-fold molar excess of Sda for 15 min in 10 μl reactions. Next, 1 μl of a 1:1 mixture of 50 mM ATP:3000μCi/μmol $\gamma$-[$^{32}$P]ATP was added and the autophosphorylation reaction was carried out for 1 hr at 25 °C. Reactions were quenched with 2.5 μl 5X SDS loading buffer with 50 mM EDTA and immediately subjected to

PAGE analysis on a 4-20% gradient tris-glycine polyacrylamide gel. Gels were dried for 1.5 hr at 55 °C and analyzed by phosphorimagery (Molecular Dynamics).

### *Phosphotransfer Assays*

KinB-CC (500 pmol) was phosphorylated for 1 hr at 25 °C as described in the autophosphorylation assays. Excess nucleotide was removed from the KinB-CC$^P$ using a 0.8 ml microspin desalting column (BioRad). The KinB-CC$^P$ was split into two equal reaction volumes. A 25-fold molar excess of Sda was added to one reaction, while the same volume of Sda storage buffer was added to the other. The reactions were then incubated at 25 °C for an additional 15 min. Spo0F (250 pmol) was then added to each reaction. At timed intervals, 10 µl aliquots were removed and mixed with 2.5 µl of 5X SDS loading buffer plus 50 mM EDTA, then snap frozen and stored in liquid nitrogen until all time points were ready for analysis by 17% SDS-PAGE. The portion of the gel below the 10 kDa molecular-weight marker was removed in order to reduce background radioactive signal. The gels were then analyzed and quantitated by phosphorimagery. To estimate the rate constants for the phosphotransfer and phosphatase reactions (Figure 2.16), the data were fit using least squares to the reaction scheme and Runge-Kutta 4th order numerical integration using the program Berkeley Madonna v8.3.21 (www.berkeleymadonna.com). The R-factors [$(R = S|C_{obs} - C_{calc}|)/SC_{obs}$, where $C_{obs}$ is the observed concentration, and $C_{calc}$ is the concentration calculated from the curve-fit parameters) for the curve-fits are: KinB-CC-P$^{-SDA}$ (cyan line in Figure 2.16), 0.22; Spo0F-P$^{-SDA}$ (red line in Figure 2.16), 0.097; KinB-CC-P$^{+SDA}$ (blue line in Figure 16), 0.050; Spo0F-P$^{+SDA}$ (orange line in Figure 2.16), 0.25.

# References

---

**Adams, P.D., Pannu, N.S., Read, R.J., and Brunger, A.T.** (1997). Cross-validated maximum likelihood enhances crystallographic simulated annealing refinement. Proceedings of the National Academy of Sciences USA *94*, 5018-5023.

**Burkholder, W.F., Kurtser, I., and Grossman, A.D.** (2001). Replication initiation proteins regulate a developmental checkpoint in Bacillus subtilis. Cell *104*, 269-279.

**Campbell, E.A., and Darst, S.A.** (2000). The anti-σ factor SpoIIAB forms a 2:1 complex with $\sigma^F$, contacting multiple conserved regions of the σ factor. Journal of Molecular Biology *300*, 17-28.

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

**Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B.r., Snoeyink, J., Richardson, J.S., *et al.*** (2007). MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Research *35*, W375-W383.

**de La Fortelle, E., Irwin, J.J., and Bricogne, G.** (1997). SHARP: A maximum-likelihood heavy-atom parameter refinement and phasing program for the MIR and MAD methods. In Crystallographic Computing, P. Bourne, and K. Watenpaugh, eds. (Boston, Kluwer Academic Publishers), pp. 1-9.

**Gonzalez-Pastor, J.E., Hobbs, E.C., and Losick, R.** (2003). Cannibalism by sporulating bacteria. Science (New York, NY *301*, 510-513.

**Hsing, W., and Silhavy, T.J.** (1997). Function of conserved histidine-243 in phosphatase activity of EnvZ, the sensor for porin osmoregulation in Escherichia coli. Journal of Bacteriology *179*, 3729-3735.

**Hunter, C.A., Singh, J., and Thornton, J.M.** (1991). Pi-pi interactions: the geometry and energetics of phenylalanine-phenylalanine interactions in proteins. J Mol Biol *218*, 837-846.

**Ireton, K., and Grossman, A.D.** (1994). A developmental checkpoint couples the initiation of sporulation to DNA replication in Bacillus subtilis. The EMBO journal *13*, 1566-1573.

**Jones, T.A., Zou, J.-Y., Cowan, S., and Kjeldgaard, M.** (1991). Improved methods for building protein models in electron denstiy maps and the location of errors in these models. Acta crystallographica *A47*, 110-119.

**Lemon, K.P., Kurtser, I., Wu, J., and Grossman, A.D.** (2000). Control of initiation of sporulation by replication initiation genes in Bacillus subtilis. J Bacteriol *182*, 2989-2991.

**Mandelstam, J., Sterlini, J.M., and Kay, D.** (1971). Sporulation in Bacillus subtilis. Effect of medium on the form of chromosome replication and on initiation to sporulation in Bacillus subtilis. The Biochemical journal *125*, 635-641.

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

**Moriya, S., Imai, Y., Hassan, A.K., and Ogasawara, N.** (1999). Regulation of initiation of Bacillus subtilis chromosome replication. Plasmid *41*, 17-29.

**Murshudov, G.N., Vagin, A.A., and Dodson, E.J.** (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallographica D Biological Crystallography *D53*, 240-255.

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

**Otwinowski, Z., and Minor, W.** (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods in Enzymology *276*, 307-326.

**Pena-Sandoval, G.R., and Georgellis, D.** The ArcB sensor kinase of Escherichia coli autophosphorylates by an intramolecular reaction. J Bacteriol.

**Perrin, S., and Gilliland, G.** (1990). Site-specific mutagenesis using asymmetric polymerase chain reaction and a single mutant primer. Nucleic Acids Research *18*, 7433-7438.

**Qin, L., Dutta, R., Kurokawa, H., Ikura, M., and Inouye, M.** (2000). A monomeric histidine kinase derived from EnvZ, an Escherichia coli osmosensor. Molecular Microbiology *36*, 24-32.

**Rost, B.** (1996). PHD: predicting one-dimensional protein structure by profile-based neural networks. Methods in Enzymology *266*, 525-539.

**Rowland, S.L., Burkholder, W.F., Cunningham, K.A., Maciejewski, M.W., Grossman, A.D., and King, G.F.** (2004). Structure and mechanism of action of Sda, an inhibitor of the histidine kinases that regulate initiation of sporulation in Bacillus subtilis. Molecular Cell *13*, 689-701.

**Russo, F.D., and Silhavy, T.J.** (1993). The essential tension: opposed reactions in bacterial two-component regulatory systems. Trends in Microbiology *1*, 306-310.

**Ruvolo, M.V., Mach, K.E., and Burkholder, W.F.** (2006). Proteolysis of the replication chechpoint protein Sda is necessary for the efficient initiation of sporulation after transient replication stress in Bacillus subtilis. Molecular Microbiology *60*, 1490-1508.

**Sharon, M., and Robinson, C.V.** (2007). The role of mass spectrometry in strucutre elucidation of dynamic protein complexes. Annual Review of Biochemistry *76*, 167-193.

**Sheldrick, G.M.** (2008). A short history of SHELX. Acta Crystallographica A Foundations of Crystallography *64*, 112-122.

**Skerker, J.M., Perchuk, B.S., Siryaporn, A., Lubin, E.A., Ashenberg, O., Goulian, M., and Laub, M.T.** (2008). Rewiring the specificity of two-component signal transduction systems. Cell *133*, 1043-1054.

**Stock, A.M., Robinson, V.L., and Goudreau, P.N.** (2000). Two-component signal transduction. Annual Review of Biochemistry *69*, 183-215.

**Strokopytov, B.V., Fedorov, A., Mahoney, N.M., Kessels, M., Drubin, D.G., and Almo, S.C.** (2005). Phase translation function revisited: structure solution of the cofilin-homology domain from yeast actin-binding protein 1 using six-dimensional searches. Acta Crystallographica D Biological Crystallography *61*, 285-293.

**Sun, P.D., and Radaev, S.** (2002). Generating isomorphous heavy-atom derivatives by a quick-soak method. Part II: Phasing of new structure. Acta Crystallographica D Biological Crystallography *D58*, 1099-1103.

**Terwilliger, T.C.** (2000). Maximum likelihood density modification. Acta Crystallographica D Biological Crystallography *D56*, 965-972.

**Terwilliger, T.C.** (2003). Automated main-chain model building by template matching and iterative fragment extension. Acta Crystallographica D Biological Crystallography *59*, 38-44.

**Trach, K.A., and Hoch, J.A.** (1993). Multisensory activation of the phosphorelay initiating sporulation in Bacillus subtilis: Identification and sequence of the protein kinase of the alternate pathway. Molecular Microbiology *8*, 69-79.

**Whitten, A.E., Jacques, D.A., Hammouda, B., Hanley, T., King, G.F., Guss, J.M., Trewhella, J., and Langley, D.B.** (2007). The structure of the KinA-Sda complex suggests an allosteric mechanism of histidine kinase inhibition. Journal of Molecular Biology *368*, 407-420.

**Yang, Y., and Inouye, M.** (1991). Intermolecular complementation between two defective mutant signal-transducing receptors of Escherichia coli. Proceedings of the National Academy of Sciences USA *88*, 11057-11061.

**Yoshikawa, H., Wake, R.G.** (1993). Initiation and termination of chromosome replication. In *Bacillus subtilis* and other gram-positive bacteria: biochemistry, physiology, and molecular genetics, A.L. Sonenshein, Hoch, J. A., Losick, R., ed. (Washington, D. C., American Society for Microbiology), pp. 507-528.

**Zapf, J., Sen, U., Madhusudan, Hoch, J.A., and Varughese, K.I.** (2000). A transient interaction between two phosphorelay proteins trapped in a crystal lattice reveals the mechanism of molecular recognition and phosphotransfer in signal transduction. Structure *8*, 851-862.

**Zhu, Y., Qin, L., Yoshida, T., and Inouye, M.** (2000). Phosphatase activity of histidine kinase EnvZ without kinase catalytic domain. Proceedings of the National Academy of Sciences USA *97*, 7808-7813.

# CHAPTER 3
## Semisynthesis of a Sporulation Histidine Kinase

## Introduction

One of the long-standing questions in the field of histidine kinase biology concerns the structural rearrangements necessary to transition an HK from the inactive to active state of autophosphorylation. Currently available crystal structures of intact HK catalytic cores (CCs, containing both the DHp and CA domains) are in the "off" conformation, where the $\gamma$-phosphoryl of the ATP bound in the CA bound domain is far removed from the phosphoacceptor His of the DHp domain (Bick et al., 2009; Marina et al., 2005). It is presumed that the autophosphorylation event is highly transient in nature, and hence difficult to trap and study using traditional methods of crystallography. Moreover, CCs separated from their signal sensing domains possess an unregulated basal level of autophosphorylation, such that the CA and DHp domains are in constant positional flux with respect to one another.

The mechanism of phosphotransfer from the HK to the RR is another unresolved topic of interest. The unstable nature of the phospho-His N-P bond, with its low free energy of hydrolysis between -12 and -14 kcal/mol, has stifled the structural investigation of HKs in their phosphorylated forms. Phospho-His has a half-life under acidic conditions on the order of seconds, though it is much more stable in alkaline conditions (Severin and Yudelovich, 1947). Simple chemical mimics of phospho-Asp (Chabre,

1990) have been used to study phosphorylated RRs (Casino et al., 2009; Gardino et al., 2003; Zhao et al., 2008), but no analogous phospho-His mimic is currently available.

Co-crystal structures of Spo0F in both the unphosphorylated form and as a phospho-Asp mimic ($BeF_3^-$) in complex with the phosphotransferase Spo0B have been used as models of the HK-RR interaction (Varughese et al., 2006; Zapf et al., 2000). Indeed, alanine scanning mutagenesis has shown that both Spo0B and KinA share the same Spo0F interacting surface (Tzeng and Hoch, 1997). However, there are several reasons to believe that these structures may not be representative of a true phospho-His HK-RR interaction. First of all, because of the directionality of the HK-RR phosphotransfer step, if anything the $BeF_3^-$-Spo0F-0B structure describes how an unphosphorylated HK recognizes a phosphorylated RR for dephosphorylation. Secondly, as mentioned in the introduction of this thesis, it is unclear whether phospho-His is an intermediate of HK induced dephosphorylation of the RR, such that phosphotransfer from Spo0F to 0B may occur by a different mechanism. Lastly, despite the fact that Spo0B has an HK-like four-helix bundle, it does not possess autokinase activity, and hence is not a real HK.

The recently reported co-crystal structure of HK 0853 from *T. maritima* in complex with its cognate RR 468 has greatly advanced our understanding of the structural determinants required for HK-RR partner specificity (Casino et al., 2009). Though the TM0853-468 complex does provide information regarding HK to RR phosphotransfer, the structure is more representative of an HK poised for RR dephosphorylation. If we are to understand the precise mechanism of phosphotransfer, the structure of a phosphorylated HK will be necessary.

Keeping the restraints of HK domain flexibility and phospho-His instability in mind, we have chosen to pursue a hybrid approach for investigating the active conformations of an HK. By employing the methods of stepwise solid phase peptide synthesis (SPPS) for the introduction of His analogs into the DHp, protein semisynthesis, and native chemical ligation (NCL), we hope to trap transient active states of KinB.

Solid phase peptide synthesis (SPPS) (Merrifield, 1986; Merrifield, 1963) has been used to incorporate non-biologically encoded amino acids into peptides at site-specific locations. In principle, any unnatural amino acid can be used in SPPS. Proteins containing amino acids with modifications to either the side chain or peptide backbone have been reported (Valiyaveetil et al., 2002; Valiyaveetil et al., 2004, 2006b), and for several cases crystal structures of these semisynthetic proteins have been solved (Chacko et al., 2004; Qin et al., 2002; Rak et al., 2003; Valiyaveetil et al., 2006a).

SPPS involves the attachment of the C-terminal amino acid to a solid support resin, and the subsequent addition of amino acids in a stepwise manner; the peptide is synthesized from C- to N-terminus (Figure 3.1). Peptides made using current SPPS technology are roughly limited to 50 amino acids or less (Dawson et al., 1997). This is because, despite the high efficiency of the coupling reaction, it is not 100%; truncated byproducts do carry through from previous coupling steps. As an example, for a 50-mer peptide, where ideally each coupling reaction is at least 99% efficient, $0.99^{50}$ would still only yield 60.5% full-length product. In practice coupling reactions are often less than 99% efficient.

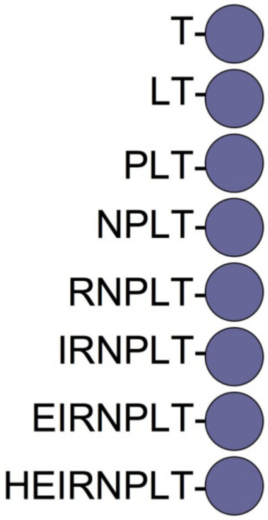


**Figure 3.1)** SPPS

Amino acids for SPPS must have their reactive functional groups protected during synthesis. This is true of the both the α-amino group and side chain functionalities, the difference being that protective groups are removed from the α-amino group after each amino acid coupling, while the side chain protecting groups are left in place until the peptide has been fully assembled. The choice of side chain protecting chemistry is entirely dependent upon which of two possible α-amino group protecting chemistries, either Boc or Fmoc, will be used for the synthesis (Carpino, 1957, 1973; Chang and Meienhofer, 1978).

Native chemical ligation allows two or more unprotected peptides to be joined by a natural peptide bond at the ligation site (Dawson et al., 1994). Peptide segments for NCL may come from either synthetic or recombinant sources. The principle of NCL lies in the chemoselective reaction that takes place between the α-cysteine of one peptide and the α-thioester of the second (Figure 3.2). The first step of the NCL reaction involves a thiol/thioester exchange, which can occur at any Cys residue within the peptide. This reaction is fully reversible under the aqueous conditions used for NCL. The second step of the reaction involves the formation of a native peptide bond via intramolecular nucleophilic attack. Unlike step one, this reaction is irreversible under the same conditions, and so over time the thiol/thioester intermediates are depleted, resulting in a highly regiospecific product (Dawson et al., 1997). It is therefore of no consequence if additional Cys residues are present in either peptide fragment.

Along those lines, any unprotected peptide may be used for NCL, as long as the chemical requirements of an α-cysteine and α-thioester at the termini of the peptides are met; the amino acid composition of the peptides is immaterial to the ligation reaction. Of course, many proteins do not possess a conveniently located Cys in their coding sequence. In such cases it is necessary to introduce a Cys by mutagenesis.

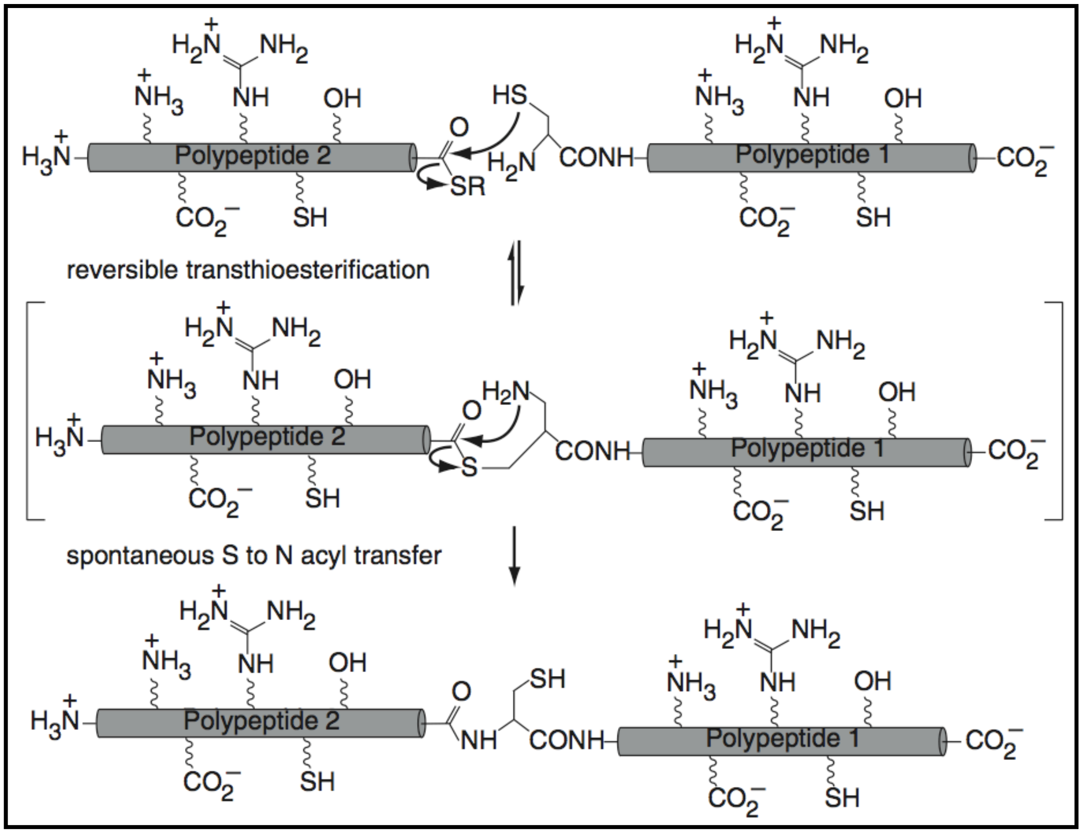


**Figure 3.2)** The mechanism of native chemical ligation. The initial step is a reversible transthioesterification reaction involving the thiol group of the N-terminal Cys-polypeptide (C-terminal fragment) and the α-thioester moiety of the N-terminal polypeptide fragment. This intermediate undergoes a spontaneous rearrangement to form a natural peptide bond at the ligation site. Taken from (Camarero and Muir, 2001).

The present chapter describes the semisynthesis of the KinB catalytic core, using stepwise SPPS and native chemical ligation to generate a full-length KinB-CC with amino acid analogs of histidine at the site of phosphorylation. The preliminary steps required (protein refolding, Cys mutagenesis) to carry out a successful semisynthetic experiment are also reported. The work described in this chapter was carried out in collaboration with the Muir Laboratory at The Rockefeller University, and in particular with Dr. Jung-Min Kee, a postdoc in the Muir lab, as well as with Professor J. David Warren, head of the Abby and Howard P. Millstein Synthetic Chemistry Core Facility at Weill Cornell Medical College. Katharina Häußermann, a master's student from Germany performing her thesis work in the Darst lab, carried out many of the peptide inhibition experiments. Only a brief description of the chemical synthesis of the amino acid analogs will be given for archiving purposes, as much of this work was performed by Dr. Kee and Dr. Warren, and is outside the expertise of the thesis author.

## Results and Discussion

### *Establishing the Native Chemical Ligation Site in KinB*

In the crystallized KinB-CC construct, the active site histidine of KinB, His 213, is conveniently located a mere 19 amino acids from the N-terminus, making it easily accessible for modification by stepwise SPPS. Because the total length of the KinB-CC is 240 amino acids, we chose to pursue a semisynthetic strategy for generating full-length KinB-CC, wherein a short peptide comprising the KinB N-terminal portion would be synthesized chemically, and the C-terminal portion would be expressed recombinantly. The KinB-CC sequence contains two naturally occurring Cys residues, the closer of which is 101 amino acids from the N-terminus. It was therefore necessary to introduce a Cys into KinB-CC at a point that was both compatible with ligation to an N-terminal

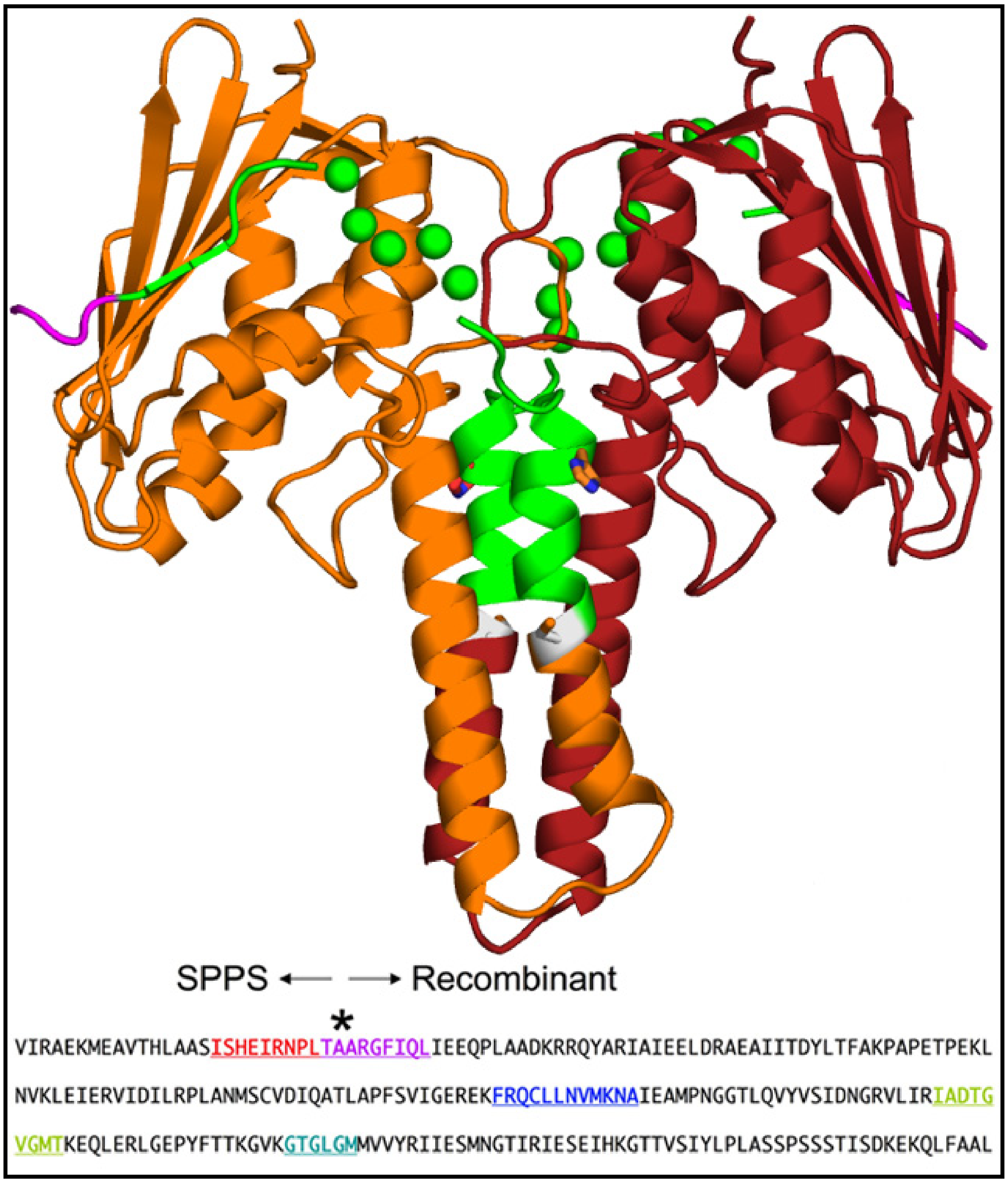


**Figure 3.3)** (Top) The KinB dimer, colored according to the semisynthetic scheme described in this chapter. The green region corresponds to residues that will be synthesized by SPPS. The remainder of the protein (orange) will be expressed recombinantly in E. coli. The region colored magenta corresponds to vector derived residues observed in the KinB/Sda crystal structure. These will not be included in the peptide synthesis. The cysteine mutant introduced to facilitate the native chemical ligation reaction is labeled grey. (Bottom) The KinB 195-CT sequence. The Ala position mutated to Cys is highlighted with an asterisk. Colored residues correspond to conserved sequence motifs: H-box (red), Sda box (purple), N-box (blue), G1-box (olive), and G2-box (cyan).

peptide of less than 50 residues, and yet downstream of His 213. His 213 is located in a region of high sequence conservation, the H-box (residues Ile 211 to Leu 219), preventing us from engineering a Cys into this motif. In addition, Sda makes contacts with KinB residues immediately following the H-box, three of which, Gly 224, Phe 225, and Leu 228, are indispensable for this interaction. We chose a conservative Ala to Cys mutation at position 221 as a trade off between mutating a residue that contacts Sda and keeping the peptide length relatively short (26-mer, KinB Val 195 to Thr 220). Position 221 of KinB is conserved as an Ala in sensor HKs, although serine and cysteine also occur, including a Cys in *B. subtilis* KinC. Ala 221 is solvent exposed in the KinB/Sda structure, occurring at roughly the midpoint of $\alpha$-helix 1 of the DHp domain (Figure 3.3). Only the backbone $\alpha$-carbon and carbonyl oxygen of Ala 221 make contacts with Sda Phe 25, so presumably the side chain is unimportant for the KinB/Sda interaction.

We first tested KinB A221C in an autophosphorylation assay to determine if the Cys mutation affected the enzymatic activity of the protein. To our surprise, rather than inhibit autophosphorylation, the A221C mutant instead showed enhanced levels of labeling when compared with wild-type (Figure 3.4). We initially reasoned that the new Cys 221 was forming a disulfide bond with Cys 317 of the CA domain, effectively locking the active conformation. In our KinB active state model (see Chapter 2), Cys 317 is within a reasonable distance to Ala 221 (Cys 221). However, when we performed the autophosphorylation assay in the presence of reducing agent (10 mM DTT), increased levels of labeling were still observed (data not shown). The fact that position 221 of KinB not only offers a convenient location for NCL, but also that a Cys at this site seems to promote autophosphorylation should prove to be advantageous for stabilizing the KinB active conformation. Sda mediated inhibition of the autophosphorylation reaction was unaffected by the Cys substitution (Figure 3.4).

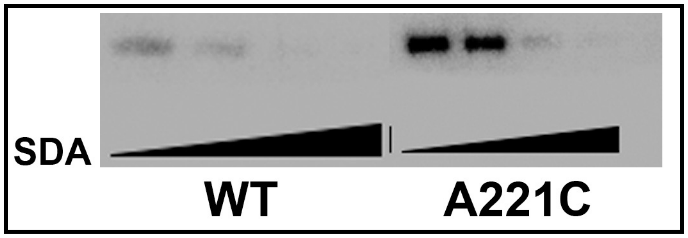


**Figure 3.4)** Autophosphorylation assay performed on KinB 195-CT wild-type (WT) and A221C. The A221C mutation does not hinder the activity of Sda.

## *KinB is Active Following Denaturation and Refolding*

Our NCL strategy is carried out under denaturing conditions. Because of this, we needed to ensure that KinB-CC could be properly refolded and still maintain its enzymatic activity. Denatured KinB, at approximately 3mg/ml in 6 M guanidine-HCl, was rapidly diluted 1:20 (final [KinB] of ~150μg/ml) with a refolding buffer (Burgess, 2009). We saw little to no signs of protein aggregation when the absorbance at 320 nm was determined on a sample of the refolded protein solution (protein aggregates tend to scatter 320 nm light (see methods) (Burgess, 2009)). Our initial refolding experiments produced an active enzyme, but we observed what appeared to be KinB multimers following an autophosphorylation assay and SDS-PAGE analysis (Figure 3.5A). We attributed this to the formation of aberrant disulfides while the protein was in the denatured state. Addition of 10mM DTT to the refolding buffer greatly reduced the KinB multimeric species (data not shown). Refolded KinB migrated as a monodisperse species on a size exclusion column (Figure 3.5C). When assayed for autophosphorylation, refolded KinB showed no discernible loss of enzymatic activity compared with native (non-denatured) KinB (Figure 3.5B).

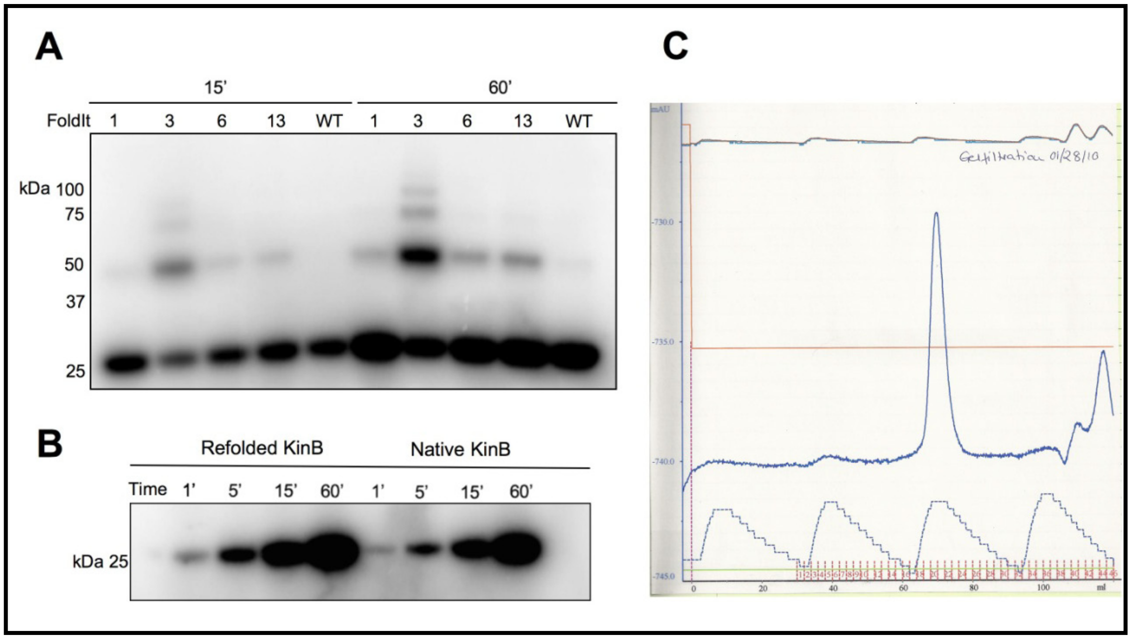


**Figure 3.5)** (A) Two time points, 15 and 60 min, from an autophosphorylation assay performed with KinB 195-CT that had been denatured in 6 M guanidine-HCl, and subsequently refolded in 4 different solutions from the FoldIt™ screen (Hampton) (see Materials and Methods). As the FoldIt™ screen has been discontinued for some time, solutions were homemade. Refolded KinB 195-CT phosphorylated to near wild-type levels, with the exception of protein refolded with solution 3. Higher molecular weight bands seen in the autoradiogram were suspected to be multimers of KinB formed by disulfides during the refolding process. (B) Time course of autophosphorylation comparing native (non-denatured) and refolded KinB 195-CT after a size exclusion chromatography step. The protein used in this assay was taken from a stock after the folding conditions on a large (~3.5 mg) scale had been optimized using FoldIt™ solution 13 (see Figure 3.5A). (C) Chromatogram of the size exclusion chromatography step used to purify the refolded protein from (B). Refolded KinB 195-CT had the expected retention time on the column, and migrated as a monodisperse species.

## *Generation of the Reagents for Native Chemical Ligation*

After we established a suitable ligation site within KinB, and showed that the KinB-CC maintains its activity after refolding, we turned our attention to the design of the necessary reagents for KinB-CC semisynthesis. KinB Val 195 to Thr 220 (26-mer) was synthesized on a 3-mercapto-propionamide-MBHA (4-Methylbenzhydrylamine) solid support resin using standard Boc-SPPS techniques (Camarero and Muir, 2001). The remaining carboxy-terminal KinB fragment, containing the Ala 221 Cys mutation, was expressed as a C-terminal PreScission™ Protease cleavable $\text{His}_6$-tagged protein. Expression in the pET21a system produces an N-terminal Met upstream of Cys 221. We designed this construct with the idea that removal of the N-Met would occur spontaneously in the cell as a result of the activity of endogenous E. coli methionyl-aminopeptidases (MetAPs), leaving a free α-Cys (Gentle et al., 2004). The efficiency of Met removal is heavily dependent upon the side chain of the penultimate amino acid. Reports vary from complete (100%) to only partial (71%) removal of N-Met when the next amino acid is Cys (Boissel et al., 1988; Hirel et al., 1989). During our original expression trials of the α-Cys fragment, in addition to removal of the N-Met we also observed partial cleavage of the Cys as judged by mass spectrometry (data not show). This was most likely due to overly active MetAPs. We also observed possible small molecule modification of the α-Cys. To combat these issues, the expression time was shortened from ~4 to 2 h to limit over-activity of the MetAPs, and free cysteine was added to the expression cultures in order to scavenge any Cys reactive small molecules in the cell (Gentle et al., 2004).

The $\alpha$-Cys protein expressed to high levels, but partitioned almost entirely to the insoluble fraction (Figure 3.6). The protein was solubilized from inclusion bodies using 6 M guanidine-HCl, and initially purified using $Ni^{2+}$ affinity chromatography under denaturing conditions. The $Ni^{2+}$ step purified the protein to greater than 80% as judged by SDS-PAGE, and from 4 L of bacterial culture yielded 73.5 mg (Figure 3.6). Electrospray ionization (ESI) mass spectrometric analysis indicated that an unidentified small molecule was still modifying the $\alpha$-Cys (Figure 3.7A). Treatment of the protein with tris(2-carboxyethyl)phosphine (TCEP) did not correct this modification (Figure 3.7B). However, when treated with O-methylhydroxylamine ($MeONH_2$), the modification was reversed, and yielded the expected protein mass within the error limits of the mass spectrometer (Figure 3.7D). Following $Ni^{2+}$ affinity purification, the $MeONH_2$ treated protein was purified to nearly 100% homogeneity via size exclusion chromatography (Figure 3.6). A band of roughly double the molecular weight of the $\alpha$-Cys was observed during the purification. We attributed this to the formation of a disulfide bond between the N-terminal Cys residues from two fragments (Figure 3.6). Treatment of the protein with TCEP abolished the putative disulfide (Figure 3.6).

An additional $\alpha$-Cys construct covering KinB residues 221 to 416, and also fused to a PreScission™ Protease cleavable $His_6$-tag, has been designed. This construct removes the 18 C-terminal residues that were disordered in the KinB/Sda crystal structure (see Chapter 2). We will likely use this construct for future ligation and crystallization experiments. A semisynthetic KinB 195-416 may crystallize more readily due to the absence of an unstructured region at its C-terminus. We have expressed and purified KinB A221C-416 using the same protocol as for KinB A221C-CT (Figure 3.6).

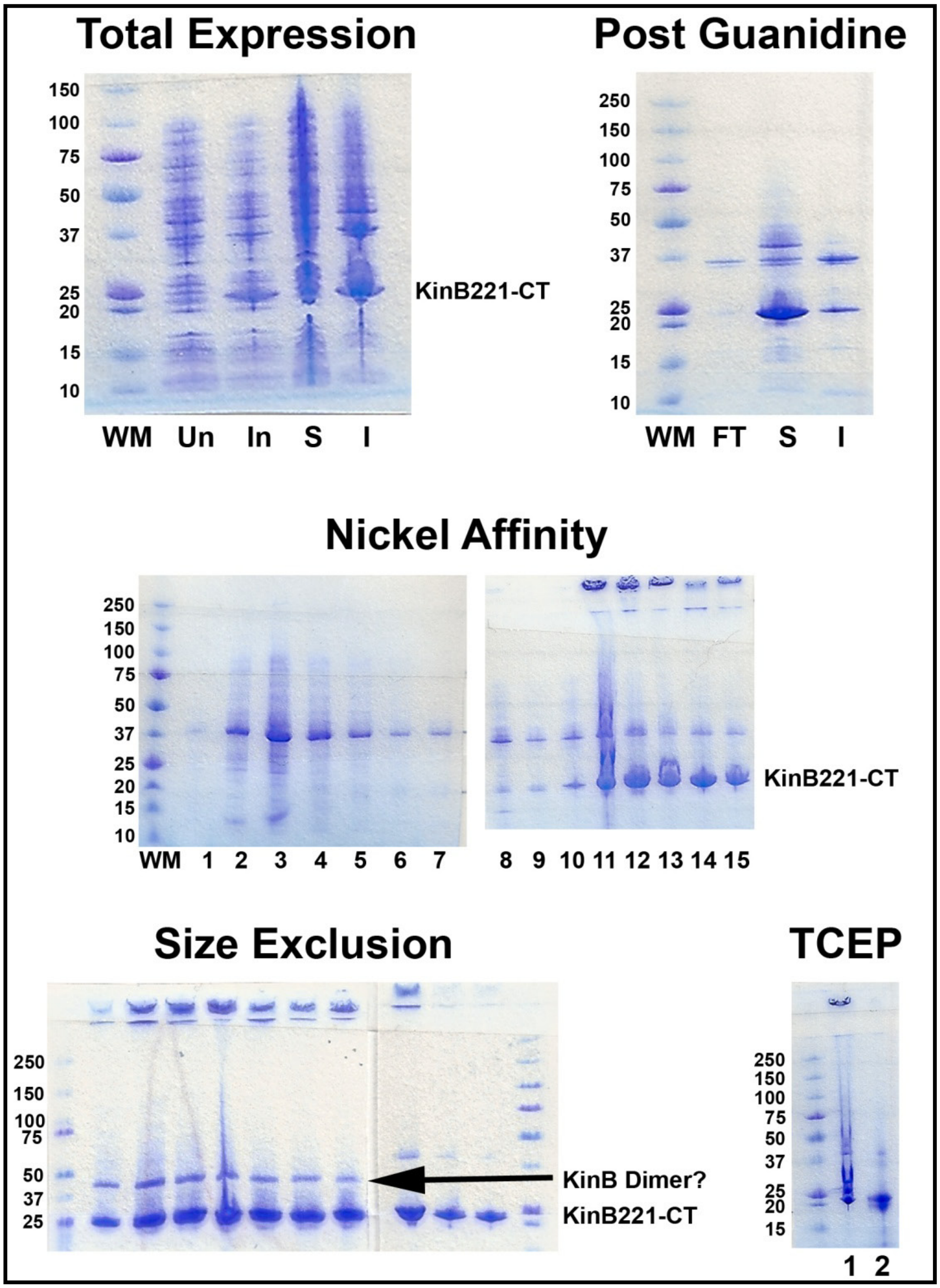


**Figure 3.6)** SDS-PAGE analysis of the various stages of the KinB α-Cys fragment purification. All chromatography steps were performed under denaturing conditions (6M guanidine-HCl). The size exclusion gel shows protein from the peak fractions. TCEP gel (1) - protein prior to TCEP treatment; (2) post TCEP treatment. WM - protein weight marker; Un - total E. coli protein levels prior to IPTG induction; In - total protein levels following IPTG induction; S - soluble fraction; I - insoluble fraction.

**Figure 3.7** (next two pages) ESI mass spectrometric analysis of the α-Cys KinB A221C-CT at various stages of the purification. (A) KinB following $Ni^{2+}$ affinity, prior to treatment with either $MeONH_2$ or TCEP. (B) KinB after the $Ni^{2+}$ column, treated with TCEP, but not with $MeONH_2$. (C) KinB after the $Ni^{2+}$ column, treated with $MeONH_2$, but not TCEP. (D) KinB after the size exclusion step ($MeOHN_2$ treated) and treated with TCEP. The expected mass of *Gst* KinB A221C-CT plus the PreScission™ Protease cleavable $His_6$-tag is 25,553 Daltons.

Figure 3.7

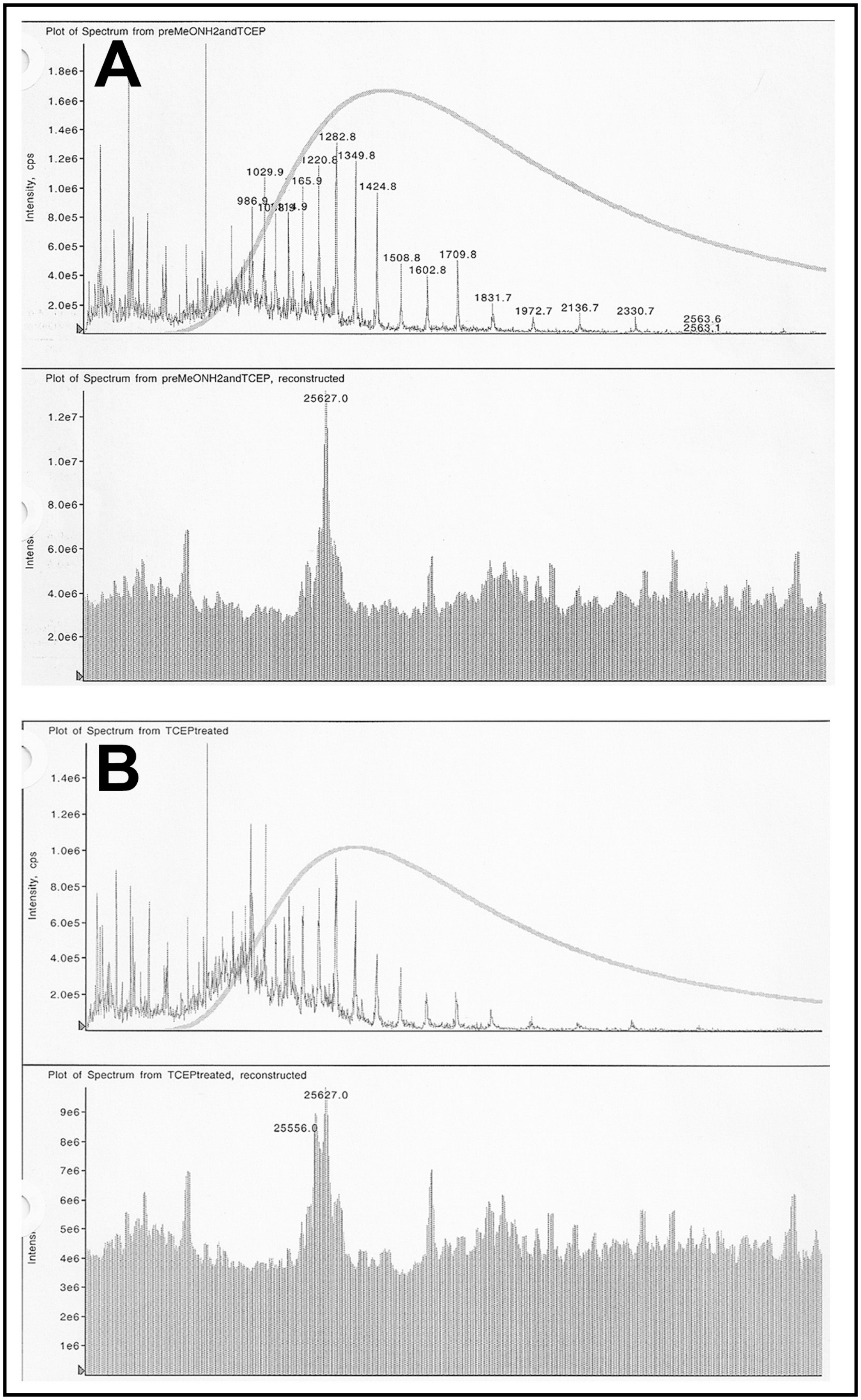
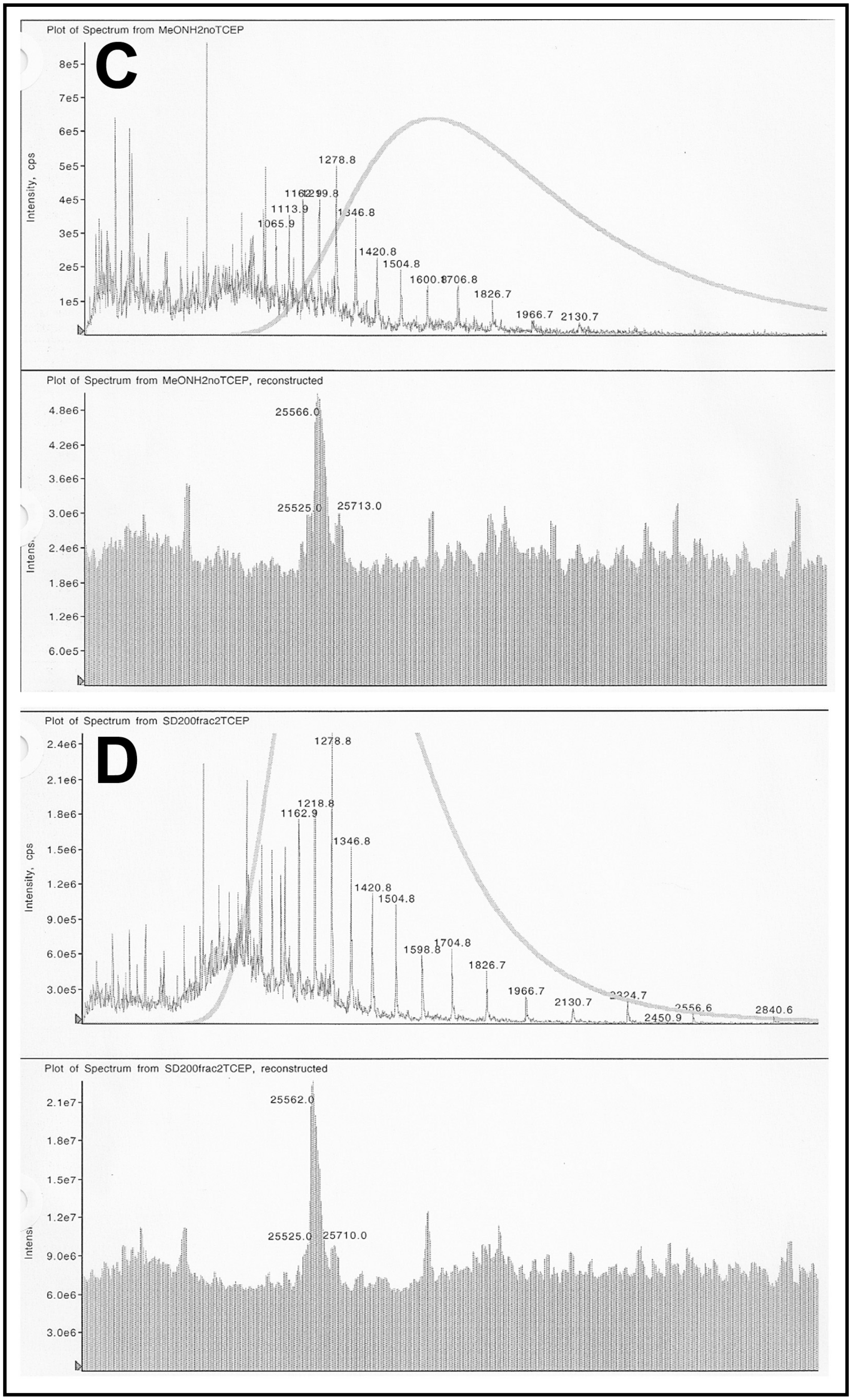

A
Plot of Spectrum from preMeONH2andTCEP
Intensity, cps
1.8e6
1.6e6
1.4e6
1.2e6
1.0e6
8.0e5
6.0e5
4.0e5
2.0e5
1282.8
1220.8
1349.8
1029.9
165.9
1424.8
986.9
1508.8
1709.8
1602.8
1831.7
1972.7
2136.7
2330.7
2563.6
2563.1
Plot of Spectrum from preMeONH2andTCEP, reconstructed
25627.0
1.2e7
1.0e7
8.0e6
6.0e6
4.0e6
2.0e6
Intens
B
Plot of Spectrum from TCEPtreated
Intensity, cps
1.4e6
1.2e6
1.0e6
8.0e5
6.0e5
4.0e5
2.0e5
Plot of Spectrum from TCEPtreated, reconstructed
25627.0
25556.0
9e6
8e6
7e6
6e6
5e6
4e6
3e6
2e6
1e6
Intens

Figure 3.7 continued



### *Synthesis of Histidine Analogs*

In order to trap the KinB active states, autophosphorylation and phosphotransfer, we synthesized stable analogs of 3-phospho-His (1-phospho-His as well) for trapping a KinB-Spo0F complex, and His covalently linked to ATP-γ-S (Figure 3.9), which will serve to bind the CA domain and lock the autophosphorylation conformation (Figure 3.10). Analysis of bisubstrate analog inhibitors of tyrosine kinases (essentially peptide substrates incorporating phospho-Tyr analogs) indicate that the tightest binding inhibitors incorporate a phospho-Tyr analog with a short spacer connecting the Tyr to the phosphate, rather than the direct linkage of phospho-Tyr (Figure 3.8) (Hines et al., 2005). This design exploits the fact that phosphoryl transfer involves a dissociative transition

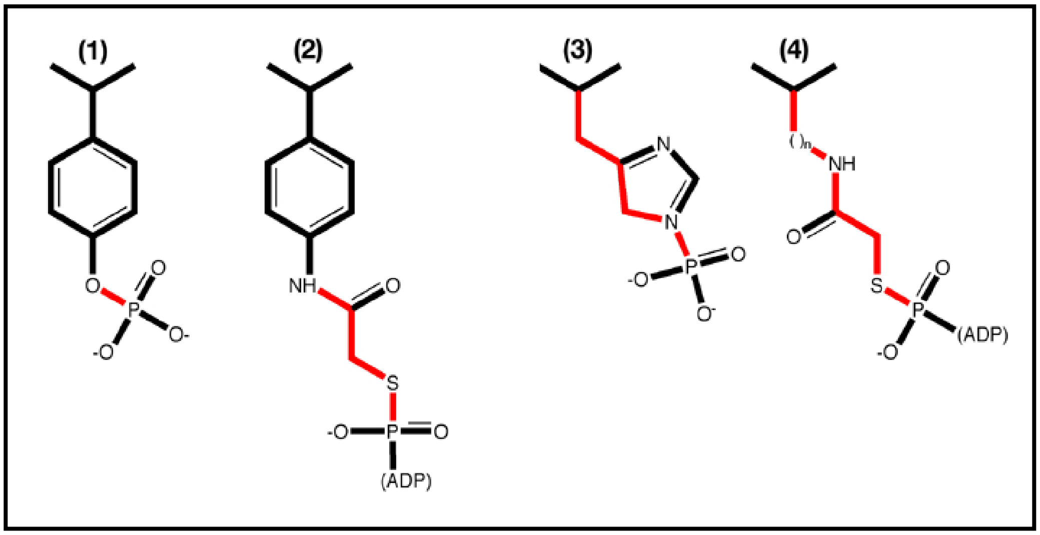


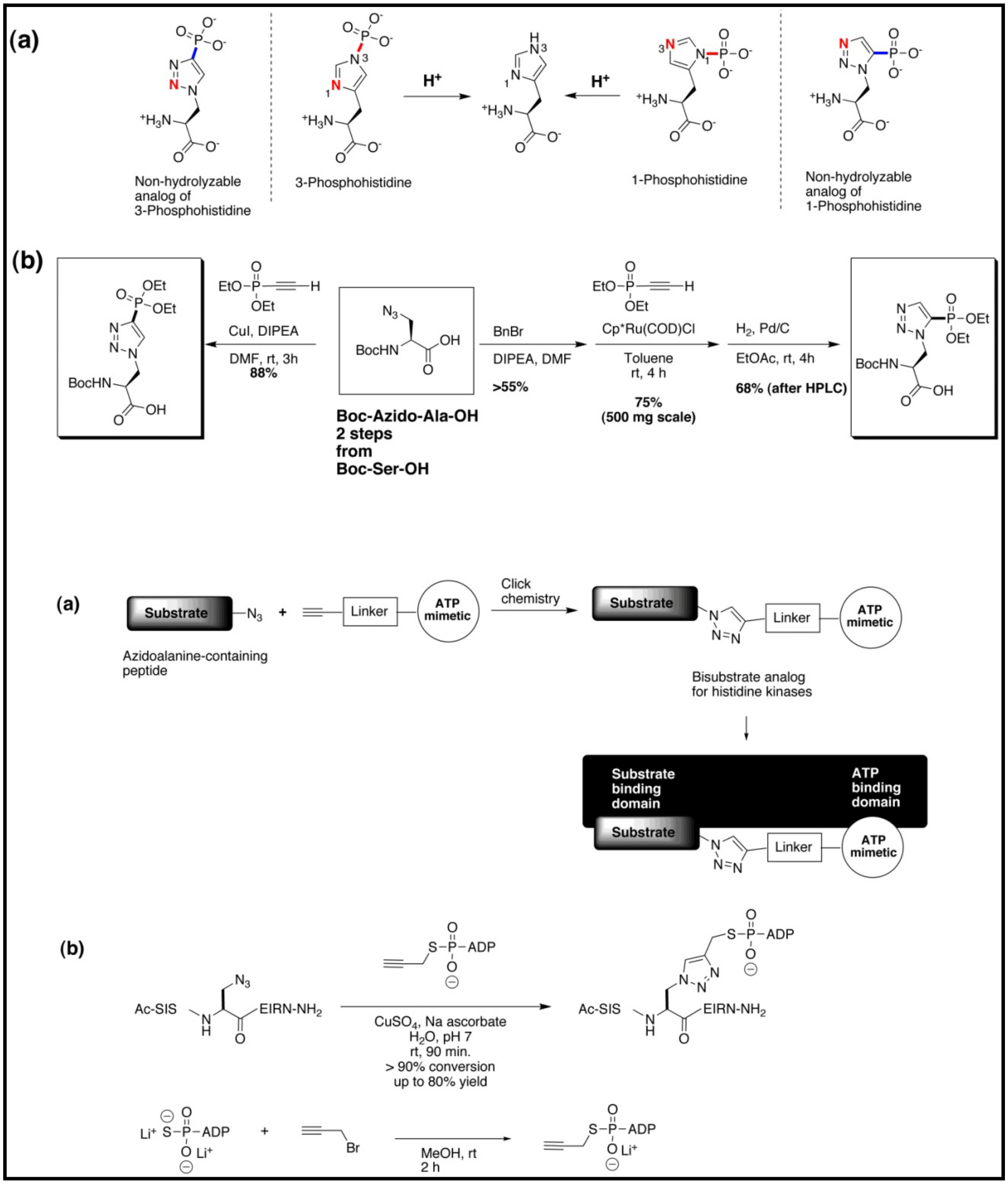
**Figure 3.8)** Comparison of ATP-γ-S conjugated tyrosine and histidine. Compound 1, phospho-Tyr; 2, Tyr analog designed to mimic the phosphorylation transition state (Parang et al., 2001); 3, phospho-His; 4, analogs designed to mimic the phosphorylation transition state of His. The final product depends on the linker length (n) of the starting amino acid, diaminopropionic acid for n=1, diaminobutanoic acid for n=2, ornithine for n=3, and lysine for n=4.

state in the case of the eukaryotic tyrosine kinases. It is unknown whether HKs use a dissociative or associative mechanism for autophosphorylation. In the associative mechanism (analogous to an $S_N2$ reaction) the phosphoryl group would adopt a pentavalent intermediate state bound to both the CA and DHp domains, before transferring completely to the DHp His. In contrast, by a dissociative mechanism the phosphoryl group would briefly dissociate from the CA domain completely to form an unstable metaphosphate intermediate (analogous to an $S_N1$ reaction) before binding to the DHp His.

Following the reasoning behind the design of phospho-Tyr analogs, we initially wanted to investigate a series of His analogs conjugated to ATP-γ-S with different linker lengths between the α-carbon and side chain amine (Figure 3.8). However, we have found it difficult to increase the number of carbons in the linker beyond what is outlined in Figure 3.9. Attachment of the alkyne with an additional carbon to ATP-γ-S did produce the correct product when analyzed by HPLC and mass spectrometry (data not shown). However, the product was unstable and we could neither purify nor characterize it. Thus, we are left with a His-ATP analog with a single carbon linker between the side chain and ATP-γ-S. This may not present a problem, as we have demonstrated that a peptide consisting of the KinB H-box motif with the histidine conjugated to ATP-γ-S by a single carbon linker (Figure 3.9) is an effective inhibitor of the autophosphorylation reaction (see below).

**Figure 3.9** (adjacent page) An overview of the synthetic schemes developed by Dr. Jung-Min Kee for the generation of phospho-His analogs (top) and a His-ATP conjugate (bottom). (Top-(a)) Stable (non-hydrolyzable) analogs of phospho-His have been designed by replacing the labile P-N bonds with stable P-C bonds. Since both 3-phospho-His and 1-phospho-His are found in biological systems, both isomers have been prepared. Significantly, both analogs are readily synthesized in preparative scale from commercially available Boc-Ser-OH in less than 5 chemical steps (Top-(b)). Boc-azidoalanine is treated with an ethynylphosphonate in the presence of copper iodide catalyst in DMF to afford the 3-phospho-His analog (Diethylphosphono-triazolylalanine). Alternatively, the same starting material (Boc-azidoalanine) can be converted to the 1-phospho-His analog by utilizing a ruthenium catalyst. Both 3-phospho-His and 1-phospho-His analogs have been successfully utilized in solid phase peptide synthesis. (Bottom-(a)) Overall synthetic scheme for introducing a His-ATP analog into a peptide. (Bottom-(b)) Detailed synthetic scheme for generating the ATP-γ-S alkyne conjugate and the His-ATP peptide. First, a peptide with an azidoalanine in place of the active His is synthesized. Commercially available ATP-γ-S is treated with propargyl bromide in methanol to give the alkyne conjugate. The alkyne conjugate is then added to the azido-peptide itself via click chemistry in the presence of a copper sulfate catalyst and sodium ascorbate, converting the azidoalanine to the His-ATP conjugate. To date, the ATP conjugate has only been introduced to a peptide comprising the KinB H-box, not the complete 26-mer that will be used for the NCL reaction outlined in this chapter.

Figure 3.9


(a)
Non-hydrolyzable analog of 3-Phosphohistidine
3-Phosphohistidine
H+
H+
1-Phosphohistidine
Non-hydrolyzable analog of 1-Phosphohistidine
(b)
CuI, DIPEA
DMF, rt, 3h
88%
Boc-Azido-Ala-OH
2 steps
from
Boc-Ser-OH
BnBr
DIPEA, DMF
>55%
Cp*Ru(COD)Cl
Toluene
rt, 4 h
75%
(500 mg scale)
H2, Pd/C
EtOAc, rt, 4h
68% (after HPLC)
(a)
Substrate
Azidoalanine-containing peptide
Linker
ATP mimetic
Click chemistry
Bisubstrate analog for histidine kinases
Substrate binding domain
ATP binding domain
(b)
Ac-SIS
EIRN-NH2
CuSO4, Na ascorbate
H2O, pH 7
rt, 90 min.
> 90% conversion
up to 80% yield
ADP
MeOH, rt
2 h

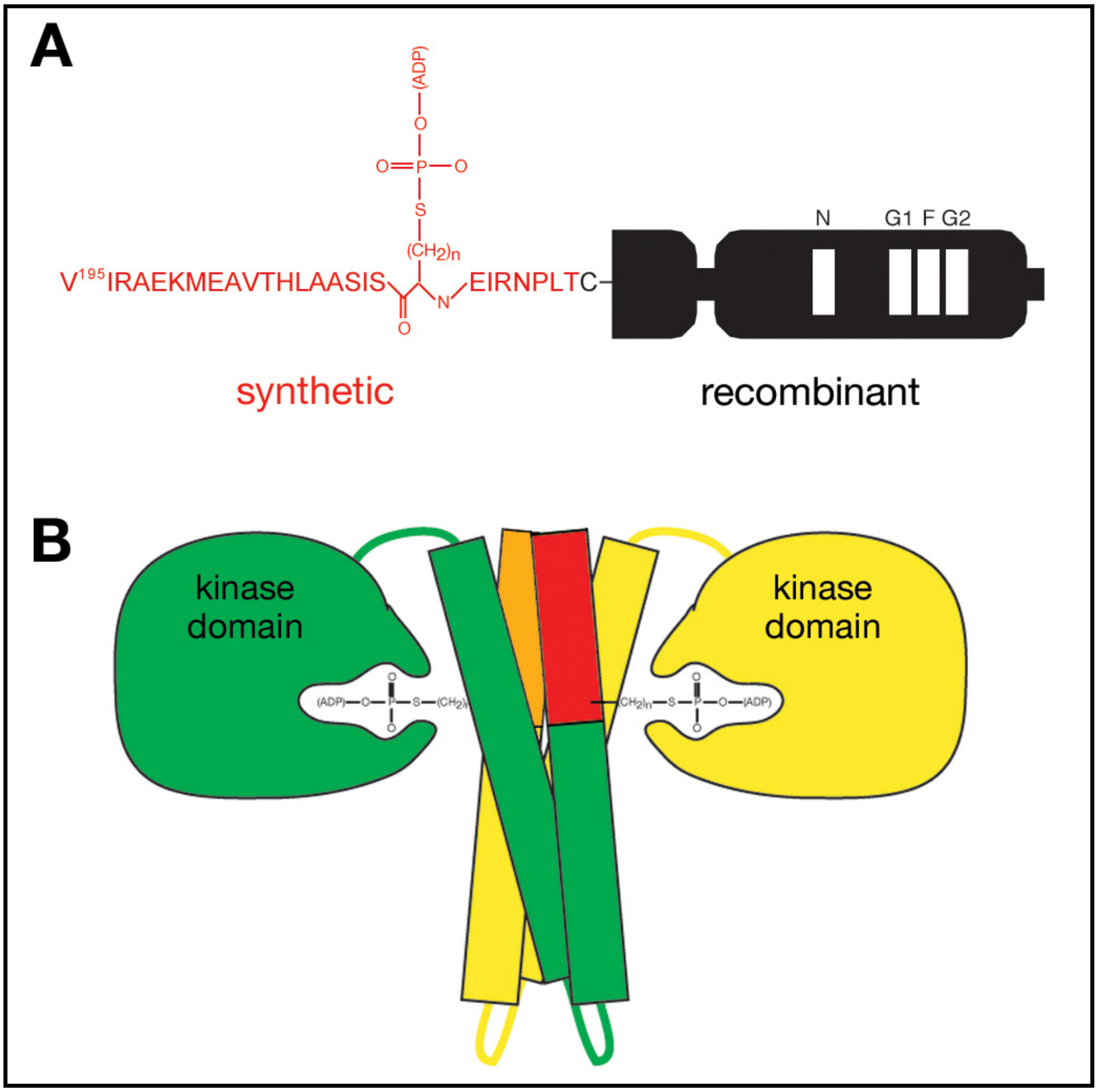


**Figure 3.10)** (A) Schematic showing the final, semisynthetic native chemical ligation products. Only the linker plus ATP-γ-S is shown in the figure (the triazole analog in place of the His is not shown). Initially, different chain lengths (n) linking residue 213 to ATP-γ-S were to be pursued, but we have experienced difficulty adding extra carbons to the linker. Red indicates the synthetic peptide. The ligated recombinant α-Cys fragment is colored black. (B) Structural cartoon illustrating the desired NCL products trapped in the act of autophosphorylation. ATP-γ-S covalently linked to His213 should bind the KinB CA domain, thereby locking its position in the active conformation. The synthetic and recombinant parts of one monomer are colored orange and yellow, respectively, and the other monomer is colored red/green. The DHp domain α-helices are illustrated as cylinders.

An additional analog of phospho-His, (4'-phosphoryl-2'-furyl)-alanine, based on the work of Schenkels et al. (Schenkels et al., 1999) has been synthesized in the laboratory of Dr. J. David Warren of Weill Cornell Medical College (Figure 3.11). Matthew Bick carried out the initial steps of the synthesis (steps 1-4). Dr. Guangtao Zhang, a postdoc in the Warren laboratory, developed the final synthetic scheme involving pseudoephedrine (Figure 3.12).

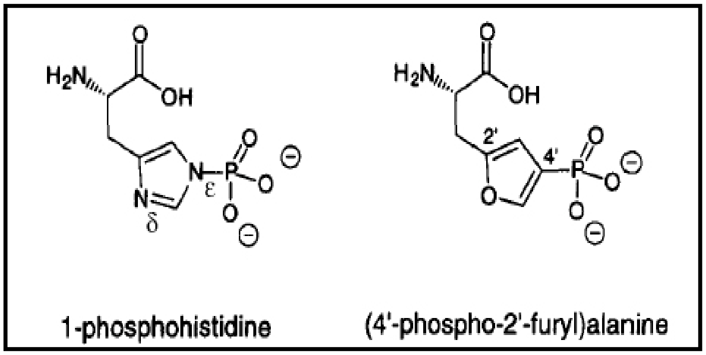


**Figure 3.11)** Comparison of 3-phospho-His (differing nomenclature leads some researchers to name the amino acid on the left 1-phospho-His) and phosphofurylalanine. Adapted from (Schenkels et al., 1999).

We had planned to use phosphofurylalanine as an analog of phospho-His in order to stabilize a complex between KinB and the response regulator Spo0F. It was assumed that after incorporation of the intermediate 8 (Figure 3.12) into a peptide, removal of the ethyl protecting groups from the phosphonate would occur spontaneously during HF cleavage of the peptide from the solid support resin. However, this has proven to be more complicated than we originally anticipated. We have in fact not been able to remove the ethyl groups from the amino acid, either by treatment with HF or with trimethylsilyl iodide, a standard deprotection method for phosphonate. Therefore, the usefulness of phosphofurylalanine as a tool for trapping an HK-RR complex is uncertain.

More likely, the 3-phospho-His analog described in Figure 3.9A, diethylphosphonotriazolylalanine, will be our analog of choice for investigating a KinB-Spo0F complex.

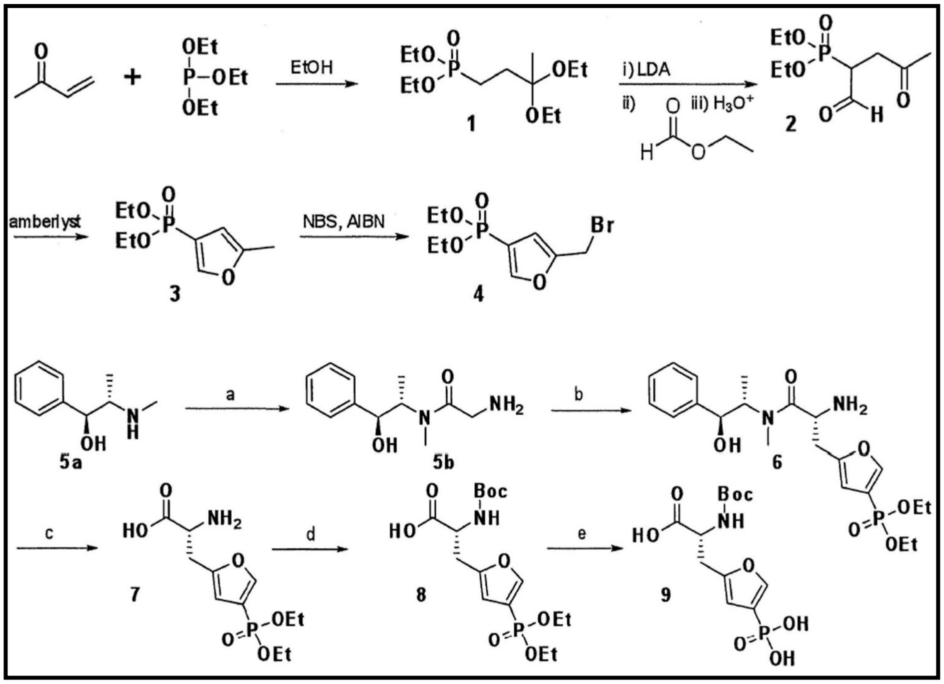


**Figure 3.12)** Synthesis of N-Boc-L-(4'phosphoryl-2'-furyl)alanine via pseudoephedrine mediated alkylation (Zhang and Warren, 2009). Synthesis of the phosphonate furyl bromide (steps 1-4) and precursors have been previously reported (Harvey, 1966; Schenkels et al., 1999). The use of pseudoephedrine (5a) to generate the intermediate precursor pseudoephedrine glycinamide (5b) for reaction with the alkyl bromide (4) (Myers et al., 1997) represents a departure from the synthetic scheme published by Schenkels et al. (Schenkels et al., 1999). As a result, the method described here avoids the tedious enzymatic resolution step reported by Schenkels at al. (Schenkels et al., 1999). The intermediate furyl phosphonate (8) has been incorporated into a peptide using Boc-based SPPS. a. LiCl, Pseudoephedrine, LiOMe, THF; b. LDA, LiCl, 5b, -78 to 0 ºC, THF; c. 0.5 M NaOH/dioxane/MeOH (v/v/v=1:1:1), Δ; d. $Boc_2O$, $NaHCO_3$, dioxane; e. TMSCl, DMF, microwave, 180 ºC.

### *KinB Native Chemical Ligation*

With the necessary reagents in hand, we next turned to working out the KinB-CC NCL reaction. We first carried out the ligation reaction with a wild-type KinB 26-mer peptide (without His analogs) in order to preserve the precious amounts of His analogs. The KinB NCL reaction conditions were established by Dr. Jung-Min Kee (see methods). Ligated semisynthetic KinB, refolded from a lyophilized state, demonstrated approximately 50% activity in an autophosphorylation assay compared with wild-type (data not shown). Optimized refolding conditions had not yet been worked out at the time of the NCL reaction. We anticipate greater recovery of active protein from future ligation reactions using optimal refolding conditions (see above). The ligation product was not resolved on HPLC from the α-Cys fragment. However, mass-spectrometric analysis of the HPLC peak confirmed the product formation. To facilitate the purification of the product from the α-Cys fragment, it was necessary to drive the reaction to completion, and hence why a 5-fold excess of peptide was used (see methods).

### *An H-box Peptide With the His-ATP Analog in Place of the Active Site His is an Effective Inhibitor of the KinB Autophosphorylation Reaction*

Crystallization experiments often require large amounts of protein in order to maximize the chance of obtaining crystals. Before scaling up the semisynthesis of KinB for crystallization trials, we first needed to establish whether our strategy for capturing the autophosphorylation active state using the His-ATP was sound. Because we have been unable to generate His-ATP analogs with varying linker lengths, our only choice for this amino acid is the one described in Figure 3.9, which has a single carbon between the triazole ring and the ATP-γ-S. In order to lock the active state conformation, the linker should be just the right length, and this length would be different depending on the mechanism of autophosphorylation, either associative or dissociative. If the linker is too long, the ATP moiety may bind the nucleotide pocket of the CA domain, but residues from the DHp domain that normally make CA contacts may not be able to reach the CA domain (Figure 3.13B). The opposite is also true. If the linker is too short, DHp residues may make CA contacts, but the ATP would not reach the back of the nucleotide pocket (Figure 3.13A).

We reasoned that a short peptide comprised of the KinB H-box sequence (SISHEIRNPL) and containing the His-ATP analog with the correct linker length could be used as an inhibitor of the autophosphorylation reaction. Presumably, a peptide with a linker that is too short would inhibit only as well as the same peptide without the His-ATP analog (unmodified peptide), and a peptide with a linker that is too long would only inhibit as well as ATP-γ-S. A peptide with the correct linker would inhibit to a greater degree than both the unmodified peptide and ATP-γ-S.

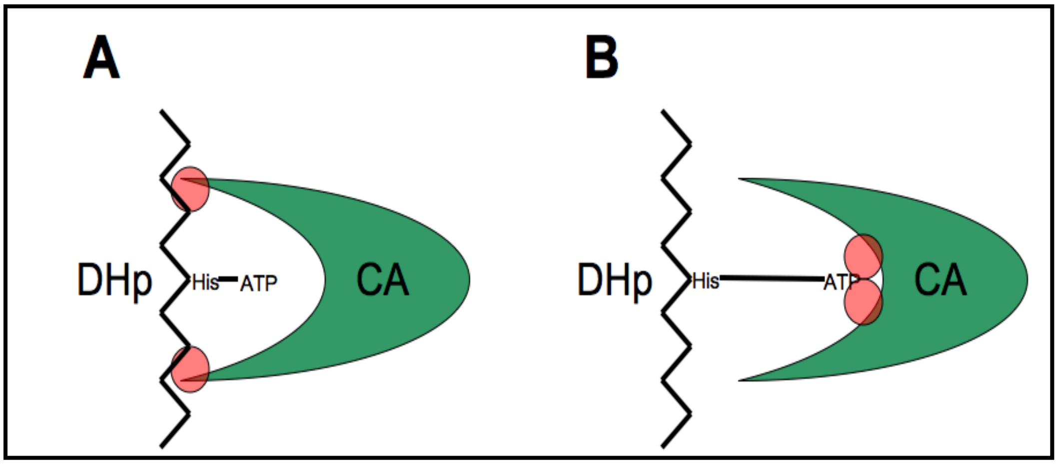


**Figure 3.13)** An exaggerated cartoon illustrating the potential effects of non-ideal linker lengths between the triazole histidine analog of the DHp domain and conjugated ATP-γ-S. Hypothetical contacts are represented by red spheres. (A) If the linker is too short, the ATP will be unable to bind the CA domain nucleotide pocket. (B) If the linker is too long, residues from the DHp domain will be unable to make contacts with the CA domain.

We carried out *in vitro* autophosphorylation assays with KinB 195-CT and varying concentrations of either unmodified H-box peptide, the ATP-γ-S alkyne conjugate (Figure 3.9, bottom), and the H-box peptide with the histidine analog conjugated to ATP-γ-S. We found that the ATP-conjugate H-box peptide is an effective inhibitor of the autophosphorylation reaction, with an $IC_{50}$ value of 5.1 μM (Figure 3.15). The ATP-alkyne inhibited the autophosphorylation of KinB with approximately 2.5-fold lower efficiency ($IC_{50}$ of 13 μM). In stark contrast, the inhibition potency of the unmodified H-box peptide (without ATP-γ-S conjugated to the histidine analog) was lower by at least three logs ($IC_{50}$ of 6.50 mM) when compared with the ATP-conjugate H-box peptide (although the data for the unmodified peptide does not fit a well defined curve, it is clear that it is not an efficient inhibitor of KinB autophosphorylation). The results of these experiments suggest that the linker length between the triazole side chain and the ATP-γ-S of our analog is at least close to optimal. This analog should prove to be competent for trapping the KinB active state of autophosphorylation.

In addition, we calculated a Hill curve, which measures the cooperativity of ligand binding to different regions of a protein, for the ATP-conjugate H-box peptide. In our case, we are measuring the influence that an ATP-conjugate H-box peptide bound to one CA domain has on the binding of an ATP-conjugate H-box peptide to the other CA domain. The result of a Hill coefficient of 1 indicated independent binding of the peptide to each KinB CA domain (Figure 3.14).

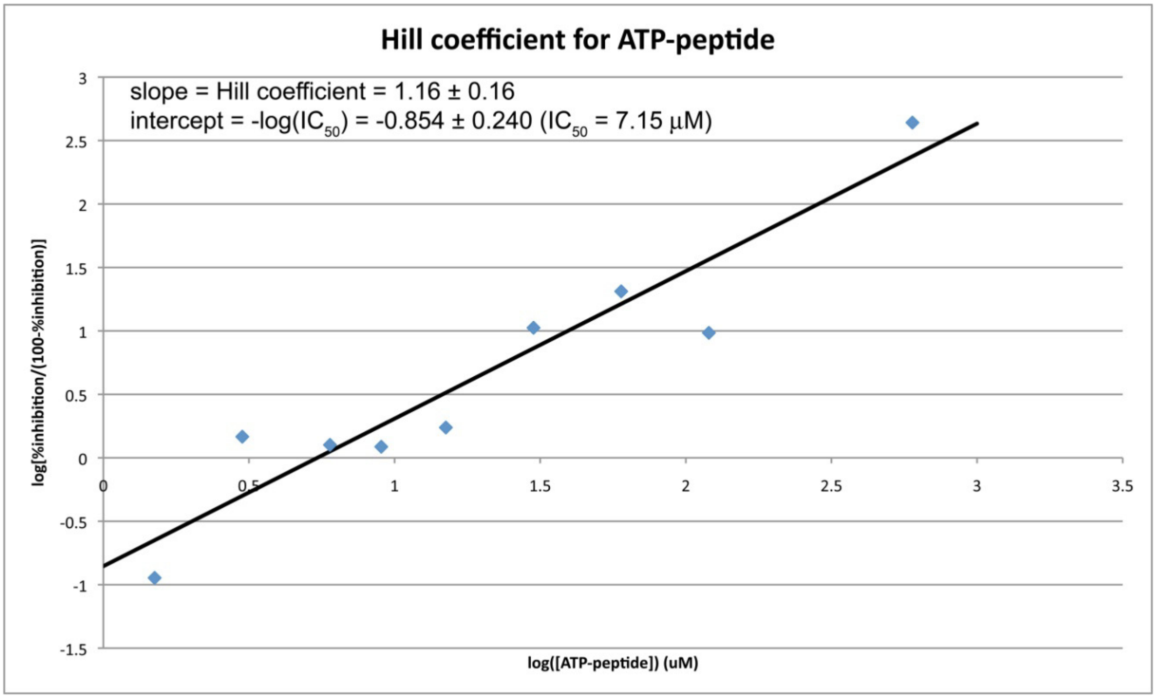
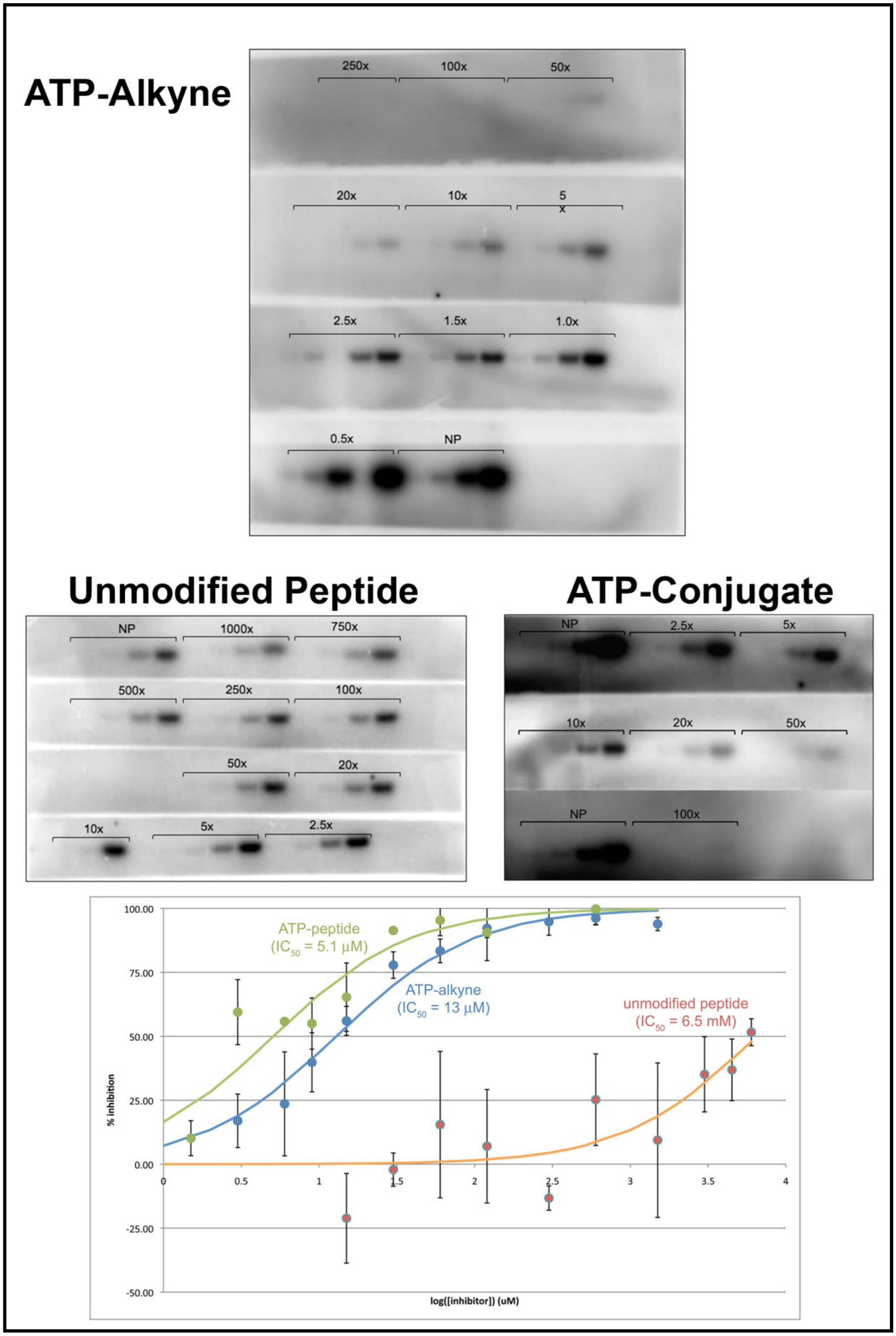


**Figure 3.14)** Hill coefficient calculated for the KinB ATP-conjugate H-box peptide. The value of 1 indicates that the peptide binds to the two CA domains independently.

**Figure 3.15** (adjacent page) Autophosphorylation time course inhibition assays performed with KinB 195-CT and either the ATP-alkyne, unmodified H-box peptide, or ATP-conjugate H-box peptide. The ATP-conjugate peptide is at least an order of magnitude more efficient as an inhibitor of the autophosphorylation reaction than the same peptide without a conjugated ATP, and is approximately 2.5 times as efficient as the ATP-alkyne alone. The protein concentration for each reaction was 6 μM. Values that bracket 4 lanes each indicate the fold inhibitor concentration above the protein concentration. Time points at 15 s, 1 min, 5 min, and 15 min were taken for each concentration. To generate the inhibition curves, bands were quantified by phosphorimagery using ImageQuant™ software (GE Life Sciences). Experiments were performed in triplicate and the standard error of the mean was calculated in order to generate error bars.

Figure 3.15



ATP-Alkyne
250x
100x
50x
20x
10x
5
x
2.5x
1.5x
1.0x
0.5x
NP
Unmodified Peptide
NP
1000x
750x
500x
250x
100x
50x
20x
10x
5x
2.5x
ATP-Conjugate
NP
2.5x
5x
10x
20x
50x
NP
100x
ATP-peptide
(IC$_{50}$ = 5.1 μM)
ATP-alkyne
(IC$_{50}$ = 13 μM)
unmodified peptide
(IC$_{50}$ = 6.5 mM)
100.00
75.00
50.00
25.00
0.00
-25.00
-50.00
% inhibition
0
0.5
1
1.5
2
2.5
3
3.5
4
log([inhibitor]) (uM)

## Conclusions and Future Directions

The reagents necessary for pursuing structural studies of a semisynthetic KinB-CC have been established. We have shown that KinB 195-CT maintains its activity when refolded from the denatured state, a crucial preliminary piece of information needed before embarking on in-depth semisynthetic studies. A 26-mer peptide composed of the N-terminal residues of KinB 195-CT can be joined by NCL to the remaining C-terminal portion of KinB to produce a functional enzyme. Analogs of phospho-His and His-ATP have been synthesized that will allow us to trap HK active state conformations, both autophosphorylation and phosphotransfer, when incorporated into a full-length semisynthetic KinB-CC. The His-ATP analog has been incorporated into a peptide comprising the KinB H-box (though not the full-length 26-mer), while the 3-phospho-His analog has been incorporated into the 26-mer. The KinB H-box peptide containing the His-ATP analog is an effective inhibitor of the KinB autophosphorylation reaction, indicating that this analog is well suited to trap the autophosphorylation state. Mutation of DHp-CA domain interface residues in the context of the His-ATP semisynthetic KinB-CC should further promote the autophosphorylation state (see below).

Our primary concern is that we will not be able to obtain enough semisynthetic KinB-CC, and at a level of purity high enough to conduct a sufficient number of crystallization trials. The presence of a C-terminal His$_6$-tag on the recombinant portion of KinB should facilitate purification of the ligated KinB-CC product away from any unliganded N-terminal peptide (the N-terminal peptide is in excess during the NCL reaction, so the presence of free $\alpha$-Cys protein should be minimal) if the full-length product cannot be resolved by HPLC. The use of automated crystallization robots will also help us conserve material by allowing us to set up sub-μl crystallization drops.

The inactive conformation observed in the TM0853 (Marina et al., 2005) structure is stabilized by a significant (1250 Å buried surface) interface between the CA and DHp domain of one monomer (Figure 3.16). Mutations designed to destabilize this interface resulted in increased rates of autophosphorylation, indicating that the interface must be broken during the autophosphorylation reaction (Marina et al., 2005). Through sequence alignment, we have identified the corresponding DHp-CA interface residues in KinB that when mutated should disrupt the DHp-CA interface and promote autophosphorylation. These mutants will be actively pursued in the context of a semisynthetic KinB.

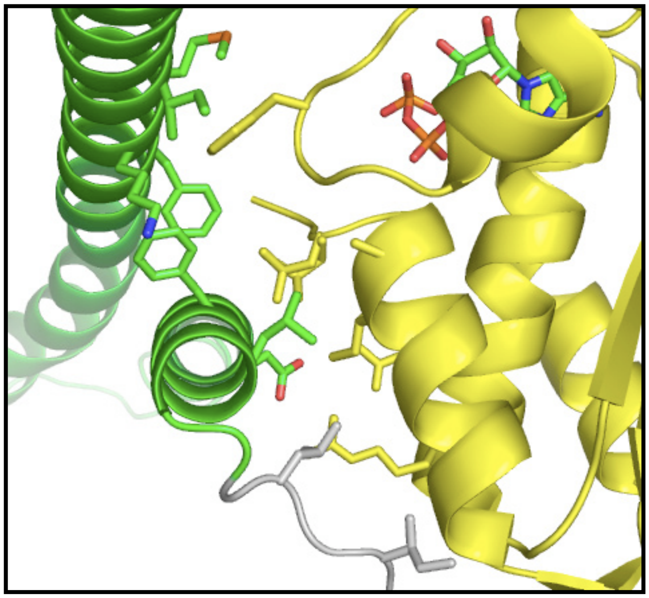


**Figure 3.16)** Close-up of the TM0853 DHp-CA domain interface (Marina et al., 2005). The DHp domain (green), the CA domain (yellow), and the linker (grey) are all from the same monomer. Side chains of residues that form the interface (mostly hydrophobic) are shown. Disruption of the DHp-CA interface through the introduction of mutant residues increased the rate of autophosphorylation in TM0853.

# Materials and Methods

## *Cloning and Protein Expression*

KinB A221C-CT was PCR (30 cycles of 95 ºC for 30 s, 55 ºC for 30 s, and 68 ºC for 55 s, followed by a final extension step of 275 s at 68 ºC) amplified from plasmid pET21a-KinB 195-CT using primers MB-44 (NdeI) and MB-94 (XhoI). MB-94 adds a PreScission™ Protease site, followed by an XhoI site, directly after KinB Leu 434. The reaction was purified with a PCR clean up kit (Qiagen), and the amplicon was digested for 2 h at 37 ºC with NdeI and XhoI. The digested DNA was ligated into NdeI/XhoI digested pET21a at 15 ºC overnight, using T4 DNA ligase (NEB), to afford the C-terminally PreScission™ Protease cleavable $His_6$-tagged KinB construct. Following digestion to remove the $His_6$-tag, the vector-derived sequence EVLFQ remains at the C-terminus of KinB. Ligation reactions were transformed and selected on LB agar supplemented with 150 μg/ml ampicillin. A nearly identical cloning strategy was used with primer MB-115 to make a C-terminally $His_6$-tagged construct truncated after KinB 416, which may be used in future experiments (residues 417 to 434 were disordered in the KinB/Sda structure. Removal of these residues may aid in crystallization).

For expression of the α-Cys fragment, pET21a-KinB A221C (C-term $His_6$-tag) was expressed in E. coli Rosetta2 (DE3) (Novagen). Cultures were grown at 37 ºC with shaking at 200 RPM. At an $OD_{600}$ = 0.6, 10 mM L-cysteine (1 M stock buffered to pH 7.0 with NaOH) was added to the cultures. Protein expression was induced at an $OD_{600}$ = 1.2 with 1 mM IPTG, and an additional 5 mM cysteine was added. Expression was carried out for 3 hr at 37 ºC with shaking at 200 RPM. Cultures were spun at 4,000 RPM for 30 min at 4 ºC. The bacterial pellet from 4 L of culture was resuspended in 25 ml of 250 mM NaCl and 100 mM sodium phosphate buffer, pH 6.0, and stored at -80 ºC.

### *Purification of the KinB α-Cys Fragment*

Cells carrying the KinB α-Cys fragment were lysed in between 12,000 and 14,000 psi at 4 ºC using an EmulsiFlex-C5™(Avestin) high-pressure homogenizer. The crude lysate was spun at 16,000 RPM, 4 ºC for 30 min in order to pellet the insoluble fraction containing KinB α-Cys. KinB α-Cys was solubilized from inclusion bodies by resuspension of the pellet in 40 ml of $Ni^{2+}$ column buffer A (250 mM NaCl, 100 mM sodium phosphate, pH 8.0, 5 mM imidazole, 6 M guanidine-HCl), and passage through a type A dounce homogenizer (10 passes each with a loose and tight fitting pestle). The sample was spun again at 16,000 RPM, 4 ºC for 30 min in order to pellet out any remaining insoluble material. The supernatant was passed over two 5 ml $Ni^{2+}$ chelating column in series, equilibrated with $Ni^{2+}$ buffer A. KinB α-Cys was eluted from the column using a step gradient to 100% $Ni^{2+}$ B (250 mM NaCl, 100 mM sodium phosphate, pH 6.0, 200 mM imidazole, 6 M guanidine-HCl). (Note that despite the presence of an N-terminal cysteine, DTT was kept from the protein buffers because of the possibility of exchange with the thioester of the N-terminal KinB peptide, which could lead to hydrolysis.) Following the $Ni^{2+}$ step, KinB α-Cys was treated overnight at room temperature, and shielded from light, with O-Methylhydroxylamine-HCl ($MeONH_2$ stock made up to 1.5 M in 250 mM NaCl, 100 mM sodium phosphate, pH 5.0, 6 M guanidine-

HCl) at a final concentration of 250 mM. Adjustment of the protein sample to pH 5.0 with HCl was required. The following day the protein was concentrated, using 15 ml 5000 MWCO Vivaspin concentrators (Vivascience), to 10 mg/ml and passed over a Superdex™ 75 (16/60) gel filtration column, equilibrated with 250 mM NaCl, 100 mM sodium phosphate, pH 6.0, 6M guanidine-HCl, at 0.3 ml/min. The protein eluted immediately after a 30 ml void volume. All protein samples for SDS-PAGE analysis were first purified from guanidine via ethanol precipitation. Samples were taken pre and post $MeONH_2$ treatment, and after treatment with TCEP on a small scale, for analysis by ESI mass spec. The protocol detailed above was also used to purify KinB A221C - 416 (C-term PreScission™ Protease $His_6$).

### *Refolding Denatured KinB 195-CT*

For initial refolding experiments, *G. stearothermophilus* KinB 195-CT in storage buffer (150 mM NaCl, 10 mM $MgCl_2$, 20 mM Tris, pH 7.0, 15% glycerol) at 1 mg/ml was concentrated to 11mg/ml using a 10,000 MWCO Vivaspin concentrator from Vivascience. The protein was diluted with 8 M guanidine-HCl to a final concentration of 2.9 mg/ml (6 M guanidine final). To 10 μl of protein, 190 μl of FoldIt™ (Hampton) solution was quickly added (16 solutions were tested in total), and the solution was mixed by gentle pipetting. The protein concentration following dilution was approximately 150 μg/ml. The protein was incubated for 60 min at RT in the refolding solution. A sample from each refolded protein solution was taken for light scattering analysis at 302 nm. Four solutions that gave the lowest absorbance values (showed the lowest levels of turbidity due to protein aggregation) were chosen for subsequent autophosphorylation activity assays. Refolded protein was concentrated to between 1 and 2 mg/ml, using 500 µl Vivaspin concentrators, and the buffer was exchanged three times on the concentrator to storage buffer (see above) plus 1 mM DTT.

The same protocol was used for large-scale (~3.5 mg of KinB 195-CT) refolding trials, except that 10 mM DTT was added to refolding solution. FoldIt™ solution 13 (55 mM MES, pH 6.5, 264 mM NaCl, 11 mM KCl, 550 mM guanidine-HCl, 1.1 mM EDTA) was chosen for the large-scale experiments, as it produced the best results from our initial trials. After refolding, the protein was concentrated to 6.3 mg/ml in a 300 µl volume. This solution was passed over an Supderdex™ 200 column (16/60) (GE Life Sciences), equilibrated with 150 mM NaCl, 20 mM Tris-Cl, pH 7.0, 10 mM $MgCl_2$, 1 mM DTT, at 0.3 ml/min. Peak fractions were collected, and the protein from these fractions was used in an autophosphorylation activity assay. The activity assays were performed the same as described in Chapter 2 of this thesis.

### *SPPS of the N-terminal (Val195-Thr220) KinB Thioester Peptide*

Matthew Bick conducted the initial peptide synthesis studies. Final optimization of the 26-mer peptide synthesis was achieved by Dr. Jung-Min Kee, PhD (Muir Laboratory, Rockefeller). The KinB thioester peptide (VIRAEKMEAVTHLAASIS<u>H</u>EIRNPLT) was synthesized according to the protocol developed by Muir and coworkers, with minor modifications (Camarero and Muir, 2001). The underlined His residue was either regular His or phospho-His analog. The steps followed for the synthesis were as follows:

1. 1 g of MBHA resin (1.0 mmol/g) was placed into the peptide synthesis vessel, and enough DMF was added to cover the dry resin (<5 ml). Incubated for 30 min.
2. The resin was drained, covered with 5% DIEA/DMF, incubated for 1 min, then drained again. This process was repeated twice, and then three 20-sec flow washes with DMF were performed.
3. 1.222 g 3-bromopropionic acid (8 mmol) were dissolved in a minimal volume of DCM (<1 to 2 ml), and then 630 µl of 99% DIPC (4 mmol) was added. The solution was shaken vigorously and left to incubate for 10 min. A white precipitate corresponding to the diisopropyl urea appeared during the activation reaction. The solution was filtered through a Pasteur pipette containing glass wool to remove the precipitate.
4. The filtered solution from step 3 was added to the MBHA resin in the peptide synthesis vessel. 800 µl of DIEA (4.5 mmol) was added plus the minimal amount of DMF (1 to 2 ml) required to achieve a uniform slurry. The coupling reaction was incubated for 30 min with occasional stirring using a glass rod.
5. The resin was drained and washed thoroughly with three 20 s DMF flow washes.
6. The coupling process (steps 3 to 5) was repeated two more times.
7. 5 ml of $Ac_2O$/DIEA/DMF solution was added to the resin and incubated for 10 min. The resin was drained and three DMF flow washes were performed, as in step 3.
8. 5 ml AcSH/DIEA/DMF solution were added to the resin and incubated for 20 min. The resin was drained and three DMF flow washes were performed. This entire process was repeated two more times.

9. 5 ml BME/DIEA/DMF solution were added to the resin and incubated for 20 min. The resin was drained and three DMF flow washes were performed.
10. Step 9 was repeated two more times.
11. 3 molar equivalents of Boc-Thr-OSu were dissolved in approximately 6 ml DMF, and this solution was added to the 3-mercaptopropionamide-MBHA resin. 714 µl of DIEA (4 molar equivalents, 4 mmol) was added and the coupling reaction was incubated for 3 to 4 hr with occasional stirring.
12. Three DMF flow washes were performed. 5 ml of $Ac_2O$/DIEA/DMF solution was added to the resin and incubated for 10 min. Next, three 20 s flow washes were performed with dichloromethane (DCM). At this stage, the resin was dried under vacuum and stored in a vacuum desiccator.
13. ~5 mg of dry resin was removed, put in a microcolumn, and deprotected with TFA for 2 min. The column was drained, and first a flow wash was performed with DMF and then with DCM. The resin was drained and dried again under vacuum.
14. A ninhydrin test was performed to check the loading on the resin, which was found to be 0.2 mmol/g.
15. The resin that had been stored was re-swollen in DMF for 1 h and the synthesis of the complete 26-mer KinE peptide was continued using the in situ neutralization/HBTU activation protocols for Boc-SPPS (Schnolzer et al., 2007).
16. Once the synthesis was complete, a flow wash with DMF and DCM was performed, as in step 12. The resin was dried under vacuum and stored in a vacuum desiccator.
17. The α-thioester resin from step 16 was cleaved by treating with 10 ml of HF:*p*-cresol solution for 1 hr at 4 °C in an HF cleavage apparatus.

18. The HF was removed under vacuum and both the peptide and resin were resuspended in <40 ml cold diethyl ether with gentle stirring for 10 min. The suspension was filtered on a glass fritted funnel under vacuum.
19. The material was washed in the filter (containing the cleaved peptide and the resin) three times, each time with 10 ml of cold diethyl ether.
20. 10 ml freshly degassed 50% acetonitrile in water containing 0.1% TFA was added to the filter to dissolve the cleaved peptide, and then filtered. This process was repeated two more times, and then the material was recovered and the filtrates were lyophilized.
21. The lyophilized propionamide α-thioester KinB peptide was purified by preparative C18 RP-HPLC (26-35% solvent B over 1 h).

### *KinB Native Chemical Ligation*

Conditions for the NCL reaction were established by Dr. Jung-Min Kee, PhD (Muir Laboratory, Rockefeller University). Gel filtration purified KinB α-Cys fragment was provided at a concentration of 880 μM. 70 μl of KinB α-Cys peptide solution (0.06 μmol in guanidine buffer) was added to a 5-fold excess (1.0 mg or 0.3 μmol) of lyophilized N-terminal α-thioester KinB 26-mer (unmodified histidine at position 213) in an eppendorf tube in order to dissolve the thioester. Mercaptophenylacetic acid (MPAA) and TCEP were added to the solution to a final concentration of 200 mM and 50 mM, respectively. At this point the pH of the solution dropped to ~ pH 2, and some MPAA remained undissolved. The solution was adjusted to pH 7.0 using 2N NaOH. The pH of the solution was monitored throughout the ligation to ensure it remained below 8.0. The atmosphere of the eppendorf was flushed with nitrogen gas and the tube was sealed and incubated at RT. Progression of the NCL reaction was monitored periodically by analytical reverse-phase HPLC (RP-HPLC, C18 Vydac column, 0-73% solvent B in solvent A gradient over 30 min). The eppendorf tube was flushed with nitrogen or argon after each sampling to maintain an oxygen-free atmosphere. After 24 h the reaction mixture was purified by semi-preparative HPLC.

### *KinB-CC Autophosphorylation Inhibition Assays*

Autophosphorylation reactions were performed *in vitro* as described (KinB-CC monomer, 6 μM; α-[$^{32}$P]ATP, 0.5 μM), with the indicated concentrations of inhibitor (NP, no inhibitor; other concentrations denote fold concentration over the KinB-CC concentration). Aliquots were taken after 15 s, 1, 5, and 15 min of incubation at 25 ºC. Reactions were halted by the addition of 5X SDS loading buffer plus 50 mM EDTA, and the $^{32}$P-phosphorylated KinB-CC was quantitated by [4-15]% SDS-PAGE and phosphorimagery (ImageQuant™). An apparent forward rate constant for each time course was determined by fitting, using least squares, to the reaction scheme:

$$\text{HK} + \text{ATP} \leftrightarrow \text{P\textasciitilde HK} + \text{ADP}$$

with Runge-Kutta 4th order numerical integration using the program Berkeley Madonna v8.3.21 (www.berkeleymadonna.com). The '% inhibition' was calculated as follows:

$$\text{'\% inhibition'} = 100 - (100 * (\text{kfNP} - \text{kf})/\text{kfNP})$$

# References

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

**Boissel, J.P., Kasper, T.J., and Bunn, H.F.** (1988). Cotranslational amino-terminal processing of cytosolic proteins. Cell-free expression of site-directed mutants of human hemoglobin. J Biol Chem *263*, 8443-8449.

**Burgess, R.R.** (2009). Refolding solubilized inclusion body proteins. Methods Enzymol *463*, 259-282.

**Camarero, J.A., and Muir, T.W.** (2001). Native chemical ligation of polypeptides. Curr Protoc Protein Sci *Chapter 18*, Unit18 14.

**Carpino, L.A.** (1957). Oxidative Reactions of Hydrazines .2. Isophthalimides - New Protective Groups on Nitrogen. J Am Chem Soc *79*, 98-101.

**Carpino, L.A.** (1973). New Amino-Protecting Groups in Organic Synthesis. Accounts Chem Res *6*, 191-198.

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

**Chabre, M.** (1990). Aluminofluoride and beryllofluoride complexes: a new phosphate analogs in enzymology. Trends Biochem Sci *15*, 6-10.

**Chacko, B.M., Qin, B.Y., Tiwari, A., Shi, G., Lam, S., Hayward, L.J., De Caestecker, M., and Lin, K.** (2004). Structural basis of heteromeric smad protein assembly in TGF-beta signaling. Mol Cell *15*, 813-823.

**Chang, C.D., and Meienhofer, J.** (1978). Solid-Phase Peptide-Synthesis Using Mild Base Cleavage of Nalpha-Fluorenylmethyloxycarbonylamino Acids, Exemplified by a Synthesis of Dihydrosomatostatin. Int J Pept Prot Res *11*, 246-249.

**Dawson, P.E., Churchill, M.J., Ghadiri, M.R., and Kent, S.B.H.** (1997). Modulation of reactivity in native chemical ligation through the use of thiol additives. J Am Chem Soc *119*, 4325-4329.

**Dawson, P.E., Muir, T.W., Clark-Lewis, I., and Kent, S.B.** (1994). Synthesis of proteins by native chemical ligation. Science *266*, 776-779.

**Gardino, A.K., Volkman, B.F., Cho, H.S., Lee, S.Y., Wemmer, D.E., and Kern, D.** (2003). The NMR solution structure of BeF(3)(-)-activated Spo0F reveals the conformational switch in a phosphorelay system. J Mol Biol *331*, 245-254.

**Gentle, I.E., De Souza, D.P., and Baca, M.** (2004). Direct production of proteins with N-terminal cysteine for site-specific conjugation. Bioconjug Chem *15*, 658-663.

**Harvey, R.G.** (1966). Reactions of triethyl phosphite with activated olefins. Tetrahedron *22*, 2561-2573.

**Hines, A.C., Parang, K., Kohanski, R.A., Hubbard, S.R., and Cole, P.A.** (2005). Bisubstrate analog probes for the insulin receptor protein tyrosine kinase: molecular yardsticks for analyzing catalytic mechanism and inhibitor design. Bioorg Chem *33*, 285-297.

**Hirel, P.H., Schmitter, M.J., Dessen, P., Fayat, G., and Blanquet, S.** (1989). Extent of N-terminal methionine excision from Escherichia coli proteins is governed by the side-chain length of the penultimate amino acid. Proc Natl Acad Sci U S A *86*, 8247-8251.

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

**Merrifield, B.** (1986). Solid phase synthesis. Science *232*, 341-347.

**Merrifield, R.B.** (1963). Solid Phase Peptide Synthesis .1. Synthesis of a Tetrapeptide. J Am Chem Soc *85*, 2149-&.

**Myers, A.G., Yang, B.H., Chen, H., McKinstry, L., Kopecky, D.J., and Gleason, J.L.** (1997). Pseudoephedrine as a practical chiral auxiliary for the synthesis of highly enantiomerically enriched carboxylic acids, alcohols, aldehydes, and ketones. J Am Chem Soc *119*, 6496-6511.

**Parang, K., Till, J.H., Ablooglu, A.J., Kohanski, R.A., Hubbard, S.R., and Cole, P.A.** (2001). Mechanism-based design of a protein kinase inhibitor. Nat Struct Biol *8*, 37-41.

**Qin, B.Y., Lam, S.S., Correia, J.J., and Lin, K.** (2002). Smad3 allostery links TGF-beta receptor kinase activation to transcriptional control. Genes Dev *16*, 1950-1963.

**Rak, A., Pylypenko, O., Durek, T., Watzke, A., Kushnir, S., Brunsveld, L., Waldmann, H., Goody, R.S., and Alexandrov, K.** (2003). Structure of Rab GDP-dissociation inhibitor in complex with prenylated YPT1 GTPase. Science *302*, 646-650.

**Schenkels, C., Erni, B., and Reymond, J.L.** (1999). Phosphofurylalanine, a stable analog of phosphohistidine. Bioorg Med Chem Lett *9*, 1443-1446.

**Schnolzer, M., Alewood, P., Jones, A., Alewood, D., and Kent, S.B.H.** (2007). In situ neutralization in boc-chemistry solid phase peptide synthesis - Rapid, high yield assembly of difficult sequences. Int J Pept Res Ther *13*, 31-44.

**Severin, S.E., and Yudelovich, R.Y.** (1947). Synthesis and properties of phosphorylated β-alanine, l-histidine, and α-alanine. Biokhimiya *12*, 105-110.

**Tzeng, Y.L., and Hoch, J.A.** (1997). Molecular recognition in signal transduction: the interaction surfaces of the Spo0F response regulator with its cognate phosphorelay proteins revealed by alanine scanning mutagenesis. J Mol Biol *272*, 200-212.

**Valiyaveetil, F.I., Leonetti, M., Muir, T.W., and Mackinnon, R.** (2006a). Ion selectivity in a semisynthetic K+ channel locked in the conductive conformation. Science *314*, 1004-1007.

**Valiyaveetil, F.I., MacKinnon, R., and Muir, T.W.** (2002). Semisynthesis and folding of the potassium channel KcsA. J Am Chem Soc *124*, 9113-9120.

**Valiyaveetil, F.I., Sekedat, M., Mackinnon, R., and Muir, T.W.** (2004). Glycine as a D-amino acid surrogate in the K(+)-selectivity filter. Proc Natl Acad Sci U S A *101*, 17045-17049.

**Valiyaveetil, F.I., Sekedat, M., MacKinnon, R., and Muir, T.W.** (2006b). Structural and functional consequences of an amide-to-ester substitution in the selectivity filter of a potassium channel. J Am Chem Soc *128*, 11591-11599.

**Varughese, K.I., Tsigelny, I., and Zhao, H.** (2006). The crystal structure of beryllofluoride Spo0F in complex with the phosphotransferase Spo0B represents a phosphotransfer pretransition state. J Bacteriol *188*, 4970-4977.

**Zapf, J., Sen, U., Madhusudan, Hoch, J.A., and Varughese, K.I.** (2000). A transient interaction between two phosphorelay proteins trapped in a crystal lattice reveals the mechanism of molecular recognition and phosphotransfer in signal transduction. Structure *8*, 851-862.

**Zhang, G., and Warren, J.D.** (2009). Asymmetric synthesis of L-(4'-phosphoryl-2'-furyl)alanine, an important analog for the study of histidine phosphorylation (New York City, Weill Cornell Medical College).

**Zhao, X., Copeland, D.M., Soares, A.S., and West, A.H.** (2008). Crystal structure of a complex between the phosphorelay protein YPD1 and the response regulator domain of SLN1 bound to a phosphoryl analog. J Mol Biol *375*, 1141-1151.

# CHAPTER 4

# Structural and Biochemical Studies of a Full-Length Sporulation Histidine Kinase

## Introduction

We now have a detailed understanding of how the catalytic and sensor domains of histidine kinases function. However, we know comparatively little about how the two domains communicate with each other, how a sensor domain transduces a perceived signal to the catalytic core to affect conformational changes and rates of autophosphorylation. Recently published structures of sensor domains have provided clues as to how signaling may occur. Evidence points towards a piston-like displacement of sensor domain helices as the mechanism of signal transduction (see the Introduction Chapter of this thesis for a more in depth discussion). However, the structure of an intact histidine kinase comprising the sensor domain, catalytic domain, and the linker region between the two will be essential if we are to fully understand how HK signaling works.

This chapter describes the steps taken towards achieving that goal. Because membrane proteins are notoriously difficult to express and crystallize, we have chosen to focus on solving the structure of a full-length cytoplasmic HK. Presumably the mechanism of signal transduction is similar, if not the same, for both membrane and cytoplasmic HKs, as the sensor and kinase domains of both are connected by a long helix. The sporulation pathway from *Bacillus* contains two cytoplasmic HKs, and therefore offers an excellent system for probing the structure of a full-length HK.

# Results and Discussion

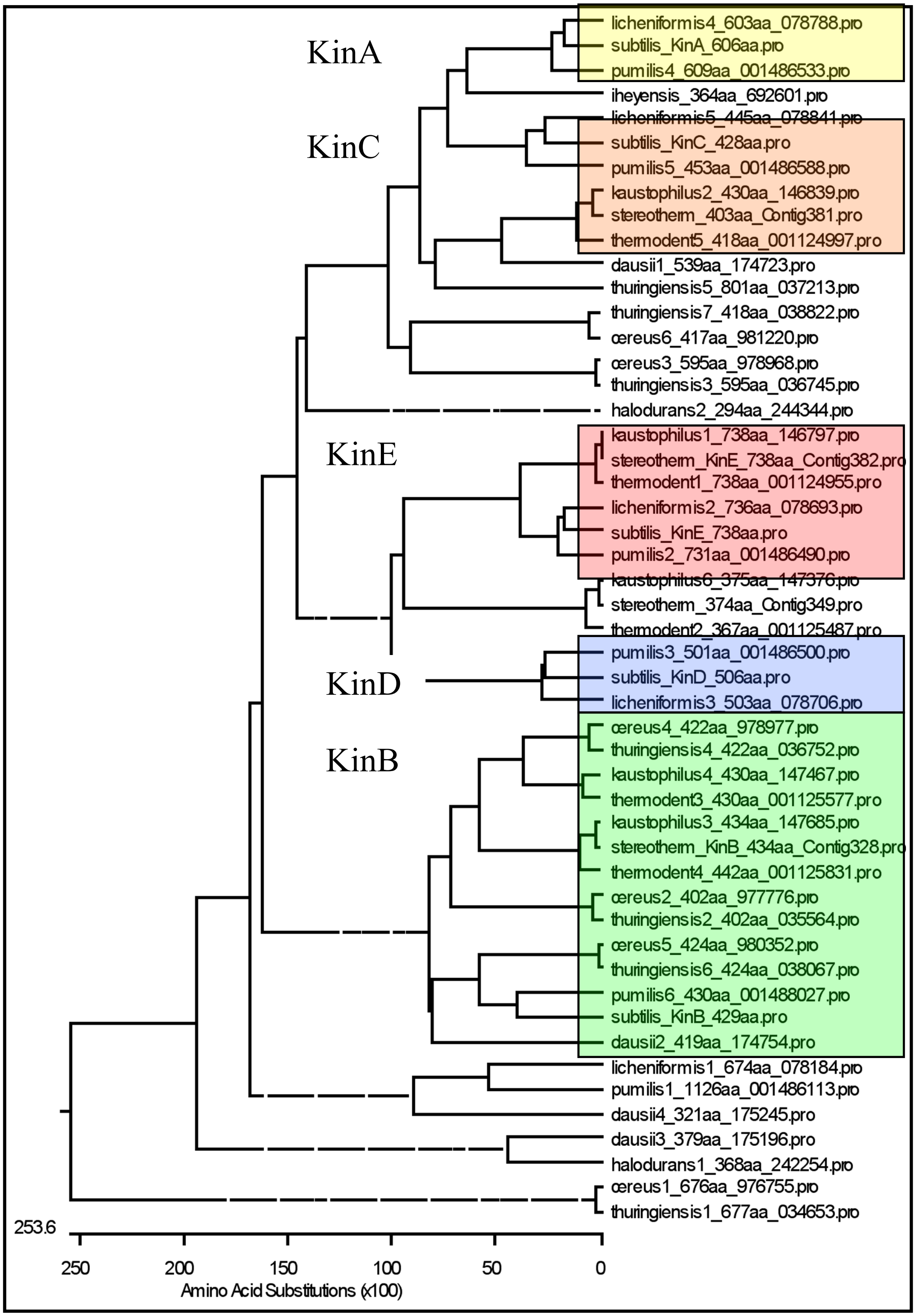
## *Search for Thermophilic Sporulation Histidine Kinases*

We initially set out to determine the structure of full-length KinA, as it is the best characterized of the sporulation kinases. We sought to crystallize a thermophilic KinA, as our experience has been that proteins from thermophilic species tend to crystallize more readily. Using the GFXXL Sda binding sequence motif and its conserved proximity to the H-Box allowed us to probe the completed genomes of *Bacillus* and *Geobacillus* species in search of sporulation HKs (Figure 4.1). Our results showed that thermophilic homologues of *B. subtilis* KinA do not exist. Two of the five sporulation kinases in *B. subtilis*, KinA and KinE, are fully cytoplasmic. *Geobacillus* only have one of these, KinE. This finding has significance in regard to the role each individual sporulation kinase plays during spore production. KinA and KinB have been identified as essential for sporulation in *B. subtilis*. Single knockouts of either KinA or KinB show a modest defect for sporulation, whereas a double knockout is unable to sporulate (Jiang et al., 2000; Trach and Hoch, 1993). But KinA can't possibly be important for species of *Geobacillus*, as they don't have KinA (they do have KinB). There are two non-mutually exclusive explanations for this. KinA may be essential for sporulation only under laboratory settings, and/or the signals to which KinA responds may not be present in the environments where *Geobacillus* species reside.

The domain structure of KinA and KinE are very similar. Both are entirely cytoplasmic, and both contain putative PAS domains in their sensor regions; KinA has

**Figure 4.1** (adjacent page) Phylogenetic tree, constructed using Proml from the Phylip suite of programs (Felsenstein, 1981), of sporulation histidine kinases from fully sequenced *Bacillus* and *Geobacillus* genomes (*Geobacillus stearothermophilus* strain 10 genome has not yet been fully sequenced). Sporulation kinases were identified by the presence of the Sda sequence motif, GFXXL, and its proximity to the H-box and then grouped based on their sensor domains. Sequences outside colored boxes could not be grouped as one of the five *B. subtilis* sporulation HKs.

Figure 4.1


KinA
KinC
KinE
KinD
KinB
licheniformis4_603aa_078788.pro
subtilis_KinA_606aa.pro
pumilis4_609aa_001486533.pro
iheyensis_364aa_692601.pro
licheniformis5_445aa_078841.pro
subtilis_KinC_428aa.pro
pumilis5_453aa_001486588.pro
kaustophilus2_430aa_146839.pro
stereotherm_403aa_Contig381.pro
thermodent5_418aa_001124997.pro
dausii1_539aa_174723.pro
thuringiensis5_801aa_037213.pro
thuringiensis7_418aa_038822.pro
cereus6_417aa_981220.pro
cereus3_595aa_978968.pro
thuringiensis3_595aa_036745.pro
halodurans2_294aa_244344.pro
kaustophilus1_738aa_146797.pro
stereotherm_KinE_738aa_Contig382.pro
thermodent1_738aa_001124955.pro
licheniformis2_736aa_078693.pro
subtilis_KinE_738aa.pro
pumilis2_731aa_001486490.pro
kaustophilus6_375aa_147376.pro
stereotherm_374aa_Contig349.pro
thermodent2_367aa_001125487.pro
pumilis3_501aa_001486500.pro
subtilis_KinD_506aa.pro
licheniformis3_503aa_078706.pro
cereus4_422aa_978977.pro
thuringiensis4_422aa_036752.pro
kaustophilus4_430aa_147467.pro
thermodent3_430aa_001125577.pro
kaustophilus3_434aa_147685.pro
stereotherm_KinB_434aa_Contig328.pro
thermodent4_442aa_001125831.pro
cereus2_402aa_977776.pro
thuringiensis2_402aa_035564.pro
cereus5_424aa_980352.pro
thuringiensis6_424aa_038067.pro
pumilis6_430aa_001488027.pro
subtilis_KinB_429aa.pro
dausii2_419aa_174754.pro
licheniformis1_674aa_078184.pro
pumilis1_1126aa_001486113.pro
dausii4_321aa_175245.pro
dausii3_379aa_175196.pro
halodurans1_368aa_242254.pro
cereus1_676aa_976755.pro
thuringiensis1_677aa_034653.pro
253.6
250
200
150
100
50
0
Amino Acid Substitutions (x100)

three putative PAS domains, while KinE has four (Figure 4.2). The fact that KinA and KinE are similar but that thermophilic species only contain KinE suggests that he appearance of KinA in mesophilic species of *Bacillus* may have been the result of a gene duplication event of KinE at some point in evolutionary history. Unlike KinA, little is known about KinE (previously ykrQ). It is the most recent of the sporulation kinases to be identified (Jiang et al., 2000).

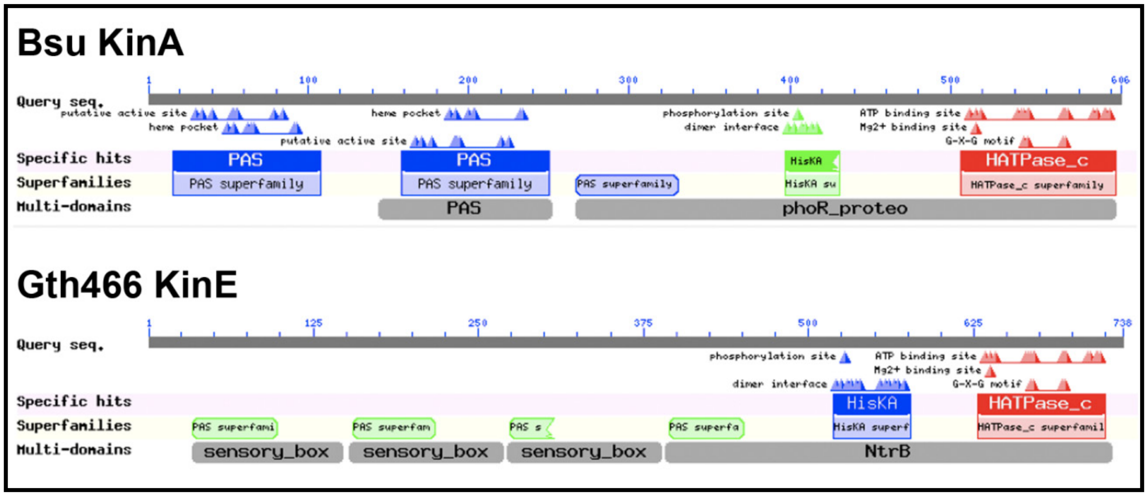


**Figure 4.2)** BLAST results outlining the domain organization of *B. subtilis* KinA and *G. thermodenitrificans* strain 466 KinE. KinA is predicted to have three PAS domains in its sensor domain, while KinE is predicted to have four.

### *Identification of Proteolytically Stable Domains of KinE*

Full-length KinE from *Gst*, when expressed on its own in *E. coli* was found to be 100% insoluble. (see below, Figure 4.4A). Several attempts were made to purify KinE from inclusion bodies under denaturing conditions using 6 M guanidine-HCl, and we were able to achieve decent levels of purity. However, we found that full-length KinE, even in the context of Sda, tended to aggregate badly during the concentration step; the highest concentration we were able to achieve with full-length KinE was 8.8 mg/ml, and this only after a full day of concentrating and after several high-speed centrifuge spins to remove aggregated material. To achieve concentrations higher than 8.8 mg/ml, 10% glycerol was required in the protein solution. We did manage to perform crystallization trials on full-length KinE/Sda from *Gth* strain 466 (Table 4.1).

**Full-length Gth 466 KinE/Sda Crystallization Trials**

| Screen | Protein Conentration and Temp |
|---|---|
| JCSG + | 6.3 mg/ml plus 5 X molar excess ADP, 22 °C |
| Classics | 3, 4, 10, and 15 mg/ml plus 5 X molar excess ADP, 22 °C |
| Cations | 3.6 and 8.5 mg/ml plus 5 X molar excess ADP, 22 °C |
| PEG II | 4.6 mg/ml 6.5 mg/ml plus 5 X molar excess ADP, 22 °C |
| MPD | 3.2 and 8.5 mg/ml plus 5 X molar excess ADP, 22 °C |
| Xtal Screen II (Hampton) | 9.7 mg/ml plus 5 X molar excess ADP, 22 °C |
| Xtal Screen I (Hampton) | 4.6 mg/ml plus 5 X molar excess ADP, 22 °C |
| PACT | 8.5 mg/ml plus 5 X molar excess ADP, 22 °C |
| Anion | 8.5 mg/ml plus 5 X molar excess ADP, 22 °C |

**Table 4.1)** Crystallization trials performed to date on full-length *Gth* KinE/Sda. Protein crystals have not been obtained from any screen.

Therefore, due to the issues of protein aggregation, we sought to identify shorter, more stable constructs of KinE that might be more amenable to crystallization. We have used this technique previously with success (Bick et al., 2009). We carried out limited trypsinolysis on full-length *Gth* 466 KinE. Stable tryptic fragments were identified via Edman sequencing (Figure 4.3). The N-terminal sequence of a prominent band, which migrated slightly below the 50 kDa weight marker, was identified as KinE residues 269-274. This region of KinE is predicted to be an α-helix that connects the second and third

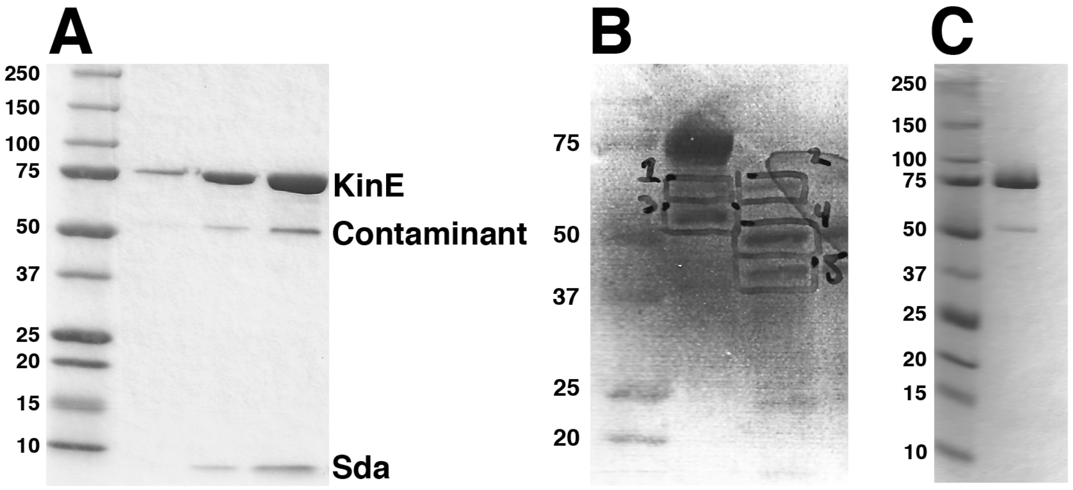


**Figure 4.3)** (A) Peak fractions from a size exclusion chromatography column of full-length KinE from *Gth* 466 co-purified with Sda. The prominent contaminant observed around 50 kDa co-purified with KinE/Sda through all chromatography steps during the purification. (B) Results of limited trypsinolysis on *Gth* 466 KinE/Sda. Lanes from left to right: Protein weight marker, uncut control sample (band 2 is the contaminant), 1:10 ratio of Trypsin to KinE. (C) Purified *Gst* KinE/Sda. A similar contaminant as the one observed during *Gth* 466 purification was also present in the *Gst* sample.

putative PAS domains (Figure 4.7). The molecular weight of the fragment indicated that in addition to the cleavage site in the sensor domain, KinE had also been cleaved at its C-terminus. Because the C-terminus of KinE encompasses the CA domain of the catalytic region, we reasoned that a construct based wholly on the trypsinolysis results, which would contain only a portion of the CA domain, would not be ideal for structural studies. We therefore chose to design expression constructs based only on the N-terminal results.

Table 4.2

| Proteolytic Fragment | N-Terminal Sequence |
|---|---|
| Contaminant (Band 3) | SEERFR (Gth KinE 272-277) |
| Tryptic Fragment 4 | LQKSEE (Gth KinE 269-274) |
| Tryptic Fragment 5 | IIHEFL (Gth KinE 149-154) |
| Contaminant (Gst gel) | LQKSE (Gst KinE 269-273) |

**Table 4.2)** N-terminal Edman sequencing results of tryptic fragments corresponding to Figure 4.3.

The contaminant observed around 50 kDa during full-length KinE purification (Figure 4.3B) was also sequenced, and identified as KinE 272-277. The fact that a naturally occurring proteolytic fragment is generated at the same location as our trypsinolysis experiments suggests that the helix connecting PAS domains 2 and 3 is solvent exposed, and highly susceptible to digestion by residual amounts of proteases carried over during the purification. Indeed, when *Gst* KinE was purified under denaturing conditions, no degradation product was observed around 50 kDa, likely because contaminating proteases had been inactivated. We decided that this region of

KinE represented an ideal location for generating truncated KinE expression constructs. In addition, the naturally occurring proteolytic fragment appeared to have a fully intact CA domain, based on its approximate molecular weight above 50 kDa on SDS-PAGE (the predicted mass of *Gth* 466 KinE 282-CT is 53,670 Daltons) (Figure 4.3B). A similar proteolytic fragment was also observed during purification of *Gst* KinE/Sda (residues 269-273 by N-terminal sequencing (Figure 4.3C).

### *Establishment of a KinE/Sda Expression System*

Full-length KinE from *Gst*, when expressed on its own at either 16, 30, or 37 ºC, was found to be 100% insoluble (Figure 4.4A). We therefore reasoned that, based on our experience with KinB (Bick et al., 2009), KinE would behave better when co-expressed with Sda. We concurrently pursued expression trials of KinE from three different thermophilic bacilli, *Gst*, *Gth* 465, and *Gth* 466. The Sda protein sequence is identical for these three strains, and therefore *Gst* Sda was used for all of our co-expression studies. Based on a combination of trypsinolysis (see above) and secondary structural prediction (Rost et al., 2004), a construct beginning at KinE residue 282, which removes the first two putative PAS domains of the sensor region, was chosen as the most promising candidate for crystallization. *Gth* 466 expressed to higher levels than both *Gth* 465 and *Gst* KinE and proved to be the most soluble, approximately 50% (Figure 4.4B). In addition, the KinE protein sequence from *Gth* 466 is more divergent than *Gth* 465 in comparison with *Gst* KinE, which proved to be the most problematic of the three.

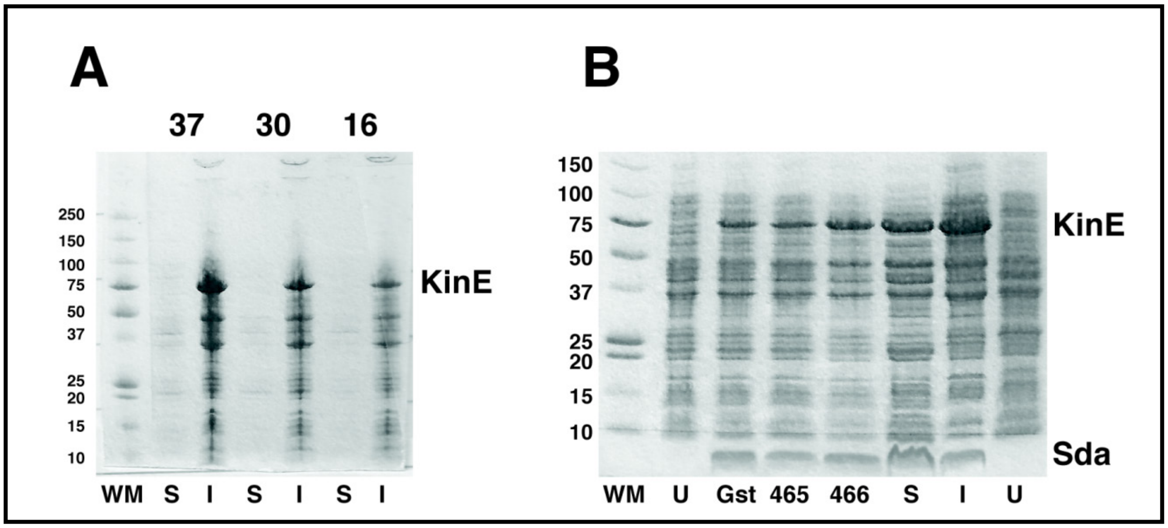


**Figure 4.4)** (A) KinE from *Gst* was 100% insoluble when in expressed in *E. coli* at three different temperatures (16, 30, and 37 ºC). (B) *Gth* strain 466 KinE expressed to the highest levels when compared with *Gst* and *Gth* strain 465 KinE. When expressed in the context of Sda, KinE was approximately 50% soluble. WM - protein weight marker. S - soluble fraction following lysis. I - insoluble fraction following lysis. U - pre-induction.

*Gth* grows at 60 ºC in its natural environment. It is common practice when heterologously expressing thermophilic proteins in *E. coli*, a mesophile, to carry out a heat-treatment purification step in order to aggregate and remove contaminating *E. coli* proteins. Proteins from thermophilic species are resistant to denaturation at higher temperatures, and hence remain soluble. However, our experience with the *Gst* KinB catalytic core suggested that the sporulation HKs might not obey this rule (KinB-CC tended to aggregate when treated at 60 ºC, perhaps because it was separated from its sensor domain, and hence from the membrane). We carried out a heat-treatment time course experiment at 60 ºC on crude *E. coli* lysate following full-length KinE/Sda expression (Figure 4.5A). 20 min of treatment at 60 ºC caused many of the soluble *E. coli* proteins to aggregate and partition to the insoluble fraction, while KinE remained approximately 50% soluble. By 40 min, however, KinE had become almost entirely insoluble. Interestingly, treatment of KinE 282-CT/Sda for 20 min at 60 ºC made the protein entirely insoluble (Figure 4.5B). Therefore, a heat treatment step was not used for KinE 282-CT (Again, as with KinB-CC, it may be that only the full-length kinase is resistant to heat treatment).

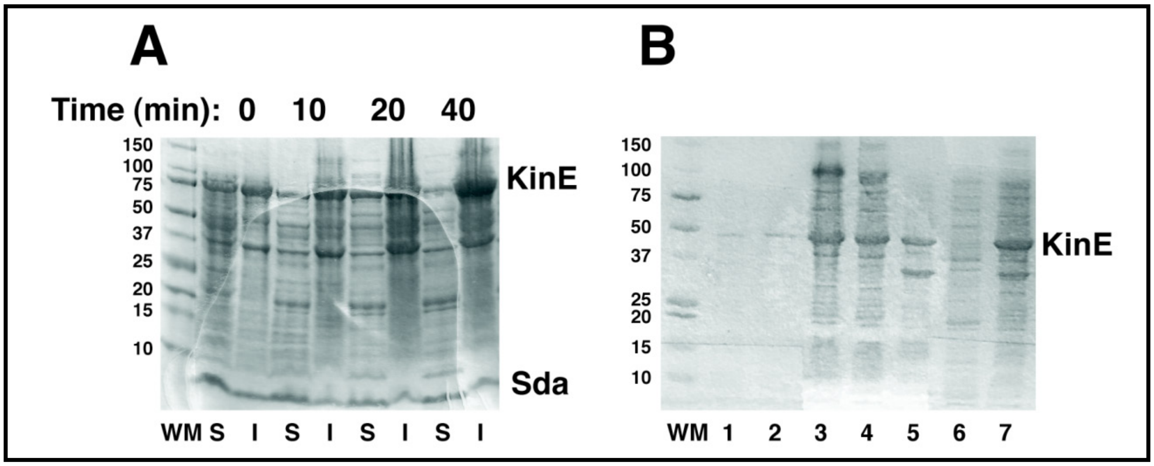


**Figure 4.5)** (A) 60 ºC heat treatment time course experiment on crude full-length KinE/Sda lysate. Heat treatment for 20 min gave the best balance of removing contaminating *E. coli* proteins while maintaining the solubility of KinE. WM - protein weight marker. S - soluble fraction following lysis. I - insoluble fraction following lysis. (B) Heat treatment of KinE 282-CT/Sda caused KinE to become insoluble. Lanes from left to right: WM- protein weight marker, 1 - *Gst* KinE 282-CT/Sda post induction, 2 - *Gth* 466 KinE 282-CT/Sda post induction, 3 - *Gth* 466 total lysate, 4 - *Gth* 466 soluble fraction following lysis, 5 - *Gth* insoluble fraction following lysis, 6 - *Gth* 466 soluble fraction following heat treatment at 60 ºC for 20 min, 7 - *Gth* insoluble fraction following heat treatment at 60 ºC for 20 min.

KinE 282-CT/Sda was purified by nickel affinity, anion exchange, and size exclusion chromatography to above 95% homogeneity, as judged by SDS-PAGE (Figure 4.6). KinE 282-CT/Sda was soluble to at least 27 mg/ml (full-length KinE/Sda was soluble to only approximately 8.8 mg/ml, and only after many hours of concentration).

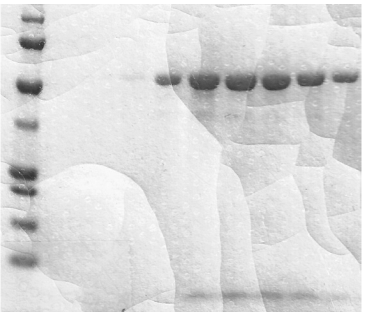


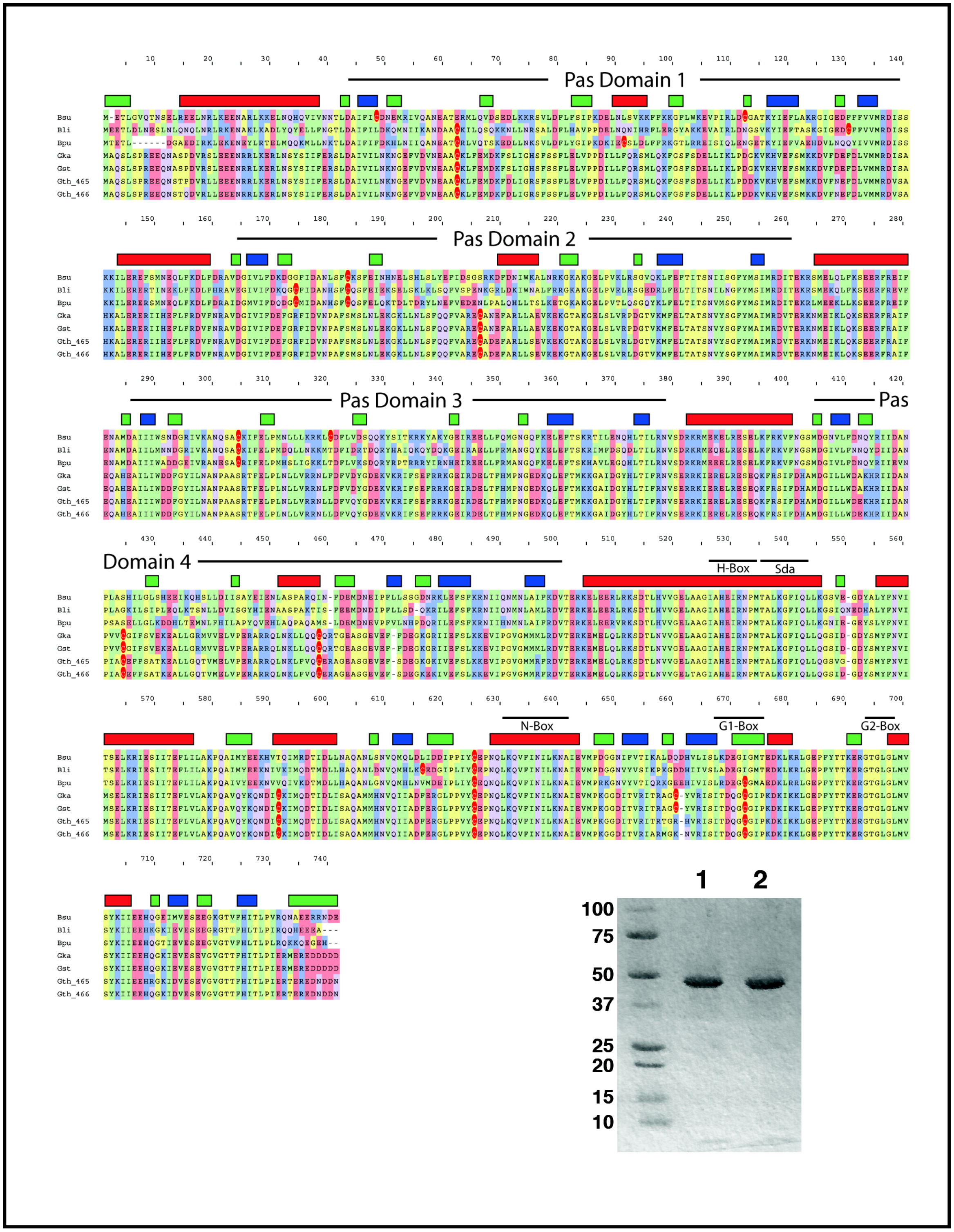
**Figure 4.6)** Peak fractions of *Gth* 466 KinE 282-CT/Sda following size exclusion chromatography.

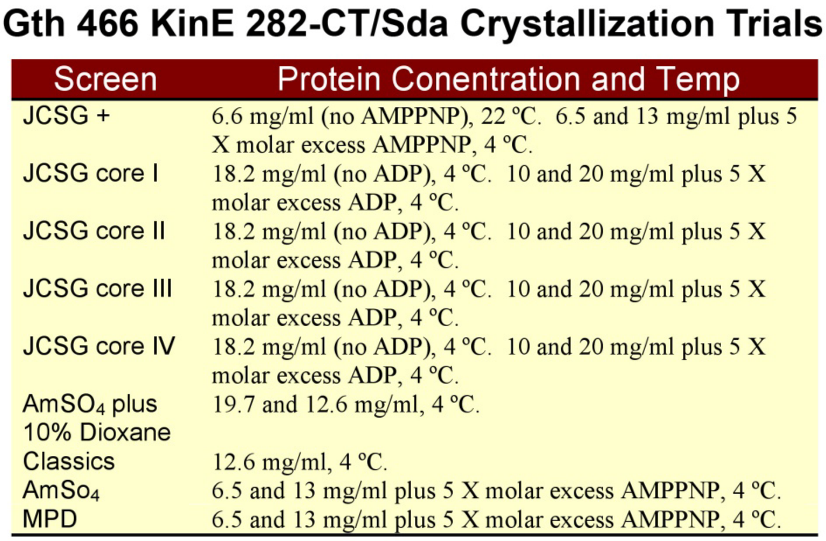
During the anion exchange chromatography step, we observed two distinct KinE species that eluted from the column at different ionic strengths, one at around 100 mM NaCl and the other at around 250-300 mM NaCl. We suspected that the Q column was retaining phosphorylated and unphosphorylated KinE with different affinities. EDTA was added to the protein solution prior to anion exchange step in order to prevent the autophosphorylation reaction (autophosphorylation is dependent upon the presence of a divalent cation). When EDTA was present in the buffer, the majority of KinE eluted from the column at 250-300 mM NaCl; the 100 mM NaCl species was greatly reduced. The opposite was observed when EDTA was not present in the Q buffer; the species shifted more towards the 100 mM NaCl species.

Despite the fact that we were able to purify KinE 282-CT/Sda to above 95% homogeneity, and that it is highly soluble, we have been unable to crystallize the complex as the ADP bound form (Table 4.3). The KinE 282-CT sequence contains 5 cysteines (7 for full-length KinE). For future crystallization trials it may prove to be advantageous to remove the non-conserved cysteines through mutagenesis (Figure 4.7). SDS-PAGE analysis of KinE/Sda under non-reducing conditions suggests that cysteines do not play a part in KinE dimerization (Figure 4.7).

**Figure 4.7** (adjacent page) Sequence alignment of KinE from *Bacillus* and *Geobacillus*. Cysteines are highlighted with red ovals. Bars above the sequence represent the results of secondary structure prediction (Rost et al., 2004), with red representing helices, blue representing strands, and green representing loops. Only residues with a confidence score of 5 or more (based on a scale from 1-10) were assigned a secondary structure. Residues spanning the putative PAS domains, based on the results of BLAST, are also labeled. (Bsu – *Bacillus subtilis*, Bli – *Bacillus licheniformis*, Bpu – *Bacillus pumilis*, Gka – *Geobacillus kaustophilus*, Gst – *Geobacillus stearothermophilus*, Gth_465 – *Geobacillus thermodenitrificans* strain 465, Gth_466 – *Geobacillus thermodenitrificans* strain 466). The gel on the bottom right represents SDS-PAGE analysis of *Gth* 466 KinE 282-CT under denaturing conditions, either with 200 mM βME (lane 1) and heating at 95 ºC for 5 min, or without βME and no heat treatment (lane 2).

Figure 4.7


Pas Domain 1
Pas Domain 2
Pas Domain 3
Pas Domain 4
H-Box
Sda
N-Box
G1-Box
G2-Box
Bsu
Bli
Bpu
Gka
Gst
Gth_465
Gth_466
1
2
100
75
50
37
25
20
15
10



Gth 466 KinE 282-CT/Sda Crystallization Trials

| Screen | Protein Conentration and Temp |
|---|---|
| JCSG + | 6.6 mg/ml (no AMPPNP), 22 °C. 6.5 and 13 mg/ml plus 5 X molar excess AMPPNP, 4 °C. |
| JCSG core I | 18.2 mg/ml (no ADP), 4 °C. 10 and 20 mg/ml plus 5 X molar excess ADP, 4 °C. |
| JCSG core II | 18.2 mg/ml (no ADP), 4 °C. 10 and 20 mg/ml plus 5 X molar excess ADP, 4 °C. |
| JCSG core III | 18.2 mg/ml (no ADP), 4 °C. 10 and 20 mg/ml plus 5 X molar excess ADP, 4 °C. |
| JCSG core IV | 18.2 mg/ml (no ADP), 4 °C. 10 and 20 mg/ml plus 5 X molar excess ADP, 4 °C. |
| $AmSO_4$ plus 10% Dioxane | 19.7 and 12.6 mg/ml, 4 °C. |
| Classics | 12.6 mg/ml, 4 °C. |
| $AmSo_4$ | 6.5 and 13 mg/ml plus 5 X molar excess AMPPNP, 4 °C. |
| MPD | 6.5 and 13 mg/ml plus 5 X molar excess AMPPNP, 4 °C. |

**Table 4.3)** Crystallization trials performed to date on *Gth* 466 KinE 282-CT/Sda. Protein crystals have not yet been obtained.

### *Sda Inhibits KinE Autophosphorylation*

Previous studies have shown that Sda inhibits CA-DHp domain cross talk of KinA and KinB, effectively shutting down the autophosphorylation reaction (Bick et al., 2009; Rowland et al., 2004). We identified three KinB residues within a sequence motif, GFXXL, that are crucial for recognizing Sda. Introduction of these amino acids by mutagenesis to an HK normally unresponsive to Sda renders that HK sensitive to Sda inhibition. We also showed that the GFXXL sequence motif is present in all five of the *B. subtilis* sporulation kinases, KinA-KinE (Bick et al., 2009). It was therefore reasonable to assume that Sda would also inhibit KinE autophosphorylation.

To confirm our hypothesis, we carried out Sda inhibition assays on KinE. From our studies on KinB, we knew that Sda in a 1:1 ratio with KinB is not sufficient to abolish autophosphorylation. So it was expected that KinE co-purified with Sda would still autophosphorylate. Indeed, this is what we found. KinE/Sda without additional Sda was labeled to significant levels in the presence of γ-[$^{32}$P]ATP (Figure 4.8A). However, with the addition of a 25 molar excess of Sda, KinE autophosphorylation was fully inhibited; no signal was observed by autoradiography (Figure 4.8B and 4.8C).

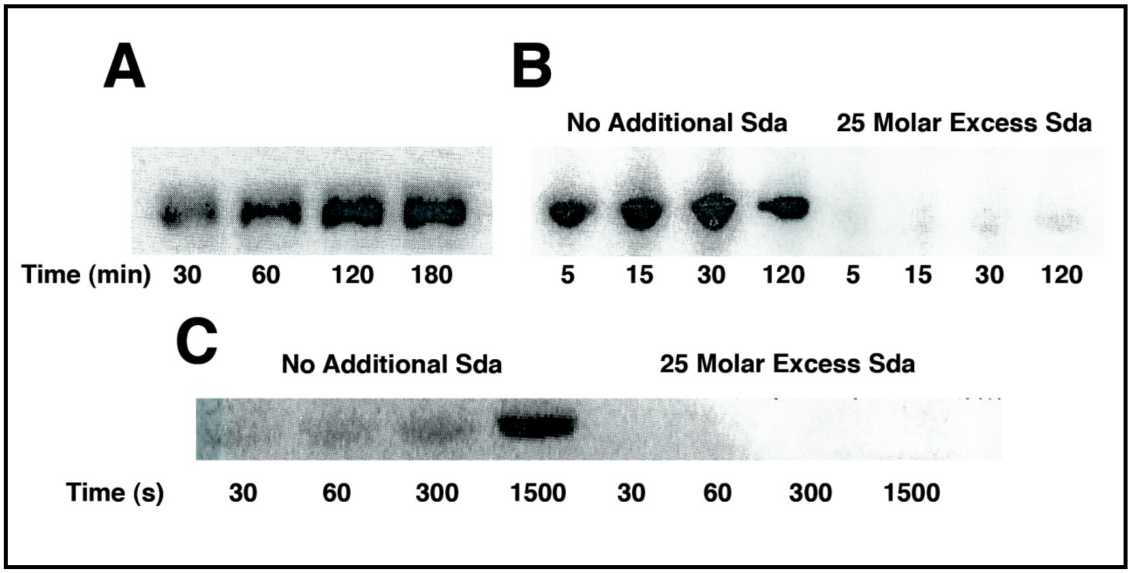


**Figure 4.8)** Autoradiography images of autophosphorylation time course experiments performed on full-length KinE/Sda in the presence of γ-[$^{32}$P]ATP and Mg$^{2+}$. (A) *Gst* KinE/Sda without the addition of excess Sda. (B) and (C) *Gth* 466 KinE/Sda with and without the addition of excess Sda.

## Conclusions and Future Directions

We have established an expression system for a full-length cytoplasmic sporulation histidine kinase. Removal of the two N-terminal PAS domains from the signaling region should not hinder the ultimate goal of the project, which is to investigate how signals are transduced to the catalytic core; the helix connecting the signaling and catalytic domains remains intact in our KinE construct. We have also shown that, as for KinB, Sda remains associated with KinE through several chromatography steps. In addition, Sda inhibits KinE autophosphorylation, a result that supports the hypothesis laid out in Chapter Two, which states that sporulation kinases containing the sequence motif GFXXL in their DHp domains are all affected by Sda. Although we have not obtained KinE/Sda crystals, we have not exhausted all possibilities. We will attempt crystallization with different nucleotides (ATP or non-hydrolyzable ATP analogs), as well as pursue KinE from different *Geobacillus* and *Bacillus* species.

# Materials and Methods

## *Cloning of KinE and KinE/Sda Double Expression Constructs*

*Geobacillus thermodenitrificans* strain DSM 466 (ATCC 29493™) (Manachini et al., 2000) was grown on nutrient agar containing peptone and beef extract (ATCC Medium 3) at 60 ºC. Isolated colonies were picked and stored in 20 µl $ddH_2O$ at -20 ºC until needed. A single colony was boiled for 10 min, and the cellular debris was separated by centrifugation for 10 min at 14,000 RPM. 10 µl of supernatant was used as a source of *Gth* genomic DNA for PCR reactions. Full-length KinE was PCR amplified (30 cycles of 30 s at 98 ºC, 30 s at 55 ºC, and 80 s at 68ºC, followed by a final extension step of 68 ºC for 6 min 40 s) using primers MB-97 (NheI) and MB-98 (BamHI). MB-97 introduces an NheI site at amino acid position 2 of KinE, and in the process changes residue 3 from a Gln to a Ser. MB-98 introduces a BamHI site directly after the KinE stop codon. Amplicons were purified with a PCR clean-up kit (Qiagen) and sequentially digested with NheI and BamHI (3 h each) at 37 ºC, with a PCR clean-up step in between enzymes. KinE was ligated into NheI/BamHI doubly digested pSKB2. Transformants were selected on LB agar supplemented with 50 µg/ml kanamycin, and colonies were screened for the presence of insert DNA by sequencing. A similar strategy was used to clone *Gth* DSM 465 and *Gst* strain 10 KinE (primers MB-95 and MB-96).

*Gst, Gth* 465, and *Gth* 466 KinE 282-CT (C-terminus) were PCR amplified (30 cycles of 95 ºC for 30 s, 55 ºC for 30 s, and 68 ºC for 87 s, followed by a final extension of 7 min 15 s at 68 ºC) from pSKB2-KinE (full-length) using primers MB-119 (*Gst*, NdeI) or MB-120 (*Gth* 465 and 466) and T7 terminator, doubly digested with NdeI and BamHI, and cloned into pSKB2.

*Gst* Sda was subcloned from pET21a-Sda by PCR (30 cycles of 95 ºC for 30 s, 55 ºC for 30 s, and 68 ºC for 20 s, followed by a final extension step of 68 ºC for 100 s) using primers Liz13 and MB-99. Liz13 introduces a BamHI site into the lac operator of pET based vectors. MB-99 changes the BamHI site in the multiple cloning site of pET21-Sda to an EagI site. The Sda amplicon was digested with BamHI and EagI at 37 ºC overnight, and purified by gel extraction (Qiagen) the following day. Sda was ligated into BamHI/EagI doubly digested pSKB2-KinE (*Gst*, *Gth* 465, and *Gth* 466, either full-length or 282-CT). The presence of Sda insert DNA was confirmed by sequencing.

### *Limited Proteolysis of Full-Length KinE*

100 pmol of *Gth* 466 KinE/Sda was mixed with trypsin (Sigma, made up in protease buffer: 100 mM Tris, pH 8.0 and 20 mM $CaCl_2$) in the following trypsin to KinE (mass:mass) ratios: 1:10 (10 pmol trypsin), 1:20 (5 pmol trypsin), 1:50 (2 pmol trypsin), and 1:100 (1 pmol trypsin), and 1:500 (0.2 pmol trypsin). Reactions were brought up to 10 µl with protease buffer and incubated at 26 ºC for 30 min. 1 µl of 100 mM PMSF was added to stop the reactions and the tubes were put in ice. 11 µl of 2X SDS loading buffer was added and the reactions were run out on 4-20% Tris-Glycine PAGE (Invitrogen).

For N-terminal sequencing of proteolytic fragments, following SDS-PAGE the protein was transferred to PVDF membrane (Bio-Rad) following the electroblotting procedure supplied by The University of Texas Medical Branch (UTMB) Protein Chemistry Laboratory (http://www.utmb.edu/proch/serv01.htm). Proteolytic fragments of interest were excised from the PVDF and sent to UTMB for N-terminal Edman sequencing.

## *Protein Expression and Purification*

*Gst*, *Gth* 465, and *Gth* 466 full-length KinE/Sda double expression vectors (Sda is from *Gst* for all double expression vectors) were transformed into *E. coli* Rosetta2™ (DE3) for protein expression. Transformed bacteria was used to inoculate a 5 ml LB culture supplemented with kanamycin at 50 µg/ml and chloramphenicol at 35µg/ml, and grown overnight at 37 ºC. The 5 ml culture was used the following day to inoculate 2 L LB cultures (kan/cam). Large-scale cultures were grown 37 ºC with shaking at 200 RPM. Protein expression was induced with 0.1 mM IPTG when the cultures reached an $OD_{600}$ of between 0.6 and 1.0. After the addition of IPTG, the temperature of the incubator was reduced to 16 ºC, and protein expression was carried out for approximately 12 h with shaking at 200 RPM. Cells were harvested by centrifugation at 4000 RPM for 30 min at 4 ºC. Pelleted bacterial from a 2 L cultures was resuspended in 35 ml nickel column buffer A (500 mM NaCl, 60 mM Tris, pH 8.0, 10 mM $MgCl_2$, 5 mM imidazole, 5% glycerol) and stored at -80 ºC.

*E. coli* harboring full-length *Gth* 466 KinE/Sda was lysed between 12,000 and 15,000 psi at 4 ºC using an EmulsiFlex-C5™(Avestin) high-pressure homogenizer, and the crude lysate was treated at 60 ºC for 25 min. 10 mM βME and 1 mM PMSF were added to the cells prior to lysis. Heat-treated lysate was clarified by centrifugation at 16,000 RPM for 30 min, 4 ºC. The supernatant was passed over a 5 ml metal chelating column (GE life sciences), charged with $NiSO_4$, at 1 ml/min, and the flow through was collected. KinE/Sda was eluted from the $Ni^{2+}$ column using a step gradient over 60 ml to 100% nickel column buffer B (same as buffer A with 200 mM imidazole) (10 ml at 20% B, 10 ml at 35%, 15 ml at 50%, and 25 ml at 100%). Twelve 5 ml fractions were collected. All fractions were analyzed by SDS-PAGE. Peak elution fractions were pooled and PreScission™ Protease (GE life sciences) was added to an approximate 1:20 mass:mass ratio with KinE/Sda in order to remove the N-terminal $His_6$-tag from KinE. KinE harbors three (GPH) vector-derived residues at its N-terminus following PreScission™ Protease digestion. Cleaved protein was dialyzed overnight at 4 ºC in 2 L of 100 mM NaCl, 20 mM Tris-Cl, pH 8.0, 10 mM $MgCl_2$, 5% glycerol and 5 mM DTT, using 3,500 MWCO dialysis tubing (Spectracor). The efficiency of the cleavage reaction was assessed by SDS-PAGE.

Following PreScission™ Protease digestion, the protein solution was centrifuged at 16,000 RPM, 4 ºC in order to remove a small amount of precipitated material. Next, KinE/Sda was passed over a 5 ml GST column in series with a 5 ml Q anion exchange column (GE life sciences) at 0.5 ml/min. The columns were equilibrated with 100 mM NaCl, 20 mM Tris-Cl, pH 8.0, 10 mM $MgCl_2$, 5 mM DTT, and 5% glycerol (Q column buffer A). After loading the protein onto the GST/Q series, the GST column was removed and KinE/Sda was eluted over 50 ml (10 column volumes) from the Q column at 1 ml/min using a linear gradient to the same buffer plus 600 mM NaCl (Q column buffer B). Twenty-five 2 ml fractions were collected. Finally, SpoIIE was passed over an SD200 (16/60) column, equilibrated with 150 mM NaCl, 20 mM Tris-Cl, pH 8.0, 10 mM $MgCl_2$, 10 mM DTT, and 10% glycerol at 0.3 ml/min. After a 30 ml void volume, forty-five 2 ml fraction were collected. Sda purified separately (see below) was used to supplement KinE/Sda after each step of the purification, assuming 20% of KinE was not bound to Sda.

*Gth* 466 KinE 282-CT/Sda was purified following the same basic protocol as full-length KinE/Sda, with the following exceptions. After PreScission™ Protease digestion, a 2nd $Ni^{2+}$ column step was frequently, though not always, added in order to remove any uncleaved KinE from the solution. The PreScission™ Protease dialysis buffer was 100 mM NaCl, 20 mM Tris-Cl, pH 8.0, 20 mM imidazole, 10 mM βME, 0.25 mM PMSF, and 5% glycerol. A mono Q column (10/100 GE life sciences) was used in place of the 5 ml HiTrap for anion exchange. Q column buffers A and B were 20 mM Tris-Cl, pH 8.0, 1 mM PMSF, 5% glycerol, 5 mM EDTA, 10 mM βME, and 100 mM NaCl for buffer A and 1 M for buffer B. A Superdex™ 200 (26/60) column was used for size exclusion, equilibrated with 150 mM NaCl, 20 mM Tris-Cl, pH 8.0, and 5 mM DTT.

Sda was expressed from a pET21a backbone at 37 ºC for 3 hr, using 1 mM IPTG. Bacteria from a 2 L culture was resuspended in 25 ml Sda buffer (150 mM NaCl, 50 mM Tris, pH 8.0, 5% glycerol) and divided into 10 aliquots, corresponding to approximately 200 ml of culture. Aliquots were stored at -80 ºC until needed. Sda was prepared each time KinE was purified. Bacteria from a 200 ml aliquot was lysed by sonication, and the crude lysate was clarified by centrifugation at 16,000 RPM for 30 min at 4 ºC. The supernatant containing soluble Sda was heated to 60 ºC for 45 min in order to aggregate the majority of contaminating soluble *E. coli* proteins. The sample was spun again at 16,000 RPM for 30 min at 4 ºC. A 5ml HiTrap Q (GE Lifesciences) anion exchange column (equilibrated with Sda buffer), loaded at 1 ml/min, was used to bind the majority of remaining contaminants; Sda did not bind the Q column. Finally, Sda was passed over a Superdex™ 200 (16/60, GE Lifesciences) size exclusion column (equilibrated with Sda buffer) at 0.3 ml/min. Forty-five 2 ml fractions were collected following a 30 ml void volume. Purified Sda was stored at 4 ºC and used to supplement KinE/Sda during its purification (see above).

### *KinE Autophosphorylation and Sda Inhibition Assays*

For KinE kinase reactions, 100 pmol of KinE/Sda was mixed with 0.1 µl *tris*(2-carboxyethyl)phosphine (TCEP) and 2 µl of a 1:1 mixture of 50 mM ATP and γ-[$^{32}$P]ATP. Reactions were brought up to 20 µl with kinase buffer (50 mM EPPS, pH 8.5, 50 mM KCl, 20 mM $MgCl_2$, and 5% glycerol). Reactions were incubated at 26 ºC. 5 µl samples (25 pmol) were removed at given time points, added to 1.25 µl 5X SDS loading buffer and flash cooled in liquid nitrogen until all time points had been taken. For inhibition assays, 200 pmol of KinE/Sda was mixed with 8 µl of a 1:1 mixture of 50 mM ATP and γ-[$^{32}$P]ATP, 0.4 µl of 100 mM DTT, and either 6 nmol of Sda or Sda storage buffer (9.2 µl). Reactions were brought up to 40 µl with kinase buffer. Reactions were incubated at 26 ºC. 10 µl samples were taken at given time points, added to 2.5 µl 5X SDS loading buffer and flash cooled in liquid nitrogen until all time points had been taken. Samples were separated by SDS-PAGE (Tris-Tricine), and the radioactive bands were imaged using a storage phosphor screen.

### *KinE/Sda Crystallization*

*Gth* 466 KinE 282/Sda was concentrated using 15 ml 5000 MWCO Vivaspin concentrators (Vivascience). Dilutions were made to various concentrations in SD200 buffer (150 mM NaCl, 20 mM Tris-Cl, pH 8.0, 5 mM DTT). Crystallization trials were set up with either KinE/Sda on its own or in the presence of 5 mM $MgCl_2$ and a 5 molar excess of ADP (relative to the protein concentration). KinE/Sda was subjected to a variety of 96-condition crystallization screens from Qiagen (Table 4.3). All crystallization trials were carried out using the vapor diffusion method and 96-well sitting drop trays (Corning). Crystallization drops consisted of a 1 μl to 1 μl ratio of protein to crystallization solution and were equilibrated over a 100 μl well volume. A similar procedure was followed for full-length *Gth* 466 KinE/Sda, except that the protein was concentrated at room temperature using an Amicon Stir Cell Concentrator under Argon pressure.

# References

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

**Felsenstein, J.** (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol *17*, 368-376.

**Jiang, M., Shao, W., Perego, M., and Hoch, J.A.** (2000). Multiple histidine kinases regulate entry into stationary phase and sporulation in Bacillus subtilis. Molecular microbiology *38*, 535-542.

**Manachini, P.L., Mora, D., Nicastro, G., Parini, C., Stackebrandt, E., Pukall, R., and Fortina, M.G.** (2000). Bacillus thermodenitrificans sp. nov., nom. rev. Int J Syst Evol Microbiol *50 Pt 3*, 1331-1337.

**Rost, B., Yachdav, G., and Liu, J.** (2004). The PredictProtein server. Nucleic Acids Res *32*, W321-326.

**Rowland, S.L., Burkholder, W.F., Cunningham, K.A., Maciejewski, M.W., Grossman, A.D., and King, G.F.** (2004). Structure and mechanism of action of Sda, an inhibitor of the histidine kinases that regulate initiation of sporulation in Bacillus subtilis. Molecular cell *13*, 689-701.

**Trach, K.A., and Hoch, J.A.** (1993). Multisensory activation of the phosphorelay initiating sporulation in Bacillus subtilis: Identification and sequence of the protein kinase of the alternate pathway. Molecular Microbiology *8*, 69-79.

# CHAPTER 5
## Towards the Structure of a SpoIIE:SpoIIAA Complex

## Introduction

Entry to sporulation is dictated by the level of Spo0A~P in the cell. However, it is not until stage II of sporulation, after polar division has occurred that the bacterium is committed to sporulate (Parker et al., 1996); the cell cannot return to vegetative growth without first becoming a full-fledged spore. Completion of the asymmetrical division septum at one pole of the cell defines stage II. The switch from medial to asymmetric division is dependent upon the redeployment of the tubulin-like protein FtsZ from the middle to the poles of the cell (Ben-Yehuda and Losick, 2002). The spore septum is composed of a much thinner layer of peptidoglycan than that of the vegetative cell, most likely because it is hydrolyzed during stage III of sporulation, when the mother cell engulfs the prespore (Perez et al., 2000).

The alternative sigma factor $\sigma^F$ is responsible for gene expression in the forespore (Losick and Stragier, 1992; Margolis et al., 1991). A complex, forespore specific regulatory network comprised of the anti-σ factor SpoIIAB (AB), the anti-anti-σ factor SpoIIAA (AA), and the membrane bound serine phosphatase SpoIIE (E), controls the activity of $\sigma^F$ (Figure 5.1). $\sigma^F$ is produced shortly after the initiation of sporulation and is present in the predivisional cell, where SpoIIAB-ATP maintains it in an inactive complex (Alper et al., 1994). The inactive AB-ATP:$\sigma^F$ complex persists in the mother cell following asymmetrical septation (Duncan and Losick, 1993). Structural analysis of an AB:$\sigma^F$ complex revealed a 2:1 stoichiometry (Figure 5.2). $\sigma^F$ makes mainly superficial contacts with one of the molecules of AB in the complex, while interactions with the second AB are more extensive. This has important implications for destabilization of the $AB_2$-ATP:$\sigma^F$ complex through the action of the SpoIIAA (see below). Binding of SpoIIAB occludes a core RNA polymerase binding site on $\sigma^F$ (Campbell et al., 2002). Like the sporulation histidine kinases, SpoIIAB is an ATP binding protein of the Bergerat fold variety. In addition to its role as an anti-$\sigma^F$, AB also functions as a serine kinase that targets the anti-anti-$\sigma$ factor SpoIIAA for phosphorylation (Min et al., 1993).

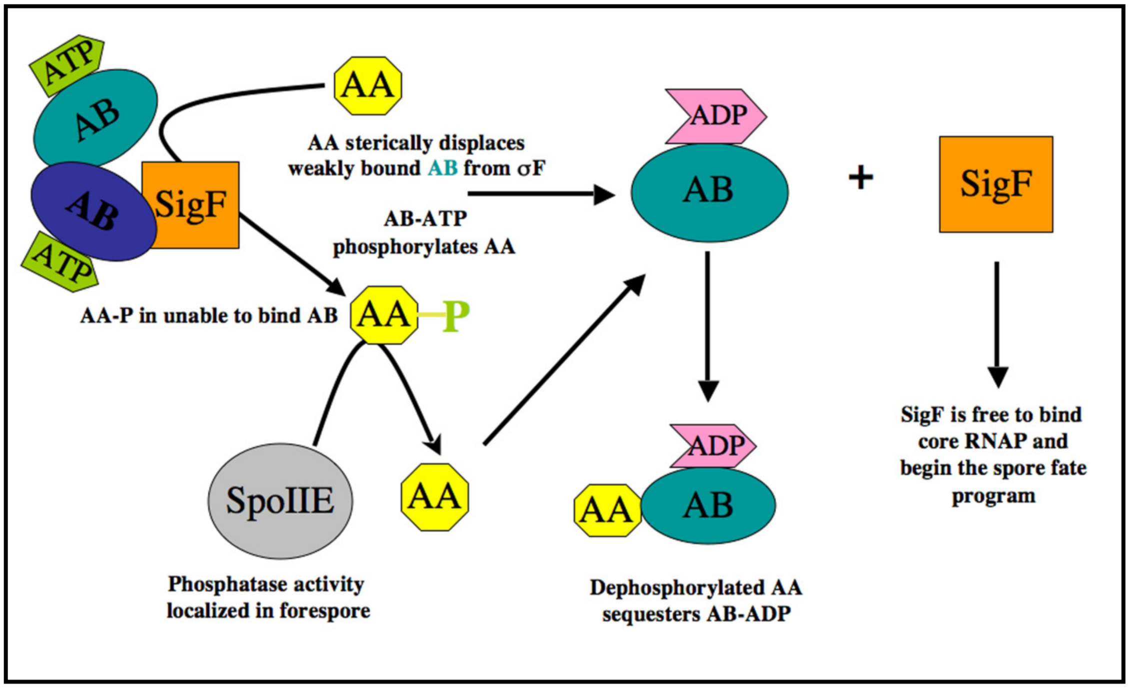


**Figure 5.1)** Schematic of the forespore specific stage II $\sigma^F$ pathway from *Bacillus subtilis*.

SpoIIAA functions as an anti-anti-$\sigma$ by interacting with the weakly bound SpoIIAB of the $AB_2$-ATP:$\sigma^F$ complex, and promoting dissociation of $\sigma^F$ from the second

AB-ATP by steric displacement (Ho et al., 2003). Crystal structures of SpoIIAA in complex with the ATP and ADP bound forms of SpoIIAB have enhanced our understanding of the mechanism of SpoIIAA induced dissociation of the $AB_2$:$\sigma^F$ complex (Figure 5.3) (Masuda et al., 2004). AB-ATP released from $\sigma^F$ targets AA for phosphorylation, rendering SpoIIAA~P, which is incompetent for recognizing AB. However, after utilizing ATP, the ADP form of AB becomes the target for an additional molecule of unphosphorylated AA, with which it makes a stable complex.

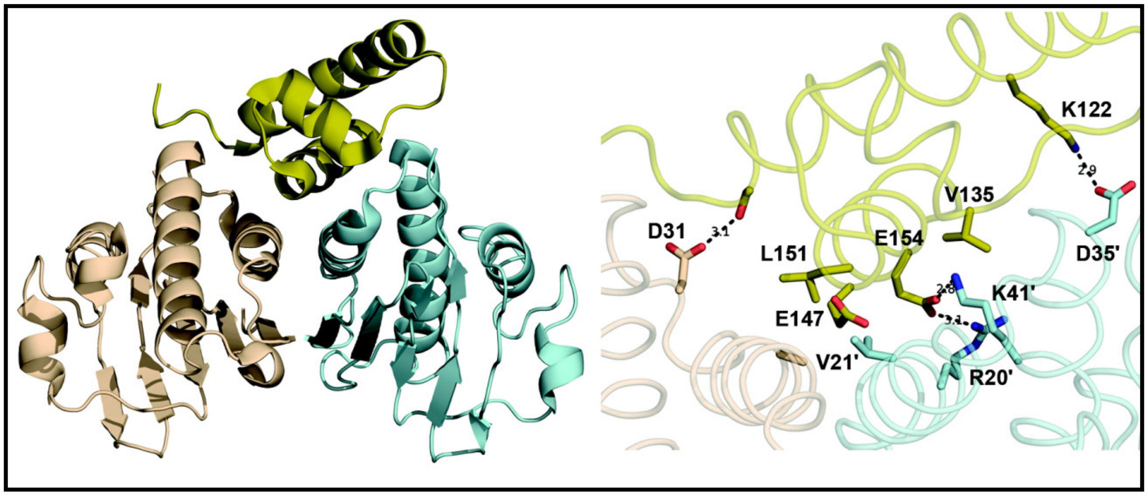


**Figure 5.2)** 2.9 Å crystal structure of the SpoIIAB$_2$:$\sigma^F$ complex from *Geobacillus stearothermophilus* (PDB code 1L0O) (Campbell et al., 2002). Only region 3.1 of $\sigma^F$ was observed in the structure (chartreuse). Two molecules of SpoIIAB (wheat and cyan) bind to $\sigma^F$ in an asymmetrical fashion. (right) Close-up of key residues, identified from the structure and genetic screens (Decatur and Losick, 1996), important for SpoIIAB:$\sigma^F$ complex formation. ‘ labeled residues are from the second SpoIIAB (cyan).

The concentration of unphosphorylated AA determines whether AB-ATP will completely lock up $\sigma^F$ (if there is more AB-ATP than AA), or whether enough AB-ADP will be complexed with AA to free $\sigma^F$ (if there is more unphosphorylated AA than AB) (Alper et al., 1994; Duncan et al., 1996). Unphosphorylated SpoIIAA is present in the predivisional cell, but not at levels sufficient to activate $\sigma^F$. Only in the forespore does AA reach the necessary threshold concentration, enough to fully sequester AB-ADP in a "sink" away from $\sigma^F$ to allow $\sigma^F$ to form holoenzyme with core RNA polymerase and carry out forespore gene expression (Carniol et al., 2004).

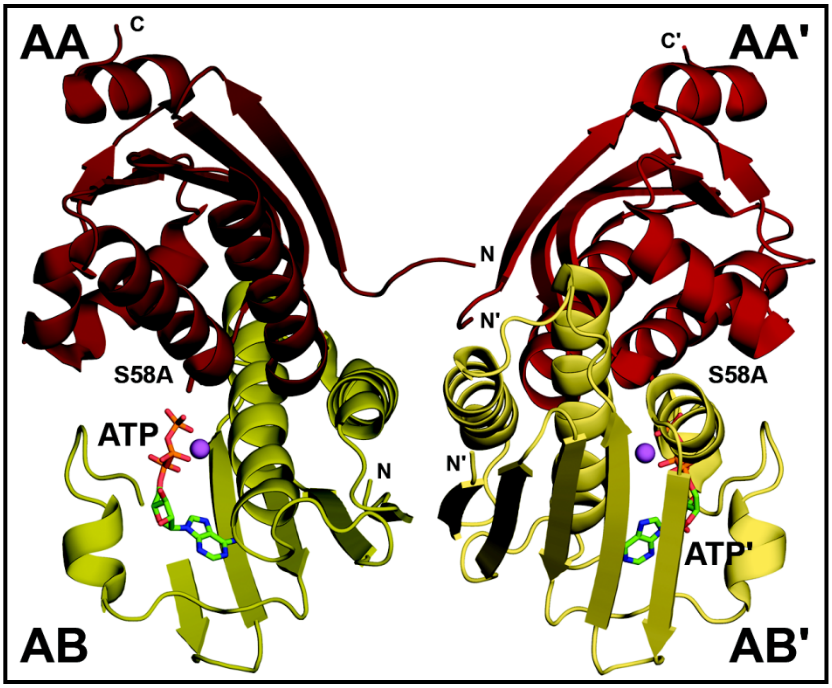


**Figure 5.3)** 2.7 Å structure of a SpoIIAB-ATP:SpoIIAA complex from *Geobacillus stearothermophilus*. SpoIIAB is poised for the kinase reaction, but is unable to phosphorylate SpoIIAA due to the Ser 58 to Ala AA mutant used for crystallization. Comparison of the AB:AA complex with the $AB_2:\sigma^F$ complex (Figure 5.2) reveals a mutually exclusive binding site on AB for AA and $\sigma^F$. ATP bound to each molecule of AB is represented as sticks. A portion (K100 - I121) of the AB model (left side) has been removed to show the close proximity of ATP to Ala 58. Magnesium ions are shown as magenta spheres. Adapted from (Masuda et al., 2004).

*B. subtilis* SpoIIAA is a 119 amino acid, single domain cytoplasmic protein. Crystal structures of SpoIIAA in the unphosphorylated and phosphorylated state from the related species *B. sphaericus* have been solved (Seavers et al., 2001). Few structural differences between unphosphorylated and phosphorylated AA exist, suggesting that covalent modification of the active site Ser functions more as a temporary negative marker rather than to promote global structural rearrangements as in many eukaryotic signaling systems. The overall fold of SpoIIAA consists of a four-stranded β-sheet with two α-helices above the sheet, one short α-helix at the side, and a short C-terminal α-helix (Figure 5.4). The conserved serine residue 58, which is the site of phosphorylation, is located at the beginning of the second α-helix.

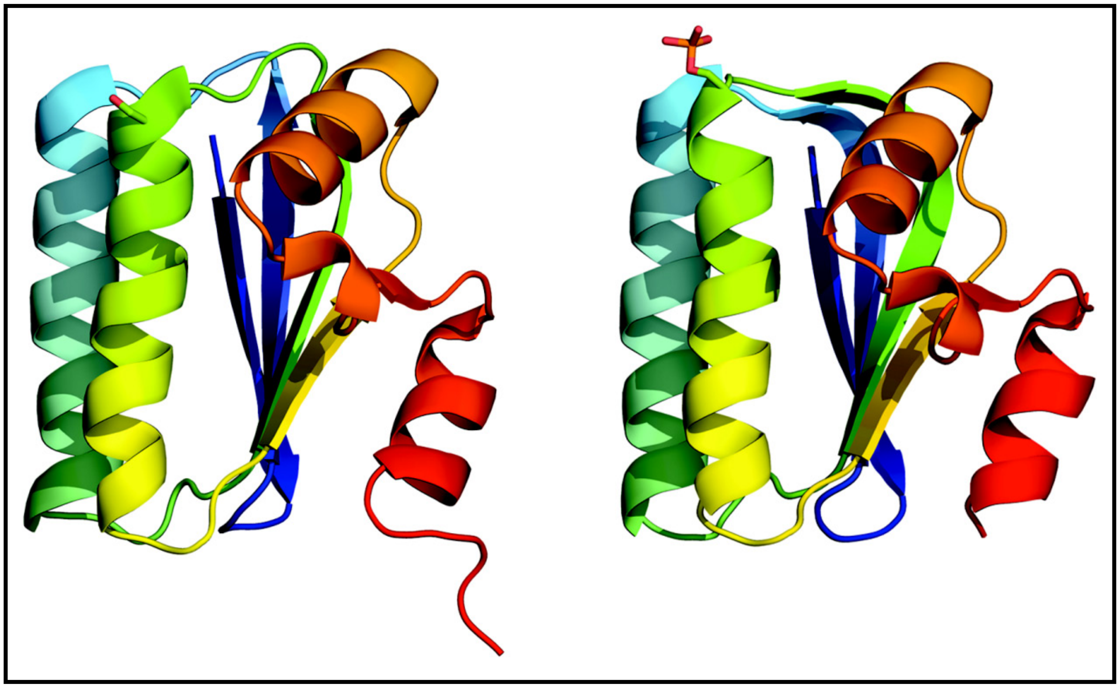


**Figure 5.4)** The 1.61 Å and 1.16 Å resolution crystal structures of *Bacillus sphaericus* SpoIIAA (left, PDB code 1H4Y) and phosphorylated SpoIIAA (right, PDB code 1H4X), respectively. Structures are colored N- to C-terminally from blue to red. The side chain of Ser57, the site of phosphorylation, is shown as sticks.

The phosphorylation state of SpoIIAA is not only contingent on the kinase activity of SpoIIAB, but also on the activity of the serine phosphatase SpoIIE. Expression of SpoIIE is directly controlled by Spo0A~P (Fawcett et al., 2000). The mechanism by which SpoIIE is able to dephosphorylate SpoIIAA in the forespore to levels sufficient for sporulation is a matter of some controversy. It has been suggested by several studies that during asymmetric division SpoIIE is targeted to the forespore side of the septum. Due to an increased surface to volume ratio, the ratio of SpoIIE to SpoIIAA within the forespore would be increased, allowing SpoIIAA to be sufficiently dephosphorylated (Duncan et al., 1995; Feucht et al., 2002; Wu et al., 1998). In a separate set of experiments, however, unequal distribution of SpoIIE was not found to be a contributing factor of $\sigma^F$ activation. Instead, a predivisional inhibitor of SpoIIE was proposed, though its discovery has remained elusive (Arigoni et al., 1999).

Genetic asymmetry, the phenomena where some genes become accessible to transcription in the forespore before others, may also contribute to elevated levels of unphosphorylated SpoIIAA in the forespore. Only one-third of the chromosome has translocated into the forespore by the time the asymmetric division septum is complete (Errington et al., 2001). The remaining two-thirds are slowly pumped in over the course of sporulation. SpoIIAB is susceptible to proteolysis at 37 ºC and is degraded with a half-life of approximately 25 min in both the mother cell and forespore (Pan et al., 2001). Because the gene for SpoIIAB is located towards the terminal end of the segregating chromosome, it will only be available for transcription long after sporulation has commenced. Thus, levels of SpoIIAB in the forespore would be temporarily reduced relative to the more stable proteins SpoIIAA and $\sigma^F$, allowing $\sigma^F$ to gain a foothold on transcription (Dworkin, 2003).

SpoIIE from *B. subtilis* is an 827 amino acid protein consisting of three domains (Figure 5.5) (Arigoni et al., 1999). Residues 33-322 comprise the highly hydrophobic domain I, consisting of 10 membrane spanning segments (residues 1-32 make up a short cytoplasmic N-terminal tail) which targets SpoIIE to the membrane. Domain II (residues 331-606) is involved in oligomerization of SpoIIE and in interactions with the tubulin-like protein FtsZ, which is required for the formation of the asymmetric septum (Ben-Yehuda and Losick, 2002; Lucet et al., 2000). Localization of SpoIIE to the septum is dependent on this interaction with FtsZ (Levin et al., 1997). Mutations within domain II result in a number of phenotypes, ranging from total sporulation deficiency to predivision activation of $\sigma^F$ (Barak and Youngman, 1996; Feucht et al., 2002). Domain III (residues 607-827) contains a serine phosphatase that has structural homology to the eukaryotic PP2C serine/threonine phosphatases involved in stress responses in both prokaryotes (i.e. sporulation) and eukaryotes (Meyer et al., 1994; Shiozaki et al., 1994). This domain has 10% sequence identity with the catalytic domain of human PP2Cα (Adler et al., 1997).

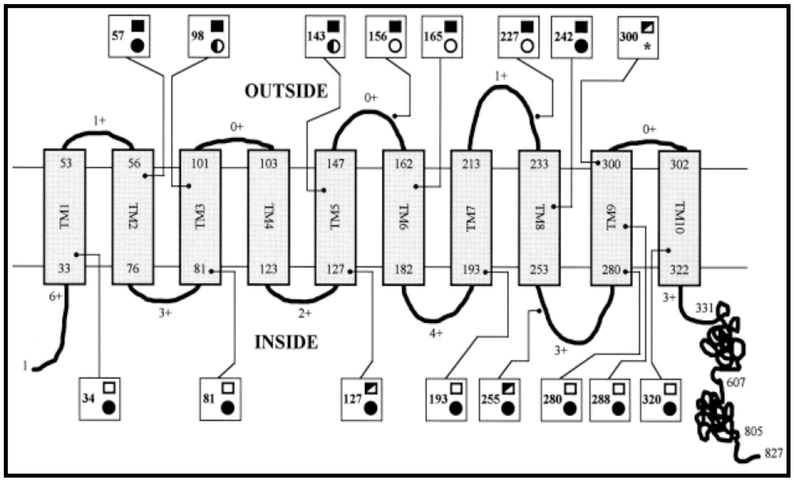


**Figure 5.5)** Predicted domain organization of SpoIIE from *Bacillus subtilis*. Transmembrane (TM) spanning regions are represented by grey rectangles. The soluble domains II and III are shown as globules. Numbers correspond to residues in the protein. The number of positively charged residues (Lys, Arg) in the loops between successive TM regions are indicated. Boxes relate to biochemical assays performed in the publication (Arigoni et al., 1999).

The crystal structure of human PP2C$\alpha$ has been solved to 2.0 Å resolution (Das et al., 1996). The N-terminal catalytic domain contains six $\alpha$-helices and 11 $\beta$-strands, a feature common to all members of the PP2C family. A central buried $\beta$-sandwich formed by the association of two anti-parallel $\beta$-sheets makes up the architecture of the catalytic center. A pair of anti-parallel $\alpha$-helices flank both $\beta$-sheets of the $\beta$-sandwich. The $\alpha$-helices are inserted between the two central $\beta$-sheets. The central $\beta$-sandwich binds two $Mn^{2+}$ ions, which coordinate the phosphate group of the substrate and provide a nucleophile for dephosphorylation. 10 invariant residues that PP2C$\alpha$ shares with SpoIIE are situated at or close to the $Mn^{2+}$ sites. A recent study by Jackson et al. identified three PP2C$\alpha$ active center aspartates (D60, D239, and D282) that when mutated to asparagine produced an enzyme with moderate to dramatic decreases in $K_{cat}$ values and dramatic increases for $K_M$ values for metal or substrate (Table 5.1) (Jackson et al., 2003). D60 bridges metals M1 and M2 (PP2C$\alpha$ utilizes either $Mn^{2+}$ or $Mg^{2+}$ for catalysis, though $Mn^{2+}$ is present in the PP2C$\alpha$ structure) in the active center, and D239 coordinates metal M2 (Figure 5.6). The third residue of importance, D282, activates a water-bound nucleophile, and thus when mutated only showed a decrease in $K_{cat}$. These active site aspartates are conserved in SpoIIE, and offer promising targets for mutagenesis and the stabilization of a SpoIIAA-E complex.

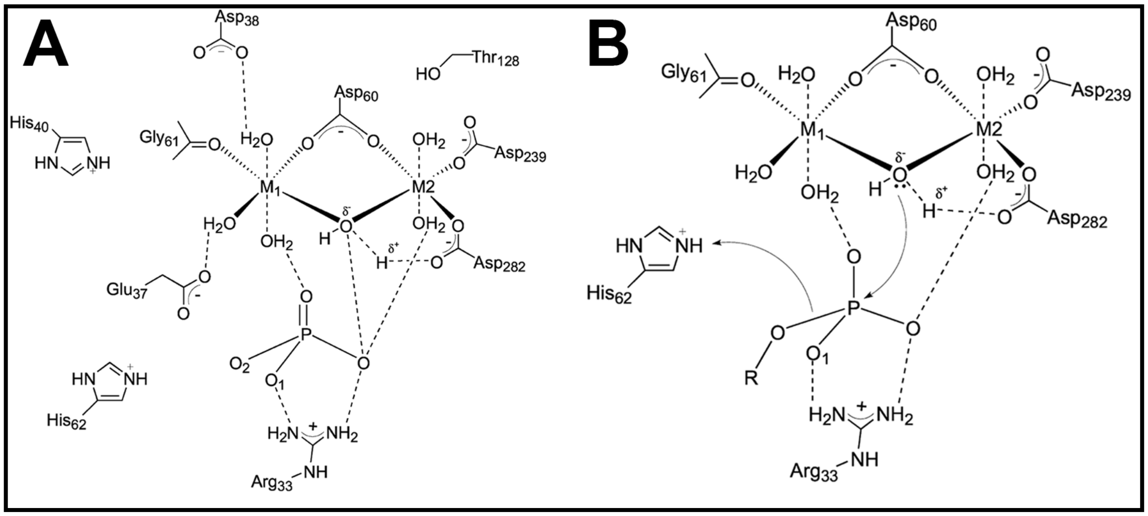


**Figure 5.6)** (A) Schematic representation of the ligand environment in human PP2Cα. Hydrogen bonding interactions coordinate the phosphate with metal centers M1 and M2 through water ligands. Asp-60 (corresponding to SpoIIE Asp-630) bridges the two metal ions. Asp-239 and Asp-282 (SpoIIE 748 and 797, respectively) make direct contacts with the metal centers. (B) Proposed mechanism for PP2Cα induced dephosphorylation. Asp-282 (SpoIIE 797) is proposed to activate the metal-associated water molecule for nucleophilic attack at the phosphorus atom by an $S_N2$ mechanism. Based on (Das et al., 1996; Jackson et al., 2003).

Table 5.1

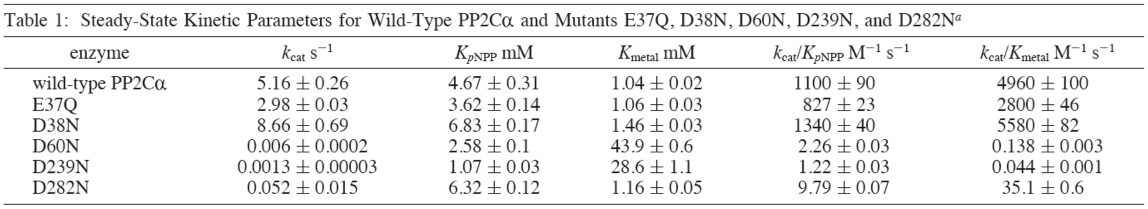


Table 1: Steady-State Kinetic Parameters for Wild-Type PP2Cα and Mutants E37Q, D38N, D60N, D239N, and D282N[a]

| enzyme | $k_{cat}$ $s^{-1}$ | $K_{pNPP}$ mM | $K_{metal}$ mM | $k_{cat}/K_{pNPP}$ $M^{-1}$ $s^{-1}$ | $k_{cat}/K_{metal}$ $M^{-1}$ $s^{-1}$ |
|---|---|---|---|---|---|
| wild-type PP2Cα | 5.16 ± 0.26 | 4.67 ± 0.31 | 1.04 ± 0.02 | 1100 ± 90 | 4960 ± 100 |
| E37Q | 2.98 ± 0.03 | 3.62 ± 0.14 | 1.06 ± 0.03 | 827 ± 23 | 2800 ± 46 |
| D38N | 8.66 ± 0.69 | 6.83 ± 0.17 | 1.46 ± 0.03 | 1340 ± 40 | 5580 ± 82 |
| D60N | 0.006 ± 0.0002 | 2.58 ± 0.1 | 43.9 ± 0.6 | 2.26 ± 0.03 | 0.138 ± 0.003 |
| D239N | 0.0013 ± 0.00003 | 1.07 ± 0.03 | 28.6 ± 1.1 | 1.22 ± 0.03 | 0.044 ± 0.001 |
| D282N | 0.052 ± 0.015 | 6.32 ± 0.12 | 1.16 ± 0.05 | 9.79 ± 0.07 | 35.1 ± 0.6 |

**Table 5.1)** Steady-State kinetic parameters for wild-type PP2Cα and mutants E37Q, D38N, D60N (SpoIIE D630), D239N (SpoIIE D748), and D282N (SpoIIE D797). From (Jackson et al., 2003).

# Results

In order to gain insights into the mechanism of SpoIIE induced dephosphorylation of SpoIIAA, we sought to determine the crystal structures of both SpoIIE on its own, and a SpoIIE-AA complex. To that end, we established preliminary expression constructs and purification protocols for both wild-type and mutant forms of SpoIIE, as well as a preliminary SpoIIAB-AA-E triple expression construct.

## *Establishing a SpoIIE Expression System*

Based on secondary structural predictions (Rost et al., 2004), *Gst* SpoIIE residue Ser 324 was identified as the first cytoplasmic residue following the membrane spanning domain I. Therefore, SpoIIE 324-CT (C-terminus) was cloned into plasmid pSKB2 and used as the basis from which subsequent SpoIIE constructs were derived. SpoIIE 324-CT encompasses the cytoplasmic domains II, which is responsible for self-oligomerization, and III, which possesses a PP2C-like serine phosphatase activity.

Due to the presence of several rare codons in the SpoIIE coding sequence, protein expression of SpoIIE 324-CT was carried out using *E. coli* Rosetta2™ (DE3) cells. The Rosetta2™ strain carries a plasmid that supplies tRNAs for 7 codons rarely used in *E. coli*. Rosetta2™ was developed primarily for the expression of eukaryotic proteins, but our experience has shown that expression levels for many non *E. coli* proteins can be greatly enhanced using this strain. When expressed at 16 ºC with 0.1 mM ITPG, SpoIIE 324-CT was estimated to be more than 50% soluble (Figure 5.7). The protein was purified to greater than 90% homogeneity using nickel affinity, anion exchange, and size

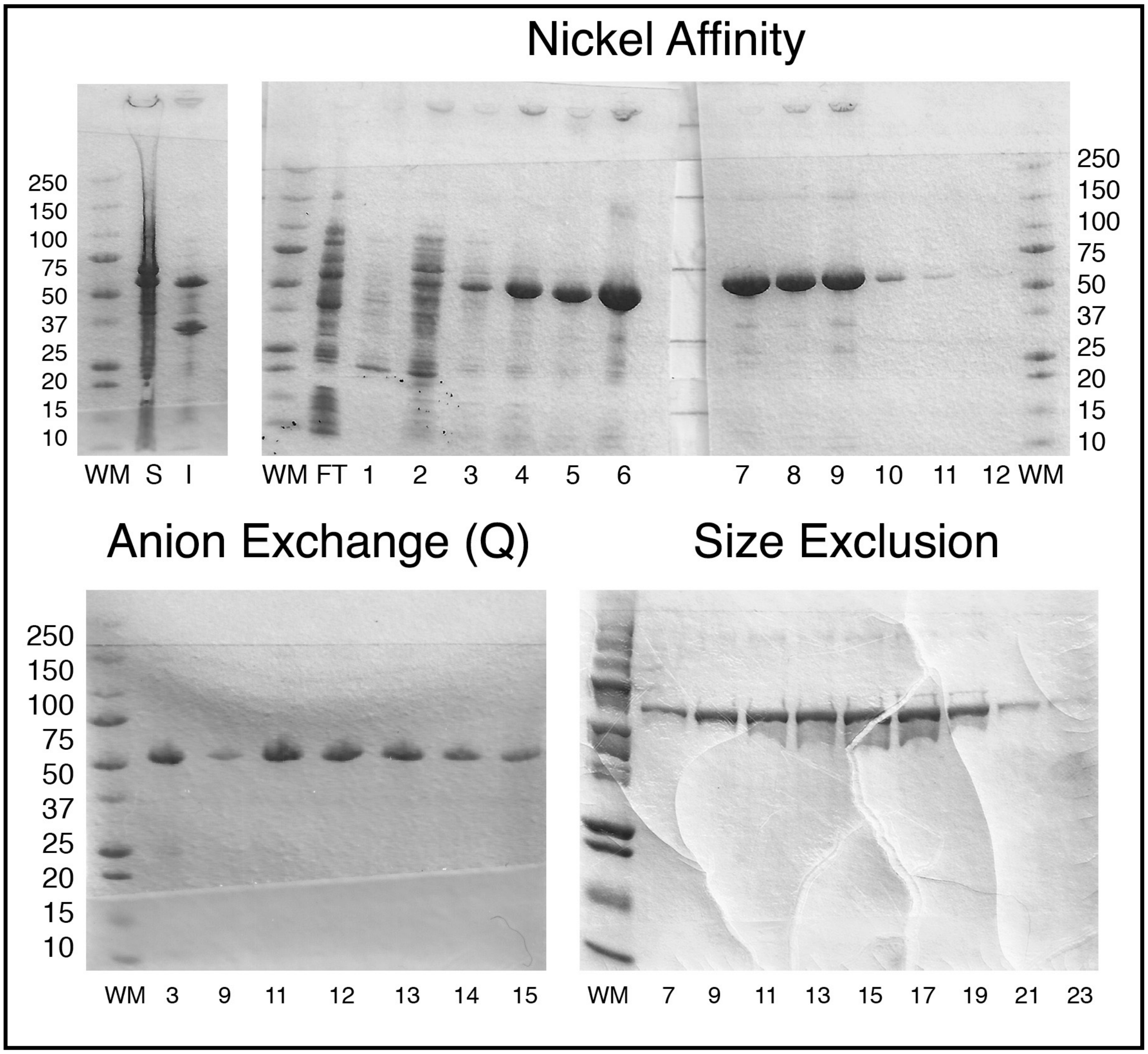


**Figure 5.7)** SDS-PAGE analysis of SpoIIE 324-CT after each stage of the purification process. Numbers below the gels indicate fractions from the corresponding chromatography step. WM - protein weight marker. S - Soluble fraction after lysis. I - Insoluble fraction after lysis. FT - Unbound protein in the flow through. For Nickel Affinity, lane numbers correspond to increasing percentage of elution buffer containing 200 mM imidazole: lanes 1 and 2 - 20% elution buffer; lanes 3 and 4 - 35%; lanes 5 through 7 - 50%; lanes 8 through 12 - 100%.

exclusion chromatography (Figure 5.7). No crystallization trials were performed with this construct because of previous observations made with *Bsu* SpoIIE that suggested any construct containing a complete domain II would suffer badly from aggregation, even in the presence of 10% glycerol and 0.1% Triton X-100. Indeed, *Gst* SpoIIE 324-CT showed signs of aggregation in a concentrator beyond 1 mg/ml.

### *Identification of Proteolytically Stable Domains of SpoIIE*

Because we suspected that SpoIIE 324-CT would not crystallize, we instead used protein from this construct to identify stable domains of SpoIIE through limited proteolysis that would not suffer from issues of aggregation. SpoIIE 324-CT was subjected to trypsin, chymotrypsin, and subtilisin limited proteolysis (Severinov et al., 1994), and N-terminal sequencing was used to identify proteolytic fragments. Due to the non-specific activity of subtilisin, N-terminal sequencing could not conclusively determine where this enzyme had cut SpoIIE. Digestion with trypsin and chymotrypsin, on the other hand, produced specific and somewhat overlapping fragments (Figure 5.8 and Table 5.2).

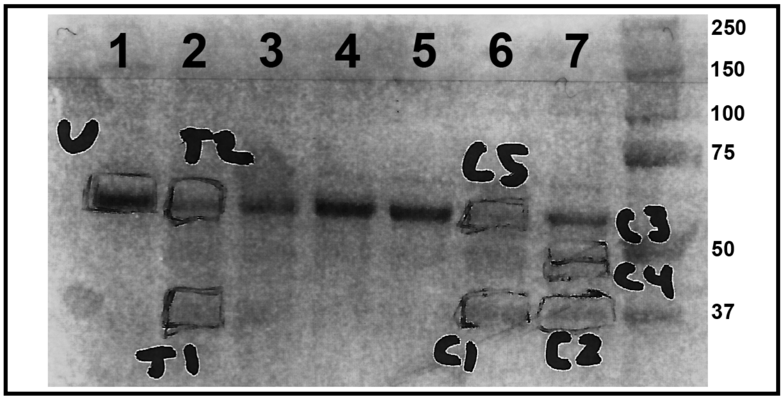


**Figure 5.8)** SDS-PAGE analysis of trypsin and chymotrypsin limited proteolysis experiments on SpoIIE 324-CT. Lane 1 - uncut control sample; Lane 2 - Trypsin:IIE 1:5; Lane 3 - Trypsin:IIE 1:10; Lane 4 - Trypsin:IIE 1:35; Lane 5 - Cymotrypsin:IIE 1:5; Lane 7 - Chymotrypsin:IIE 1:10.

Table 5.2

| Proteolytic Fragment | N-Terminal Sequence |
|---|---|
| Uncut (U) | GPHMS (Vector & SpoIIE 324) |
| Trypsin 1 (T1) | KLVAE (SpoIIE 479-483) |
| Trypsin 2 (T2) | ???PGTAEYI (SpoIIE 331-339) |
| Chymotrypsin 1 (C1) | ?ANIL (SpoIIE 465-469) |
| Chymotrypsin 4 (C4) | ?YM?? (likely SpoIIE 419-423) |

**Table 5.2)** N-terminal Edman sequencing results on the proteolytic fragments from Figure 5.8.

Based on the proteolysis experiments, we constructed SpoIIE 334-CT and SpoIIE 479-CT. SpoIIE 479-CT contains the C-terminal half of domain II and all of domain III. This construct was expressed in the same manner as 324-CT, and purified using the same chromatography steps of nickel affinity, anion exchange, and size exclusion (Figure 5.9). The protein expressed to very high levels (185 mg was recovered after the first nickel column step), and proved to be very soluble to at least 62 mg/ml without the need for glycerol or detergent. This finding suggests that a crucial determinant(s) for SpoIIE oligomerization resides in the N-terminal half of domain II.

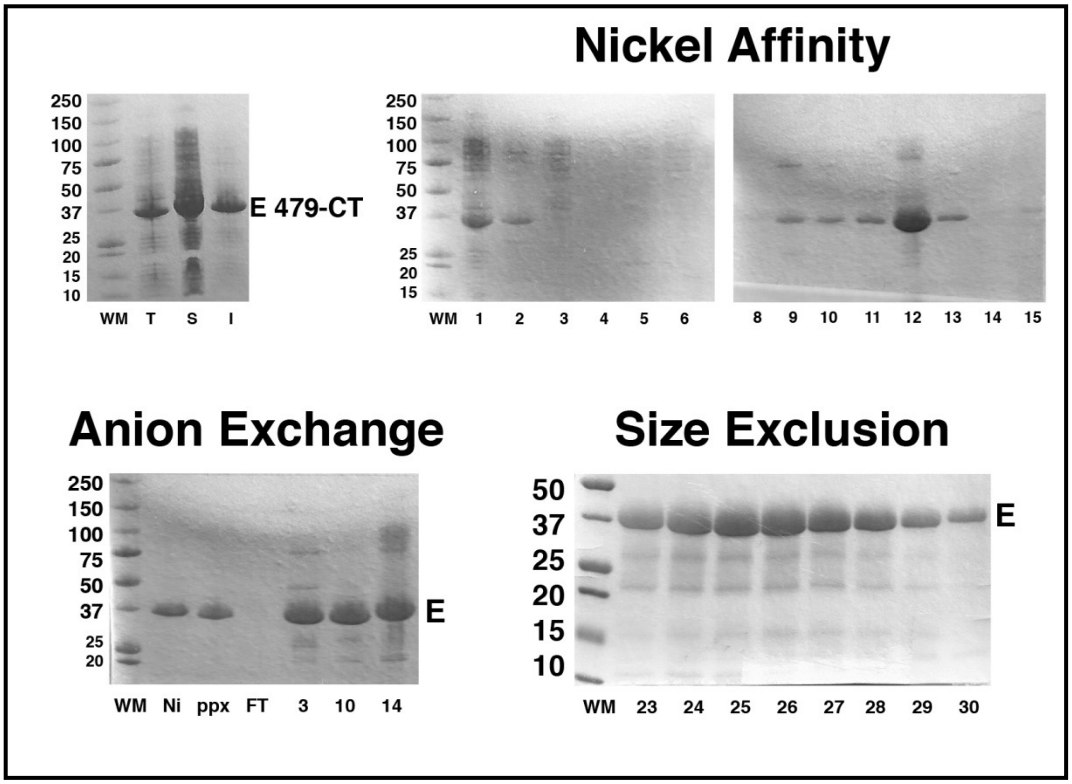
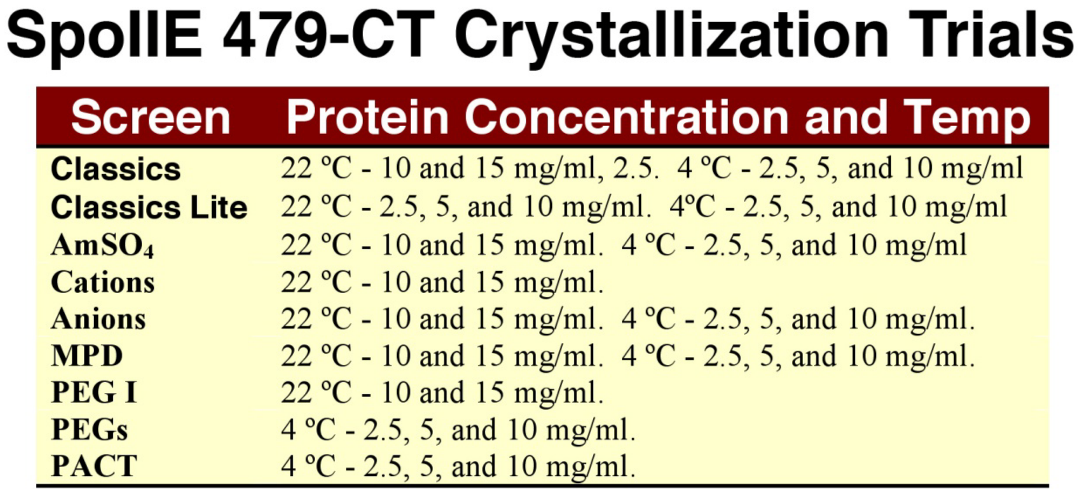


**Figure 5.9)** SDS-PAGE analysis of SpoIIE 479-CT after each stage of the purification process. Numbers below the gels indicate fractions from the corresponding chromatography step. WM - protein weight marker. S - Soluble fraction after lysis. I - Insoluble fraction after lysis. FT - Unbound protein in the flow through. Ni - pooled nickel fractions prior to PreScission™ Protease cleavage. PPX - Protein after digestion with PreScission™ Protease. For Nickel Affinity, lane numbers correspond to increasing percentage of elution buffer containing 200 mM imidazole: lane 1 - soluble fraction prior to nickel affinity; lane 2 - insoluble fraction prior to nickel affinity; lane 3 - unbound protein in the flow through; lanes 4 and 5 - 20% elution buffer; lanes 6 and 7 - 35%; lanes 8 through 10 - 50%; lanes 11 through 15 - 100%.

## *Crystallization of SpoIIE*

The purity of SpoIIE 479-CT was estimated by SDS-PAGE to be only about 90% (Figure 5.9, size exclusion). Nevertheless, we used this protein to screen for preliminary crystallization conditions (Table 5.3). To date, we have not observed protein crystals in any condition (trials were performed in the fall of 2007). For future crystallization studies it may be fruitful to further optimize the purification scheme and/or add an additional chromatography step.

Table 5.3



**SpoIIE 479-CT Crystallization Trials**

| Screen | Protein Concentration and Temp |
|---|---|
| **Classics** | 22 °C - 10 and 15 mg/ml, 2.5. 4 °C - 2.5, 5, and 10 mg/ml |
| **Classics Lite** | 22 °C - 2.5, 5, and 10 mg/ml. 4°C - 2.5, 5, and 10 mg/ml |
| **$AmSO_4$** | 22 °C - 10 and 15 mg/ml. 4 °C - 2.5, 5, and 10 mg/ml |
| **Cations** | 22 °C - 10 and 15 mg/ml. |
| **Anions** | 22 °C - 10 and 15 mg/ml. 4 °C - 2.5, 5, and 10 mg/ml. |
| **MPD** | 22 °C - 10 and 15 mg/ml. 4 °C - 2.5, 5, and 10 mg/ml. |
| **PEG I** | 22 °C - 10 and 15 mg/ml. |
| **PEGs** | 4 °C - 2.5, 5, and 10 mg/ml. |
| **PACT** | 4 °C - 2.5, 5, and 10 mg/ml. |

**Table 5.3)** Crystallization trials performed to date on SpoIIE 479-CT.

## *Active Site Mutations of SpoIIE*

Domain III, which begins approximately at residue 607, is a PP2C-like serine phosphatase (Adler et al., 1997). Jackson et al. identified key active site Asp residues in PP2Cα that when mutated to Asn dramatically decreased values for $K_{cat}$, but either maintained or increased the values of $K_m$ for metal ($Mn^{2+}$) or substrate (Jackson et al., 2003). Because these residues are conserved, using similar logic we sought to introduce the corresponding Asp to Asn mutations into SpoIIE. SpoIIE harboring these mutations would presumably still bind phosphorylated SpoIIAA (no decrease in $K_m$), but dephosphorylation of AA would be abolished (decreased $K_{cat}$), and in turn the E-AA complex would be stabilized.

Each mutation, D630N, D748N, and D797N, was first introduced into the SpoIIE 334-CT construct. A construct encompassing the complete domains II and III of SpoIIE was chosen because SpoIIE might be prevented from multimerizing in the context of a stable E-AA complex. After the Asp to Asn mutants were generated, each was incorporated into an AB-AA-E triple expression vector. The triple expression construct was designed under the assumption that when expressed together in the cell, the kinase activity of SpoIIAB would produce sufficient levels of phosphorylated SpoIIAA, which would in turn promote complex formation between mutant SpoIIE and AA. SDS-PAGE analysis showed that when co-expressed, AB, AA, and E all expressed to high levels (Figure 5.10). However, while AB and AA partitioned to the soluble fraction, E was 100% insoluble, even in the presence of 1% Triton (Figure 5.10). This could be explained in one of two ways. Either the mutant SpoIIE proteins are unable to recognize AA, or the E-AA interaction is unchanged by single Asp to Asn substitutions and remains a transient event. Clearly the E-AA interaction is not strong enough to either pull E into the soluble fraction, or AA into the insoluble fraction.

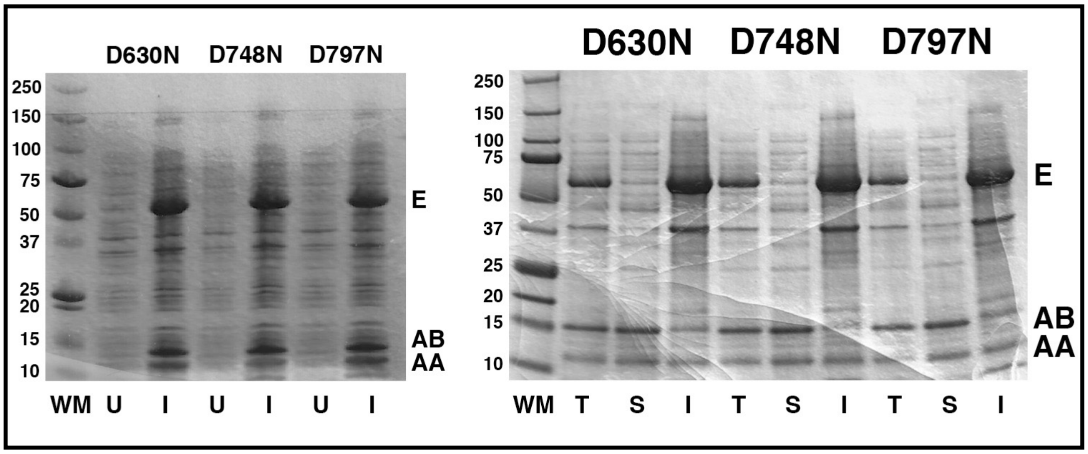


**Figure 5.10)** SpoIIE Asp to Asn mutants express to high levels, but are insoluble, even when expressed in the context of SpoIIAB and SpoIIAA. WM - Protein weight marker, U - pre expression, I (left gel) - post IPTG induction, T - total protein from bacterial lysate, S - soluble fraction following lysis, I (right gel) - insoluble fraction following lysis.

SpoIIE 334-CT is also 100% insoluble when expressed on its own (data not shown). It is interesting to note that inclusion of merely an extra 10 residues (SpoIIE 324-CT vs. 334-CT) to the construct renders the protein more than 50% soluble. We therefore sought to introduce the catalytic site Asp to Asn mutations into SpoIIE 324-CT. The decision was made to bypass the single mutants, and instead focus on generating a D630N, D748N, D797N triple mutant. Due to changing thesis priorities, to date only the pSKB2-SpoIIE 324-CT D748N D797N double mutant has been made. The SpoIIE triple mutant has not yet been constructed, and the double mutant has not yet had its expression tested, either by itself or in the context of an AB-AA-E triple expression construct.

# Materials and Methods

All oligonucleotide primers for PCR were resuspended to 100 pmol/µl in TE buffer, pH 8.0, unless otherwise noted. All PCR reactions were carried out with Accuprime™ pfx *Taq* DNA polymerase (2.5 units/µl), and the corresponding 10 X buffer containing dNTPs and $Mg^{2+}$ (Invitrogen). All PCR reactions were carried out in 100 µl reactions consisting of 0.5 µl of forward and reverse primers, 10 µl of 10 X pfx buffer, 0.5 µl DNA template, 0.5 µl of pfx polymerase, and 88.5 µl of $ddH_2O$, unless otherwise noted. All PCR reactions had an initial DNA melting step of 120 s at 95 ºC. All vector DNA was treated with calf intestinal phosphatase (CIP) for 1 h at 37 ºC prior to ligation. All ligation reactions were carried out at 15 ºC for approximately 16 h using T4 DNA ligase (NEB). All transformations for cloning steps were carried out with freshly prepared $CaCl_2$ competent *E. coli*, strain CC118.

## *Cloning SpoIIE from Genomic* Geobacillus stearothermophilus *DNA.*

*Gst* strain DSM 13240 (German Collection of Microorganisms and Cell Culture), also known as strain 10 (University of Oklahoma), was grown on trypticase soy agar overnight (~16 h) at 65 °C. Isolated colonies were picked and resuspended in 20 µl of $ddH_2O$, and frozen at -20 °C until needed. For full-length SpoIIE amplification, a single colony in 20 µl of $ddH_2O$ was boiled in a water bath for 10 min and spun at 14,000 RPM in a microfuge to pellet out the cellular debris. SpoIIE was amplified using primers MB-13 (NheI) and MB-14 (BamHI) and the following PCR parameters: 30 cycles of 15 s at

95 °C, 30 s at 54 °C, and 150 s at 68 °C, followed by a final extension step of 68 °C for 12.5 min. Amplicons were analyzed by agarose gel electrophoresis.

Full-length SpoIIE was separated on a 1% agarose gel and extracted using a gel extraction kit (Qiagen) following the manufacturer's protocol. Gel purified SpoIIE was sequentially digested with NheI and BamHI in 50 µl reactions at 37 °C, with a PCR clean up kit (Qiagen) step in between the NheI and BamHI digestions. SpoIIE was ligated into NheI/BamHI digested pSKB2 (Kan$^r$). pSKB2 is a derivative of pET28a with a PreScission™ Protease site in place of the thrombin cleavage site. Ligation reactions were transformed and selected by overnight growth at 37 °C on LB agar supplemented with 50 µg/ml kanamycin. Colonies were picked into 5 ml of LB supplemented with kanamycin and grown overnight at 37 °C, with shaking at 200 RPM. Plasmid DNA from 1.5 ml of overnight cultures was prepared using a miniprep kit (Qiagen) following the manufacturer's protocol. Plasmids containing SpoIIE insert DNA were confirmed by analytical NheI/BamHI digestion and agarose gel electrophoresis. Maxipreped (Qiagen) DNA of the pSKB2-SpoIIE construct was used for all subsequent cloning. (It was later discovered through sequencing of the pSKB2-SpoIIE plasmid that primer MB-13 had in fact introduced a frame shift at the beginning of the coding sequence. Therefore, pSKB2-SpoIIE is not suitable for protein expression, but is perfectly usable for subcloning).

### *Subcloning of SpoIIE 324-CT, SpoIIE 334-CT, and SpoIIE 479-CT*

SpoIIE 324-CT (C-terminus), 334-CT, and 479-CT were all PCR amplified from pSKB2-SpoIIE full-length using primers MB-69 (NdeI), and MB-33 (NdeI), and MB-77 (NdeI), respectively, and the T7 terminator. MB-69, MB-33, and MB-77, all introduce an NdeI site immediately 5' to the first codon of their respective amplicons. The PCR parameters for each insert were as follows: SpoIIE 324-CT: 30 cycles of 30 s at 95 °C, 30 s at 55 °C, and 120 s at 68 °C, followed by a final extension step of 10 min at 68 ºC. SpoIIE 334-CT: 30 cycles of 30 s at 95 ºC, 30 s at 60 ºC, 100 s at 68 ºC, followed by a final extension of 7 min at 68 ºC. SpoIIE 479-CT: 30 cycles of 30 s at 95 ºC, 30 s at 55 ºC, and 90 s at 68 ºC, followed by a final extension of 5 min, 50 s at 68 ºC. All amplicons were purified using a PCR clean up kit and doubly digested with NdeI and BamHI overnight at 37 °C. Digested products were gel purified and ligated into NdeI/BamHI digested pSKB2. Ligation reactions were transformed and selected as for full-length pSKB2-SpoIIE. Candidate constructs were confirmed by sequencing with the T7 promoter and terminator primers.

### *SpoIIE Asp to Asn Mutagenesis*

All mutations were introduced using the megaprimer method, with slight modifications (Sarkar and Sommer, 1990). Primers for each Asp to Asn mutant (MB-46 for D630N, MB-48 for D748N, and MB-50 for D797N) were first used with the T7 terminator primer to generate a megaprimer by PCR. All megaprimers were gel purified and resuspended in $ddH_2O$. SpoIIE 334-CT was amplified from pSKB2-SpoIIE full-length using the megaprimer (final concentration of 0.5 μM) containing the bases coding for the each specific mutation, and MB-33 as the second primer (30 rounds of PCR amplification: 95 °C for 30 s, 55 °C for 60 s, 68 °C for 120 s, followed by a final extension step of 10 min at 68 ºC). SpoIIE 334-CT mutant amplicons were then doubly digested with NdeI and BamHI, and subsequently ligated back into NdeI/BamHI pSKB2.

### *SpoIIAB-AA Double Expression Construct*

Both *Gst* SpoIIAB C145S and SpoIIAA were NdeI/BamHI digested from plasmids pLC6 and pLC5a, respectively (pLC6 and pLC5a were provided by Elizabeth Campbell). SpoIIAB with a C145S mutation has previously been used for structural studies (Campbell et al., 2002). Insert DNA was gel purified and ligated into NdeI/BamHI digested pET21a. For construction of the pET21a-AB-AA double expression vector, SpoIIAA was subcloned from pET21a using primers Liz13 and MB-27 with the following PCR conditions: 95 ºC for 120 s, 30 cycles of 30 s at 95 ºC, 30 s at 56 ºC, and 35 s at 68 ºC, followed by a final extension step of 68 ºC for 175 s. Liz 13 introduces a BamHI site into the lac operon of pET based vectors. MB-27 abolishes the BamHI site from the multiple cloning site of pET21a. BamHI/EcoRI digested SpoIIAA was ligated into BamHI/EcoRI digested pET21a-SpoIIAB to afford the pET21a-AB-AA double expression construct.

### *SpoIIAB-AA-E Triple Expression Construct*

SpoIIE 334-CT Asp to Asn mutants were amplified from their respective plasmids with primers Liz13 Eco and MB-29 with the following PCR conditions: 30 cycles of 30 s at 95 ºC, 30 s at 55 ºC, and 120 s at 68 ºC, followed by a final extension step of 68 ºC for 10 min. Liz13 Eco introduces an EcoRI site into the Lac operon of pET based vectors. MB-29 changes the SalI site in the multiple cloning site of pET based vectors to a NotI site. Amplicons were purified with a PCR clean up kit (Qiagen) and doubly digested with EcoRI and NotI overnight at 37 ºC. Digested insert DNA was gel purified and ligated into EcoRI/NotI digested pET21a-SpoIIAB-AA to afford the pET21a-SpoIIAB-AA-E triple expression constructs. Transformants were selected on LB agar supplemented with 150 μg/ml ampicillin and clones were confirmed by DNA sequencing.

### *SpoIIE 324-CT and 479-CT Protein Expression*

pSKB2-SpoIIE 324-CT was transformed into freshly made chemically competent *E. coli* Rosetta2 (DE3) cells (Novagen). Transformed cells were used to inoculate a 5 ml LB culture supplemented with 50 μg/ml kanamycin and 35 μg/ml chloramphenicol. The culture was grown overnight at 37 ºC with shaking at 200 RPM. The 5 ml culture was used to inoculate a 300 ml LB (Kan/Cam) culture, which was grown overnight at 37 ºC, again at 200 RPM. For large-scale protein expression, 50 ml of the 300 ml culture was used to inoculate four 2 L LB (Kan/Cam) cultures. 2 L cultures were grown at 37 ºC, 200 RPM until an $OD_{600}$ of 1.0. Protein expression was induced by the addition of IPTG to a final concentration of 0.1 mM. The temperature was then reduced to 16 ºC, and the cultures were incubated overnight at 200 RPM. Cells were pelleted at 4000 RPM, 4 ºC for 30 min and the supernatant was discarded. 2 L of cells were resuspended in 30 ml of nickel column buffer A (20 mM Tris-Cl, pH 7.4, 0.5 M NaCl, 10 mM $MgCl_2$ 5 mM imidazole, 1% Triton X-100, 5% glycerol) and frozen at -80 ºC.

All chromatography steps for protein purification were carried out on an ÄKTA fast performance liquid chromatography (FPLC) system (GE life sciences). 1 mM each of PMSF and β-ME was added to 2 L of cells prior to lysis. Cells were lysed between 12,000 and 15,000 psi at 4 ºC using an EmulsiFlex-C5™(Avestin) high-pressure homogenizer, and the crude lysate was clarified by centrifugation at 16,000 RPM for 30 min at 4 ºC. The supernatant was passed over a 5 ml metal chelating column (GE life sciences), charged with $NiSO_4$, at 1 ml/min, and the flow through was collected. Protein was eluted from the $Ni^{2+}$ column using a step gradient over 60 ml to 100% nickel column buffer B (same as buffer A with 200 mM imidazole) (10 ml at 20% B, 10 ml at 35%, 15 ml at 50%, and 25 ml at 100%). Twelve 5 ml fractions were collected. All fractions were analyzed by SDS-PAGE.

Peak elution fractions were pooled and PreScission™ Protease (GE life sciences) was added to an approximate 1:150 mass:mass ratio with SpoIIE 324-CT in order to remove the N-terminal $His_6$-tag. SpoIIE324-CT harbors four vector-derived residues at its N-terminus following PreScission™ Protease digestion. Cleaved protein was dialyzed overnight at 4 ºC in 2 L of 100 mM NaCl, 20 mM Tris-Cl, pH 7.9, 10 mM $MgCl_2$, 10% glycerol, 1% Triton X-100, and 1 mM DTT, using 12,000 - 14,000 MWCO dialysis tubing (Spectracor). The efficiency of the cleavage reaction was assessed by SDS-PAGE.

Next, SpoIIE 324-CT was passed over a 5 ml GST column in series with a 5 ml Q anion exchange column (GE life sciences) at 0.5 ml/min. The columns were equilibrated with 100 mM NaCl, 20 mM Tris-Cl, pH 7.9, 10 mM $MgCl_2$, 1 mM DTT, and 10% glycerol (Q column buffer A) in an attempt to remove Triton from the protein. After loading the protein onto the GST/Q series, the GST column was removed and SpoIIE 324-CT was eluted over 50 ml (10 column volumes) from the Q column at 1 ml/min using a linear gradient to the same buffer plus 700 mM NaCl (Q column buffer B). Twenty-five 2 ml fractions were collected. Two peaks were observed, the first at fraction 3 (~10% high salt buffer), and the second around fraction 1 (~300 mM NaCl). It was assumed that the first peak was aggregated SpoIIE, and hence was not used for further chromatography steps.

Finally, SpoIIE was passed over an SD200 (16/60) column, equilibrated with the Q column buffer A, at 0.3 ml/min. After a 30 ml void volume, forty-five 2 ml fraction were collected. Two partially overlapping peaks were observed, one corresponding to presumably multimeric species of SpoIIE, the other to monomeric SpoIIE. All fractions were analyzed by SDS-PAGE. Aliquots of monomeric SpoIIE 324-CT were snap frozen in liquid nitrogen and stored at -80 ºC.

SpoIIE 479-CT was expressed essentially the same as SpoIIE 324-CT, and followed the same sequence of chromatography steps, but with the following exceptions. Nickel column buffers A and B were 150 mM NaCl, 20 mM Tris-Cl, pH 7.0, 10 mM $MgCl_2$, 5 mM imidazole (or 200 mM imidazole for buffer B), and 5% glycerol. PreScission™ Protease dialysis buffer was 100 mM NaCl, 20 mM Tris-Cl, pH 7.0, 10 mM $MgCl_2$, 1 mM DTT. The GST/Q column buffers A and B and gel filtration buffer were the same as the PreScission™ Protease dialysis buffer (except 700 mM NaCl for Q column buffer B, and 2 mM DTT for the gel filtration buffer).

## *Limited Proteolysis of SpoIIE 324-CT*

58 pmol of SpoIIE 324-CT was mixed with either trypsin or α-chymotrypsin (Sigma, made up in protease buffer: 20mM Tris, pH 7.9, 50 mM NaCl, 5% glycerol, 0.1 mM EDTA, 1mM DTT, 10 mM $CaCl_2$) in the following protease to SpoIIE (mass:mass) ratios: 1:5.8 (10 pmol protease), 1:11.6 (5 pmol protease), 1:36 (1.6 pmol protease), and 1:53 (1.1 pmol protease). Reactions were brought up to 15 μl with protease buffer and

incubated at 26 ºC for 30 min. 0.5 µl of 100 mM PMSF was added to stop the reactions and the tubes were put in ice. 10 µl of each reaction plus 2.5 µl of 5X SDS loading buffer was run out on 4-20% Tris-Glycine PAGE (Invitrogen).

For N-terminal sequencing of proteolytic fragments, following SDS-PAGE the protein was transferred to PVDF membrane (Bio-Rad) following the electroblotting procedure supplied by The University of Texas Medical Branch (UTMB) Protein Chemistry Laboratory (http://www.utmb.edu/proch/serv01.htm). Proteolytic fragments of interest were excised from the PVDF and sent to UTMB for N-terminal Edman sequencing.

### *SpoIIE 479-CT Crystallization*

Gel filtration purified SpoIIE 479-CT (in 100 mM NaCl, 20 mM Tris-Cl, pH 7.0, 10 mM $MgCl_2$, 2 mM DTT) was concentrated using 10,000 MWCO Vivaspin concentrators (Vivascience). Dilutions were made using the same buffer and the protein was subjected to a variety of 96-condition crystallization screens from Qiagen (Table 5.2). All crystallization trials were carried out using the vapor diffusion method and 96-well sitting drop trays (Corning). Crystallization drops consisted of a 1 µl to 1 µl ratio of protein to crystallization solution and were equilibrated over a 100 µl well volume.

# References

**Adler, E., Donella-Deana, A., Arigoni, F., Pinna, L.A., and Stragler, P.** (1997). Structural relationship between a bacterial developmental protein and eukaryotic PP2C protein phosphatases. Mol Microbiol *23*, 57-62.

**Alper, S., Duncan, L., and Losick, R.** (1994). An adenosine nucleotide switch controlling the activity of a cell type-specific transcription factor in B. subtilis. Cell *77*, 195-205.

**Arigoni, F., Guerout-Fleury, A.M., Barak, I., and Stragier, P.** (1999). The SpoIIE phosphatase, the sporulation septum and the establishment of forespore-specific transcription in Bacillus subtilis: a reassessment. Mol Microbiol *31*, 1407-1415.

**Barak, I., and Youngman, P.** (1996). SpoIIE mutants of Bacillus subtilis comprise two distinct phenotypic classes consistent with a dual functional role for the SpoIIE protein. J Bacteriol *178*, 4984-4989.

**Ben-Yehuda, S., and Losick, R.** (2002). Asymmetric cell division in B. subtilis involves a spiral-like intermediate of the cytokinetic protein FtsZ. Cell *109*, 257-266.

**Campbell, E.A., Masuda, S., Sun, J.L., Muzzin, O., Olson, C.A., Wang, S., and Darst, S.A.** (2002). Crystal structure of the Bacillus stearothermophilus anti-sigma factor SpoIIAB with the sporulation sigma factor sigmaF. Cell *108*, 795-807.

**Carniol, K., Eichenberger, P., and Losick, R.** (2004). A threshold mechanism governing activation of the developmental regulatory protein sigma F in Bacillus subtilis. J Biol Chem *279*, 14860-14870.

**Das, A.K., Helps, N.R., Cohen, P.T., and Barford, D.** (1996). Crystal structure of the protein serine/threonine phosphatase 2C at 2.0 A resolution. EMBO J *15*, 6798-6809.

**Decatur, A.L., and Losick, R.** (1996). Three sites of contact between the Bacillus subtilis transcription factor sigmaF and its antisigma factor SpoIIAB. Genes Dev *10*, 2348-2358.

**Duncan, L., Alper, S., Arigoni, F., Losick, R., and Stragier, P.** (1995). Activation of cell-specific transcription by a serine phosphatase at the site of asymmetric division. Science *270*, 641-644.

**Duncan, L., Alper, S., and Losick, R.** (1996). SpoIIAA governs the release of the cell-type specific transcription factor sigma F from its anti-sigma factor SpoIIAB. J Mol Biol *260*, 147-164.

**Duncan, L., and Losick, R.** (1993). SpoIIAB is an anti-sigma factor that binds to and inhibits transcription by regulatory protein sigma F from Bacillus subtilis. Proc Natl Acad Sci U S A *90*, 2325-2329.

**Dworkin, J.** (2003). Transient genetic asymmetry and cell fate in a bacterium. Trends Genet *19*, 107-112.

**Errington, J., Bath, J., and Wu, L.J.** (2001). DNA transport in bacteria. Nat Rev Mol Cell Biol *2*, 538-545.

**Fawcett, P., Eichenberger, P., Losick, R., and Youngman, P.** (2000). The transcriptional profile of early to middle sporulation in Bacillus subtilis. Proc Natl Acad Sci U S A *97*, 8063-8068.

**Feucht, A., Abbotts, L., and Errington, J.** (2002). The cell differentiation protein SpoIIE contains a regulatory site that controls its phosphatase activity in response to asymmetric septation. Mol Microbiol *45*, 1119-1130.

**Ho, M.S., Carniol, K., and Losick, R.** (2003). Evidence in support of a docking model for the release of the transcription factor sigma F from the antisigma factor SpoIIAB in Bacillus subtilis. J Biol Chem *278*, 20898-20905.

**Jackson, M.D., Fjeld, C.C., and Denu, J.M.** (2003). Probing the function of conserved residues in the serine/threonine phosphatase PP2Calpha. Biochemistry *42*, 8513-8521.

**Levin, P.A., Losick, R., Stragier, P., and Arigoni, F.** (1997). Localization of the sporulation protein SpoIIE in Bacillus subtilis is dependent upon the cell division protein FtsZ. Mol Microbiol *25*, 839-846.

**Losick, R., and Stragier, P.** (1992). Crisscross regulation of cell-type-specific gene expression during development in B. subtilis. Nature *355*, 601-604.

**Lucet, I., Feucht, A., Yudkin, M.D., and Errington, J.** (2000). Direct interaction between the cell division protein FtsZ and the cell differentiation protein SpoIIE. EMBO J *19*, 1467-1475.

**Margolis, P., Driks, A., and Losick, R.** (1991). Establishment of cell type by compartmentalized activation of a transcription factor. Science *254*, 562-565.

**Masuda, S., Murakami, K.S., Wang, S., Anders Olson, C., Donigian, J., Leon, F., Darst, S.A., and Campbell, E.A.** (2004). Crystal structures of the ADP and ATP bound forms of the Bacillus anti-sigma factor SpoIIAB in complex with the anti-anti-sigma SpoIIAA. J Mol Biol *340*, 941-956.

**Meyer, K., Leube, M.P., and Grill, E.** (1994). A protein phosphatase 2C involved in ABA signal transduction in Arabidopsis thaliana. Science *264*, 1452-1455.

**Min, K.T., Hilditch, C.M., Diederich, B., Errington, J., and Yudkin, M.D.** (1993). Sigma F, the first compartment-specific transcription factor of B. subtilis, is regulated by an anti-sigma factor that is also a protein kinase. Cell *74*, 735-742.

**Pan, Q., Garsin, D.A., and Losick, R.** (2001). Self-reinforcing activation of a cell-specific transcription factor by proteolysis of an anti-sigma factor in B. subtilis. Mol Cell *8*, 873-883.

**Parker, G.F., Daniel, R.A., and Errington, J.** (1996). Timing and genetic regulation of commitment to sporulation in Bacillus subtilis. Microbiology (Reading, England) *142 (Pt 12)*, 3445-3452.

**Perez, A.R., Abanes-De Mello, A., and Pogliano, K.** (2000). SpoIIB localizes to active sites of septal biogenesis and spatially regulates septal thinning during engulfment in bacillus subtilis. J Bacteriol *182*, 1096-1108.

**Rost, B., Yachdav, G., and Liu, J.** (2004). The PredictProtein server. Nucleic Acids Res *32*, W321-326.

**Sarkar, G., and Sommer, S.S.** (1990). The "megaprimer" method of site-directed mutagenesis. Biotechniques *8*, 404-407.

**Seavers, P.R., Lewis, R.J., Brannigan, J.A., Verschueren, K.H., Murshudov, G.N., and Wilkinson, A.J.** (2001). Structure of the Bacillus cell fate determinant SpoIIAA in phosphorylated and unphosphorylated forms. Structure *9*, 605-614.

**Severinov, K., Fenyo, D., Severinova, E., Mustaev, A., Chait, B.T., Goldfarb, A., and Darst, S.A.** (1994). The sigma subunit conserved region 3 is part of "5'-face" of active center of Escherichia coli RNA polymerase. J Biol Chem *269*, 20826-20828.

**Shiozaki, K., Akhavan-Niaki, H., McGowan, C.H., and Russell, P.** (1994). Protein phosphatase 2C, encoded by ptc1+, is important in the heat shock response of Schizosaccharomyces pombe. Mol Cell Biol *14*, 3742-3751.

**Wu, L.J., Feucht, A., and Errington, J.** (1998). Prespore-specific gene expression in Bacillus subtilis is driven by sequestration of SpoIIE phosphatase to the prespore side of the asymmetric septum. Genes Dev *12*, 1371-1380.

# CHAPTER 6

# Crystal Structures of the Sulfotransferase Teg12 in a Complex with the Teicoplanin Aglycone

## Introduction

During the 20th century, widespread use of antibiotics significantly reduced the threat of many once lethal infectious diseases. However, the success of these wonder drugs may soon become their Achilles' heel. Bacterial pathogens that have developed resistance to most widely used antibiotics are now regularly seen in clinical settings. Vancomycin and teicoplanin are glycopeptide antibiotics used in the treatment of many gram-positive bacterial infections, including methicillin-resistant *Staphylococcus aureus* (MRSA). With the appearance of vancomycin resistant *Enterococci* in the late 1980s and resistant *Staphylococci* in the early 1990s, these traditional antibiotics of last resort are in danger of becoming clinically compromised (Hiramatsu et al., 1997; Linden, 2008; Uttley et al., 1988). As with many bacterial natural products, the discovery of additional glycopeptide congeners that might combat the growing problem of antibiotic resistance has slowed as it has become increasingly difficult to identify new biodiversity from which novel molecules might be characterized.

The vast majority of bacteria present in the environment remain recalcitrant to culturing (Rappe and Giovannoni, 2003). This uncultured majority no doubt contains previously inaccessible glycopeptide biosynthetic gene clusters, many of which could encode the biosynthesis of novel glycopeptide congeners. Although metabolites produced by bacteria that are difficult to culture in the laboratory cannot be characterized using standard microbiological methods, it is possible to extract DNA directly from environmental samples and then analyze this DNA for sequences that might encode the biosynthesis of new natural products. In a recent analysis of DNA extracted directly from desert soil, we uncovered a new glycopeptide biosynthetic gene cluster (the TEG gene cluster) that is predicted to encode the biosynthesis of the first polysulfated glycopeptide congeners (Banik and Brady, 2008). The TEG gene cluster contains three closely related 3'-phosphoadenosine 5'-phosphosulfate (PAPS) dependent sulfotransferases (Teg12, 13 and 14). In vitro, these three sulfotransferase finishing enzymes sulfate teicoplanin-like glycopeptides at three unique sites and, in combination, can be used to produce seven different glycopeptide sulfation patterns (Figure 6.1).

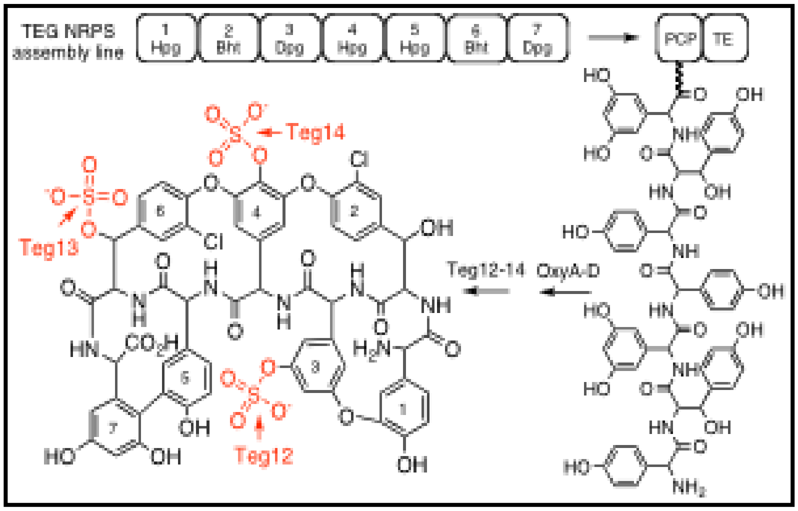


**Figure 6.1)** Predicted structures for the glycopeptide congeners produced by the TEG gene cluster. Seven nonribosomal peptide synthetase modules encoded by the TEG cluster are predicted to biosynthesize a heptapeptide that is oxidatively cross-linked by OxyA-D homologs and then sulfated by Teg12, 13, and 14.

Vancomycin- and teicoplanin-like glycopeptides are structurally defined by the presence of an oxidatively cross-linked heptapeptide core (Hubbard and Walsh, 2003; Kahne et al., 2005). The peptide core is initially produced as a linear polymer by nonribosomal peptide synthetases, which is then oxidatively cross-linked into either three or four large macrocycles by conserved cytochrome P450 oxidases. During their biosynthesis, each member of this family of antibiotics is functionalized by a unique collection of finishing enzymes that includes glycosyl transferases, halogenases, acyl transferases and sulfotransferases. Although glycopeptides show some variation in the sequence of the heptapeptide core, the bulk of the structural diversity seen within this class of antibiotics arises from the functionality added by finishing enzymes. Over 150 different glycosylated, halogenated, and alkylated glycopeptide congeners have been characterized from cultured bacteria (Nicolaou et al., 1999). Only three naturally occurring sulfated congeners have been identified to date from studying this same pool of bacteria (Boeck and Mertz, 1986; Nicolaou et al., 1999). While anionic glycopeptides have rarely been reported as natural products, telavancin, a semisynthetic phosphono congener that proved to be a very effective antibiotic in clinical trials, was recently approved for use in humans by the FDA (Higgins et al., 2005; Stryjewski et al., 2005). Increasing the hydrophilicity with the addition of the negatively charged phosphono group was found to significantly improve the adsorption, distribution, metabolism, and excretion profile of this class of antibiotics (Leadbetter et al., 2004). The enzymatic synthesis of anionic glycopeptides may provide a facile means to access additional anionic congeners with improved pharmacological properties.

In this chapter we report a series of Teg12 sulfotransferase structures, including an apo structure, a binary structure complexed with the teicoplanin aglycone substrate, as well as a ternary structure containing both PAP and the teicoplanin aglycone. In the binary and ternary structures, the glycopeptide substrate is observed bound at two different locations. Teg12 appears to undergo a series of conformational rearrangements during glycopeptide recruitment, binding and catalysis. These studies provide insight into the sulfotransferase mechanism, as well as insights into how this rarely seen class of finishing enzymes might be engineered to produce novel anionic glycopeptides.

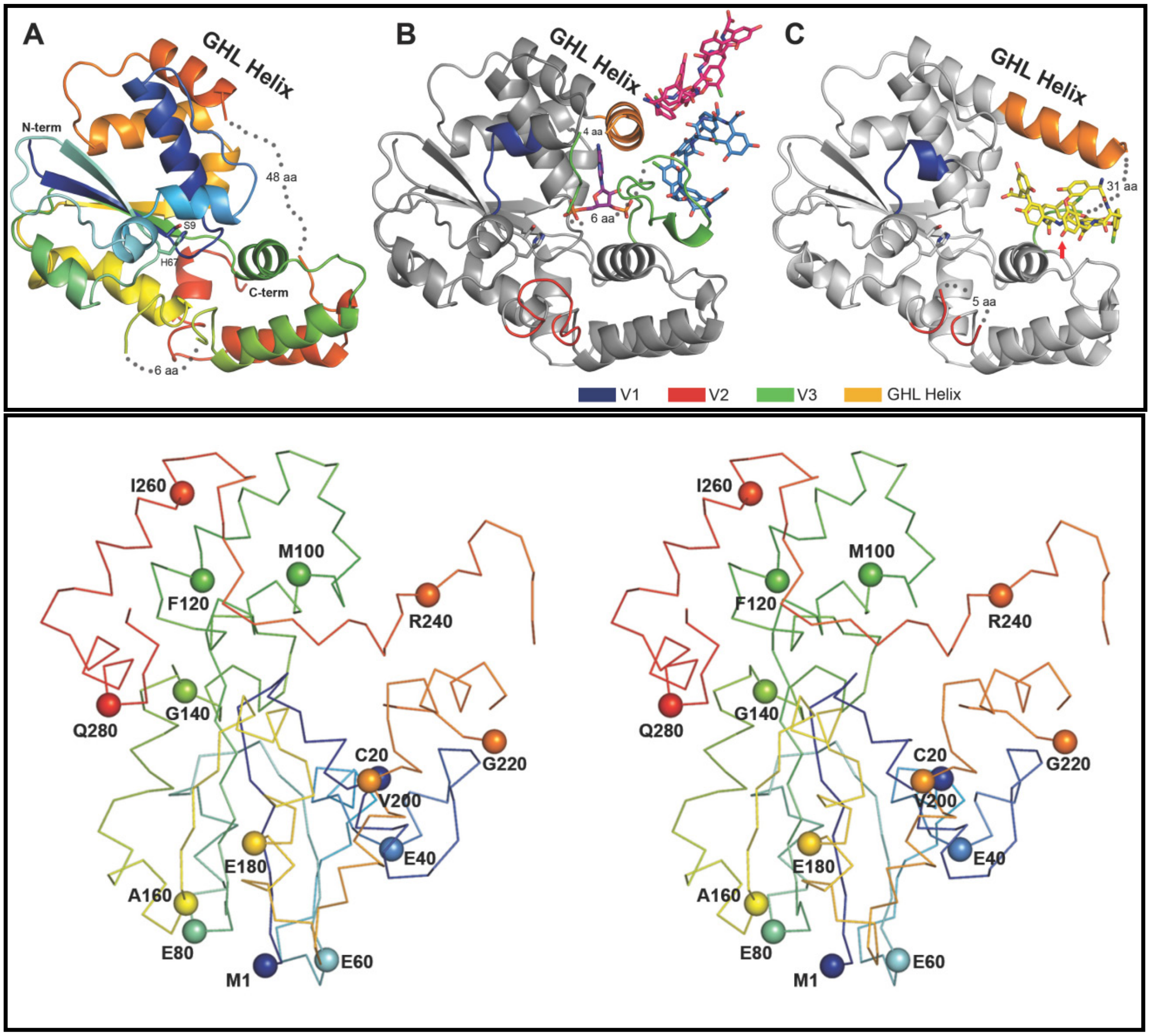
## Results

### *Teg12 Apo Structure*

N-terminally $His_6$-tagged Teg12 was affinity purified using nickel NTA resin and crystallized without the need for further purification. The apo structure was solved to 2.91 Å resolution by molecular replacement using StaL (PDB code 2OV8) as a search model (Shi et al., 2007). StaL is a related sulfotransferase involved in the biosynthesis of the monosulfated glycopeptide congener A47934 (Boeck and Mertz, 1986; Lamb et al., 2006). StaL has 52.9% sequence identity with Teg12. The Teg12 structure superimposes well on StaL (rmsd of 1.05 Å over 210 $C\alpha$ atoms). Teg12 crystallizes as a dimer and the overall structure resembles that of StaL. Running Teg12 through the DALI server shows that its closest eukaryotic sulfotransferase relative is the human cytosolic sulfotransferase 1C1 (SULT1C1, PDB code 1ZHE), with an rmsd of 2.8 Å over all $C\alpha$ atoms.

**Figure 6.2** (adjacent page) The Three Teg12 structures are presented. Only a single monomer from the dimer is shown for clarity. (A) Teg12-Apo is colored using a rainbow scheme, from blue (N-term) to red (C-term). Regions of disorder are connected with grey dots, with the number of disordered residues indicated. The flexible GHL Helix is shown in orange. Side chains for the proposed active site residues His67 and Ser9 are also shown. (B) Teg12-ternary complex containing PAP and the teicoplanin aglycone. The protein has been colored grey in order to accentuate regions of the structure that differ from Teg12-Apo. Specifically, these regions are the variable loops V1-V3, colored blue, red, and green, respectively. Also shown is a molecule of PAP in the active site, a molecule of teicoplanin aglycone that interacts with the V3 loop (colored sky blue), and an additional molecule of teicoplanin aglycone involved in crystal packing interactions (colored hot pink). (C) Teg12-binary structure complexed with teicoplanin aglycone in the active site cavity. Again, the protein has been colored grey with V1-V3 colored as in the ternary structure. The active site is shown as a stereo close-up image in Figure 6.6. Teg12 sulfates the hydroxyl of residue 3 of the teicoplanin aglycone, which has been indicated in the figure with a red arrow. Omitted from the binary structure is a second molecule of teicoplanin aglycone bound to the outside of the protein on the opposite side of the GHL helix. (bottom panel) Stereo view of a Cα trace of the Teg12-ternary structure represented as a ribbon diagram with the Cα carbon of every 20th amino acid shown as a sphere.

Figure 6.2



(Dombrovski et al., 2006). As with other sulfotransferases in this family, Teg12 consists of a single globular α/β domain composed of a parallel beta sheet core surrounded by alpha helices (Figure 6.2A). The beta sheet core of both StaL and Teg12 contains four strands. The dimer interface, which resembles that seen in StaL, consists of a symmetrical interaction between a short helix-loop motif from one monomer and the same helix-loop motif from the other monomer. In addition to the hydrophobic contacts that exist between the two helices there is a hydrogen bond between the carbonyl oxygen of Gly51 at the end of the helix from one monomer and the backbone nitrogen of Val74 of the other monomer. The active site cavity of each monomer faces the dimer interface.

Minor differences exist between the two monomers of the apo structure. The pET28a vector derived N-terminal tail, including the $His_6$, thrombin cleavage site and T7 tags, from monomer A is ordered and involved in crystal packing interactions. The arrangement of these residues within the crystal causes distortions in the region Gly28 through Ser41 when compared with StaL, and subsequent Teg12 complex structures. In monomer B, residues Gly28 through Ser41 could not be modeled in the electron density map, further underscoring the conformational flexibility of this region. From our co-crystal structures and that of StaL, it is known that residues Glu216 through Asp250 constitute a largely disordered loop. In the Teg12 apo structure, this disorder extends even further N-terminally into what is observed as a helix in the Teg12 binary and ternary structures and StaL. Electron density falls off in monomer A after Ser203, but extends through Glu210 in monomer B. Ser203 corresponds to Cys196 in StaL. In StaL Cys196 makes a disulfide bond with Cys20 (residue 20 is also a cysteine in Teg12). The absence of a disulfide in Teg12 could impart a greater flexibility in this region and explain the additional observed disorder. In subsequent Teg12 co-crystal structures, portions of this helix-loop region (Thr204-Asp250) play key roles in binding the glycopeptide substrate. Because of its role in binding the glycopeptide, we have termed this entire conformationally flexible region the GHL (glycopeptide-helix-loop).

ClustalW alignment of the three TEG sulfotransferases shows that most of the sequence variability seen within this family of sulfotransferases is concentrated on three short loops (variable regions V1, V2, V3) that surround the predicted active site (Figure

6.3). The large disordered loop from the GHL corresponds to the longest of these three variable regions, V3. V2 encompasses residues Gly127 through Gly137. Much of V2 (Ala130 to Gly135 in monomer A and Asn129 through Gly136 in monomer B) is disordered in the Teg12 apo structure. V1, the shortest of the three variable sequences, corresponds to an ordered region (though disordered in monomer B) that spans Ile37 through Thr43.

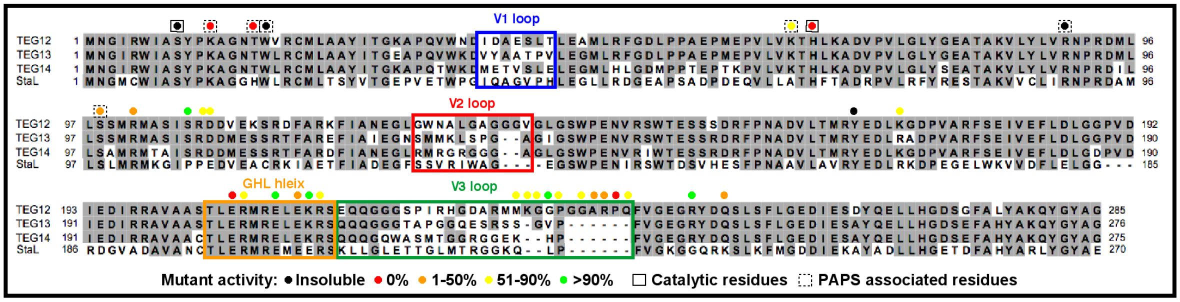


**Figure 6.3)** ClustalW alignment of glycopeptide sulfotransferases. Highly variable sequences (V1, V2 and V3) and the flexible part of the GHL helix appear in colored boxes that match the coloration seen in Figures 6.2B and 6.2C. Results from alanine exchange mutagenesis experiments are color coded by percent activity. Results for catalytic and PAPS associated residues are shown in solid or dashed boxes, respectively. All other mutated residues are thought to interact with the glycopeptides. The H67A mutant was insoluble and therefore H67Q data is reported.

Electron density was observed in the PAPS binding site of monomer A (this binding site was empty in monomer B). However, neither PAP nor PAPS could be modeled into the density. We were able to model the dipeptide aspartame (N-(L-a-aspartyl)-L-phenylalanine,1-methyl ester) into this density (Figure 6.4). Aspartame is a component of the Silver Bullets™ screen (Hampton) used during crystallization. The orientation of aspartame in the structure mimics that of PAP. Its phenylalanine ring stacks with Trp17, as does the adenine ring of PAP, and its N-terminal nitrogen hydrogen bonds with the hydroxyl of Ser98. Eukaryotic sulfotransferases are known to tightly bind ribose, adenine and other nucleotides in the PAPS binding pocket (Lin and Yang, 2000).

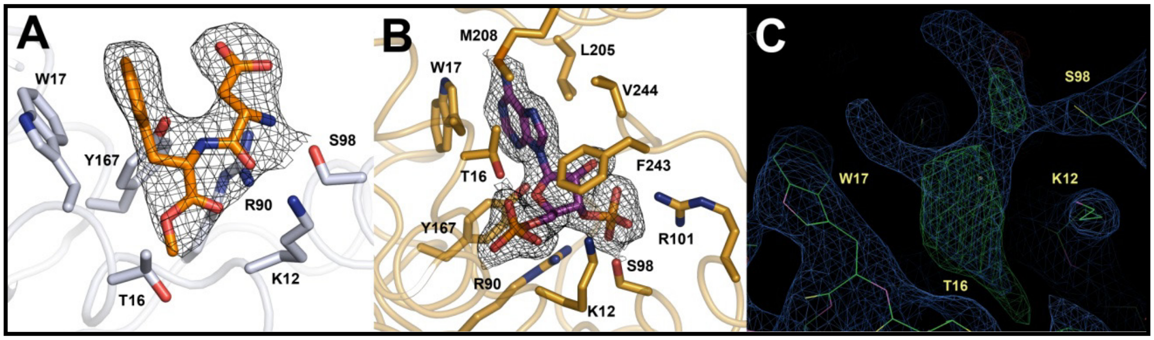


**Figure 6.4)** Comparison of the ligands bound in the PAPS binding sites from Teg12-Apo and Teg12-ternary. Electron density from the $2|F_o| - |F_c|$ map surrounding each ligand and contoured at $1\sigma$ is shown. (A) The PAPS binding site of monomer A from Teg12-Apo contained a molecule of the dipeptide aspartame (N-(L-a-aspartyl)-L-phenylalanine,1-methyl ester). Several Teg12 residues important for coordinating PAP also make contacts with aspartame. Their side chains are shown. (B) PAP bound in the Teg12-ternary complex. Leu205 and Met208 from the GHL helix and Phe243 and Val244 from the V3 loop make significant Van der Waals contacts with PAP. (C) Electron density in the PAPS binding site of the Teg12-apo monomer A, before the addition of aspartame to the model is shown. Both the $2|F_o| - |F_c|$ (contoured at $1\sigma$) (blue) and the $|F_o| - |F_c|$ (contoured at $3\sigma$) (green) maps are shown. Aspartame was added to the model only after the protein had been fully built, at the second to last refinement step (before the addition of waters).

## *Teg12 Co-crystal Structures*

The TEG gene cluster is predicted to encode the biosynthesis of a heptapeptide (hydroxyphenylglycine (Hpg)-betahydroxytyrosine (Bht)-dihydroxyphenylglycine (Dpg)-Hpg-Hpg-Bht-Dpg) that is oxidatively cross-linked into the four macrocycles seen in teicoplanin-like glycopeptides (Figure 6.1). This TEG derived, oxidatively cross-linked heptapeptide skeleton only differs from the teicoplanin aglycone by the substitution of Bht for Tyr at the 2nd position in the peptide core. All three TEG sulfotransferases can use the teicoplanin aglycone as a substrate. We therefore used this molecule in Teg12 co-crystallization experiments. The $V_{max}$ and $K_m$ values for the teicoplanin aglycone were determined to be 215.3 ± 25.2 nmol/min/mg and 59.6 ± 1.1 μM, respectively. Attempts to obtain co-crystals by soaking Teg12 apo crystals with either the teicoplanin aglycone or the co-substrate PAP were unsuccessful. The addition of PAP to drops containing apo crystals caused the crystals to rapidly dissolve into the mother liquor. Therefore, we co-crystallized Teg12 and the teicoplanin aglycone with and without PAP. PAP is the desulfated byproduct of PAPS from the sulfonation reaction. Our co-crystallization experiments led to a Teg12 binary structure bound to teicoplanin, and a ternary structure bound to PAP and teicoplanin. These structures represent the first examples of sulfotransferase structures containing glycopeptide substrates.

## *Teg12 Ternary Structure*

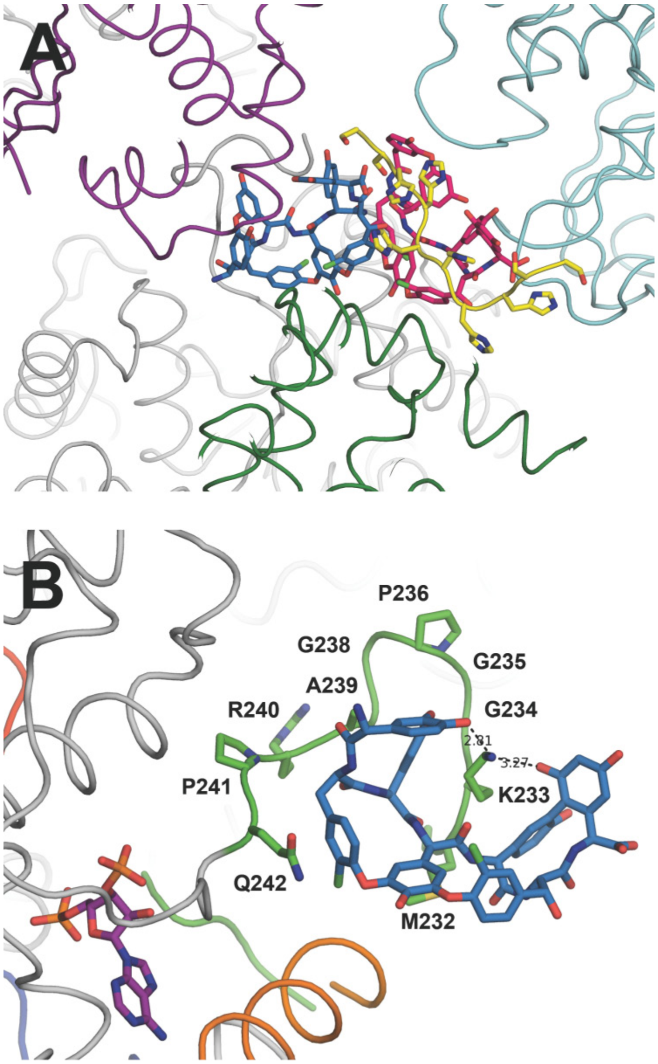
We obtained a 2.05 Å structure of a Teg12-PAP-teicoplanin aglycone ternary complex by co-crystallization (Figure 6.2B and 6.2 bottom panel). Several regions of the Teg12 sequence not seen in the other Teg12 structures could be modeled into the electron density. Loop region Gly127 to Val137, which is primarily disordered in the other Teg12 structures, is fully ordered in one of the monomers of the ternary complex. This short loop contains the V2 region, one of the three regions that are highly variable within the three TEG sulfotransferase sequences. The V2 loop in Teg12 is extended by four amino acids compared with StaL, and by two amino acids when compared with Teg13 and Teg14. A much larger portion of the GHL helix and loop could be modeled in the ternary structure. The GHL helix extends through Ser15 in one monomer. In the opposite monomer the helix extends to Gln218 and transitions uninterrupted to the GHL loop.

Both monomers of the ternary structure contain a single molecule of PAP bound in the predicted active site location. Three highly conserved structural motifs in sulfotransferases are known to play a role in binding PAPS, all of which are seen in Teg12. Lys12 from the 5'-phosphate binding loop (PSB) (Pro11-Thr16) hydrogen bonds with the 5'-phosphate of PAP, Ser98 from the 3'-phosphate binding loop (PB) (Val89-Ser99) hydrogen bonds with the 3'-phosphate of PAP, and Trp17 parallel stacks with the adenine base. A number of additional residues from the PSB loop region, the PB loop region and the conformationally flexible GHL region are within hydrogen bonding distance of PAP (Figure 6.4B). Thr16 is only ~2.6 Å from the 5' phosphate, to which it hydrogen bonds. This residue corresponds to a histidine in StaL, although threonine is the more common side chain at this position in eukaryotic sulfotransferases (Shi et al., 2007). A general hydrogen bond donor at position 16 is likely to be important for properly coordinating PAPS in the active site. Indeed, an alanine replacement mutant of Thr16 was inactive for the sulfation reaction. A K12A mutant was also inactive, while S98A was partially active (27% compared with wild-type). A W17A mutant did not express as soluble protein, indicating a role not only in binding PAP, but also in protein folding and stability.

PAPS dependent sulfotransferases are believed to use a histidine as a general base to activate the hydroxyl or amine used in the sulfate exchange reaction. Teg12 contains three histidine residues, His67, His226 and His269. Of these, only His67 is located near the co-substrate binding site. The 5'-phosphate of PAP is ~5 Å from the imidazole of His67, and when PAPS is modeled in its place, the sulfate is positioned directly adjacent to His67 (Figure 6.7). In this model, both Lys12 and Lys65 are close enough to coordinate with the sulfate. In place of the sulfate from PAPS, a water molecule occupies this space in the ternary structure. Ser9, which is within hydrogen bonding distance of the imidazole ring, was proposed to help activate the catalytic histidine in StaL. Attempts to mutate the predicted catalytic histidine in StaL resulted in an insoluble protein and therefore its role in catalysis had not been previously confirmed (Shi et al., 2007). Alanine exchange mutants were generated for His67, Ser9, and Lys65. H67A and S9A did not express as soluble proteins, while K65A was partially active (65%). Additional glutamate and glutamine mutants of His67 were generated to potentially provide better expression. H67E also expressed as an insoluble protein. However, H67Q, which was soluble, produced a completely inactive sulfotransferase.

### *Teicoplanin Aglycone Bound in the Ternary Structure*

One of the monomers in the ternary structure, in addition to PAP, contains two molecules of teicoplanin aglycone. Both molecules of the aglycone are located outside of the predicted active site pocket, and appear to be organized primarily by crystallographic



forces (Figure 6.5A). One of the molecules of aglycone makes numerous contacts with the GHL loop (Figure 6.5B). Beginning with the N-terminal section of the GHL loop, residues Gly220 through Ile224 fold back towards the dimer interface. Pro223 stacks with Trp134 and helps to stabilize this region. The subsequent six amino acids (Arg 225-Arg230) are disordered, before density picks up again at Met231. However, these residues occupy a region of space where they could easily make

**Figure 6.5)** (A) Two molecules of teicoplanin aglycone were modeled in the Teg12-ternary complex, colored sky blue and hot pink. Their location within the crystal appears to mediate the packing of several Teg12 monomers, as shown in purple, cyan, dark green, and grey. The two teicoplanins belong to the grey Teg12 in the crystal. Also shown in yellow is the $His_6$ portion of the N-terminal derived tag from the pET28a expression vector. The tag-aglycone interaction is proposed to mimic the binding of the aglycone to its cellular target, D-Ala-D-Ala of the bacterial cell wall. (B) A close-up of the GHL loop-aglycone complex from the Teg12-ternary structure. Side chains for all alanine mutants are shown. Lys233 intercalates into the glycopeptide and makes hydrogen bond contacts with residues 1 and 7 of the aglycone. For orientation purposes, the molecule of PAP is shown in purple. Also shown in orange is the GHL helix.

additional contacts with the glycopeptides. Lys233 through Gln242 closely trace the outer surface of the N-terminal three amino acids of the aglycone. These ten amino acids immediately precede the five residues (Phe243-Gly247) that interact with PAP. Lys233 intercalates between the Dpg at position 3 and the Hpg at position 5, and it hydrogen bonds with hydroxyls from the Hpg at position 1 and the Dpg at position 7. The carboxylate from Asp271 of a symmetry mate makes several contacts with backbone nitrogens in the concave cleft of the glycopeptides.

The carboxylate from Asp271 of a symmetry mate makes several contacts with backbone nitrogens in the concave cleft of the glycopeptides. The C-terminal end of the teicoplanin aglycone extends out from the surface of the protein and interacts with a second glycopeptide molecule that is bound to the His$_6$-tag from an adjacent monomer. Despite the assumption that these aglycone-GHL interactions are mediated primarily by crystal packing, we generated a series of mutants in GHL loop residues that contact the glycopeptides. All mutants in this loop region, with the exception of G235A, had deleterious effects on the rate of substrate conversion, supporting the notion that the V3 loop interacts with the substrate and is involved in recruitment. The P241A mutation abolished Teg12's activity completely. While the observed glycopeptide-GHL interactions observed in the ternary structure are clearly stabilized by crystal packing interactions, our mutagenesis experiments make it so we cannot rule out the possibility that this represents a snapshot of the substrate recruitment path. A summary of all mutations made in Teg12 and their corresponding activities is given in Table 6.3.

### *Teg12 Binary Structure and the Teg12 Active Site*

We solved a Teg12-teicoplanin aglycone structure to 2.27 Å resolution (Figure 6.2C). A single monomer constitutes the asymmetric unit and the dimer is generated by crystallographic symmetry. The active site contains a single molecule of teicoplanin aglycone. Two sets of side chain hydrogen bonds play key roles in creating the glycopeptide binding cavity. The base of the binding pocket is largely defined by the outstretched side chain of Arg107 coordinating with the side chain of Gln251. The back of the cavity is largely defined by two outstretched arginine side chains (Arg214 and Arg248) that interact with each other via a mediating water molecule. Numerous hydrogen bonds exist between amino acid side chains from residues found in the GHL helix (Glu206, Arg207, Glu210 and Lys213) and the glycopeptide (Figure 6.6). Several of the residues that make contacts with PAP also interact with the aglycone (Lys12, Arg90, Arg101, and Tyr167). Side chains from Tyr167 and Lys12 coordinate the C-terminal carboxylate, while a backbone carbonyl oxygen from Lys213 and water hydrogen bonded to Glu216 coordinate the N-terminal amine.

The GHL helix and the helix immediately N-terminal (residues Ile193 to Ser203) form a single contiguous helix that runs along the top surface of the aglycone (Figure 6.2C). In the ternary structure with an empty active site pocket, these two helices adopt a bent conformation that results from unwinding of the long helix at Thr204 (compare orange GHL helix, Figures 6.2B and 6.2C). This bent conformation is also observed in StaL. In the bent conformation, the helices are at a nearly perpendicular angle to one another. The long, largely disordered loop region of the GHL that interacts with PAP in the ternary structure is completely displaced by the bound glycopeptide in the binary structure. This disordered loop is now positioned at the back of Teg12, behind the bound glycopeptide. The front surface of the glycopeptide, where the sulfotransferase reaction is predicted to occur on residue 3 (see red arrow in Figure 6.2C), is largely exposed to solvent. With the exception of an additional small compensatory movement of a short loop (Gln31 through Ile37) into the void left by the straightening of the flexible helix from the GHL, the remainder of this Teg12 binary structure is essentially identical to that seen in the other Teg12 structures.

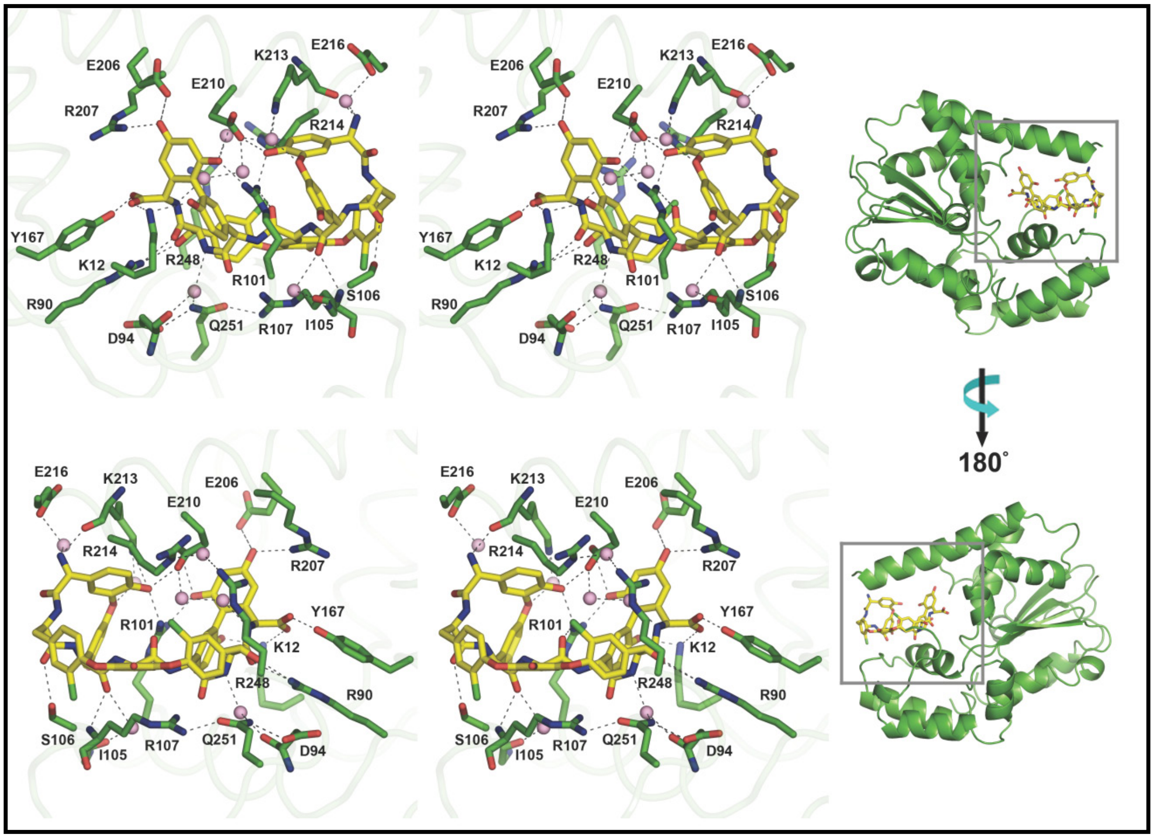


**Figure 6.6)** Stereo views of the front and back sides of the active site of the Teg12-binary structure. Protein residues are colored green, and the aglycone is colored yellow. Hydrogen bonds are indicated with dashed lines. Water molecules that make important hydrogen bonds are shown as pink spheres.

The C-terminus of the glycopeptide is buried in the PSB loop from the strand-loop-helix motif and the helix from the PB loop containing the strand-turn-helix motif runs along the under side of the glycopeptide. While most residues only contact the outer surface of the glycopeptide, one residue, Arg101, extends into the core of the molecule. The side chain of Arg101 intercalates between the Dpg at position 3 and the Hpg at position 5. The guanidinium from Arg101 hydrogen bonds with the backbone carbonyl from the Hpg at position 4 as well as with hydroxyls from the Dpg at position 3 and the Dpg at position 7. In this respect, Arg101 behaves in much the same way as Lys233 in the ternary structure. We generated a series of alanine replacement mutants for residues contacting the aglycone to determine which are important for substrate conversion. The results are summarized in Table 6.3. R101A showed a reduction of activity to 27% compared with wild-type. Interestingly, an E212A mutation also showed a reduction of activity to 26%. Glu212 is part of the GHL helix, but does not make contacts with the aglycone in the binary structure. Instead, this residue is found on the upper, solvent exposed face of the helix. Arg248, along with Arg214 forms the back of the active site cavity. An alanine placed at position 248 resulted in a two-fold increase in sulfation. Removing the bulky side chain at this position may impart more flexibility on the active site, allowing the glycopeptide to shift more easily towards His67 (see discussion).

Although the glycopeptide substrate interacts with different sets of GHL residues in the binary and ternary structures, the general glycopeptide binding motif seen in these structures is very similar. In both cases, a long positively charged amino acid side chain (Lys233 in the ternary, Arg101 in the binary) extends into the core of the glycopeptide, where it coordinates with the same set of side chain hydroxyls. All other glycopeptide-protein contacts observed in these structures involve the outer surface of the glycopeptide.

In addition to the glycopeptide bound in the active site cavity, there is a second glycopeptide molecule bound on the surface of the protein (not shown in Figure 6.2C). The location of the second teicoplanin aglycone in the crystal is likely not biologically relevant. Instead, it appears to mediate crystal contacts between symmetry partners. A stretch of electron density, large enough to contain four residues, is located in the D-ala-D-ala binding cleft of the second aglycone. Although it is presumed that this stretch of residues comes from the N-terminal vector derived $His_6$-tag of an adjacent symmetry mate, we were not able to determine its sequence, and hence it was modeled as polyalanine.

## Discussion

Teg12 sulfates the hydroxyl on the dihydroxyphenyl glycine at position 3 of the teicoplanin aglycone (Figure 6.2C, red arrow). In an alignment of the binary and ternary structures, this hydroxyl is 16.4 Å from where the sulfate of PAPS would be and there is significant overlap between the C-terminus of the glycopeptide and PAP (Figure 6.7). Before the sulfonation reaction can proceed, there must be a substantial conformational rearrangement in the active site. In the binary structure, the interaction of the glycopeptide with Arg 101 appears to preclude the substrate from shifting to accommodate PAPS or from its shifting into the vicinity of His67. Arg101 makes

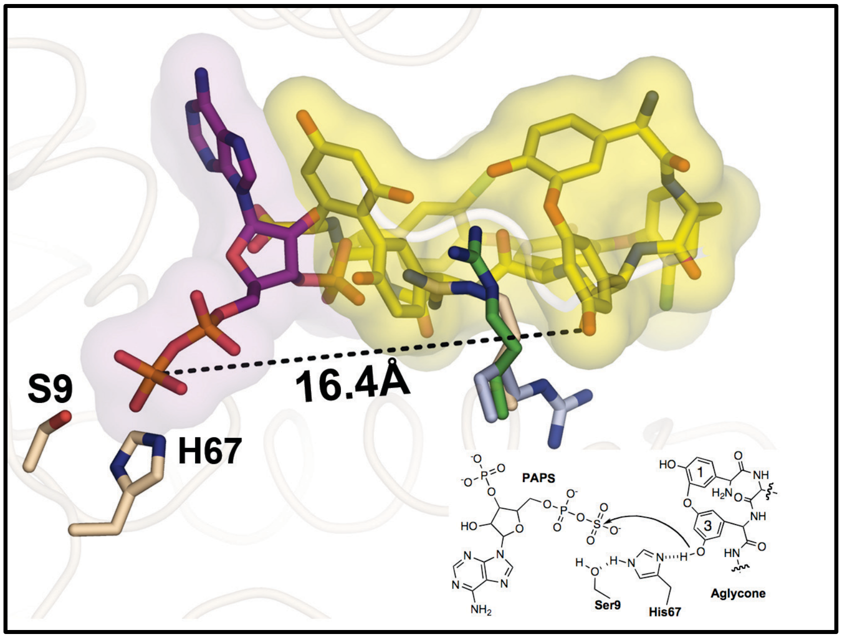


**Figure 6.7)** A composite of the active site from the Teg12 binary and ternary structures. PAPS was modeled into the ternary structure in place of PAP. In this aligned view, PAPS and teicoplanin aglycone would not be able to occupy their respective positions simultaneously, as there would be strong clashes between the two substrates. Residue 3 of the teicoplanin aglycone, the site of sulfation, is more than 16 angstroms from the sulfate of PAPS when the ternary and binary structures are overlapped. A substantial rearrangement of teicoplanin aglycone within the active site would have to occur in order for sulfation to proceed by the proposed in-line attack mechanism. The positions of Arg101 in the binary (green) and ternary (tan) structures appear to preclude the movement of teicoplanin aglycone towards His67. Arg101's position in the context of the apo structure (grey) creates a more open active site, where the aglycone could pivot towards His67 more easily.

mutually exclusive contacts with the substrates present in the binary and ternary structures. In the binary structure it is inserted into the peptide core and hydrogen bonds with residues 1 and 7, while in the ternary structure it forms a key salt bridge with the 3-phosphate of PAP. If, when both substrates are present in the active site, Arg101 were to adopt the conformation seen in the apo structure, the glycopeptide would largely be free to move out of the PAPS binding pocket and towards His67. By pivoting towards His67, the glycopeptide would occupy the same general location that substrates bind in eukaryotic sulfotransferases (Allali-Hassani et al., 2007), and it would also be positioned so that it could interact with all three variable loops.

The movement the glycopeptide must undergo to reach the PAPS sulfate traces a path similar to that which the GHL helix undergoes when it bends. This conformational rearrangement may serve not only as a mechanism to allow large glycopeptide substrates into the active site, but it could also play a role in positioning a glycopeptide for catalysis once it enters the active site. Alanine exchange mutagenesis of residues found on both the lower glycopeptide associated face (E206A, R207A, and R214A) and the upper solvent exposed face of the helix (E212A) led to decreases in substrate conversion. Residues around the entire helix likely make key contacts with the glycopeptide at different points in time as a result of the rotation in the helix as it bends. In fact, all residues that contact the glycopeptide in the binary structure face out from the active site cavity in the ternary structure. One alanine exchange mutant, R248A, resulted in a two-fold increase in activity. R248, along with R214 of the GHL helix, and a mediating water molecule, forms a cage-like enclosure over the top of the glycopeptide (Figure 6.6). Introduction of an alanine side chain at this position would open up the back side of the active site to the solvent, and could provide substrates in the active site more room for positional rearrangement. The sequence of the GHL helix is conserved throughout the TEG sulfotransferases and a large portion of it is also conserved in StaL. The mechanics of the conformational change from the straight to bent structures may therefore be conserved throughout this family of finishing enzymes.

Sequence and structural similarities between eukaryotic and prokaryotic sulfotransferases indicate that all members have likely arisen from one common ancestor. In eukaryotic structures the loop that would correspond to the V3 loop of the GHL behaves in much the same way as it does in Teg12; it is disordered in apo structures and associated with PAP and the substrate in co-crystal structures (Allali-Hassani et al., 2007). The open active site conformation that results from the straightening of the GHL helix was not observed in previous StaL structures. While a GHL-like helix-loop region is present in eukaryotic sulfotransferases, the helix from these structures does not appear to exhibit the same conformational rearrangements from a single straight helix to two helices with a bent conformation during substrate binding as it does in Teg12. At over 1000 Daltons, glycopeptides are significantly larger than the substrates used by most eukaryotic sulfotransferases. The increased conformational plasticity that results from the ability of the GHL helix to easily flex may help TEG-like sulfotransferases accommodate larger substrates.

In addition to the conformational flexibility seen in the V3 loop, both the V1 and V2 loops adopt different conformations in this series of structures, suggesting they could easily reorganize to accommodate an incoming glycopeptide substrate. The sequence differences seen between Teg12, 13, and 14 suggest that these three short variable regions likely control the selection and orientation of the glycopeptide substrate bound in the active site. Systematically altering the residues found in the three TEG variable regions may provide a means to generate new glycopeptide finishing enzymes that sulfate a broader collection of glycopeptide congeners than is currently possible with the small number of sulfotransferases that have been identified naturally. The sulfated teicoplanin aglycone derivatives produced by the native TEG sulfotransferase retain potent *in vitro* antibacterial activity (Banik and Brady, 2008).

The cloning and characterization of biosynthetic gene clusters derived from uncultured bacteria provides a means to access both novel small molecules and new biosynthetic enzymes. Teg12 is one of the first enzymes discovered using culture independent methodologies to be characterized structurally, and the Teg12-teicoplanin aglycone co-crystal structures are the first examples of a substrate complexed with a member of this family of glycopeptide finishing enzymes. This series of Teg12 structures provides key insights into how sulfotransferases might be engineered to generate additional anionic glycopeptides that could be evaluated against clinically relevant drug resistant bacteria.

# Materials and Methods

## *Teg12 Expression and Purification*

Teg12 was cloned and expressed as previously described (Banik and Brady, 2008). Briefly, teg12 was amplified (30 cycles of 95 °C for 30 s, 60 °C for 30 s, and 72 °C for 90 s; FailSafe system from Epicentre) from eDNA cosmid clone D30 using the following primers: Teg12FWD(BclI):GCGCTGATCAATGAACGGAATTCGATGG, Teg12REV(HindIII):GCGCAAGCTTTCCTTAACCGGCATACCCGTA. Restriction enzyme sites used for cloning are shown in bold. The resulting product was doubly digested with BclI and HindIII and subsequently ligated into pET28a, which had been BamHI/HindIII doubly digested. The resulting construct was then transformed into E. coli BL21(DE3) for protein expression. Expression cultures were grown to $OD_{600}$=0.6, followed by IPTG induction, and overnight growth at 20 °C. The culture was pelleted by centrifugation (3,200 x g for 30 min), the supernatant was discarded, and the cell pellet was resuspended in 40 mL lysis buffer (50 mM HEPES, pH 7.5, 0.5 M NaCl, 5% (vol/vol) glycerol, 20 mM imidazole, pH 8, 10 mM β-mercaptoethanol and 0.5% (vol/vol) Triton X-100). The resuspended cell pellet was lysed by sonication, and the insoluble portion was removed by centrifugation (15,000 x g for 30 min). The cleared cell lysate was incubated with 1 mL Ni-NTA resin for 15 min. The slurry was loaded onto a column, allowed to empty by gravity flow, washed with 40 mL lysis buffer, and finally washed with 40 mL wash buffer (50 mM HEPES, pH 7.5, 0.5 M NaCl, 5% (vol/vol) glycerol, 20 mM imidazole, pH 8.0, and 10 mM β-mercaptoethanol). The protein was eluted by the addition of 15 mL of elution buffer (50 mM HEPES, pH 7.5, 0.5 M NaCl, 5% (vol/vol) glycerol, 125 mM imidazole, pH 8, and 10 mM β-mercaptoethanol). No attempt was made to remove the vector derived His$_6$-tag, resulting in a Teg12 protein plus 34 additional residues N-terminal to the start methionine. Protein was concentrated using Vivascience Vivaspin 30,000 MWCO ultrafiltration concentrators, and was buffer exchanged 3 times into protein buffer (200 mM NaCl, 20 mM HEPES, pH 7.5, 5% glycerol, and 1 mM DTT).

### *Teg12 Crystallization*

Concentrated protein was centrifuged at 14,000 rpm (4 °C, 30 min) in a microcentrifuge to remove any insoluble material prior to crystallization. All crystals were grown using the hanging drop vapor diffusion method. Initial Teg12-apo crystals were obtained by mixing 1 µl of protein (7.5 mg/ml in protein buffer) with 1 µl of reservoir solution (1.0 M sodium citrate, 0.1 M sodium cacodylate, pH 6.5, JCSG core III-48, Qiagen) over a 500 µl reservoir. Blade-like crystals grew overnight at 22 °C and reached a maximal size of 400 µm X 50 µm X 20 µm in approximately one week. To improve crystal thickness, Teg12-apo crystals were optimized by microseeding, in addition to mixing 1 µl of protein at 7.5 mg/ml with 0.5 µl of reservoir and 0.5 µl of Silver Bullets™ reagent 29 (Hampton Research). A component of the Silver Bullets™ screen, aspartame, was modeled into the PAPS binding site of one of the monomers of the Teg12 dimer. Crystals were soaked in a 20 µl drop containing reservoir solution plus 10% ethylene glycol. The drop was allowed to dehydrate by exposure to open air at room temperature for approximately 5 hours before flash cooling the crystals in liquid ethane.

Teg12-ternary crystals were co-crystallized at 4 °C in the presence of 2 mM PAP and 1 mM teicoplanin aglycone. Teg12 was first concentrated to 20 mg/ml in protein buffer. The protein was then diluted 1:1 with 50 mM CHES, pH 9.1, 2 mM teicoplanin aglycone, 4 mM PAP, achieving a final concentration of 10 mg/ml Teg12 in 0.5X protein buffer, 25 mM CHES, pH 9.1, 1 mM teicoplanin aglycone, 2 mM PAP. 1 μl of protein was mixed 1:1 with reservoir solution (0.2 M ammonium acetate and 20% w/v PEG 3350, JCSG, core I-25, Qiagen). Crystals appeared in 2-3 days and grew to a maximal size of 100 μm X 50 μm X 50 μm in approximately 1 week. These crystals were of an irregular chunk-like morphology and had cracks throughout. Crystals were cryo-protected by quickly dunking in reservoir solution plus 15% ethylene glycol and were flash cooled in liquid nitrogen.

Teg12-binary crystals were co-crystallized at 4 °C in the presence of 1 mM teicoplanin aglycone. Similar to Teg12-ternary crystallization, the protein was first concentrated to 20 mg/ml in protein buffer, then diluted to 10 mg/ml with 50 mM CHES, pH 9.1, 2 mM aglycone (final 0.5X protein buffer, 25 mM CHES, 1 mM teicoplanin aglycone). 1 μl of protein solution was mixed with 1 μl of reservoir solution (2.0 M sodium formate, 0.1 M sodium acetate, pH 4.6, JCSG core III-85, Qiagen). Cubic crystals grew between 2 and 3 weeks, and were approximately 50 μm X 50 μm X 50 μm. Crystals were soaked in 6.0 M sodium formate, 0.1 M sodium acetate, pH 4.6, 1 mM teicoplanin aglycone overnight, prior to flash cooling in liquid nitrogen.

### *Data Collection and Structure Solving*

All data sets were reduced and scaled using the HKL2000 package (Otwinowski and Minor, 1997). Data for Teg12-apo crystals were collected at the NSLS, beamline X29A. All but one of the crystals screened diffracted poorly to approximately 4 Å resolution. The crystal from which the 2.91 Å dataset was collected rotated briefly out of the cryostream, and thereby had gone through a room temperature annealing cycle of several seconds. Diffraction from this crystal was dramatically improved compared with other crystals taken from the same drop. Data for Teg12-apo was reduced and scaled in space group $C222_1$. Phase information was obtained by molecular replacement using the program Phaser and StaL (GenBank accession number AAM80529, PDB code 2OV8), devoid of all flexible loops, as the search model (McCoy et al., 2007). The initial molecular replacement model was refined against the Teg12-apo dataset using rigid body refinement in Refmac (Murshudov et al., 1997). Additional features of the map were enhanced through density modification, and 2-fold ncs averaging in CNS (Brunger, 2007; Brunger et al., 1998). The model was rebuilt manually using the program Coot (Emsley and Cowtan, 2004). Full restrained refinement was carried out using the translation/libration/screw model in Refmac, with the addition of hydrogen atoms, converging to a final $R_{work}$ and $R_{free}$ of 21.96 and 27.19, respectively (Winn et al., 2001). NCS restraints were not used during refinement. The final model comprises residues 1-129, 136-203, and 251-285 for monomer A, and 1-27, 42-128, 137-210, 247-285 for monomer B. The Teg12-apo model was used as a molecular replacement model for all subsequent structures.

Teg12-ternary and Teg12-binary data sets were collected at the APS, microfocus beamline 24-IDE. Teg12-ternary data was reduced and scaled in space group $P2_12_12_1$. Teg12-binary was scaled in space group $I2_12_12_1$. The crystal structure of glycopeptide aglycone A-40926 was used as a starting point to generate a restraint definition file for teicoplanin aglycone using the program Phenix Elbow (Schafer et al., 1996). Geometry optimization was achieved using the semi-empirical quantum mechanical AM1 method. Teg12-binary and Teg12-ternary models were refined using the translation/ libration/screw model in Phenix Refine to a final $R_{work}$ and $R_{free}$ of 17.30 and 22.61, and 17.12 and 22.47, respectively (Adams et al., 2002). The final Teg12-binary model comprises residues 1-129, 135-216, and 247-285. The final Teg12-ternary model comprises residues 1-215, 220-224, and 231-285 for monomer A, and residues 1-129, 136-224, and 240-285. All structures were validated using the Molprobity server from the Richardson laboratory at Duke University (Davis et al., 2007).

### *Site-Directed Mutagenesis*

Teg12 point mutants were generated using the "megaprimer" method, with slight modifications (Sarkar and Sommer, 1990). Oligonucleotide primers were designed for each mutant (Table 6.2), and a megaprimer was generated by PCR amplification from the Teg12/pET28a construct using the Pfx Accuprime System (Invitrogen), the relevant mutant oligonucleotide primer, and either the T7 promoter (for mutations at residues 9-108) or the T7 terminator (for mutations at residues 167-251) as the second oligonucleotide primer, (30 rounds of amplification: 95 °C for 30 s, 55°C for 30 s, 68 °C for 30 s). The full length mutant Teg12 gene was amplified from the Teg12/pET28a construct, using the megaprimer, which then contained the bases that code for the specific mutant residue, and either the T7 terminator (for mutations at residues 9-108) or the T7 promoter (for mutations at residues 167-251) as the second oligonucleotide primer (30 rounds of amplification: 95 °C for 30 s, 55 °C for 30 s, 68 °C for 80 s). Full-length mutant amplicons were then sequentially digested with BamHI and HindIII, and subsequently ligated into BamHI/HindIII doubly digested pET28a. Ligated constructs were transformed into E. coli EC100 (Epicentre), and sequenced to identify successfully mutated constructs. Mutant constructs containing the desired point mutation were then transformed into *E. coli* BL21 (DE3) for protein expression.

### *Mutant Teg12 Expression and Purification*

Mutant proteins were expressed and purified in a manner similar to the native Teg12, except on a reduced scale. 100 mL overnight expression cultures were pelleted and resuspended in 4 mL lysis buffer. After sonication to lyse the cells, the crude lysates were centrifuged to remove insoluble material (10 min at 15,000 x g). The cleared lysates were incubated with 100 μl Ni-NTA resin for 15 min. The slurry was then loaded onto a column, allowed to empty by gravity flow, washed with 4 mL lysis buffer, followed by a second wash with 4 mL wash buffer. The protein was eluted by the addition of 1.5 mL elution buffer. All Teg12 mutants used in activity assays appeared to be homogeneous by polyacrylamide gel electrophoresis.

### *Teg12 Activity Assays*

All soluble Teg12 mutants were assayed for activity using the teicoplanin aglycone as a substrate. 50 µL reactions were run in duplicate, as follows: 15 mM HEPES, pH 7.5, 1 mM 3'-phosphoadenosine-5'-phosphosulfate (PAPS), 0.1 mM DTT, 1.2 mM teicoplanin aglycone (in DMSO), and 500 ng purified protein in elution buffer. Reactions were carried out at 30 °C for each of the four time points (10, 15, 20, 25 min), followed by heat inactivation at 99 °C for 10 min, and a further 10 min in an ice water bath. $V_{max}$ and $K_m$ values were determined under the same reaction conditions using the teicoplanin aglycone as substrate (5 mM to 100 mM). 25 µL of each reaction was run on a Waters analytical HPLC system (C18 (4.6 x 150 mm)). A linear gradient (1.5 ml/min) was run from an initial condition of 95:5 20 mM ammonium acetate:acetonitrile to 70:30 20 mM ammonium acetate:acetonitrile over twenty minutes. The area under the UV peak (Diode Array, 240 nm-400 nm) was determined for both the monosulfated product and the teicoplanin aglycone substrate at each time point. The percent substrate conversion for duplicate time points was averaged. The slope of the graph derived from the four time points for Teg12 and each mutant was then determined. Relative activity of each mutant is reported as a percent of the slope for wild-type Teg12.

Table 6.1

| | Teg12 apo | Teg12 ternary | Teg12 binary |
|---|---|---|---|
| **Data collection** | | | |
| Space group | $C222_1$ | $P2_12_12_1$ | $I2_12_12_1$ |
| Cell dimensions | | | |
| a, b, c (Å) | 79.39, 126.09, 145.13 | 77.12, 78.56, 100.54 | 66.12, 80.15, 132.94 |
| α, β, γ (º) | 90, 90, 90 | 90, 90, 90 | 90, 90, 90 |
| No. reflections (observed/unique) | 55,339/15,372 | 177,647/38,619 | 66,588/16,647 |
| Resolution (Å) | 50.00 – 2.93<br>(3.03 – 2.93) | 50.00 – 2.05<br>(2.12 – 2.05) | 50.00 – 2.28<br>(2.36 – 2.28) |
| $R_{merge}/R_{sym}$ | 0.063 (0.396) | 0.084 (0.464) | 0.079 (0.542) |
| I/σI | 19.10 (2.04) | 16.43 (2.62) | 15.50 (2.17) |
| Completeness (%) | 95.5 (98.0) | 97.7 (96.6) | 99.9 (100.0) |
| Redundancy | 3.6 (3.6) | 4.6 (4.6) | 4.0 (4.1) |
| Refinement | | | |
| Resolution (Å) | 72.57 – 2.91 | 32.95 – 2.04 | 34.32 – 2.27 |
| No. reflections | 14,562 | 36,951 | 15,761 |
| $R_{work}$ / $R_{free}$ | 21.96 / 27.19 | 17.12 / 22.47 | 17.30 / 22.61 |
| No. atoms | 3,628 | 4,820 | 2,222 |
| Protein atoms | 3,550 | 4,247 | 1,945 |
| Ligand/ion | 44 | 233 | 170 |
| Waters | 34 | 340 | 95 |
| *B* factors | 36.2 | 35.7 | 43.0 |
| Protein | 35.9 | 35.6 | 40.3 |
| Ligand/ion | 61.7 | 30.1 | 69.5 |
| Water | 39.1 | 40.5 | 47.4 |
| Solvent Content | 52.4 | 40.9 | 46.5 |
| Rms bond (Å) | 0.014 | 0.007 | 0.006 |
| Rms angle (°) | 1.495 | 1.183 | 1.038 |
| Rotamer Outlier (%) | 3.6 | 0.7 | 1.0 |
| Ramachandran (favored/outlier) (%) | 94.6 / 0.6 | 98.7 / 0.0 | 100.0 / 0.0 |
| PDB code | 3MGC | 3MGB | 3MG9 |

**Table 6.1)** Data collection and refinement statistics for the Teg12 structures.

Table 6.2

| Mutant | Primer (5’-3’) |
|---|---|
| S9A | GTTTCCAGCCTTTGGATACGCTGCGATCCATCGAATTCC |
| K12A | CCACGTGTTTCCAGCCGCTGGATACGATGC |
| T16A | CACCTGACCCACGCGTTTCCAGCCTTTGG |
| W17A | CAACATGCACCTGACCGCCGTGTTTCCAGC |
| K65A | CATCGGCCTTGAGGTGCGTCGCCACCAGCACCGGTTC |
| H67A | CACATCGGCCTTGAGGGCCGTCTTCACCAG |
| R90A | CATATCCCGCGGGTTCGCCACGAGATAGAG |
| S98A | GGCCATGCGATCGAGGCGAGCAGCATATC |
| R101A | TATCGAGGCCATGGCCATCGAGCTGAG |
| S106A | CTACGTCGTCGCGCGCTATCGAGGCCATGCG |
| R107A | GCTTTTTTCTACGTCGTCGGCCGATATCGAGGC |
| D108A | GCTTTTTTCTACGTCGGCGCGCGATATCGAGGCC |
| Y167A | GTGCTGACGATGCGTGCTGAGGACCTGAAGGGC |
| K171A | CGTTATGAGGACCTGGCGGGCGATCCGGTCGCACGG |
| E206A | GCTGCCTCCACGCTGGCGCGGATGCGTGAACTG |
| R207A | GCTGCCTCCACGCTGGAGGCGATGCGTGAACTG |
| E210A | GAGCGGATGCGTGCACTGGAGAAACGGAG |
| E212A | CGGATGCGTGAACTGGCGAAACGG |
| K213A | CGGATGCGTGAACTGGAGGCACGGAGCGAGCAGCAG |
| R214A | CGGATGCGTGAACTGGAGAAAGCGAGCGAGCAGCAG |
| M232A | GGTGATGCGAGAATGGCGAAAGGG |
| K233A | GATGCGAGAATGATGGCAGGGGG |
| G234A | GCGAGAATGATGAAAGCGGGACC |
| G235A | AGAATGATGAAAGGGGCACCTGGTG |
| P236A | ATGATGAAAGGGGGAGCTGGTGG |
| G238A | AAAGGGGGACCTGGTGCCGCGAGG |
| A239R | AAAGGGGGACCTGGTGGCCGGAGG |
| R240A | GGACCTGGTGGCGCGGCGCCCCAG |
| P241A | CCTGGTGGCGCGAGGGCCCAGTTC |
| Q242A | GGTGGCGCGAGGCCCGCGTTCGTG |
| R248A | CAGTTCGTGGGCGAGGGCGCGTACGACCAGTCCCTG |
| Q251A | GAGGGCAGGTACGACGCGTCCCTGTCCTTCTTG |

**Table 6.2)** Oligonucleotide primers for site-directed mutagenesis.

Table 6.3

| Catalytic or PAP binding | Activity (%) | V3 aglycone binding | Activity (%) | Non-V3 aglycone binding | Activity (%) |
|---|---|---|---|---|---|
| S9A | N/A | M232A | 74 | R101A | 29 |
| K12A | 0 | K233A | 79 | S106A | 92 |
| T16A | 0 | G234A | 82 | R107A | 82 |
| W17A | N/A | G235A | 100 | D108A | 82 |
| K65A | 61 | P236A | 73 | Y167A | N/A |
| H67A | N/A | G238A | 56 | K171A | 69 |
| H67E | N/A | A239R | 37 | E206A | 0 |
| H67Q | 0 | R240A | 48 | R207A | 79 |
| R90A | N/A | P241A | 0 | E210A | 97 |
| S98A | 27 | Q242A | 64 | E212A | 26 |
| | | | | K213A | 99 |
| | | | | R214A | 67 |
| | | | | R248A | 196 |
| | | | | Q251A | 41 |

**Table 6.3)** Relative activity of Teg12 mutants. N/A indicates an insoluble mutant.

# References

**Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Read, R.J., Sacchettini, J.C., Sauter, N.K., and Terwilliger, T.C.** (2002). PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr D Biol Crystallogr *58*, 1948-1954.

**Allali-Hassani, A., Pan, P.W., Dombrovski, L., Najmanovich, R., Tempel, W., Dong, A., Loppnau, P., Martin, F., Thornton, J., Edwards, A.M., *et al.*** (2007). Structural and chemical profiling of the human cytosolic sulfotransferases. PLoS Biol *5*, e97.

**Banik, J.J., and Brady, S.F.** (2008). Cloning and characterization of new glycopeptide gene clusters found in an environmental DNA megalibrary. Proc Natl Acad Sci U S A *105*, 17273-17277.

**Boeck, L.D., and Mertz, F.P.** (1986). A47934, a novel glycopeptide-aglycone antibiotic produced by a strain of Streptomyces toyocaensis taxonomy and fermentation studies. J Antibiot (Tokyo) *39*, 1533-1540.

**Brunger, A.T.** (2007). Version 1.2 of the Crystallography and NMR system. Nat Protoc *2*, 2728-2733.

**Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., *et al.*** (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr *54*, 905-921.

**Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B., 3rd, Snoeyink, J., Richardson, J.S., *et al.*** (2007).

MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res *35*, W375-383.

**Dombrovski, L., Dong, A., Bochkarev, A., and Plotnikov, A.N.** (2006). Crystal structures of human sulfotransferases SULT1B1 and SULT1C1 complexed with the cofactor product adenosine-3'- 5'-diphosphate (PAP). Proteins *64*, 1091-1094.

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

**Higgins, D.L., Chang, R., Debabov, D.V., Leung, J., Wu, T., Krause, K.M., Sandvik, E., Hubbard, J.M., Kaniga, K., Schmidt, D.E., Jr.*, et al.*** (2005). Telavancin, a multifunctional lipoglycopeptide, disrupts both cell wall synthesis and cell membrane integrity in methicillin-resistant Staphylococcus aureus. Antimicrob Agents Chemother *49*, 1127-1134.

**Hiramatsu, K., Aritaka, N., Hanaki, H., Kawasaki, S., Hosoda, Y., Hori, S., Fukuchi, Y., and Kobayashi, I.** (1997). Dissemination in Japanese hospitals of strains of Staphylococcus aureus heterogeneously resistant to vancomycin. Lancet *350*, 1670-1673.

**Hubbard, B.K., and Walsh, C.T.** (2003). Vancomycin assembly: nature's way. Angew Chem Int Ed Engl *42*, 730-765.

**Kahne, D., Leimkuhler, C., Lu, W., and Walsh, C.** (2005). Glycopeptide and lipoglycopeptide antibiotics. Chem Rev *105*, 425-448.

**Lamb, S.S., Patel, T., Koteva, K.P., and Wright, G.D.** (2006). Biosynthesis of sulfated glycopeptide antibiotics by using the sulfotransferase StaL. Chem Biol *13*, 171-181.

**Leadbetter, M.R., Adams, S.M., Bazzini, B., Fatheree, P.R., Karr, D.E., Krause, K.M., Lam, B.M., Linsell, M.S., Nodwell, M.B., Pace, J.L.*, et al.*** (2004). Hydrophobic vancomycin derivatives with improved ADME properties: discovery of telavancin (TD-6424). J Antibiot (Tokyo) *57*, 326-336.

**Lin, E.S., and Yang, Y.S.** (2000). Nucleotide binding and sulfation catalyzed by phenol sulfotransferase. Biochem Biophys Res Commun *271*, 818-822.

**Linden, P.K.** (2008). Vancomycin resistance: are there better glycopeptides coming? Expert Rev Anti Infect Ther *6*, 917-928.

**McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J.** (2007). Phaser crystallographic software. J Appl Crystallogr *40*, 658-674.

**Murshudov, G.N., Vagin, A.A., and Dodson, E.J.** (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr *53*, 240-255.

**Nicolaou, K.C., Boddy, C.N.C., Brase, S., and Winssinger, N.** (1999). Chemistry, biology, and medicine of the glycopeptide antibiotics. Angew Chem Int Ed Engl *38*.

**Otwinowski, Z., and Minor, W.** (1997). Processing of X-ray Diffraction Data Collected in Oscillation Mode. Methods in Enzymology *276A*, 307-326.

**Rappe, M.S., and Giovannoni, S.J.** (2003). The uncultured microbial majority. Annu Rev Microbiol *57*, 369-394.

**Sarkar, G., and Sommer, S.S.** (1990). The "megaprimer" method of site-directed mutagenesis. Biotechniques *8*, 404-407.

**Schafer, M., Schneider, T.R., and Sheldrick, G.M.** (1996). Crystal structure of vancomycin. Structure *4*, 1509-1515.

**Shi, R., Lamb, S.S., Bhat, S., Sulea, T., Wright, G.D., Matte, A., and Cygler, M.** (2007). Crystal structure of StaL, a glycopeptide antibiotic sulfotransferase from Streptomyces toyocaensis. J Biol Chem *282*, 13073-13086.

**Stryjewski, M.E., O'Riordan, W.D., Lau, W.K., Pien, F.D., Dunbar, L.M., Vallee, M., Fowler, V.G., Jr., Chu, V.H., Spencer, E., Barriere, S.L.,** ***et al.*** (2005). Telavancin versus standard therapy for treatment of complicated skin and soft-tissue infections due to gram-positive bacteria. Clin Infect Dis *40*, 1601-1607.

**Uttley, A.H., Collins, C.H., Naidoo, J., and George, R.C.** (1988). Vancomycin-resistant enterococci. Lancet *1*, 57-58.

**Winn, M.D., Isupov, M.N., and Murshudov, G.N.** (2001). Use of TLS parameters to model anisotropic displacements in macromolecular refinement. Acta Crystallogr D Biol Crystallogr *57*, 122-133.

# CHAPTER 7
## Crystal Structure of the Sulfotransferase Teg14

## Introduction

Vancomycin and teicoplanin are glycopeptide antibiotics used clinically to treat many gram-positive bacterial infections, including methicillin resistant *Staphylococcus aureus* (MRSA) (Nicolaou et al., 1999; Wolter et al., 2007). During their biosynthesis each member of this family of antibiotics is functionalized by a unique set of finishing enzymes that includes glycosyl, acyl, methyl and sulfotransferases (Nicolaou et al., 1999; Wolter et al., 2007). The discovery of new glycopeptide congeners has slowed in recent years as it has become increasingly difficult to identify new biodiversity from which novel molecules might be characterized. The vast majority of bacteria present in nature remain recalcitrant to culturing and therefore they represent a potentially novel source of small molecules (Rappe and Giovannoni, 2003). Although metabolites produced by these bacteria cannot be characterized using standard microbiological methods, it is possible to extract DNA directly from environmental samples and analyze this DNA for sequences that encode the biosynthesis of novel natural products (Brady et al., 2009). In a survey of DNA extracted from desert soil, we uncovered a biosynthetic gene cluster (the TEG gene cluster) that is predicted to encode the biosynthesis of the first polysulfated glycopeptide congeners (Banik and Brady, 2008).

The TEG gene cluster contains three closely related sulfotransferases (Teg12, 13, and 14) that sulfate teicoplanin like glycopeptides at three unique sites. Over 150 different glycosylated, halogenated, and alkylated glycopeptide congeners have been characterized from cultured bacteria, yet only three sulfated congeners have been identified from studying these same microbes (Nicolaou et al., 1999). Although anionic glycopeptides have rarely been reported as natural products, the semisynthetic phosphono congener, telavancin, recently received FDA approval (Higgins et al., 2005; Stryjewski et al., 2005). The enzymatic synthesis of anionic glycopeptides could provide a facile means to access additional anionic congeners.

In the previous chapter we presented three crystal structures of the sulfotransferase Teg12 in its apo form, in a binary complex bound to the teicoplanin aglycone, and in a ternary complex bound to teicoplanin aglycone and the co-substrate byproduct, 3'-phosphoadenosine 5'-phosphate (PAP). These studies provided insight into the sulfotransferase mechanism employed by this class of finishing enzymes. In the current chapter we present an apo structure of the third sulfotransferase from the TEG pathway, Teg14. Teg14 displays many of the structural characteristics found in other members of this class of enzymes. We discuss the Teg14 structure itself and make comparisons with other sulfotransferases.

# Results

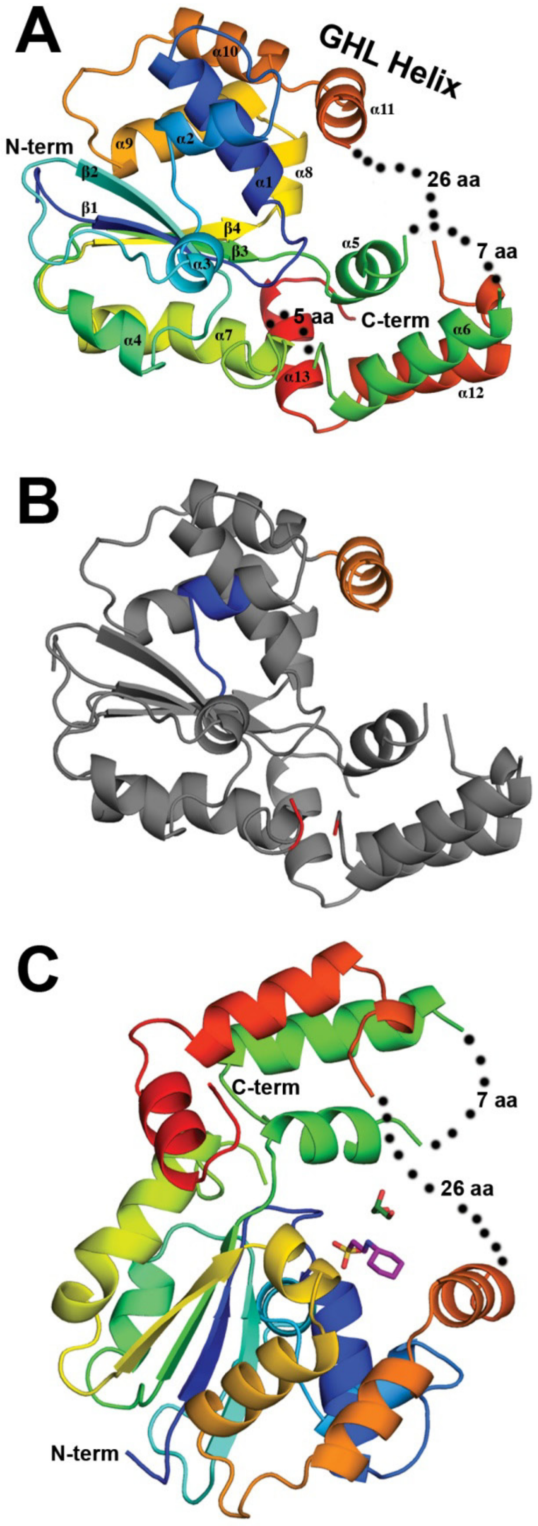
---

## *Overall Structure and Comparison with Teg12 and StaL*

N-terminally $His_6$-tagged Teg14 was affinity purified using nickel NTA resin and crystallized without the need for further purification. The structure was solved to 2.7 Å resolution by molecular replacement using the Teg12 ternary structure as a search model. The overall structure closely resembles that of Teg12 (82.9 % sequence identity) and of StaL (52.2% sequence identity) (for a sequence alignment, see Chapter 6, Figure 6.2). StaL is a related sulfotransferase involved in the biosynthesis of the monosulfated glycopeptide congener A47934 (Shi et al., 2007). Sulfotransferases from this family are biologically active as dimers. Teg14 dimerizes in the same manner as Teg12 and StaL. The Teg14 crystallographic asymmetric unit is comprised of a single monomer; the dimer is reconstituted by crystallographic symmetry. The dimer interface consists of a symmetrical interaction between a short helix-loop motif from one monomer and the same helix-loop motif from the other monomer. As seen in Teg12, there is a key hydrogen bond between the carbonyl oxygen of Gly51 at the end of the helix from one monomer and the backbone nitrogen of Val74 of the other monomer. The side chain of Glu65 from one monomer is also within hydrogen bonding distance of the side chains of Thr66 and Lys65 of the opposite monomer. Additional hydrophobic contacts exist between the two interface helices and help to stabilize the Teg14 dimer. In the Teg14 dimer, the active site cavity of each monomer faces the dimer interface.

**Figure 7.1** (adjacent page) (A) Overview of the 2.7 Å crystal structure of Teg14. The model is colored N- to C-terminally from blue to red. Regions of disorder are represented with grey dots, with the number of disordered residues indicated. (B) The same view as A, but with the protein colored grey in order to accentuate the regions of sequence variability, V1 (blue) and V2 (red). V3, which is encompassed by the 29 amino acid disordered region, was not visible in the structure. Also shown in orange is the GHL helix, which adopts the bent conformation as in the Teg12 ternary and StaL structures. (C) An alternate view of (A). The molecules of CHES and glycerol in the active center are shown as sticks.

As with other sulfotransferases in this family, Teg14 consists of a single α/β globular domain composed of a parallel beta sheet core surrounded by alpha helices (Figure 7.1). The beta sheet core contains four strands. Teg14 contains the three sequence variable loop regions V1 (Met37-Glu43), V2 (Arg127-Gly134), and V3 (Gln215-Pro233) we identified in Teg12. V1, along with V2 and V3, has been hypothesized to play a role in substrate recruitment and orientation of the substrate within the active site cavity. The region Gly28-Glu43, which encompasses V1, adopts three highly divergent conformations in the Teg12 structures, underscoring the conformational flexibility of these residues. However, the observed conformation of this region in the current Teg14 structure, where a greater proportion of the region is helical (Trp34-Val40), may be preferred, as it is also observed in the Teg12 ternary



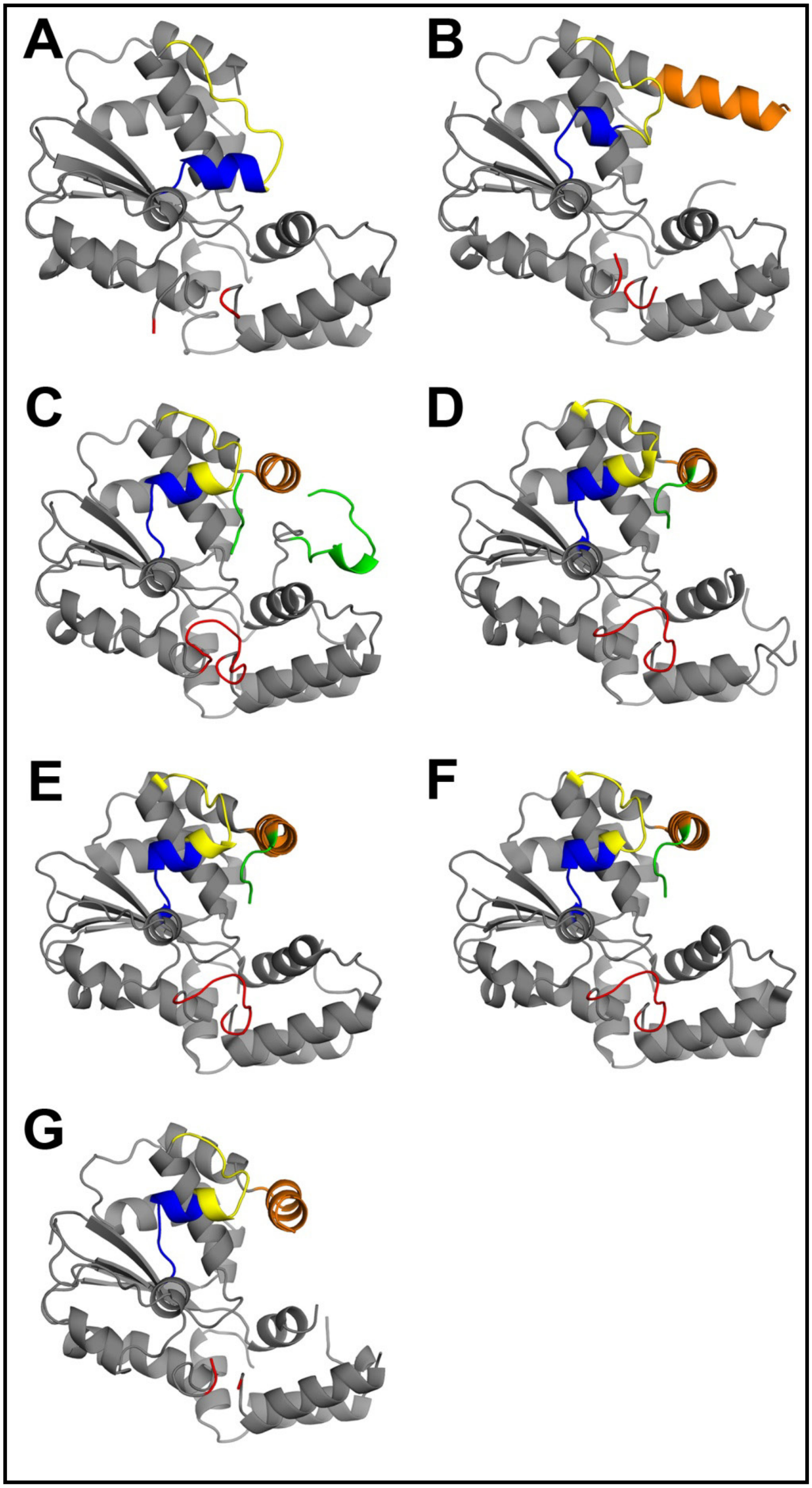
complex, as well as in the three different structures of StaL (Figure 7.2).

Met128-Gly132 of V2 and a large region (Gln215-Asp240) that fully encompasses V3 could not be modeled in the electron density map and are presumed to be disordered. The V3 loop is part of a larger motif we have termed the GHL (glycopeptide helix loop), and is believed to be the main determinant of substrate specificity. N-terminal to V3 is the helical portion of the GHL. The Teg14 GHL helix adopts the bent conformation observed in the Teg12 ternary structure and StaL structures, where there is also an absence of a glycopeptide in the active site (Figure 7.2). We have proposed that unwinding of the GHL helix from a straight to bent conformation may help to bring substrate into the active site pocket. Cys201 of the GHL helix makes a disulfide cross-link to Cys20 of helix 1. This disulfide occurs exactly at the point where the GHL helix unwinds from the straight to the bent conformation (compare the Teg12 binary structure with the other structures in Figure 7.2). The portion of the GHL helix that displays conformational diversity is directly C-terminal to Cys201. It is unclear whether this disulfide plays a functional role, as Cys201 corresponds to Ser203 in Teg12. However, we have suggested that the lack of a disulfide in Teg12 could explain why the region C-terminal to Ser203 was disordered in the Teg12 apo structure, as the lack of a disulfide could impart more flexibility to the protein. A disulfide is also present at this position in StaL (Cys20-Cys196).

.

**Figure 7.2** (adjacent page) Comparison of Teg12, Teg14, and StaL crystal structures. (A) Teg12 apo, (B) Teg12 binary complexed with the teicoplanin aglycone, (C) Teg12 ternary complexed with PAP and the teicoplanin aglycone, (D) StaL apo, (E) StaL complexed with PAP, (F) StaL sulfate, (G) Teg14 apo. Regions of sequence variability are colored blue (V1), red (V2), and green (V3). The flexible GHL helix is colored orange. Residues 28 to 43 (Teg12, Teg14, and StaL) encompass V1 (blue). The remainder of these residues, which also show structural variability, are colored yellow.

Figure 7.2



The loop region (Met102 to Asp108) connecting helices 5 and 6 showed a lack of well ordered electron density, and we were unable to place residues in this portion of the model. These residues were ordered in our Teg12 structures, ordered in the StaL-PAP and StaL-sulfate structures, but disordered in the StaL-apo structure. Electron density picked up again at Asp109, though the region Asp109 to Phe116 displayed a high degree of flexibility, as evidenced by the unusually high temperature factors (above 100) and lack of side chain information for several residues. The N-terminal vector derived-tag, portions of which could be modeled in all three Teg12 structures, is entirely disordered in Teg14.

### *The Teg14 Active Site*

Teg14 was crystallized in its apo form. However, during refinement two regions of prominent electron density were observed in the active site cavity. One of these regions was easily assignable to a molecule of CHES, a buffer component of the crystallization solution. The CHES molecule is oriented in much the same way as PAP is in the Teg12 ternary and StaL-PAP structures. The majority of the residues important for binding PAP also coordinate CHES (Figure 7.3A). Interactions between CHES and the protein are primarily mediated through Van der Waals contacts, though there are several key hydrogen bonds. The hydroxyl oxygen of Tyr165 (Tyr167 in Teg12 and Tyr163 in StaL) hydrogen bonds with the nitrogen of CHES. This tyrosine is in the same general vicinity in the Teg12 and StaL structures, but is located too far from PAP to hydrogen bond with either the 3’-phosphate or oxygen from the ribose. Mutation of Tyr167 in Teg12 renders the protein insoluble, implicating a role for it in protein stability.

Backbone nitrogens from Ala13, Gly14, Asn15, and Thr16 of the 5'-phosphosulfate-binding motif (5'-PSB), and from Trp17 are all within contact distance of the sulfate from CHES. The cyclohexyl ring roughly parallel stacks with Trp17, mimicking how the adenine of PAP stacks with Trp17 in Teg12 and StaL. Leu203 and Met206 (Leu205 and Met208 in Teg12 and Leu198 and Met201 in StaL) of the GHL helix cover the top of the cyclohexyl ring of CHES, as they do with the adenine of PAP in the Teg12 ternary and StaL-PAP structures.

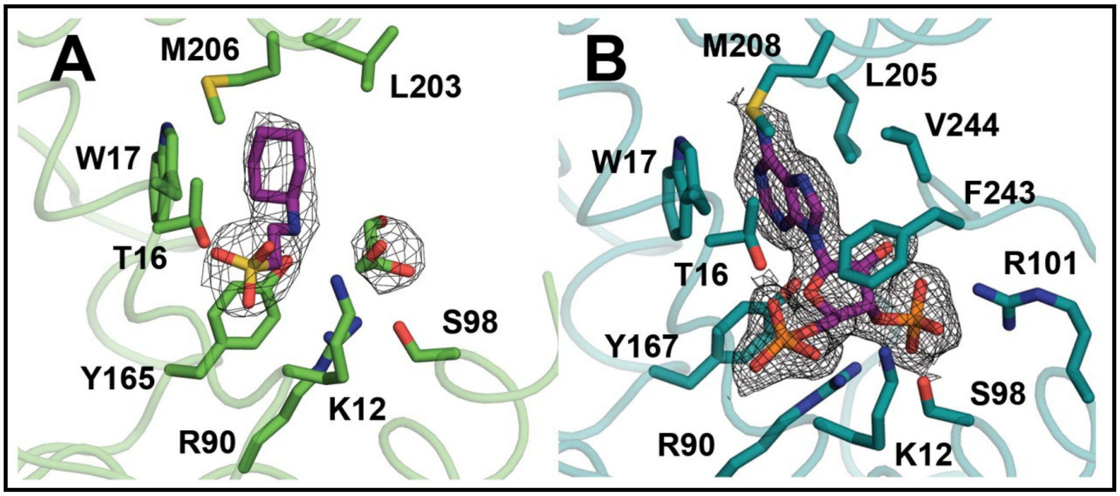


**Figure 7.3)** A comparison of the Teg14 (A) and Teg12-PAP (B) active sites. Several of the key residues that contact PAP in Teg12 also coordinate CHES in Teg14.

The second region of electron density that was enhanced during refinement proved more difficult to assign. This density is in the immediate vicinity of the hydroxyl of Ser98, a residue which is part of the 3'-phosphate-binding motif (3'-PB) and important for binding PAPS. Several small molecules and ions were modeled into the density and subjected to refinement in order to identify the best fit. We initially thought that sulfate could be correct, as the density was somewhat spherical and sulfate could have been present in the crystallization buffer from the hydrolysis of CHES. However, when sulfate was placed in the model and refined, the resulting b-factors were in the 140 range. The b-factor for the adjacent side chain oxygen from Ser98 is 90.5. Imidazole was also attempted and gave b-factors around 90, but did not fit the density well. A chloride ion gave the lowest b-factor of 82, but did not satisfy all of the density. Ultimately glycerol was chosen as the ligand that best fit the density, and gave refined b-factors of around 90 (Figure 7.3A). Only three residues make significant contacts with the putative glycerol, Lys12, Arg90, and Ser98. Of these, only Ser98 is within hydrogen bonding distance. A large portion of the glycerol is exposed to the solvent region. Although a water molecule is within 3.3 Å of a glycerol oxygen, the solvent overall in this region is poorly ordered. This fact could explain the high temperature factors observed for glycerol. In its position around Ser98, the glycerol appears to act as a surrogate for the 3'-phosphate of PAPS. A StaL-sulfate structure has a sulfate modeled in the location where we have modeled glycerol.

Teg-like sulfotransferases are thought to catalyze the sulfonation reaction through an in-line attack mechanism by utilizing histidine as a general base. In the Teg14 structure, His67 is positioned approximately 7 Å from the sulfate of the ordered molecule of CHES (Figure 7.4). A well ordered water molecule occupies the intermediate space where the sulfate from PAPS would be, and is within hydrogen bonding distance to His67. Water also occupies this space in the Teg12 binary and ternary structures (not in Teg12-apo), and in the StaL apo structure (not in StaL-PAP). A sulfate takes the place of the His67 water in the StaL-sulfate structure. In all three Teg12 structures, all three StaL structures, and the Teg14 structure, the basic ε nitrogen of the imidazole ring from His67 is positioned facing Ser9 and within hydrogen bonding distance. His67 is thought to abstract a proton from Ser9 during the sulfonation reaction.

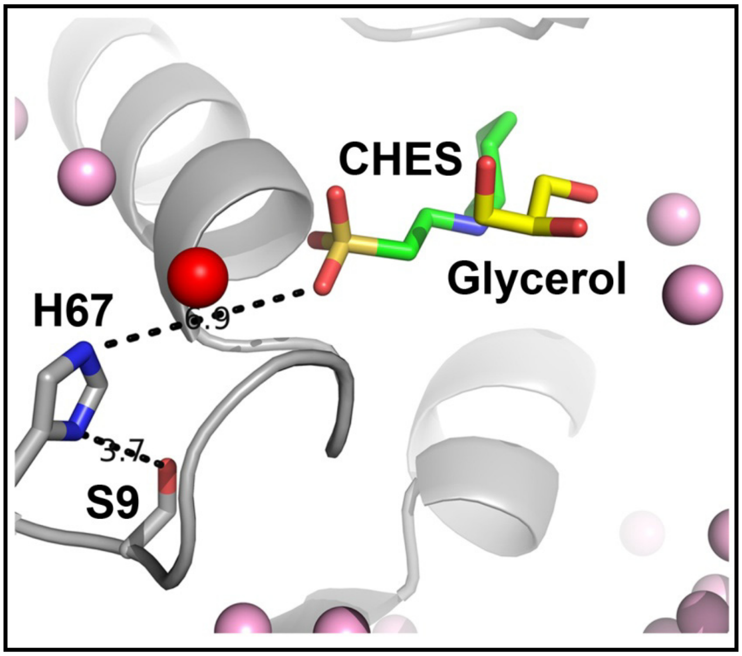


**Figure 7.4)** The relation of CHES to the proposed active residues His67 and Ser9 in the Teg14 active site. The side chain of His67 is approximately 7 Å from the sulfate of CHES. Water molecules are shown as pink spheres. The water molecule that resides where the sulfate of PAPS would be is colored red.

# Discussion

One of the goals of synthetic biology is to engineer metabolic pathways and individual proteins in order to generate small molecules that are not accessible to synthesis through traditional organic chemistry (Keasling, 2008). We initially embarked on structural studies of the TEG pathway sulfotransferases not only to improve our understanding of the sulfation mechanism, but also with the idea of eventually engineering this group of enzymes to accept a broader range of glycopeptide scaffolds and/or co-substrates. The TEG gene cluster is predicted to encode the biosynthesis of a heptapeptide (hydroxyphenylglycine (Hpg)-betahydroxytyrosine (Bht)-dihydroxyphenylglycine (Dpg)-Hpg-Hpg-Bht-Dpg) that is oxidatively cross-linked into the four macrocycles seen in teicoplanin-like glycopeptides. This TEG derived, oxidatively cross-linked heptapeptide skeleton only differs from the teicoplanin aglycone by the substitution of Bht for Tyr at the 2nd position in the peptide core. We have already shown that the Teg sulfotransferases can utilize the teicoplanin aglycone as a substrate (Banik and Brady, 2008). Each Teg enzyme sulfates teicoplanin at one of three different positions to produce a unique monosulfated teicoplanin derivative. In combination, two Teg enzymes can produce 3 different disulfated teicoplanins, and when all three Teg enzymes are used in conjunction, a trisulfated teicoplanin is produced (see Chapter 6, Figure 6.1). It is possible that the Teg sulfotransferases can accept, or through precise mutagenesis be engineered to accept glycopeptide scaffolds other than their predicted substrate and teicoplanin. The Teg12 binary structure shows that a large portion of teicoplanin is exposed to solvent. One could envision a teicoplanin-like derivative, with additional prosthetic groups attached to either the Dpg at position 3 or the Hpg at position 4, that Teg12 could still accommodate in the active site (Figure 7.5).

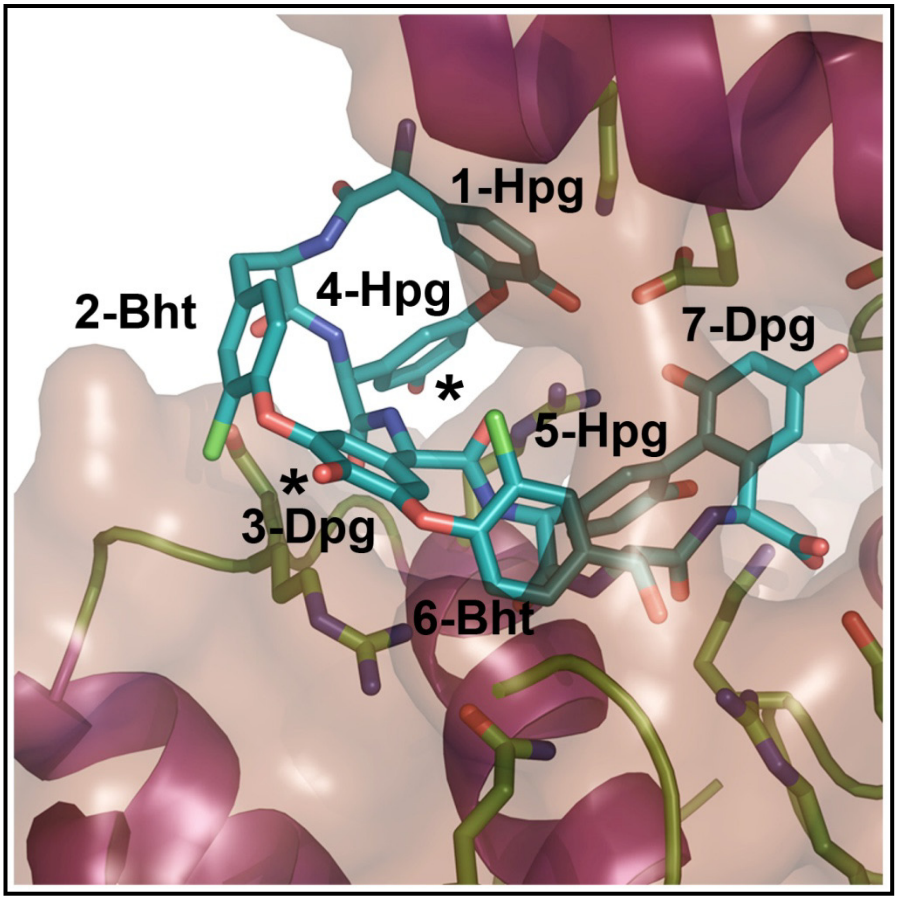


**Figure 7.5)** A surface rendering of the Teg12 binary structure complexed with the teicoplanin aglycone. Residues of the aglycone are labeled, and the * indicate residues whose side chains face the open solvent. Active site residues of Teg12 are shown as sticks and colored yellow.

Regarding the co-factor, from our structural studies of Teg12 and Teg14, it is clear that a number of co-substrates in addition to PAPS are potentially accessible to the active site (aspartame in the case of Teg12 and CHES in the case of Teg14). Whether co-substrates containing anionic substituents other than sulfate could be used to add functional groups to glycopeptide scaffolds will require experimentation.

Attempts to obtain co-crystals of Teg14 with the teicoplanin aglycone, the monosulfated teicoplanin products of Teg12 and Teg13, or the disulfated teicoplanin product of Teg12 and Teg13 have to date been unsuccessful. Therefore, at present we do not know the precise orientation of teicoplanin (either the non-sulfated or various sulfated forms) in the Teg14 active site. However, we do know that Teg14 sulfates the hydroxyl from the Hpg at position 4 of teicoplanin (Banik and Brady, 2008), and therefore that residue 4 must at some point in the reaction cycle be in the direct vicinity of the active site His67.

Another possibility, although unlikely, is that the co-substrate PAPS and teicoplanin do not occupy the active site simultaneously during the sulfonation reaction. The orientation of teicoplanin in the active site of the Teg12 binary structure is most likely representative of a biological relevant conformation. This interpretation of the structure is supported by alanine scanning mutagenesis experiments, which demonstrated that residues contacting the aglycone, when mutated to alanine, have a deleterious effect on product conversion. However, although the binary conformation is most likely biologically relevant, a composite of the Teg12 binary and ternary (bound to PAP in the active site) active site indicate that PAP and the aglycone can not occupy their crystallized conformations simultaneously, as there would be considerable clashes between PAP and the C-terminus of the aglycone. If PAPS were first to donate a sulfuryl group to the protein, producing a sulfo-His intermediate as in the case of bacterial arylsulfate sulfotransferases, for example (Malojčić et al., 2008), and then dissociate from the protein prior to recruitment of the aglycone to the active site, then the conformations of the Teg12 binary and ternary structures would not be forced to co-exist.

# Materials and Methods

## *Cloning Teg14 From an Environmental DNA Library*

Teg14 (Sulf3) was pcr amplified (30 cycles of 95° C for 30s, 60° C for 30s, and 72° C for 90s; FailSafe system from Epicentre) from eDNA clone D30 with the following forward and reverse primers: Sulf3For (BclI): GCGCTGATCAATGAACGGTATTCG ATGGATC, and Sulf3Rev (HindIII): GCGCAAGCTTACAATCCGCCCGTTAGCC GGC. The amplified product was doubly digested with BclI/HindIII and ligated into BamHI/HinDIII digested pET28a (Banik et al., 2008).

## *Protein Expression and Purification*

pET28a-Teg14 was transformed into E. coli BL21 (DE3) for protein expression. Small-scale overnight cultures were used to inoculate (1:1,000 dilution) 1-L cultures of LB plus 50μg/ml kanamycin. The cultures were grown at 37° C until the density reached an $OD_{600}$ of 0.6, at which point the temperature was reduced to 20 °C. After 1 h, protein expression was induced with 0.5mM IPTG. After 14-16 h the cultures were harvested by centrifugation (3,200 x g for 30 min). The cell pellet was resuspended in 40 ml of lysis buffer (50 mM HEPES, pH 7.5, 0.5 M NaCl, 5% (vol/vol) glycerol, 20 mM imidazole, pH 8, 10 mM β-mercaptoethanol, and 0.5% Triton X-100) and the cells were lysed by sonication. Crude cell lysates were centrifuged at 25,000 x g for 30 min, and the supernatants were then incubated for 15 min (24 °C) with 1 mL of Ni-NTA resin (equilibrated with lysis buffer). After 15 min the slurry was loaded onto a column, washed with 40 mL of lysis buffer, followed by 40 mL of wash buffer (50 mM HEPES, pH 7.5 0.5 M NaCl, 5% (vol/vol) glycerol, 20 mM imidazole, pH 8, and 10 mM β-mercaptoethanol), and finally eluted with 15 mL of elution buffer (50 mM HEPES, pH 7.5, 0.5 M NaCl, 5% (vol/vol) glycerol, 125 mM imidazole, pH 8, and 10 mM β-mercaptoethanol). The eluted protein was collected into 7 ~2 mL fractions. The highest concentrated 2 to 3 fractions, as determined by absorbance at 280 nm, were pooled for use in crystallization. The protein was concentrated to 20mg/mL and buffer exchanged 3 times into crystallization buffer (20 mM HEPES, pH 7.5, 5% (vol/vol) glycerol, 200 mM NaCl, 5 mM DTT) using a 20 mL 30,000 MWCO Vivaspin concentrator (Vivascience). No attempt was made to remove the N-terminal His$_6$-tag, resulting in an additional 34 vector derived amino acids appended to the N-terminus of the Teg14 protein sequence.

### *Crystallization and X-ray diffraction*

Teg14 at 20mg/mL was screened at 22 °C for crystals using the JCSG suites (cores I-IV) from Qiagen. All screening was performed in 24 well trays (Qiagen) using the hanging drop method. Crystallization drops contained a 1μl:1μl mixture of protein and precipitant solutions over a 500 μl reservoir volume. Three crystals appeared overnight in condition JCSG core II, A3 (1.0 M Sodium citrate, 0.1 M CHES, pH 9.5), and grew to a maximal size of approximately 300 μM x 300 μM x 300 μM in 4 days. Diffraction quality from two of the crystals was impaired while searching for suitable cryo conditions. The third crystal, which was harvested from the crystallization drop using a cryoloop (Hampton Research), briefly (<1 min) soaked in reservoir solution plus 30% glycerol, and flash-cooled directly in a liquid nitrogen cryo stream at 100K, diffracted well beyond 2.5 Å resolution. These crystals have proven to be irreproducible. Diffraction data from the Teg14 crystal was measured using an R-AXIS IV++ area detector connected to a Rigaku/MSC MicroMax-007 HF Cu rotating-anode generator (wavelength 1.5418 Å) equipped with Varimax confocal Max-Flux optics at The Rockefeller University. The crystal to detector distance was 200mm. 99, 10 min exposure images with 0.5° oscillations per frame were collected. Diffraction data was processed and scaled to 2.7 Å using the HKL-2000 package (Otwinowski & Minor, 1997). The default values for profile fitting radius, box size, spot radius were used during integration. For scaling, 10 resolution zones were used with an error scale factor of 1.3 (default). The error model was adjusted for the individual zones, from zone 1 to 10 as follows: 0.05, 0.07, 0.09, 0.10, 0.12, 0.12, 0.12, 0.11, 0.11, and 0.11. The Teg14 crystal belongs to the trigonal space group $P3_121$, with unit-cell parameters a = b = 112.19 Å, c = 74.81 Å, $\alpha = \beta = 90.00°$, $\gamma = 120.00°$. A single Teg14 monomer, including the 34 vector derived residues, two molecules of glycerol and a single molecule of CHES (N-Cyclohexyl-2-aminoethanesulfonic acid) buffer (total molecular weight of 34,822 Daltons) in the asymmetric unit gave a packing density, VM (Matthews, 1968), of 3.90 Å3 Da-1, corresponding to a solvent content of 68.43%. Diffraction data analysis statistics are given in Table 7.1.

### *Structure Determination and Refinement*

The Teg14 structure was solved by molecular replacement with the program Phaser (Storoni et al., 2004; McCoy et al., 2005). The recently reported Teg12 ternary structure (Bick et. al., 2010), devoid of all flexible loops and small molecule substrates, was used as a search model. The ternary structure was chosen as the search model because it is the highest resolution structure from the group of three Teg12 structures reported. Teg12 has a sequence identity of 82.9% with Teg14. Phaser gave a single clear solution with a rotational Z score of 13.3, translational Z score of 33.0, and log-likelihood gain of 1,278. No clashes were observed in the solution. A single Teg14 monomer was identified in the asymmetric unit. Rigid body refinement of the molecular replacement solution using Phenix.refine (Afonine et al., 2005) gave an R and Rfree of 0.410 and 0.415, respectively. Manual rebuilding of the model using the resulting $2|F_o| - |F_c|$ and $|F_o| - |F_c|$ maps was carried out in Coot (Emsley & Cowtan, 2004). Multiple rounds of restrained refinement were carried out using Phenix.refine, employing the translation-libration-screw (TLS) protocol and 2 atomic displacement parameter (ADP) groups per residue (Winn et al., 2001; Afonine et al., 2007). Teg14 was partitioned into 15 TLS groups as follows: group1-residues 0-12, 83-90, 57,67, and 157-165 (the beta sheet core), group2-residues 166-170, group 3-residues 171-186, group 4-residues 187-203, group 5-residues 204-214, group 6-residues 13-30 and the active site CHES, group 7-residues 31-42, group 8-residues 43-56, group 9-residues 68-82, group 10-residues 91-101, group 11-residues 109-127, group 12-residues 133-156, group 13, residues 241-248, group 14-residues 249-261, and group 15-residues 262-275 (Figure 7.6). Side chain atoms for which there was no observable density were removed from the model. After all protein residues were assigned, the water picking protocol in Phenix.refine was used to identify density corresponding to waters. Several waters were then added manually. Coot identified five of these waters as “unusual”, however manual inspection suggested they were reasonably assigned. Both X-ray to stereochemistry (“optimize wxc”) and X-ray to

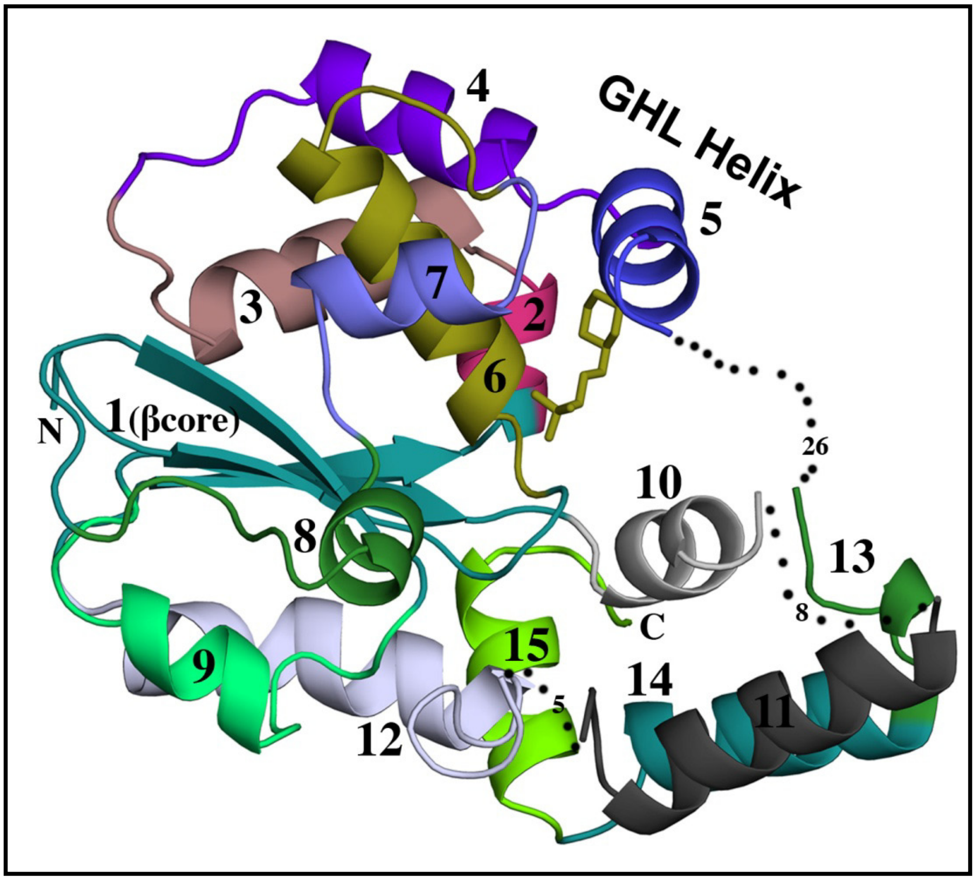


**Figure 7.6)** Teg14 colored according to the groups used for TLS refinement in Phenix.refine. The groups are numbered from 1 to 15 (see the methods section for the residues for each group). Regions of disorder are connected with grey dots, and the number of disordered residues is indicated.

ADP ("optimize wxu") weights were optimized during the final round of refinement. The model converged to a final $R_{work}$ and $R_{free}$ of 19.55 and 23.56, respectively. The structure was validated using the Molprobity server at Duke University (Davis et al., 2007), the results of which along with refinement statistics are presented in Table 7.1. All figures of the protein model were generated using the program PyMol (Delano Scientific LLC, http://www.pymol.org).

## Table 7.1

| | |
|---|---|
| Data Collection | |
| Space Group | $P3_121$ |
| Unit-cell parameters (Å, °) | $a = b = 112.03$, $c = 74.72$, $\alpha = \beta = 90.0$, $\gamma = 120.0$ |
| Temperature | 100K |
| Matthews coefficient (Å$^3$ Da$^{-1}$) | 3.89 |
| Solvent content (%) | 68.38 |
| Resolution (Å) | 25.00 – 2.70 (2.80 – 2.70) |
| Wavelength (Å) | 1.5418 |
| $R_{\mathrm{merge}}$ (%) | 0.073 (0.494) |
| $I/\sigma(I)$ | 14.6 (2.5) |
| Completeness (%) | 95.8 (93.1) |
| Reflections (observed) | 39,851 |
| Reflections (unique) | 14,590 |
| Redundancy | 2.7 (2.2) |
| Refinement | |
| Resolution (Å) | 22.79 – 2.70 |
| Reflections | 13,643 |
| $R_{\mathrm{work}}/R_{\mathrm{free}}$ (%) | 19.55/23.56 |
| No. monomers/AU | 1 |
| No. non-H atoms | 1,876 |
| Protein | 1,797 |
| Ligand | 25 |
| Water | 54 |
| $B$ factors (Å$^2$) | 59.31 |
| Protein | 59.19 |
| Ligand | 73.02 |
| Water | 57.19 |
| R.m.s. deviations | |
| Bond lengths (Å) | 0.004 |
| Bond angles (°) | 0.805 |
| Dihedral angles (°) | 14.59 |
| Molprobity validation | |
| Clashscore (All-atom contacts) | 20.8 (82nd percentile) |
| Poor rotamers | 0.57% (Arg127) |
| Ramachandran outliers | 0.00% |
| Ramachandran favored | 97.39% |
| C$\beta$ deviations > 0.25Å | 0 |
| Molprobity score | 1.93 (98th percentile) |
| Residues with bad bonds | 0.00% |
| Residues with bad angles | 0.42% (Pro58) |

**Table 7.1)** Data collection and refinement statistics for Teg14.

# References

**Banik, J.J., and Brady, S.F.** (2008). Cloning and characterization of new glycopeptide gene clusters found in an environmental DNA megalibrary. Proc Natl Acad Sci U S A *105*, 17273-17277.

**Brady, S.F., Simmons, L., Kim, J.H., and Schmidt, E.W.** (2009). Metagenomic approaches to natural products from free-living and symbiotic organisms. Nat Prod Rep *26*, 1488-1503.

**Higgins, D.L., Chang, R., Debabov, D.V., Leung, J., Wu, T., Krause, K.M., Sandvik, E., Hubbard, J.M., Kaniga, K., Schmidt, D.E., Jr.*, et al.*** (2005). Telavancin, a multifunctional lipoglycopeptide, disrupts both cell wall synthesis and cell membrane integrity in methicillin-resistant Staphylococcus aureus. Antimicrob Agents Chemother *49*, 1127-1134.

**Keasling, J.D.** (2008). Synthetic biology for synthetic chemistry. ACS Chem Biol *3*, 64-76.

**Malojčić, G., Owen, R. L., Grimshaw, J. P. A., Brozzo, M. S., Dreher-Teo, H, Glockshuber, R.** (2008). A structural and biochemical basis for PAPS-independent sulfuryl transfer by aryl sulfotransferase from uropathogenic *Escherichia coli*. Proc Natl Acad Sci U S A *105*, 19217-19222.

**Nicolaou, K.C., Boddy, C.N.C., Brase, S., and Winssinger, N.** (1999). Chemistry, biology, and medicine of the glycopeptide antibiotics. Angew Chem Int Ed Engl *38*.

**Rappe, M.S., and Giovannoni, S.J.** (2003). The uncultured microbial majority. Annu Rev Microbiol *57*, 369-394.

**Shi, R., Lamb, S.S., Bhat, S., Sulea, T., Wright, G.D., Matte, A., and Cygler, M.** (2007). Crystal structure of StaL, a glycopeptide antibiotic sulfotransferase from Streptomyces toyocaensis. J Biol Chem *282*, 13073-13086.

**Stryjewski, M.E., O'Riordan, W.D., Lau, W.K., Pien, F.D., Dunbar, L.M., Vallee, M., Fowler, V.G., Jr., Chu, V.H., Spencer, E., Barriere, S.L., *et al.*** (2005). Telavancin versus standard therapy for treatment of complicated skin and soft-tissue infections due to gram-positive bacteria. Clin Infect Dis *40*, 1601-1607.

**Wolter, F., Schoof, S., and Süssmuth, R.** (2007). Synopsis of Structural, Biosynthetic, and Chemical Aspects of Glycopeptide Antibiotics. In Glycopeptides and Glycoproteins, pp. 143-185.