Predicting and testing determinants of histidine-kinase functions by leveraging protein sequence information

by

Orr Ashenberg

A.B. Chemistry and Chemical Biology
Harvard College, Cambridge, MA, 2006

SUBMITTED TO THE PROGRAM OF
COMPUTATIONAL AND SYSTEMS BIOLOGY IN PARTIAL
FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

SEPTEMBER 2012

Signature of Author:_____
Program of Computational and Systems Biology
September 06, 2012

Certified by: _____
Amy E. Keating
Associate Professor of Biology
Thesis Supervisor

_____
Michael T. Laub
Associate Professor of Biology
Thesis Supervisor

Accepted by:_____
Christopher B. Burge
Professor of Biology and Biological Engineering
Director of Computational and Systems Biology Ph.D. program

Predicting and testing determinants of histidine-kinase functions by leveraging protein sequence information

by

Orr Ashenberg

Submitted to the Program of Computational and Systems Biology
on September 06, 2012 in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Computational and Systems Biology
at the Massachusetts Institute of Technology

Abstract

All cells sense and respond to their environments using signal transduction pathways. These pathways control a sweeping variety of cellular processes across the domains of life, but the pathways are often built from a small, shared set of protein domains. At the core of tens of thousands of signal transduction networks in bacteria is a pair of proteins, a histidine kinase and a response regulator. Upon receiving an input signal, a histidine kinase autophosphorylates and then catalyzes transfer of its phosphoryl group to a cognate response regulator, which often activates a transcriptional response. Bacteria typically encode dozens of kinases and regulators, and the kinases function as dimers in all known examples. This dimeric state raises two functional questions. Do histidine kinases specifically form dimers? Once a kinase has dimerized, does a chain in the dimer phosphorylate itself (*cis*) or its partner chain (*trans*)? Specific kinase dimerization is likely important to avoid detrimental crosstalk between separate signaling pathways, and how autophosphorylation occurs is central to kinase activity. In my thesis, I have taken biochemical and evolutionary approaches to identify molecular determinants for both dimerization specificity and autophosphorylation.

To study dimerization specificity, I developed an *in vitro* binding assay to measure kinase dimerization, and I then showed that a paralogous pair of kinases from *E. coli* specifically formed homodimers over heterodimers. Residues important for dimerization specificity were predicted by measuring amino acid coevolution within kinases, which leverages the enormous amount of sequence information available for the kinase family. Experimental verification of these predictions showed that a set of residues at the base of the kinase dimerization domain was sufficient to establish homospecificity. This same region of the kinase, in particular the loops at the base of the kinase dimer, was also important for determining autophosphorylation mechanism. Recent work showed that kinases could autophosphorylate either *in cis* or *in trans*, and I found that a *trans* kinase could be made to autophosphorylate *in cis* by replacing its loop with the loop from a *cis* kinase. I also found that two sets of orthologs, despite having significantly diverged loop sequences, had conserved their autophosphorylation mechanisms. This raised the possibility that kinase loops may be under selection to maintain the same autophosphorylation mechanism.

Thesis supervisors: Amy E. Keating, Michael T. Laub
Titles: Associate Professor of Biology, Associate Professor of Biology

Acknowledgements

Many people contributed to help make the work in this thesis possible. I would first like to thank my advisors, Amy Keating and Michael Laub. Of all the properties that made them excellent mentors, the best was their enthusiasm for science. Amy was supportive at all stages of my various projects, but also gave me room to figure things out on my own. She helped develop my scientific thinking and communication abilities. I also really enjoyed following the breadth of projects in her lab, and it is hard to imagine someone else balancing the computational and experimental interests of the lab. Thanks to Mike for taking me on in his lab, and letting me fiddle around with experiments when I was beginning. I learned a lot from the clarity and rigor he looked for when both discussing, and writing science, and his strong interest in evolution was very influential for me. I also learned more about *C. crescentus* than I ever imagined I would.

Thanks to my committee members Bob Sauer and Bruce Tidor. They helped steer my projects and got me to focus on the bigger scientific questions. It was fun, and often challenging, to hear their different viewpoints during committee meetings. Bob and Bruce both were also excellent teachers for biochemistry, computational biology literature discussion, and algorithms in computational biology. I would also like to thank Bruce for getting me interested in MIT's computational and systems biology program. Thanks to Penny Beuning for taking some time to think about my research.

The past and present members of the Laub and Keating labs created fun and scientifically stimulating work environments. Barrett Perchuk patiently taught me how to survive in the wet lab. The interaction specificity people in the Laub lab, Emily Capra, Emma Lubin, and Anna Podgornaia provided great feedback on my projects. I thank my many bay mates, Celeste Peterson, Emma Lubin, Kasia Gora, Andy Yuan, Anna Podgornaia, Chris Aakre, and Josh Modell. I particularly appreciated the sense of humor of Josh Modell and Christos Tsokos.

From the Keating lab, I would especially like to thank Aaron Reinke and Scott Chen. Aaron and Scott both love discussing science, and it was a pleasure learning so much from them. Thanks also to Chris Negron, who I had many fun adventures with, and Jennifer Kaplan, who endured daily serenades. I would also like to thank the wise postdocs in the lab. Sanjib Dutta took the time to guide me early on, and Evan Thompson has some wonderful life stories. My undergraduate student, Kate Rozen-Gagnon, was hard-working, and successfully got her summer project to work.

Thanks to friends outside of lab, Danny Park, Vasudev Vadlamudi, Joshua Tesoriero, Tracy Washington, Stian Ueland, and Andrew Nager. Also thanks to several friendly administrators, Linda Earle, Luke McNeill, Sally MacGillivray, Bonnie Lee Whang, and Darlene Ray, the late-night custodial staff of Doug, Raoul, Brian, and Richie, and Debby Pheasant, the director of the BIF.

Finally I would like to thank my parents, who are actually both doctors as well. My dad got me into the sciences, and encouraged me to work hard and smart at all things science. My mom balanced this by encouraging me to leave lab sometimes and get outside.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

CA- catalytic and ATP-binding

DHp- dimerization and histidine phosphotransfer

EGFR- epidermal growth factor receptor

FRET- fluorescence resonance energy transfer

HAMP- histidine kinases, adenylyl cyclases, methyl-accepting chemotaxis proteins, and phosphatases

HK- histidine kinase

MI- mutual information

PAS- Per-Arnt-Sim

RR- response regulator

RTK- receptor tyrosine kinase

TCST- two-component signal transduction

Y2H- yeast-two-hybrid

Chapter 1

Introduction

Signal transduction pathways are used to control critical cellular processes across all domains of life. Signaling pathways are central to diverse processes like metabolism, cell differentiation, cell migration, and cell cycle control, and they allow cells to sense and respond to their environments (Lemmon and Schlessinger, 2010; Stock et al., 2000). Deregulation within signaling pathways is commonly associated with disease and cancer states (Blume-Jensen and Hunter, 2001). The information flow through signal transduction pathways is most often mediated by protein kinases and protein phosphatases, which allow for reversible protein phosphorylation (Hunter, 2000). The importance of protein kinases and the diversity of processes they control are reflected in the fact that there are at least 518 human kinases, making up ~1.7% of all human genes (Manning et al., 2002).

Kinase signaling overview

Kinases and the signaling networks they make up can be studied at the molecular, systems, and evolutionary levels. At the molecular level, all kinases carry out two common functions: receiving an input signal that modulates the balance between an inactive and active state, and recognizing and phosphorylating a substrate upon activation. Kinases may be activated through a variety of input signals including small molecule ligands, protein-protein interactions, post-translational modification, and cellular localization (Endicott et al., 2012; Gao and Stock, 2009). These signals lead to changes in tertiary or quaternary structure either within or between kinase domains. A full understanding of these changes requires solving structures for both the inactive and active states, but this is often difficult in practice, especially because many kinases are membrane-bound receptors.

Similarly, understanding kinase-substrate recognition is difficult because relatively few

structures of a kinase interacting with its substrate exist. Such interactions are usually weak and transient, and often involve only a small number of residues in a less-ordered region of the substrate (Endicott et al., 2012). One experimental approach that identifies substrate consensus sequences is the screening of randomized peptide libraries against purified kinases (Songyang et al., 1994). A complementary *in vivo* approach, quantitative phosphoproteomic mass spectrometry, maps the phosphorylation sites within a cell but does not directly reveal signaling network connectivity, because the kinases responsible for different modifications are not known (Macek et al., 2008; Tedford et al., 2008).

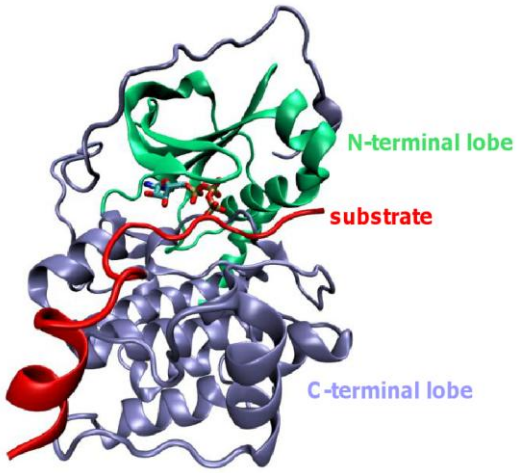


Figure 1.1 Eukaryotic protein kinase domain.
The kinase domain from protein kinase A (PDB ID 1ATP) in its active state and in complex with its substrate (red). The kinase substrate binds in the cleft between the N- and C-terminal lobes, positioning its serine or threonine phosphorylation site near the kinase catalytic residues and the γ-phosphate of ATP.

The following discussion uses epidermal growth factor receptor (EGFR) as an illustrative example of how molecular, systems, and evolutionary approaches can provide insights into kinase function. EGFR has been studied in great molecular detail and exemplifies many of the principles discussed above. EGFR is a receptor tyrosine kinase (RTK) belonging to the larger eukaryotic protein kinase superfamily, which consists of serine/threonine- and tyrosine-specific protein kinases (Taylor and Kornev, 2011). These kinases catalyze transfer of the γ-phosphoryl

group of ATP to an alcohol side chain on a protein substrate. The protein kinase catalytic domain is composed of a small N-terminal lobe and a large C-terminal lobe (Fig 1.1). The kinase active site sits in a cleft between these two lobes and accommodates binding of both ATP and the substrate. Inactive kinases are structurally diverse whereas active kinases share a common overall structure (Noble et al., 2004). Transition from the inactive to active state often involves rearrangement of an activation segment in the C-terminal lobe (Endicott et al., 2012).

To achieve signal responsiveness, the cytoplasmic protein kinase domain of EGFR is linked through a transmembrane helix to extracellular ligand-binding domains (Lemmon and Schlessinger, 2010). Binding of epidermal growth factor to the ligand-binding domains causes homodimerization of EGFR. EGFR, along with three other receptors in the same family, can also form heterodimers thereby enhancing signaling diversity (Alroy and Yarden, 1997). Dimerization leads to kinase activation through disruption of *cis*-autoinhibitory interactions in the protein kinase domain followed by *trans* autophosphorylation on tyrosine residues (Honegger et al., 1989; Lemmon and Schlessinger, 2010; Low-Nam et al., 2011). A small set of protein kinases may also undergo *cis* autophosphorylation on their activation segments (Lochhead, 2009). Autophosphorylation of EGFR leads to recruitment of the Grb2/Sos complex and Ras, which initiates downstream signaling (Fig 1.2). Grb2 recognizes EGFR through its SH2 domain, which binds specific sequences containing a phosphorylated tyrosine (Lowenstein et al., 1992). SH2 domains are highly amplified, with 120 occurrences in humans, and they show distinct sequence binding specificities (Huang et al., 2008; Jones et al., 2006).

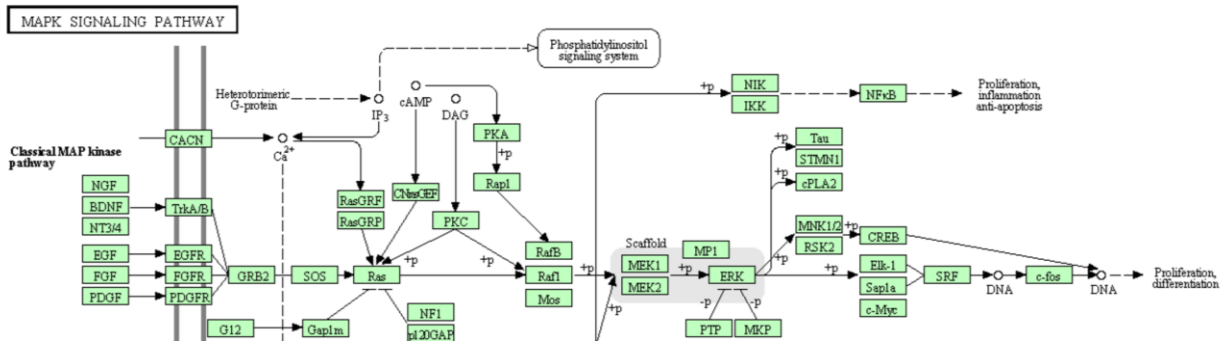


Figure 1.2 EGFR signaling pathway.
Wiring diagram of an EGFR signaling pathway from the KEGG Module database (Kanehisa and Goto, 2000). Binding of epidermal growth factor (EGF) to EGF receptor leads to receptor dimerization. This activates the receptor's cytoplasmic protein kinase domain causing autophosphorylation of tyrosine residues in this domain. The Grb2/SOS complex recognizes the phosphotyrosine residues, and then recruits RAS, which activates the downstream MAPK signaling cascade of Raf, MEK, and ERK. ERK then phosphorylates several targets, which change activities of transcription factors controlling processes like differentiation.

In the systems level view of kinase signaling, the focus is on how interactions between proteins in a pathway collectively carry out that pathway's function. Experimental parameters for all members of a signaling pathway are measured and incorporated into a mathematical model. This model maps pathway inputs to final outputs, which typically involve changes in gene expression or protein-protein interactions. Accurate reconstruction of the spatial and temporal dynamics of signaling requires measurement of the network connectivity along with rate constants and concentrations (Aldridge et al., 2006). Often, not all these parameters can be measured, and the model can be very complicated due to intrinsic complexity in the signaling network's architecture (Fig 1.2). Many pathways have multiple inputs, branch points, and feedback, and these arrangements can lead to functionally important dynamic behaviors. For example, the MAPK/ERK signaling cascade downstream of EGFR gives ultrasensitive switch-like responses to input stimuli (Huang and Ferrell, 1996). Such a mechanism allows for all-or-none decisions, which are appropriate in processes like mitogenesis.

At the evolutionary level, studies of kinase signaling have become more powerful due to significant improvements in genome sequencing technologies allowing identification of kinases from thousands of organisms across the tree of life. As a result, kinase families can now sometimes contain thousands of homologs, and this makes it possible to more confidently address broad questions, like what is the phylogenetic distribution of kinases, and specific questions, like how does new function arise in a kinase family. The phylogenetic distribution of kinases displays strong trends. Serine, threonine, and tyrosine phosphorylation, mediated by the protein kinase superfamily, is widespread across eukaryotes, whereas histidine phosphorylation, mediated by histidine kinases, is common across prokaryotes but only rarely seen in plants and lower eukaryotes (Capra and Laub, 2012; Endicott et al., 2012). Receptor tyrosine kinases such as EGFR may have been one of the innovations leading to multi-cellularity in metazoans, as these kinases are found only in metazoans and a few sister groups of the metazoans (Miller, 2012).

Although kinase families are highly conserved within kingdoms, the number of diverse functions carried out by members of a given kinase family is enormous. Gene duplication is one of the main mechanisms for generating new signaling pathways from highly conserved kinase domains (Zhang, 2003). A duplicated kinase from a signaling pathway must incorporate new input signals and substrate partners while not disrupting existing signaling pathways. Comparative studies examining sequence variation across homologs in signaling pathways show that neofunctionalization is often accomplished through a combination of domain rearrangements and adaptive mutations (Alm et al., 2006; Capra et al., 2012). For example, a newly duplicated EGFR may respond to a different input by acquiring a new extracellular sensory domain, whereas an

SH2 domain may recognize a new phosphotyrosine peptide substrate through mutations in its surface loops (Kaneko et al., 2011; Manning et al., 2002).

Overall, molecular, systems, and evolutionary based approaches feed back upon one another in our understanding of kinases. In this thesis, I use molecular and evolutionary approaches to learn how sequence encodes two diverse molecular functions in histidine kinases: dimerization specificity and autophosphorylation. In both problems, I ultimately test hypotheses about histidine kinase function using targeted biochemical experiments.

Two-component signal transduction

Histidine kinases (HK), together with response regulators (RR), make up two-component signal transduction (TCST) pathways, a predominant signaling modality in bacteria (Stock et al., 2000). Two-component signaling pathways allow bacteria to sense and respond to changes in both external and internal environmental conditions. These pathways regulate an extensive range of processes including chemotaxis, cell cycle, sporulation, and virulence. In a typical two-component pathway, a sensor histidine kinase receives an input signal, undergoes autophosphorylation onto a conserved histidine residue, and then specifically transfers the phosphoryl group to a conserved aspartate residue on the response regulator (Fig 1.3). Most response regulators contain DNA-binding output domains and are activated to effect changes in gene expression upon phosphorylation, allowing a response to the environmental input (Galperin, 2010). Several histidine kinases are bifunctional and have phosphatase activity towards their cognate response regulators, resetting the signaling system to its initial state (Stock et al., 2000).

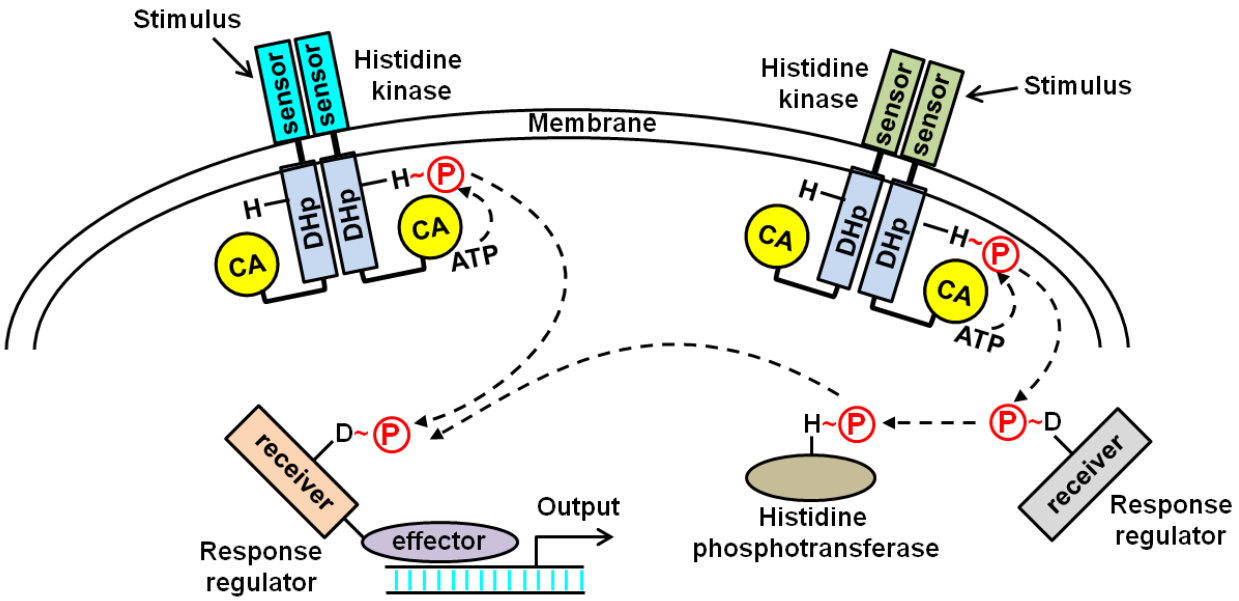


Figure 1.3 Two-component signal transduction pathways.
Signaling in an orthodox phosphotransfer pathway (left) and a phosphorelay pathway (right). In an orthodox pathway, the extracytoplasmic sensory domain of the histidine kinase receives an input stimulus, resulting in autophosphorylation of a histidine in the kinase cytoplasmic transmitter region. The transmitter region consists of a DHp domain, which mediates dimerization, and a CA domain, which binds ATP and catalyzes histidine phosphorylation. The phosphoryl group on the histidine is transferred to an aspartate on the receiver domain of the response regulator. This phosphorylation alters the activity of the response regulator effector domain leading to changes in gene expression. A phosphorelay pathway has two additional phosphotransfer steps relative to an orthodox phosphotransfer pathway. The phosphoryl group from the histidine kinase is transferred to a response regulator, and then to a histidine phosphotransferase, before reaching a terminal response regulator.

As shown in the next sections, TCST systems share several functional parallels with eukaryotic receptor tyrosine kinases like EGFR. RTKs and transmembrane receptor histidine kinases have similar domain architectures and face a common problem of receiving and transmitting a signal across a cell membrane. In both systems, kinase homodimers and heterodimers can form, interaction specificity is important, and *trans* and *cis* autophosphorylation mechanisms have been observed. A significant advantage in studying TCST systems is that their sheer numbers make them well-suited for sequence-based studies, especially molecular coevolution. In addition, the ease of identifying kinases and their substrate regulators in the genome greatly simplifies the problem of predicting substrates and phosphorylation sites relative to eukaryotic kinases. This,

combined with the often-simpler wiring in TCST pathways and the genetic tractability of bacteria, makes building and experimentally testing bacterial signal transduction models more feasible.

Histidine kinases

Histidine kinases are usually transmembrane receptors but some are soluble cytoplasmic proteins. In both forms, the kinases are modular, with highly conserved catalytic transmitter regions linked to diverse sensory domains (Fig 1.3). A receptor histidine kinase links its sensory domain to its cytoplasmic transmitter region through a transmembrane α-helix. The transmitter region consists of two domains: an N-terminal DHp (*d*imerization and *h*istidine *p*hosphotransfer) domain and a C-terminal CA (*c*atalytic and *A*TP-binding) domain (Fig 1.4) (Stock et al., 2000). Histidine kinases function as homodimers, with a few exceptions, and their dimerization is mediated by the DHp domain, which forms a four-helix bundle (Goodman et al., 2009). Unlike eukaryotic EGFR, histidine kinases exist as dimers in the absence of ligand binding to their sensory domains (Gao and Stock, 2009; Scheu et al., 2010). Each chain of the DHp dimer consists of two antiparallel helices, α1 and α2, linked by a poorly-conserved loop (Marina et al., 2005; Tomomori et al., 1999). Midway along α1 is the conserved H-box region, which contains the solvent-exposed histidine phosphorylation site.

The CA domain catalyzes ATP hydrolysis and transfer of the γ-phosphoryl group to the histidine in the H-box of the DHp domain (Fig 1.4). Based on a small number of studies, autophosphorylation within the histidine kinase dimer appears to occur exclusively either in *cis* or in *trans* (Casino et al., 2009; Ninfa et al., 1993; Yang and Inouye, 1991). *In cis*, a chain in the kinase dimer phosphorylates itself, whereas *in trans*, a chain in the kinase dimer phosphorylates

its partner chain. The physiological significance though of this difference and what determines mechanism preference is unknown. To phosphorylate the histidine, the CA domain must move ~15 Å relative to the conformation captured in the crystal structure (Fig 1.4) (Albanesi et al., 2009; Marina et al., 2005). This requires significant changes in the interface between the CA and DHp domain. In contrast to the two-lobe structure of eukaryotic protein kinases, the CA domain forms an α/β-sandwich fold with a highly conserved ATP-binding cavity. Conserved sequence motifs known as the N, G1, F, and G2 boxes make up this binding cavity along with a flexible lid that covers the nucleotide (Tanaka et al., 1998).

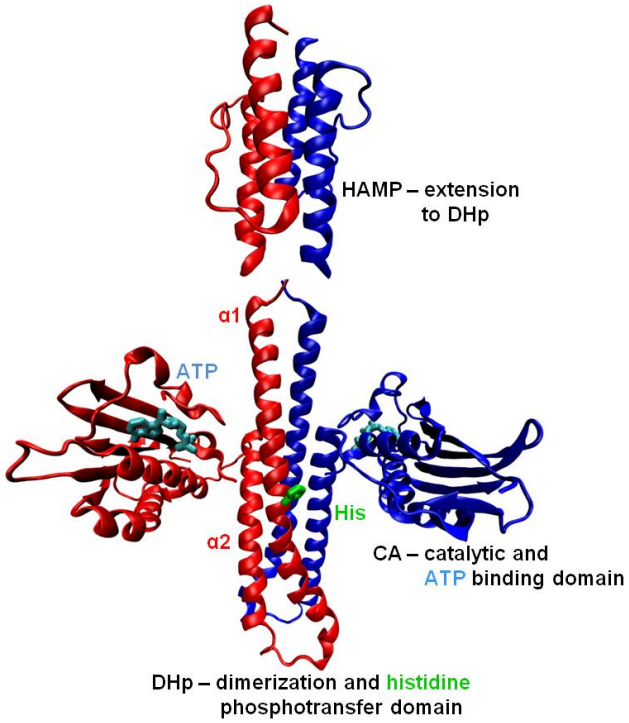


Figure 1.4 Structure of cytoplasmic transmitter region of histidine kinase.
The transmitter region homodimer from *T. maritima* HK853 (PDB ID 2C2A) is shown. The DHp domain homodimer forms a four-helix bundle, with the conserved histidine phosphorylation site (green) on the face of the first helix, α1, and the CA domain forms an ATP-binding cavity occupied by an ATP analog (cyan). Many kinase transmitter regions have a coiled-coil HAMP domain N-terminal to the DHp domain. In the HAMP homodimer from *A. fulgidus* receptor Af1503 (PDB ID 2ASW), the two-helices adopt an unusual knobs-to-knobs packing, and they are connected by an extended linker.

The transmitter region must undergo large conformational changes to switch between enzymatic states and this is driven by signal transduced from upstream sensory domains. In contrast to the catalytic transmitter regions of the histidine kinase, the extracellular sensory domains vary significantly, presumably because they have adapted to different stimuli (Stock et al., 2000). Upon stimulus, these domains must transduce the signal through a transmembrane helix to the transmitter region of the kinase. The Per-Arnt-Sim (PAS) domain is the most common kinase sensory domain (found in ~33% of kinases) and is responsive to small ligands or changes in light or oxygen (Gao and Stock, 2009). In kinases, this domain forms parallel dimers, mediated by a helix bundle, and the N- and C-termini both point to the cell membrane (Moglich et al., 2009b). In a comparison of the citrate-bound structure of the PAS domain from kinase CitA to the citrate-free state, citrate binding led to large structural rearrangements including the β-sheet at the C-terminus being pulled-upward (Fig 1.5) (Sevvana et al., 2008). This piston-like movement is likely transduced through the rigid transmembrane helix to the transmitter region. Piston-like mechanisms, as well as rotations, have also been observed in other kinase sensory domains as well as the bacterial chemoreceptors (Cheung and Hendrickson, 2009; Falke and Hazelbauer, 2001; Neiditch et al., 2006).

Although in many membrane receptor kinases the transmembrane helix links the sensory domain directly to the cytoplasmic transmitter region, a significant fraction of kinases (~31%) have an intervening cytoplasmic HAMP (*h*istidine kinases, *a*denylyl cyclases, *m*ethyl-accepting chemotaxis proteins, and *p*hosphatases) domain between the transmembrane helix and transmitter region (Fig 1.4) (Gao and Stock, 2009). HAMP domains form parallel, homodimeric four-helix coiled coils, but they have at least two different conformations that are thought to

correspond to different signaling states (Hulko et al., 2006). One conformation is tightly packed but includes atypical side-chain packing like knobs-to-knobs, whereas the other conformation is more loosely packed with the helices splaying apart from one another (Airola et al., 2010). The equilibrium between the two states is modulated by the incoming signal from the sensory domain and transmembrane helix, and conversion between states occurs through vertical helix displacements combined with helix rotations and tilts. The most pronounced changes are in the end of the C-terminal helix, which forms a contiguous helix with the first helix of the DHp domain (Ferris et al., 2012; Ferris et al., 2011).

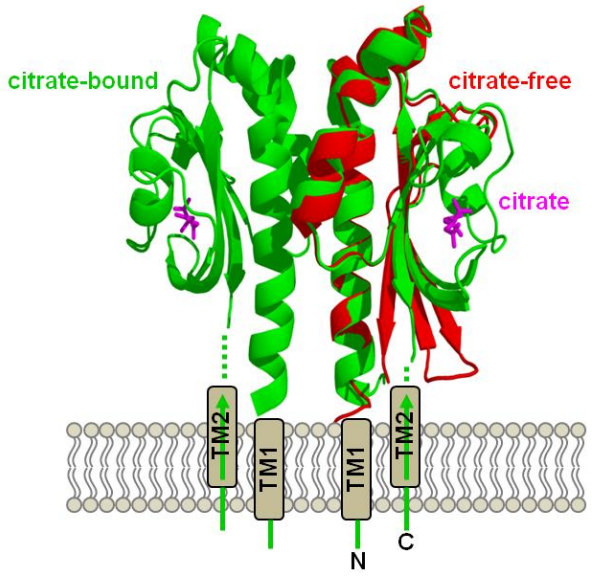


Figure 1.5 PAS domain from *K. pneumoniae* CitA in citrate-free and citrate-bound state.
One chain from the citrate-free PAS domain (PDB ID 2V9A) of CitA is aligned to the citrate-bound PAS homodimer (PDB ID 2J80). Citrate-binding leads to a flexing upwards of the main β-sheet, which also pulls the C-terminus of the PAS domain upward. This may cause a piston-like displacement in transmembrane helix 2 (TM2), which links the C-terminus of the PAS domain to the cytoplasmic kinase transmitter region. Adapted from Figure 6 in (Sevvana et al., 2008).

The incoming signal from the upstream transmembrane helix or HAMP domain alters the equilibrium between the kinase and phosphatase states of the catalytic transmitter region. Recent crystal structures of *B. subtilis* DesK and *T. maritima* HK853 suggest that the main

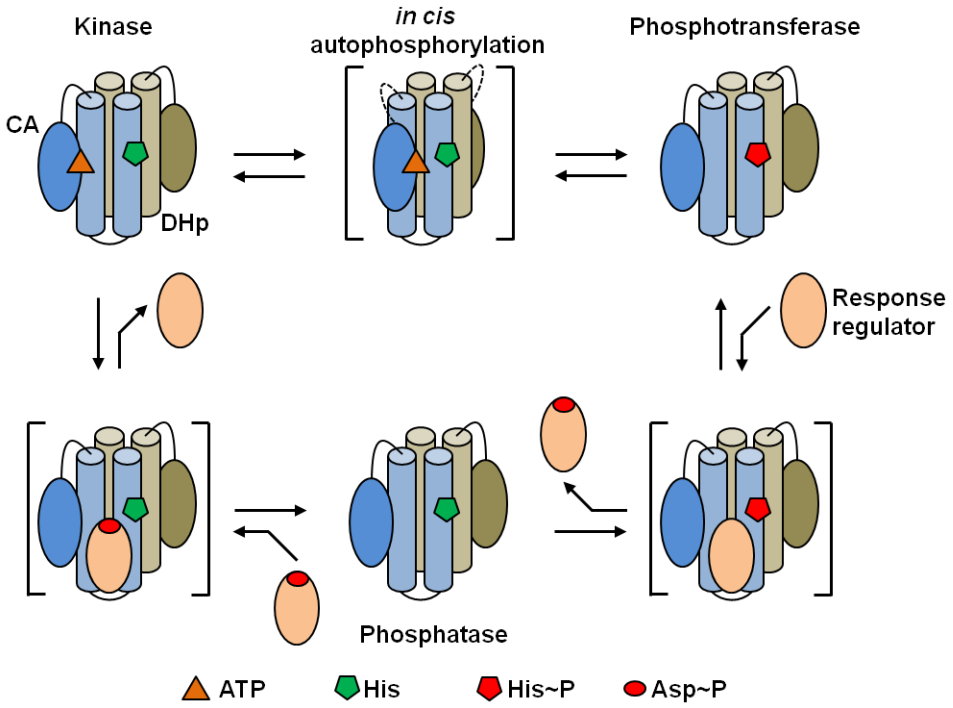


Figure 1.6 Different signaling states of histidine kinases.
Cartoon depiction of a histidine kinase switching between three enzymatic states: kinase, phosphotransferase, and phosphatase. The square brackets denote transition states in the reactions. In the kinase state, the CA domain binds ATP and orients itself near the histidine residue on the DHp domain. *In cis* autophosphorylation is depicted. In the phosphotransferase state, the response regulator binds the phosphorylated histidine kinase and accepts the phosphoryl group onto an aspartate residue. In the phosphatase state, the histidine kinase can bind the phosphorylated response regulator and catalyze hydrolysis of the phosphoryl group in Asp~P. With these latter two states, the CA domain is positioned away to allow the response regulator to dock against the DHp domain. Adapted from Figure 3 in (Gao and Stock, 2009).

conformational differences among the kinase and phosphatase states are positioning of the CA domain and the packing of the membrane-proximal half of the DHp helix bundle, where the signal is first received (Fig 1.6) (Albanesi et al., 2009; Casino et al., 2009). In the phosphatase states for both DesK and HK853, the CA domain is sequestered away from the phosphorylation site histidine on the DHp domain, preventing further autophosphorylation cycles and allowing space for a phosphorylated response regulator to bind. The CA domain is sequestered when small rotations in the DHp helix bundle expose hydrophobic core residues that can interact with hydrophobic residues in the lid covering the CA domain's ATP-binding cavity. In the kinase

state, these DHp core residues are not exposed and the interactions are not made, allowing the CA domain more mobility to approach the phosphorylation site. Signal from upstream domains can promote either the kinase or phosphatase states. For example, increased osmolarity switches on EnvZ's kinase activity, and high phosphate switches on PhoR's phosphatase activity (Carmany et al., 2003; Igo et al., 1989). Considering that dozens of different sensory domains are seen linked to the conserved catalytic transmitter region, histidine kinases must exhibit significant modularity. This modularity is further underscored by a few rationally engineered kinase chimeras where foreign sensory domains were introduced. Membrane receptor kinase EnvZ, normally activated by osmolytes, has been made responsive to aspartate and to light (Levskaya et al., 2005; Utsumi et al., 1989). Aspartate sensitivity was introduced through fusing the sensory domain and transmembrane helices from the aspartate chemoreceptor Tar to the transmitter region of EnvZ, and light sensitivity was introduced by fusing the phytochrome from Cph1. Similar activity has also been achieved in soluble cytoplasmic kinases by fusing a light-sensitive PAS domain to the transmitter region of kinase FixL (Moglich et al., 2009a). The kinase's ability to accommodate so many different sensory domains suggests there are a few common principles governing signaling between the sensory and transmitter regions.

Response regulators

Response regulators consist of a conserved N-terminal receiver domain and usually a diverse C-terminal effector domain. The receiver domain forms an α/β fold with a parallel β sheet surrounded by α helices, and the phosphoryl-acceptor aspartate is found on a solvent-exposed loop (Stock et al., 2000). The first kinase-regulator complex structure showed that two receiver domains dock against the lower half of the DHp domain four-helix bundle, on opposite sides

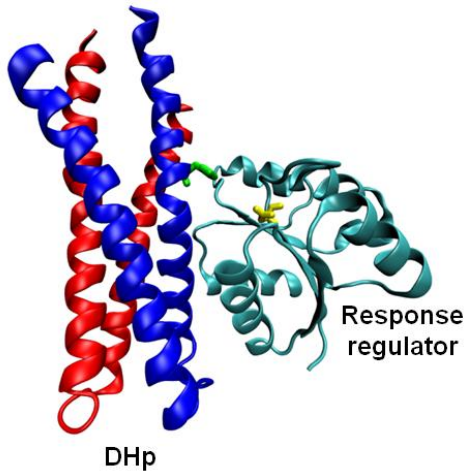


Figure 1.7 Histidine kinase in complex with response regulator.
Crystal structure of the DHp domain of *T. maritima* HK853 in complex with its cognate response regulator RR468 (PDB ID 3DGE). The CA domains of the kinase, and the second response regulator, are omitted for clarity. Helix α1 and a neighboring loop of the response regulator pack against the lower half of the four-helix bundle in the kinase DHp domain. This positions the phosphoryl-donor histidine (green) of the kinase near the phosphoryl-acceptor aspartate (yellow) of the regulator.

(Casino et al., 2009). Helix α1 and a nearby loop of the receiver domain fit between the α1 and α2 helices of the DHp domain, thereby forming a six-helix bundle (Fig 1.7). This complex is thought to represent the phosphatase state as the distance between the histidine of the DHp domain and the aspartate of the receiver domain is great enough (7.9 Å) to accommodate the water molecule necessary for hydrolysis of a phosphorylated aspartate.

Phosphorylation of the receiver domain favors its active state over its inactive state, and one common activation mechanism is dimerization of the receiver domain. For example, fourteen *E. coli* response regulators from the common OmpR/PhoB family all dimerize upon phosphorylation (Gao et al., 2008). Similarly to eukaryotic protein kinases, the inactive states of these regulators are structurally different, but the active state dimers are structurally similar (Gao et al., 2007). The receiver domain can act to positively or negatively regulate activity of the effector domain. Effector domains most commonly bind DNA, but others are protein-binding or

catalytic (Galperin, 2010).

Two-component signaling pathways

A model kinase for TCST pathways is the transmembrane receptor EnvZ from *E. coli* and I will focus on it in chapters 2 and 3. The first solved structures of a DHp domain and a CA domain were from EnvZ (Tanaka et al., 1998; Tomomori et al., 1999). EnvZ consists of a periplasmic sensory domain linked by a transmembrane helix to a HAMP domain and catalytic transmitter region. Increased osmolarity activates EnvZ, although the exact mechanism remains unclear (Jung et al., 2001; Leonardo and Forst, 1996). Upon autophosphorylation, EnvZ phosphorylates its cognate response regulator, OmpR, leading to OmpR dimerization. OmpR acts as a transcription factor through its DNA-binding effector domain and differentially regulates expression of outer-membrane porin genes *ompC* and *ompF*, thus allowing the cell to regulate its osmotic pressure (Pratt and Silhavy, 1995).

Although most TCST pathways perform a single phosphotransfer step like EnvZ-OmpR, a few systems, known as phosphorelays, have expanded to include an additional two phosphotransfer steps (Fig 1.3) (Stock et al., 2000). Perhaps the most understood phosphorelay directs the entry to sporulation in *B. subtilis* (Buelow and Raivio, 2010). Environmental signals promote autophosphorylation on the histidine in several sensor kinases (KinA-KinE). These kinases phosphorylate the aspartate in response regulator Spo0F, which then transfers the phosphoryl group to a histidine in histidine phosphotransferase Spo0B. In the final transfer, the phosphoryl group from Spo0B passes to the aspartate in response regulator Spo0A, which controls many sporulation genes. This extended pathway allows extra points for signal control. Rap and Spo0E family proteins act as phosphatases towards regulators Spo0F and Spo0A, respectively. KipI and

Sda act as autophosphorylation inhibitors against KinA and KinB, respectively, by binding their DHp domains. These extra regulators are controlled by signals relevant to the sporulation decision, such as competence, blocks in DNA replication, or nutrient availability (Buelow and Raivio, 2010).

Up to this point I have discussed individual TCST pathways, but these pathways are highly amplified, with histidine kinases and response regulators making up the largest family of bacterial signaling proteins (Galperin, 2005). A typical bacterial cell contains dozens, and sometimes more than a hundred, of these pathways. For example *Geobacter sulfurreducens* encodes 92 kinases (2.7% of its genes) and 112 regulators (Galperin, 2005). Presumably there is pressure for pathway insulation, as cross-talk between pathways would lead to an input stimulus for one pathway activating an inappropriate cellular response from another pathway (Laub and Goulian, 2007). To avoid cross-talk, the proteins in the pathways must form specific protein-protein interactions, but this is challenging as kinases and regulators are structurally similar paralogs. Although specificity in principle could be achieved through a contextual strategy where proteins that would otherwise interact are kept apart in space and time, increasing experimental evidence demonstrates that specificity is achieved through molecular recognition (Laub and Goulian, 2007; Ubersax and Ferrell, 2007).

Molecular recognition in two-component signaling pathways

Molecular recognition has been demonstrated across a wide set of protein families including eukaryotic bZIP transcription factors, SH2 domains, PDZ domains, and Bcl-2 family proteins (Chen et al., 2005; Jones et al., 2006; Newman and Keating, 2003; Stiffler et al., 2007). bZIPs contain leucine-zipper coiled coils with a characteristic seven-residue heptad repeat, and the

coiled coils mediate formation of homodimers or heterodimers. Using protein arrays, all possible pairings in 49 human bZIPs were measured and significant protein interaction specificity was observed (Newman and Keating, 2003). This specificity arises through combinations of hydrophobic and electrostatic residues at the dimerization interfaces (Vinson et al., 2006).

Similarly, molecular recognition has been established in both histidine kinase-response regulator phosphotransfer and in response regulator dimerization. In a bacterial cell, there are dozens of response regulators a given histidine kinase could phosphorylate, however typically only a single response regulator is targeted. Phosphotransfer can be assayed *in vitro* by purifying histidine kinases and response regulators and tracking protein phosphorylation status using [γ-^{32}P]ATP. Measuring phosphotransfer from a histidine kinase to all possible response regulators in the genome demonstrated that kinases have a clear kinetic preference for phosphotransfer to their *in vivo* cognate regulators (Skerker et al., 2005). For example, *E. coli* EnvZ preferentially phosphorylated OmpR over the other 31 response regulators found in *E. coli*. Response regulator dimerization shows a similar degree of interaction specificity; although many possible heterodimers could form, only a few do and there is significant homodimerization specificity. Dimerization can be measured by purifying regulators with N-terminal fusions of CFP and YFP, and monitoring fluorescence resonance energy transfer (FRET) between regulator pairs (Gao et al., 2008). Measurements of all possible dimers among the 14 OmpR/PhoB response regulators in *E. coli* demonstrated each regulator specifically formed homodimers.

Given the extent of molecular recognition found in TCST pathways, histidine kinase dimerization might also be expected to be highly specific. Cross-regulation between different TCST pathways has been observed in only a limited number of cases, suggesting that histidine

kinases should preferentially form homodimers over heterodimers (Laub and Goulian, 2007). Histidine kinases have been shown to homodimerize *in vitro* and *in vivo,* but whether these interactions are specific has not been addressed (Hidaka et al., 1997; Scheu et al., 2010). If kinase dimerization were not specific *in vitro*, this would suggest a contextual strategy could be in effect. If dimerization were specific, which parts of the kinase contribute to specificity would be an open question. Presumably, dimerization specificity determinants would be located at the dimerization interface, but this interface is large and can extend across several domains from the DHp to upstream HAMP or PAS domains. Specificity determinants could be spread along the dimerization interface, or localized. Knowing the number and location of specificity determinants could guide our understanding of how recently diverged paralogs achieve pathway insulation. The biochemical identification of these determinants can be directed by evolutionary sequence analyses of the tens of thousands available histidine kinase sequences.

Autophosphorylation mechanism in histidine kinases

Autophosphorylation in histidine kinase dimers is a central function, but whether phosphorylation of one DHp chain versus another is functionally relevant is unclear. Both *cis* and *trans* mechanisms lead to the same product, a phosphorylated histidine. In addition, the structural determinants that specify one mechanism over the other are unknown. One proposed determinant is the loops connecting the helices in the DHp domain, which could set the position of the CA domain relative to the DHp domain (Casino et al., 2009). Depending on whether this connection is left-handed or right-handed, the CA domain in a subunit of the dimer could either phosphorylate the DHp domain of the same subunit (*cis*) or the opposite subunit (*trans*). Loop sequences even within histidine kinase orthologs are quite diverse so one consequence of the

loop connection model may be autophosphorylation mechanism diversity across orthologs. The presence of such diversity in orthologs would argue against significant functional differences between the two mechanisms. Alternatively, conservation in orthologs could suggest selection against changing autophosphorylation mechanism.

Identifying determinants of molecular function

Identifying the determinants of dimerization specificity or autophosphorylation in histidine kinases requires methods to interrogate protein function. Determining the function of a protein and how such function is encoded is a challenging problem. Even in a well-studied model organism like yeast, ~25% of genes remain unannotated, and this is most certainly an underestimate of our functional knowledge as many genes have more than one function (Hibbs et al., 2009). The main approach to function prediction is looking for homology to proteins of known function, either at the sequence or structural level (Lee et al., 2007). The performance of sequence homology methods is dictated by the sequence identity between the protein of known function and the unknown protein. For example, confident enzyme annotation requires 40-60% sequence identity (Tian and Skolnick, 2003). However there can be significant exceptions to such guidelines. TriA from *Pseudomonas* sp. strain NRRL B-12227 is 98% identical to AtzA from *Pseudomonas* sp. strain ADP, but the former catalyzes deamination whereas the latter catalyzes dehalogenation (Seffernick et al., 2001).

Sequence homology methods can often predict a protein's general function by assigning it to a protein family. For example, the Pfam database can usually identify a new histidine kinase as such because the protein will be predicted to be a member of the HisKA family (DHp domain) and the HATPase_c family (CA domain) (Punta et al., 2012). However within a protein

family, the specific details of molecular function, such as substrate specificity, almost always vary. For example, EnvZ and its paralog, RstB, are both histidine kinases, but each phosphorylates a different response regulator. Sequence homology methods are less successful at predicting molecular details such as substrate specificity or protein partner because less functional annotation exists at these more detailed levels, and also there are typically fewer residues needed to specify such functions than there are needed to specify the overall family. In the next section, I give a non-comprehensive overview of methods for dissecting molecular function within a protein family, with a focus on biochemical and evolutionary approaches.

Biochemical approaches

Below, four biochemical approaches for identifying molecular function determinants within a protein family are discussed: (1) mutagenesis, (2) characterizing chimera function, (3) screening mutant libraries, and (4) characterizing ortholog function. All these methods, except for the third, are used in the later chapters to study kinase dimerization specificity and autophosphorylation.

Mutagenesis has been used to probe every protein function, from protein folding to enzymatic activity and protein interactions. One folding study characterized the extent to which mutations could be tolerated in the hydrophobic core of the N-terminal domain of λ repressor (Lim et al., 1992). Guided by structure, core positions were mutated to change steric complementarity (V→L), buried volume (V→F), or polarity (V→D). All of the mutants that retained the hydrophobic nature of the core were still able to fold, although there were conformational changes as several mutants had weakened binding to an antibody against the repressor. In contrast, the polar mutants likely formed molten globules. This study demonstrated that mutations in the protein core disrupt folding to different extents, with polar mutations being most

detrimental. A corollary is that when such mutations are observed in natural sequences, they may need to be accompanied by compensatory mutations at neighboring sites.

Alanine scanning has been used extensively to either discover the location of protein-protein interfaces or isolate residues that significantly contribute to the binding interface. An alanine scan of the response regulator Spo0F was used to predict residues contributing to its interaction with histidine phosphotransferase Spo0B, and a subsequent structure of the complex verified many of these predictions (Tzeng and Hoch, 1997; Zapf et al., 2000). An alanine scan was also conducted on 33 solvent-exposed residues in the DHp domain of EnvZ to identify residues important for autophosphorylation or for phosphotransfer (Capra et al., 2010). Although several mutations weakened autophosphorylation, only one mutation significantly weakened phosphotransfer to the cognate regulator OmpR. This indicates mutations at several residues are likely needed to change phosphotransfer specificity.

In another form of mutagenesis, regions from two different proteins, often homologs, are combined together and the resulting chimera's function is determined. By comparing the chimera's function to that of its two parent proteins, the protein region controlling the function of interest is identified. In an early study on interaction specificity, the chimera approach was used to identify specificity determinants in the heterodimer formed between bZIPs Fos and Jun (O'Shea et al., 1992). Two types of chimeras were made in the sequence context of bZIP homodimer GCN4. The "inside" chimeras replaced the core residues of GCN4 with the corresponding core residues from Jun or Fos. The "outside" chimeras replaced the surface residues of GCN4 with the corresponding surface residues from Jun or Fos. The two inside chimeras specifically heterodimerized, like Fos and Jun, whereas the two outside chimeras

showed no preference for forming a heterodimer over homodimers. This demonstrated the crucial role of core residues in specificity, and further chimeras showed that the electrostatic core residues in particular contributed greatly to specificity. A similar chimera approach to look at phosphotransfer specificity in histidine kinases showed that the DHp domain, rather than the CA domain, determined substrate partnering (Skerker et al., 2008).

Screening a library of protein variants for a desired function allows exploration of much larger sequence spaces relative to directed mutagenesis. This approach becomes especially informative when combined with high-throughput sequencing. In one case, a library of ~600,000 variants of a human WW domain was built, and proteins able to bind a peptide partner were selected using phage display (Fowler et al., 2010). The central 33 residues in the WW domain were varied using chemical DNA synthesis, and the initial library included 83% of all possible single mutants and 27% of all possible double mutants, according to library deep sequencing. After six rounds of moderate selection for peptide binding, 16% of the initial variants remained. To classify a site as permissive vs. intolerant to mutation, the frequency of observing an amino acid at a given site after selection was compared to the frequency before selection. The resulting enrichment ratios gave data analogous to an all-residue scan. For example, mutation to proline was selected against across the entire sequence, whereas mutation to positively charged residues was favored in the second loop. In general, mutations were not tolerated in the β sheet that binds the peptide, but were permitted in the loops connecting the strands. This type of study informs both protein engineering efforts and efforts to understand what sequence space evolving proteins may occupy.

Instead of directly introducing mutations and tracking function changes, an orthogonal approach is to examine how natural sequence variation across homologous proteins is related to changes in

function. This approach was used to study how protein kinase R (PKR) in some primates can avoid interaction with the vaccinia poxvirus protein K3L (Elde et al., 2009). During viral infection, PKR senses double-stranded RNA and phosphorylates the translation initiation factor eIF2α, stopping protein synthesis and spread of virus. Viral K3L is a mimic of eIF2α and can bind PKR, allowing for viral propagation. To prevent interaction with K3L, PKR's interaction interface must evolve to avoid binding K3L while maintaining recognition of eIF2α. The interactions between 20 primate PKR orthologs and vaccinia K3L were measured to identify possible strategies for resisting K3L mimicry. PKR orthologs from Old World and New World monkeys interacted with K3L, but all orthologs from hominoids, except from gibbon, did not. The close relation between gibbon and other hominoid PKR orthologs narrowed down and highlighted the sequence differences responsible for the K3L interaction changes between gibbon and other hominoids.

Evolutionary approaches

Analysis of a protein's evolutionary record can serve as a powerful complement to experimental approaches. Evolutionary methods, in particular comparative sequence analysis, have become increasingly powerful as protein sequence databases rapidly expand (Yang, 2005). There can be many sequence differences across homologous proteins, resulting from a combination of purifying selection, positive selection, and genetic drift. Residues in a protein experiencing selection will be important for protein function, so methods of identifying such residues are needed. Below, we discuss methods for identifying protein residues that are (1) conserved, (2) under positive selection, or (3) coevolving with other residues in the protein.

For most proteins, several functions are usually under strong purifying selection, including

folding and maintenance of a particular structure, the ability to carry out an enzymatic activity, and the ability to interact with various protein partners. Amino acids critical for these functions are often highly conserved in homologous proteins, as substitutions at these positions will often disrupt function and be eliminated by natural selection. For example, the histidine in the H-box of the kinase DHp domain is conserved across virtually all histidine kinases because mutation of this residue eliminates kinase activity. Conserved residues will not be informative, however, when considering how molecular function varies within a protein family. Residues contributing to protein interaction specificity, or autophosphorylation mechanism specificity, must vary in order to distinguish family member functions.

Homologous protein sequences may vary due to positive selection. Genes under positive selection are most commonly involved in host-pathogen interactions, molecular recognition during fertilization, or neofunctionalization after gene duplication (Yang, 2005). Positive selection between a pair of genes is typically measured as the ratio of the rates of non-synonymous (dN) and synonymous (dS) substitutions, which reflects the balance of selection and genetic drift. $dN/dS > 1$ indicates positive selection, or relaxed purifying selection, and this ratio can be calculated between a pair of genes, within specific lineages of a phylogeny, or at individual sites in a gene. Measurement of positive selection in PKR was used to complement the ortholog characterization studies discussed above (Elde et al., 2009). Many of the sites in PKR experiencing the greatest positive selection were at the interface with eIF2α, and therefore likely to contact viral K3L. In a chimera mutation experiment, these sites were mutated in gibbon PKR to their corresponding identities in human PKR, and the resulting protein no longer interacted with vaccinia K3L. In the converse experiment, the same sites were mutated in human PKR to

their corresponding identities in gibbon PKR, and that chimera then interacted with K3L. Only one to three sites needed to be mutated to achieve these interaction changes, showing that positive selection can quickly rewire protein interfaces. Unfortunately, the signatures of positive selection can be obscured in genes that diverged a long time ago, due to purifying selection taking place after the adaptive changes (Capra et al., 2012). This may be an issue with histidine kinase families, many of which have diverged for hundreds of millions of years.

Patterns of correlated substitutions, or amino acid covariation, can also provide insight into functionally important and related sets of residues (Fitch and Markowitz, 1970; Korber et al., 1993). Functionally related residues in a protein will often coevolve, and the extent of this coevolution can be quantified from a sufficiently large multiple sequence alignment of homologous proteins. The identification of residues that coevolve in two interacting proteins can highlight residues important for complex formation; if a mutation in one of the proteins destabilizes the complex, an interaction can often be restored through a compensatory mutation in the other protein. Scenarios in which an initial mutation is neutral, rather than deleterious, but creates a favorable context for an otherwise deleterious mutation have also been suggested (Ortlund et al., 2007).

The most strongly covarying pairs of residues within a protein, or in a pair of interacting proteins, are often in physical contact (Morcos et al., 2011). Consider the interaction between two proteins A and B involving a glycine at position 2 and a phenylalanine at position 5 (Fig. 1.8). A mutation of the glycine in protein A to phenylalanine would result in a steric clash, thereby weakening the interaction; however, a subsequent mutation in protein B from a phenylalanine to a glycine may compensate and restore the interaction. Covariation between

these positions can be detected by examining the identity of residues at these positions in homologs of proteins A and B.

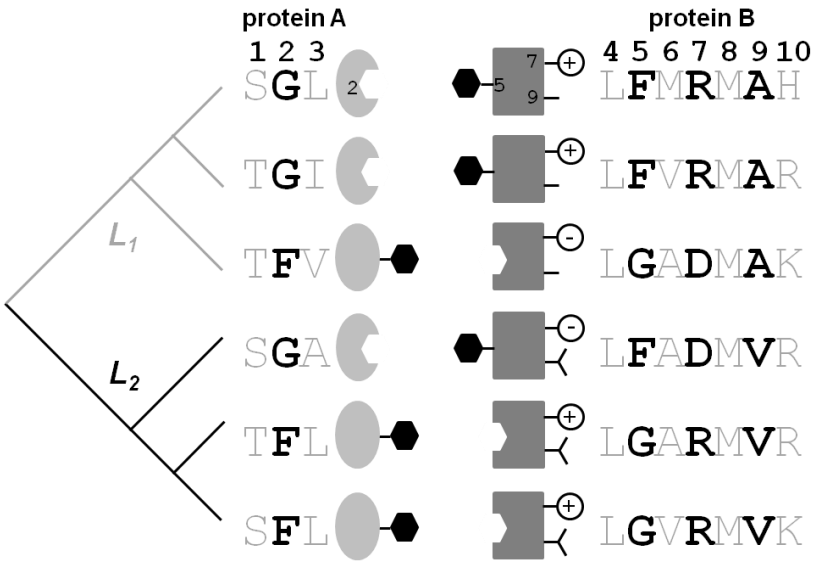


Figure 1.8 Covariation between interacting proteins.
Protein A interacts with Protein B. Each row of the multiple sequence alignment consists of an interacting pair of protein A and protein B homologs from a single species. The phylogenetic tree indicates that the protein sequences come from species in two different lineages: L_1 (grey branches) and L_2 (black branches). Positions 2 and 5 covary, thereby preserving a hydrophobic packing interaction at the protein-protein interface, whereas positions 2 and 7 are independent of one another. Positions 5 and 9 also covary, but this signal comes from shared phylogeny in protein B homologs.

Many different scoring metrics for covariation exist, with one of the most widely-used being mutual information (MI) (Korber et al., 1993). To calculate MI between columns *i* and *j* in a multiple sequence alignment, both single and joint probability distributions of amino acids are used. Let p_x^i and p_y^j be the probabilities (or frequencies) of amino acid *x* in column *i* and amino acid *y* in column *j*, respectively, and let $p_{xy}^{i,j}$ be the frequency of simultaneously finding residue *x* in column *i* and residue *y* in column *j* of an alignment. Given *n* different types of amino acids in column *i* and *m* different types of amino acids in column *j*,

$$MI(i,j) = \sum_{x=1}^{n}\sum_{y=1}^{m} p_{xy}^{i,j} \log \frac{p_{xy}^{i,j}}{p_x^i p_y^j} \quad (1)$$

Mutual information measures how much the uncertainty at one position is reduced given information about another position. For example, consider the alignment in Figure 1.8 where the amino acid distribution at position 2 is $p_G^2 = 0.5$ and $p_F^2 = 0.5$, at position 5 is $p_G^5 = 0.5$ and $p_F^5 = 0.5$, and at position 7 is $p_R^7 = 0.67$ and $p_D^7 = 0.33$. If position 2 is a G, position 5 is always an F, whereas if position 2 is an F, position 5 is always a G. This covariation is reflected in the joint probability distribution term $p_{xy}^{i,j}$, where $p_{GF}^{2,5} = 0.5$, $p_{FG}^{2,5} = 0.5$, $p_{GG}^{2,5} = 0$ and $p_{FF}^{2,5} = 0$. Knowing the residue at position 2 removes all uncertainty at position 5, resulting in a large MI score for this pair. In contrast, consider the MI score for positions 2 and 7. Knowing whether position 2 is an F or G does not reduce the uncertainty as to whether R or D occupies position 7. Consequently the MI score is zero as the joint probability distribution $p_{xy}^{2,7}$ is simply the product of probability distributions p_x^2 and p_y^7.

In many sequence alignments, there will often be significant covariation signal even when two columns in a sequence alignment are functionally independent of one another. This background signal can arise from statistical noise and, sometimes, from the shared ancestry of the sequences (Wollenberg and Atchley, 2000). The statistical noise, a form of sampling bias, stems from the fact that columns in an alignment are being used to estimate an amino acid probability distribution; hence, too few sequences in the alignment can produce inaccurate estimates of the true probability distribution. The shared ancestry, or phylogenetic signal, arises because no pair of positions in a protein can ever be truly independent as they are ultimately derived from a

common ancestor and have been inherited as a unit. As an example, consider position 9 in the alignment of Figure 1.8, which is not functionally coupled to the other positions, but due to drift is an A in lineage L_1 and a V in lineage L_2. This phylogenetic pattern results in a relatively high MI score between positions 5 and 9, even though no functional coupling exists between this pair of positions. Many methods have been proposed to correct for the signals arising from statistical and phylogenetic noise (Brown and Brown, 2010; Dunn et al., 2008; Weigt et al., 2009).

Analyses of amino acid coevolution in large sets of cognate kinase-regulator pairs led to the identification of a small set of strongly covarying residues (Capra et al., 2010; Skerker et al., 2008) (Fig. 1.9A,B). The majority of these residue pairs were in direct contact at the kinase-regulator interaction interface, suggesting that the pairs were phosphotransfer specificity determinants (Casino et al., 2009) (Fig. 1.9C). The coevolving residues were further validated as critical determinants of specificity by rationally rewiring the specificity of model two-component signaling proteins. The three strongest covarying positions in kinase EnvZ were mutated to match the residues found at the equivalent positions in the kinase RstB (Skerker et al., 2008). Strikingly, this triple mutant of EnvZ specifically phosphorylated RstA, the cognate substrate of RstB, rather than OmpR, the cognate substrate of wild-type EnvZ, indicating that phosphotransfer specificity had been completely rewired (Fig. 1.9D). Similar rewiring experiments demonstrated that substituting the highly covarying residues in OmpR with those in RstA produced a regulator that was preferentially phosphorylated by RstB rather than EnvZ (Capra et al., 2010). These mutagenesis experiments clearly demonstrated that the covarying pairs function as phosphotransfer specificity determinants.

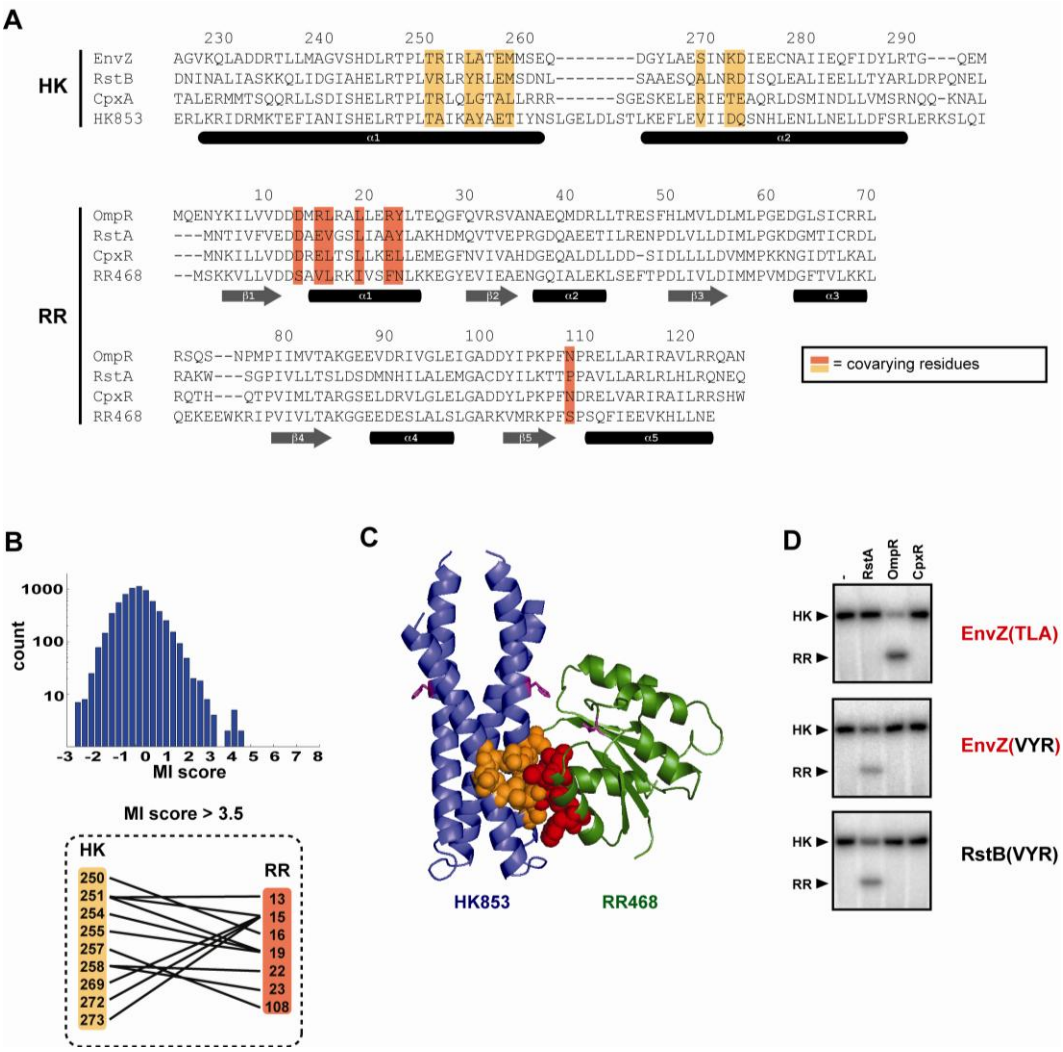


Figure 1.9. Amino-acid coevolution analysis.
(A)-(D) Coevolution analysis applied to the problem of identifying phosphotransfer specificity determinants in two-component signal transduction proteins. (A) A set of cognate histidine kinases (HK) and response regulators (RR) was retrieved by searching across all available prokaryotic genomes with HMMER. These sequences were aligned using HMMER, and the resulting alignment was filtered to eliminate pairwise sequence redundancy and gapped columns. Covariation was then measured using adjusted mutual information. The multiple sequence alignment shown includes a subset of cognate kinase-regulator pairs and highlights significantly coevolving residues in the DHp domains of histidine kinases and the receiver domains of response regulators. Secondary structure elements of the kinases and regulators are labeled underneath the sequences and assigned using the *T. maritima* kinase-regulator co-crystal structure. Significant covarying pairs are identified in the tail of the covariation score distribution and then grouped into clusters (B). Pairs are mapped onto the kinase-regulator co-crystal structure (C) with covarying residues highlighted as in (A) and shown in space-filling form. (D) Rewiring phosphotransfer specificity through rational mutagenesis. Phosphotransfer specificity was measured by autophosphorylating each kinase with [γ^{32}P]-ATP and then incubating the phosphorylated kinase with a response regulator (RstA, OmpR, or CpxR). EnvZ, labeled TLA based on the identities of the three specificity residues T250, L254, and A255, specifically phosphotransfers to OmpR (top panel) whereas RstB, labeled VYR based on the identities of its three specificity residues, specifically phosphotransfers to RstA (bottom panel). Mutating the specificity residues in EnvZ to match those found in RstB rewires phosphotransfer specificity from OmpR to RstA (middle panel).

In the following chapters, I apply many of the biochemical and evolutionary approaches discussed above to answer questions about molecular function in histidine kinases. I focus on how homodimerization specificity is encoded, and how a *cis* vs. *trans* autophosphorylation mechanism is specified. In chapter 2, I develop an *in vitro* binding assay and demonstrate that histidine kinase paralogs EnvZ and RstB specifically homodimerize. By measuring amino acid coevolution within kinases, I predict and then experimentally verify a small region of the kinase where dimerization specificity is encoded. In chapter 3, I show that changing the sequence of loops in the DHp domain of the kinase can switch the autophosphorylation mechanism between *trans* and *cis*. Furthermore, despite quite divergent loop sequences, I find that autophosphorylation mechanism is conserved within EnvZ orthologs and within PhoR orthologs, despite these orthologs having quite divergent loop sequences. This suggests the autophosphorylation mechanism may be under selection.

References

Airola, M. V., Watts, K. J., Bilwes, A. M., and Crane, B. R. (2010). Structure of concatenated HAMP domains provides a mechanism for signal transduction. Structure **18,** 436-48.

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. Proc Natl Acad Sci U S A **106,** 16185-90.

Aldridge, B. B., Burke, J. M., Lauffenburger, D. A., and Sorger, P. K. (2006). Physicochemical modelling of cell signalling pathways. Nat Cell Biol **8,** 1195-203.

Alm, E., Huang, K., and Arkin, A. (2006). The evolution of two-component systems in bacteria reveals different strategies for niche adaptation. PLoS Comput Biol **2,** e143.

Alroy, I., and Yarden, Y. (1997). The ErbB signaling network in embryogenesis and oncogenesis: signal diversification through combinatorial ligand-receptor interactions. FEBS Lett **410,** 83-6.

Blume-Jensen, P., and Hunter, T. (2001). Oncogenic kinase signalling. Nature **411,** 355-65.

Brown, C. A., and Brown, K. S. (2010). Validation of coevolving residue algorithms via pipeline sensitivity analysis: ELSC and OMES and ZNMI, oh my! PLoS One **5,** e10779.

Buelow, D. R., and Raivio, T. L. (2010). Three (and more) component regulatory systems - auxiliary regulators of bacterial histidine kinases. Mol Microbiol **75,** 547-66.

Capra, E. J., and Laub, M. T. (2012). Evolution of Two-Component Signal Transduction Systems. Annu Rev Microbiol.

Capra, E. J., Perchuk, B. S., Lubin, E. A., Ashenberg, O., Skerker, J. M., and Laub, M. T. (2010). Systematic dissection and trajectory-scanning mutagenesis of the molecular interface that ensures specificity of two-component signaling pathways. PLoS Genet **6,** e1001220.

Capra, E. J., Perchuk, B. S., Skerker, J. M., and Laub, M. T. (2012). Adaptive Mutations that Prevent Crosstalk Enable the Expansion of Paralogous Signaling Protein Families. Cell **150,** 222-32.

Carmany, D. O., Hollingsworth, K., and McCleary, W. R. (2003). Genetic and biochemical studies of phosphatase activity of PhoR. J Bacteriol **185,** 1112-5.

Casino, P., Rubio, V., and Marina, A. (2009). Structural insight into partner specificity and phosphoryl transfer in two-component signal transduction. Cell **139,** 325-36.

Chen, L., Willis, S. N., Wei, A., Smith, B. J., Fletcher, J. I., Hinds, M. G., Colman, P. M., Day, C. L., Adams, J. M., and Huang, D. C. (2005). Differential targeting of prosurvival Bcl-2 proteins by their BH3-only ligands allows complementary apoptotic function. Mol Cell **17,** 393-403.

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

Dunn, S. D., Wahl, L. M., and Gloor, G. B. (2008). Mutual information without the influence of phylogeny or entropy dramatically improves residue contact prediction. Bioinformatics **24,** 333-40.

Elde, N. C., Child, S. J., Geballe, A. P., and Malik, H. S. (2009). Protein kinase R reveals an evolutionary model for defeating viral mimicry. Nature **457,** 485-9.

Endicott, J. A., Noble, M. E., and Johnson, L. N. (2012). The structural basis for control of eukaryotic protein kinases. Annu Rev Biochem **81,** 587-613.

Falke, J. J., and Hazelbauer, G. L. (2001). Transmembrane signaling in bacterial chemoreceptors.

Trends Biochem Sci **26,** 257-65.
Ferris, H. U., Dunin-Horkawicz, S., Hornig, N., Hulko, M., Martin, J., Schultz, J. E., Zeth, K., Lupas, A. N., and Coles, M. (2012). Mechanism of regulation of receptor histidine kinases. Structure **20,** 56-66.
Ferris, H. U., Dunin-Horkawicz, S., Mondejar, L. G., Hulko, M., Hantke, K., Martin, J., Schultz, J. E., Zeth, K., Lupas, A. N., and Coles, M. (2011). The Mechanisms of HAMP-Mediated Signaling in Transmembrane Receptors. Structure **19,** 378-85.
Fitch, W. M., and Markowitz, E. (1970). An improved method for determining codon variability in a gene and its application to the rate of fixation of mutations in evolution. Biochem Genet **4,** 579-93.
Fowler, D. M., Araya, C. L., Fleishman, S. J., Kellogg, E. H., Stephany, J. J., Baker, D., and Fields, S. (2010). High-resolution mapping of protein sequence-function relationships. Nat Methods **7,** 741-6.
Galperin, M. Y. (2005). A census of membrane-bound and intracellular signal transduction proteins in bacteria: bacterial IQ, extroverts and introverts. BMC Microbiol **5,** 35.
Galperin, M. Y. (2010). Diversity of structure and function of response regulator output domains. Curr Opin Microbiol **13,** 150-9.
Gao, R., Mack, T. R., and Stock, A. M. (2007). Bacterial response regulators: versatile regulatory strategies from common domains. Trends Biochem Sci **32,** 225-34.
Gao, R., and Stock, A. M. (2009). Biological insights from structures of two-component proteins. Annu Rev Microbiol **63,** 133-54.
Gao, R., Tao, Y., and Stock, A. M. (2008). System-level mapping of *Escherichia coli* response regulator dimerization with FRET hybrids. Mol Microbiol **69,** 1358-72.
Goodman, A. L., Merighi, M., Hyodo, M., Ventre, I., Filloux, A., and Lory, S. (2009). Direct interaction between sensor kinase proteins mediates acute and chronic disease phenotypes in a bacterial pathogen. Genes Dev **23,** 249-59.
Hibbs, M. A., Myers, C. L., Huttenhower, C., Hess, D. C., Li, K., Caudy, A. A., and Troyanskaya, O. G. (2009). Directing experimental biology: a case study in mitochondrial biogenesis. PLoS Comput Biol **5,** e1000322.
Hidaka, Y., Park, H., and Inouye, M. (1997). Demonstration of dimer formation of the cytoplasmic domain of a transmembrane osmosensor protein, EnvZ, of *Escherichia coli* using Ni-histidine tag affinity chromatography. FEBS Lett **400,** 238-42.
Honegger, A. M., Kris, R. M., Ullrich, A., and Schlessinger, J. (1989). Evidence that autophosphorylation of solubilized receptors for epidermal growth factor is mediated by intermolecular cross-phosphorylation. Proceedings of the National Academy of Sciences **86,** 925-929.
Huang, C. Y., and Ferrell, J. E., Jr. (1996). Ultrasensitivity in the mitogen-activated protein kinase cascade. Proc Natl Acad Sci U S A **93,** 10078-83.
Huang, H., Li, L., Wu, C., Schibli, D., Colwill, K., Ma, S., Li, C., Roy, P., Ho, K., Songyang, Z., Pawson, T., Gao, Y., and Li, S. S. (2008). Defining the specificity space of the human SRC homology 2 domain. Mol Cell Proteomics **7,** 768-84.
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,** 929-40.
Hunter, T. (2000). Signaling--2000 and beyond. Cell **100,** 113-27.
Igo, M. M., Ninfa, A. J., and Silhavy, T. J. (1989). A bacterial environmental sensor that

functions as a protein kinase and stimulates transcriptional activation. Genes Dev **3,** 598-605.

Jones, R. B., Gordus, A., Krall, J. A., and MacBeath, G. (2006). A quantitative protein interaction network for the ErbB receptors using protein microarrays. Nature **439,** 168-74.

Jung, K., Hamann, K., and Revermann, A. (2001). K+ stimulates specifically the autokinase activity of purified and reconstituted EnvZ of *Escherichia coli*. J Biol Chem **276,** 40896-902.

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res **28,** 27-30.

Kaneko, T., Sidhu, S. S., and Li, S. S. (2011). Evolving specificity from variability for protein interaction domains. Trends Biochem Sci **36,** 183-90.

Korber, B. T., Farber, R. M., Wolpert, D. H., and Lapedes, A. S. (1993). Covariation of mutations in the V3 loop of human immunodeficiency virus type 1 envelope protein: an information theoretic analysis. Proc Natl Acad Sci U S A **90,** 7176-80.

Laub, M. T., and Goulian, M. (2007). Specificity in two-component signal transduction pathways. Annu Rev Genet **41,** 121-45.

Lee, D., Redfern, O., and Orengo, C. (2007). Predicting protein function from sequence and structure. Nat Rev Mol Cell Biol **8,** 995-1005.

Lemmon, M. A., and Schlessinger, J. (2010). Cell signaling by receptor tyrosine kinases. Cell **141,** 1117-34.

Leonardo, M. R., and Forst, S. (1996). Re-examination of the role of the periplasmic domain of EnvZ in sensing of osmolarity signals in *Escherichia coli*. Mol Microbiol **22,** 405-13.

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., and Voigt, C. A. (2005). Synthetic biology: engineering *Escherichia coli* to see light. Nature **438,** 441-2.

Lim, W. A., Farruggio, D. C., and Sauer, R. T. (1992). Structural and energetic consequences of disruptive mutations in a protein core. Biochemistry **31,** 4324-33.

Lochhead, P. A. (2009). Protein kinase activation loop autophosphorylation in cis: overcoming a Catch-22 situation. Sci Signal **2,** pe4.

Low-Nam, S. T., Lidke, K. A., Cutler, P. J., Roovers, R. C., van Bergen en Henegouwen, P. M., Wilson, B. S., and Lidke, D. S. (2011). ErbB1 dimerization is promoted by domain co-confinement and stabilized by ligand binding. Nat Struct Mol Biol **18,** 1244-9.

Lowenstein, E. J., Daly, R. J., Batzer, A. G., Li, W., Margolis, B., Lammers, R., Ullrich, A., Skolnik, E. Y., Bar-Sagi, D., and Schlessinger, J. (1992). The SH2 and SH3 domain-containing protein GRB2 links receptor tyrosine kinases to ras signaling. Cell **70,** 431-42.

Macek, B., Gnad, F., Soufi, B., Kumar, C., Olsen, J. V., Mijakovic, I., and Mann, M. (2008). Phosphoproteome analysis of *E. coli* reveals evolutionary conservation of bacterial Ser/Thr/Tyr phosphorylation. Mol Cell Proteomics **7,** 299-307.

Manning, G., Whyte, D. B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002). The protein kinase complement of the human genome. Science **298,** 1912-34.

Marina, A., Waldburger, C. D., and Hendrickson, W. A. (2005). Structure of the entire cytoplasmic portion of a sensor histidine-kinase protein. Embo J **24,** 4247-59.

Miller, W. T. (2012). Tyrosine kinase signaling and the emergence of multicellularity. Biochim Biophys Acta **1823,** 1053-7.

Moglich, A., Ayers, R. A., and Moffat, K. (2009a). Design and signaling mechanism of light-regulated histidine kinases. J Mol Biol **385,** 1433-44.

Moglich, A., Ayers, R. A., and Moffat, K. (2009b). Structure and signaling mechanism of Per-ARNT-Sim domains. Structure **17,** 1282-94.

Morcos, F., Pagnani, A., Lunt, B., Bertolino, A., Marks, D. S., Sander, C., Zecchina, R., Onuchic, J. N., Hwa, T., and Weigt, M. (2011). Direct-coupling analysis of residue coevolution captures native contacts across many protein families. Proc Natl Acad Sci U S A **108,** E1293-301.

Neiditch, M. B., Federle, M. J., Pompeani, A. J., Kelly, R. C., Swem, D. L., Jeffrey, P. D., Bassler, B. L., and Hughson, F. M. (2006). Ligand-induced asymmetry in histidine sensor kinase complex regulates quorum sensing. Cell **126,** 1095-108.

Newman, J. R., and Keating, A. E. (2003). Comprehensive identification of human bZIP interactions with coiled-coil arrays. Science **300,** 2097-101.

Ninfa, E. G., Atkinson, M. R., Kamberov, E. S., and Ninfa, A. J. (1993). Mechanism of autophosphorylation of *Escherichia coli* nitrogen regulator II (NRII or NtrB): trans-phosphorylation between subunits. J Bacteriol **175,** 7024-32.

Noble, M. E. M., Endicott, J. A., and Johnson, L. N. (2004). Protein Kinase Inhibitors: Insights into Drug Design from Structure. Science **303,** 1800-1805.

O'Shea, E. K., Rutkowski, R., and Kim, P. S. (1992). Mechanism of specificity in the Fos-Jun oncoprotein heterodimer. Cell **68,** 699-708.

Ortlund, E. A., Bridgham, J. T., Redinbo, M. R., and Thornton, J. W. (2007). Crystal structure of an ancient protein: evolution by conformational epistasis. Science **317,** 1544-8.

Pratt, L. A., and Silhavy, T. J. (1995). Porin regulon of *Escherichia coli*. *In* "Two-component signal transduction" (J. A. Hoch and T. J. Silhavy, Eds.), pp. 105-127. ASM Press, Washington, D. C.

Punta, M., Coggill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., Heger, A., Holm, L., Sonnhammer, E. L., Eddy, S. R., Bateman, A., and Finn, R. D. (2012). The Pfam protein families database. Nucleic Acids Res **40,** D290-301.

Scheu, P. D., Liao, Y. F., Bauer, J., Kneuper, H., Basche, T., Unden, G., and Erker, W. (2010). Oligomeric sensor kinase DcuS in the membrane of *Escherichia coli* and in proteoliposomes: chemical cross-linking and FRET spectroscopy. J Bacteriol **192,** 3474-83.

Seffernick, J. L., de Souza, M. L., Sadowsky, M. J., and Wackett, L. P. (2001). Melamine deaminase and atrazine chlorohydrolase: 98 percent identical but functionally different. J Bacteriol **183,** 2405-10.

Sevvana, M., Vijayan, V., Zweckstetter, M., Reinelt, S., Madden, D. R., Herbst-Irmer, R., Sheldrick, G. M., Bott, M., Griesinger, C., and Becker, S. (2008). A ligand-induced switch in the periplasmic domain of sensor histidine kinase CitA. J Mol Biol **377,** 512-23.

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,** 1043-54.

Skerker, J. M., Prasol, M. S., Perchuk, B. S., Biondi, E. G., and Laub, M. T. (2005). Two-component signal transduction pathways regulating growth and cell cycle progression in a bacterium: a system-level analysis. PLoS Biol **3,** e334.

Songyang, Z., Shoelson, S. E., McGlade, J., Olivier, P., Pawson, T., Bustelo, X. R., Barbacid, M., Sabe, H., Hanafusa, H., Yi, T., and et al. (1994). Specific motifs recognized by the SH2 domains of Csk, 3BP2, fps/fes, GRB-2, HCP, SHC, Syk, and Vav. Mol Cell Biol **14,**

2777-85.
Stiffler, M. A., Chen, J. R., Grantcharova, V. P., Lei, Y., Fuchs, D., Allen, J. E., Zaslavskaia, L. A., and MacBeath, G. (2007). PDZ domain binding selectivity is optimized across the mouse proteome. Science **317,** 364-9.
Stock, A. M., Robinson, V. L., and Goudreau, P. N. (2000). Two-component signal transduction. Annu Rev Biochem **69,** 183-215.
Tanaka, T., Saha, S. K., Tomomori, C., Ishima, R., Liu, D., Tong, K. I., Park, H., Dutta, R., Qin, L., Swindells, M. B., Yamazaki, T., Ono, A. M., Kainosho, M., Inouye, M., and Ikura, M. (1998). NMR structure of the histidine kinase domain of the *E. coli* osmosensor EnvZ. Nature **396,** 88-92.
Taylor, S. S., and Kornev, A. P. (2011). Protein kinases: evolution of dynamic regulatory proteins. Trends Biochem Sci **36,** 65-77.
Tedford, N. C., White, F. M., and Radding, J. A. (2008). Illuminating signaling network functional biology through quantitative phosphoproteomic mass spectrometry. Brief Funct Genomic Proteomic **7,** 383-94.
Tian, W., and Skolnick, J. (2003). How well is enzyme function conserved as a function of pairwise sequence identity? J Mol Biol **333,** 863-82.
Tomomori, C., Tanaka, T., Dutta, R., Park, H., Saha, S. K., Zhu, Y., Ishima, R., Liu, D., Tong, K. I., Kurokawa, H., Qian, H., Inouye, M., and Ikura, M. (1999). Solution structure of the homodimeric core domain of *Escherichia coli* histidine kinase EnvZ. Nat Struct Biol **6,** 729-34.
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,** 200-12.
Ubersax, J. A., and Ferrell, J. E., Jr. (2007). Mechanisms of specificity in protein phosphorylation. Nat Rev Mol Cell Biol **8,** 530-41.
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 **245,** 1246-9.
Vinson, C., Acharya, A., and Taparowsky, E. J. (2006). Deciphering B-ZIP transcription factor interactions in vitro and in vivo. Biochim Biophys Acta **1759,** 4-12.
Weigt, M., White, R. A., Szurmant, H., Hoch, J. A., and Hwa, T. (2009). Identification of direct residue contacts in protein-protein interaction by message passing. Proc Natl Acad Sci U S A **106,** 67-72.
Wollenberg, K. R., and Atchley, W. R. (2000). Separation of phylogenetic and functional associations in biological sequences by using the parametric bootstrap. Proc Natl Acad Sci U S A **97,** 3288-91.
Yang, Y., and Inouye, M. (1991). Intermolecular complementation between two defective mutant signal-transducing receptors of *Escherichia coli*. Proc Natl Acad Sci U S A **88,** 11057-61.
Yang, Z. (2005). The power of phylogenetic comparison in revealing protein function. Proc Natl Acad Sci U S A **102,** 3179-80.
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,** 851-62.
Zhang, J. (2003). Evolution by gene duplication: an update. Trends in Ecology & Evolution **18,** 292-298.

Chapter 2

Determinants of homodimerization specificity in histidine kinases

Reproduced with permission of Elsevier B.V. from

Ashenberg, O., Rozen-Gagnon, K., Laub, M. T., and Keating, A. E. (2011). Determinants of homodimerization specificity in histidine kinases. J Mol Biol **413,** 222-35.

O.A, M.T.L., and A.E.K designed the study. O.A. performed all computational and experimental work. K. G. helped develop the competition FRET assay. O.A, M.T.L., and A.E.K analyzed data and wrote the manuscript.

Abstract

Two-component signal transduction pathways consisting of a histidine kinase and a response regulator are used by prokaryotes to respond to diverse environmental and intracellular stimuli. Most species encode numerous paralogous histidine kinases that exhibit significant structural similarity. Yet in almost all known examples, histidine kinases are thought to function as homodimers. We investigated the molecular basis of dimerization specificity, focusing on the model histidine kinase EnvZ and RstB, its closest paralog in *Escherichia coli*. Direct binding studies showed that the cytoplasmic domains of these proteins each form specific homodimers *in vitro*. Using a series of chimeric proteins, we identified specificity determinants at the base of the four-helix bundle in the dimerization and histidine phosphotransfer domain. Guided by molecular coevolution predictions and EnvZ structural information, we identified sets of residues in this region that are sufficient to establish homospecificity. Mutating these residues in EnvZ to the corresponding residues in RstB produced a functional kinase that preferentially homodimerized over interacting with EnvZ. EnvZ and RstB likely diverged following gene duplication to yield two homodimers that cannot heterodimerize, and the mutants we identified represent possible evolutionary intermediates in this process.

Introduction

Protein-protein interactions are central to most biological processes and typically must be highly selective; that is, proteins must interact preferentially with functionally relevant, or cognate, partners while minimizing interactions with non-cognate proteins. Understanding the basis of interaction specificity is a challenge, especially when the interacting proteins are members of large paralogous families that share similar sequences and structures. Biological strategies for ensuring specificity can be either contextual or intrinsic. In a contextual strategy, proteins that would otherwise interact are kept apart by spatial or temporal localization. In an intrinsic strategy, biophysical properties of the proteins themselves are sufficient to ensure that they interact preferentially with their cognate partners. Intrinsic specificity has been demonstrated in a growing number of systems, such as eukaryotic and viral bZIP transcription factors, Bcl-2 family proteins, PDZ domains, and bacterial two-component signaling proteins (Chen et al., 2005; Newman and Keating, 2003; Reinke et al., 2010; Skerker et al., 2005; Stiffler et al., 2007).

Two-component signal transduction pathways allow prokaryotes to sense and respond to diverse environmental and intracellular stimuli (Stock et al., 2000). These pathways typically pair a sensor histidine kinase with a cognate response regulator. Upon activation, a histidine-kinase dimer undergoes autophosphorylation and then catalyzes transfer of the phosphoryl group to its cognate response regulator, which typically activates a transcriptional response. Two-component signaling pathways are present in most bacteria and have been extensively studied because of their functional importance and because they are a good model for studying fundamental mechanisms of signal transduction. These pathways are also an excellent system in which to explore mechanisms of interaction specificity. Although most bacteria encode dozens of

paralogous histidine kinases and response regulators, cross-talk between pathways is limited (Laub and Goulian, 2007).

Recent work has shown that histidine kinases harbor a strong preference *in vitro* for phosphorylating their *in vivo* cognate response regulators (Skerker et al., 2005). Similarly, response regulators show an intrinsic preference for homodimerization *in vitro* (Gao et al., 2008a). These findings indicate that interaction specificity in these signaling pathways is established largely at the level of molecular recognition and that contextual or cellular strategies are not necessary to insulate different pathways. Because histidine kinases dimerize to autophosphorylate, kinase heterodimerization represents another possible source of cross-talk that presumably must be minimized (Scheu et al., 2010; Yang and Inouye, 1991). Many histidine kinases have been shown to homodimerize *in vitro*, but have rarely been tested for heterodimerization (Bilwes et al., 1999; Hidaka et al., 1997; Marina et al., 2005; Ninfa et al., 1993). There are only a few examples in which heteroassociations between histidine kinases have been reported, including the ethylene receptors ETR1 and ERS2 in *Arabidopsis thaliana* (Gao et al., 2008b; Grefen et al., 2008) and the kinases GacS and RetS in *Pseudomonas aeruginosa* (Goodman et al., 2009). It thus remains unresolved whether histidine kinases homodimerize specifically and, if so, how specificity is determined.

Here, we investigate the homodimerization specificity of the model histidine kinase EnvZ from *E. coli* (Forst et al., 1989). The closest paralog of EnvZ is RstB, which shares a common domain organization and has ~30% sequence similarity to EnvZ in the cytoplasmic dimerization domain. If there were cross-talk in *E. coli*, RstB is the histidine kinase most likely to heterodimerize with EnvZ. Both EnvZ and RstB function as integral membrane homodimers. Periplasmic sensory

domains in each protein are linked to highly conserved cytoplasmic domains through a transmembrane α helix (Cheung and Hendrickson, 2010). The cytoplasmic portion of each kinase consists of a parallel, dimeric four-helix bundle HAMP (*h*istidine kinases, *a*denylyl cyclases, *m*ethyl-accepting chemotaxis proteins, and *p*hosphatases) domain (Airola et al., 2010; Hulko et al., 2006), a parallel, dimeric four-helix bundle DHp (*d*imerization and *h*istidine *p*hosphotransfer) domain containing a conserved histidine that is the site of autophosphorylation (Albanesi et al., 2009; Bick et al., 2009; Casino et al., 2009; Marina et al., 2005; Tomomori et al., 1999), and an α/β-sandwich CA (*c*atalytic and *A*TP-binding) domain that catalyzes autophosphorylation (Tanaka et al., 1998). An NMR structure of the DHp domain of EnvZ (Tomomori et al., 1999) shows both similarities and interesting differences compared to a crystal structure of the DHp domain of TM0853, a histidine kinase from *Thermotoga maritima* (Marina et al., 2005). The two DHp domains share the same overall four-helix bundle architecture but differ in the ordering of the helices. The HAMP domains in histidine kinases appear to be linked to DHp domains through a continuous α helix that comprises the last α helix of the HAMP domain and the first α helix of the DHp domain. Both the DHp and HAMP domains likely dimerize in EnvZ, but whether either or both domains contribute to homodimerization specificity is unclear.

Computational analysis of amino-acid coevolution, combined with knowledge of protein structures, has proven very useful for elucidating the specificity determinants of histidine kinase-response regulator interactions (Capra et al., 2010; Skerker et al., 2008). This approach looks for amino-acid covariation at pairs of positions within large multiple sequence alignments, and is particularly powerful for examining prokaryotic protein families as there are often thousands, or even tens of thousands, of homologous sequences in genome databases. Highly covarying pairs

can indicate that two residues interact to maintain protein structure or function. If histidine kinases must homodimerize specifically to suppress cross-talk, then specificity determining residues could show statistically significant covariation. This was shown to be the case for residues at the histidine kinase-response regulator interface and for residues mediating polyketide synthase protein interactions (Skerker et al., 2008; Thattai et al., 2007). For the former, subsequent analyses have demonstrated how kinase-regulator specificity can be rewired with minimal sequence changes localized to regions originally identified using covariation approaches.

Here, we took a biochemical approach to show that the cytoplasmic domains of EnvZ and RstB specifically self-associate *in vitro*. Guided by an analysis of amino-acid coevolution within histidine kinases, we identified a small number of residues in the DHp domain of EnvZ that can be substituted with the corresponding residues from RstB to create a functional homodimer that preferentially self-associates rather than associating with EnvZ. Our results suggest that histidine-kinase dimerization specificity is, like kinase-regulator interaction and response-regulator dimerization, hardwired at the level of molecular recognition.

Results

EnvZ and RstB specifically homodimerize

To assess whether the cytoplasmic domains of EnvZ and RstB homodimerize specifically, we used fluorescence resonance energy transfer (FRET) to examine interactions between $EnvZ_{HDC}$ and $RstB_{HDC}$ fused to CFP and YFP. The domains present in each kinase construct are denoted with subscripts where H, D, and C indicate HAMP, DHp, and CA, respectively (for details, see Methods). If each kinase specifically homodimerizes, we expected that a mixture of CFP- and

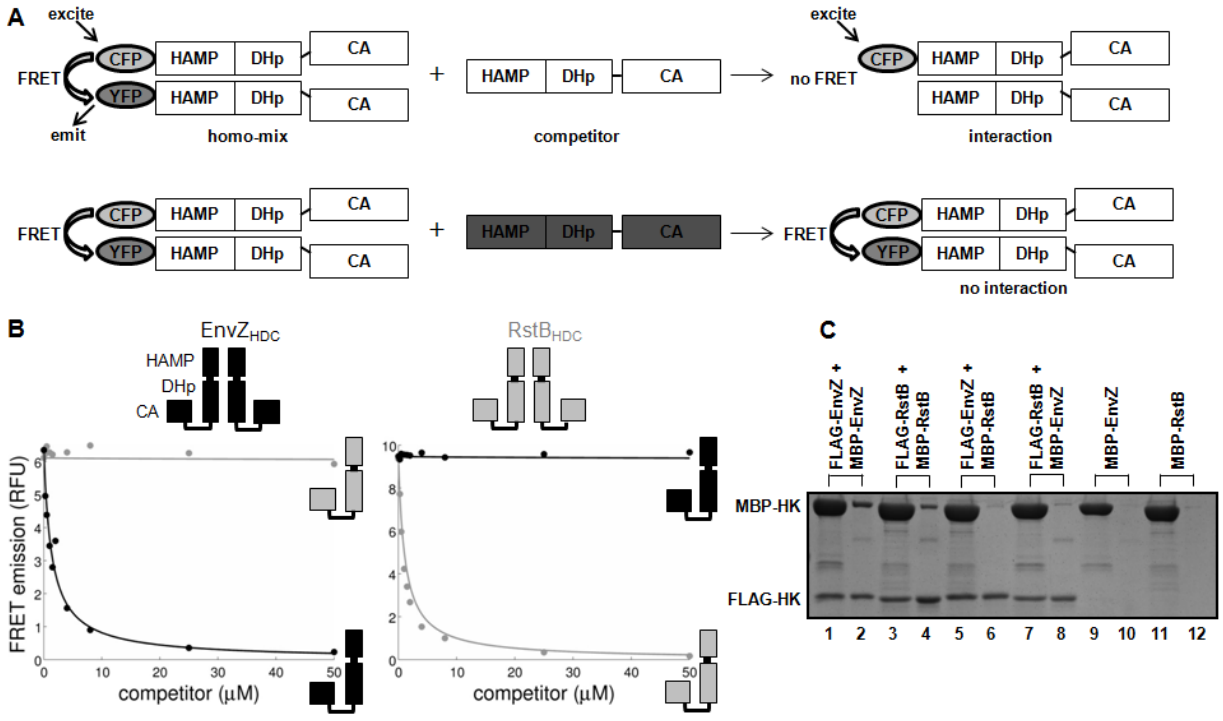


Figure 2.1. The cytoplasmic domains of *E. coli* histidine kinases EnvZ and RstB specifically homodimerize. (A) Schematic of *in vitro* FRET competition assay. FRET signal from a complex of CFP-kinase A and YFP-kinase A is reduced by interaction with an unlabeled competitor kinase. (B) FRET competition assay for homo-association and hetero-association of $EnvZ_{HDC}$ and $RstB_{HDC}$. Protein concentrations were 0.5 μM CFP-$EnvZ_{HDC}$ and 0.5 μM YFP-$EnvZ_{HDC}$ (left panel) or 0.5 μM CFP-$RstB_{HDC}$ and 0.5 μM YFP-$RstB_{HDC}$ (right panel). The curves are fit as described in the methods. Cartoons of EnvZ and RstB show the HAMP, DHp and CA domains. (C) Pull-down assay with purified $EnvZ_{HDC}$ and $RstB_{HDC}$. A mixture of FLAG-labeled kinase and MBP-labeled kinase was incubated, and complexes were isolated using anti-FLAG beads. Odd lanes show inputs and even lanes show elutions. The following mixtures were assayed: FLAG-$EnvZ_{HDC}$ + MBP-$EnvZ_{HDC}$ (lane 2), FLAG-$RstB_{HDC}$ + MBP-$RstB_{HDC}$ (lane 4), FLAG-$EnvZ_{HDC}$ + MBP-$RstB_{HDC}$ (lane 6), FLAG-$RstB_{HDC}$ + MBP-$EnvZ_{HDC}$ (lane 8). Non-specific binding to beads was assessed for MBP-$EnvZ_{HDC}$ (lane 10) and MBP-$RstB_{HDC}$ (lane 12).

YFP-labeled subunits would produce a FRET signal that could be inhibited by increasing concentrations of an unlabeled copy of the same kinase, but not the other kinase (Fig. 2.1A). All FRET experiments described below were performed in this competition format.

A mixture of CFP-$EnvZ_{HDC}$ and YFP-$EnvZ_{HDC}$ produced a FRET signal that was reduced in a concentration-dependent manner by the addition of unlabeled $EnvZ_{HDC}$ (Fig. 2.1B). Similarly, the FRET signal from a mixture of CFP-$RstB_{HDC}$ and YFP-$RstB_{HDC}$ decreased with the addition of unlabeled $RstB_{HDC}$ (Fig. 2.1B). These data confirmed the expected homodimerization of our $EnvZ_{HDC}$ and $RstB_{HDC}$ constructs. We next tested whether EnvZ and RstB could heteroassociate by adding unlabeled $RstB_{HDC}$ to the mixture of CFP-$EnvZ_{HDC}$ and YFP-$EnvZ_{HDC}$. Even at concentrations up to 100 times higher than the labeled EnvZ concentration, unlabeled $RstB_{HDC}$ did not significantly inhibit homodimerization. Similarly, unlabeled $EnvZ_{HDC}$ did not disrupt the $RstB_{HDC}$ homodimer (Fig. 2.1B). To quantify the relative stabilities of interactions between kinases, we fit homodimer dissociation constants and heterodimer dissociation constants, as detailed in the methods (Table 2.1, 2.2, Fig. 2.2, 2.3). The homodimer K_d for $EnvZ_{HDC}$ was 0.4 μM, less than the previously reported value of 10 μM measured using a pull-down assay with His_6-labeled $EnvZ_{HDC}$ (Hidaka et al., 1997). Similarly, the homodimer K_d of $RstB_{HDC}$ was 0.3 μM, whereas the $EnvZ_{HDC}$-$RstB_{HDC}$ heterodimer K_d was > 50 μM. Note that when competition in the FRET assay is very weak, the sensitivity of this competition assay is limited and any K_d weaker than 50 μM fits the $EnvZ_{HDC}$-$RstB_{HDC}$ data equally well (Fig. 2.4). Overall, these experiments indicate $EnvZ_{HDC}$ and $RstB_{HDC}$ homodimers are each more stable than the $EnvZ_{HDC}$-$RstB_{HDC}$ heterodimer.

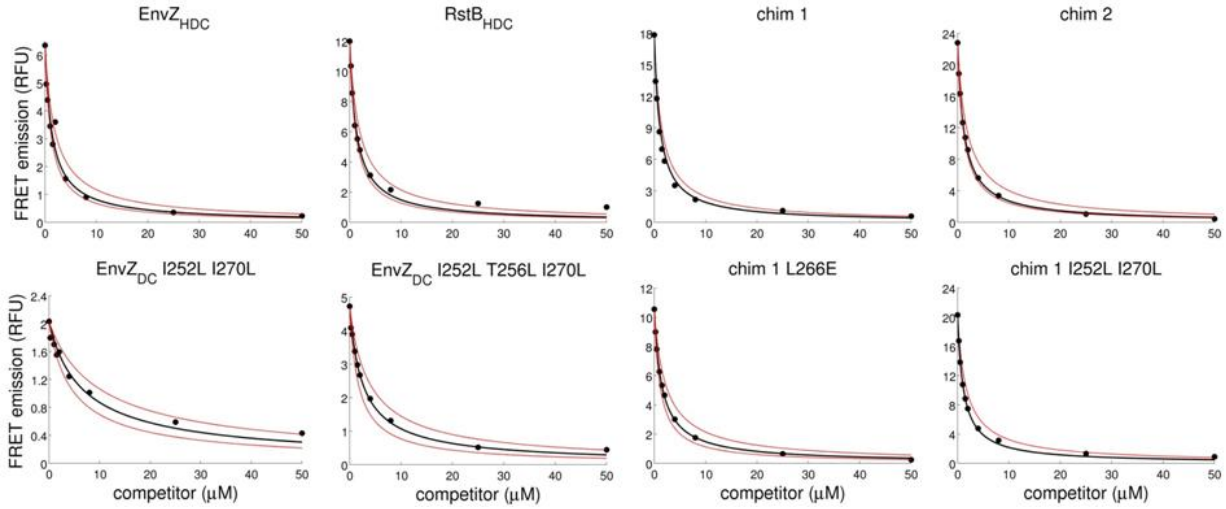


Figure 2.2. Fitting homodimer dissociation constants.
As described in the methods, K_d values for homodimerization were fit in MATLAB (black line) for each histidine kinase using the experimental FRET competition data (black circles). The decrease in corrected FRET emission with increasing competitor was fit by simulating the equilibria that govern the experiment for a given K_d. The lower and upper limits of K_d values that fit the data with $0.95R^2_{max}$ are shown in red. For each kinase, a 0.5 μM equimolar mixture of CFP-kinase A and YFP- kinase A was mixed with unlabeled kinase-A competitor.

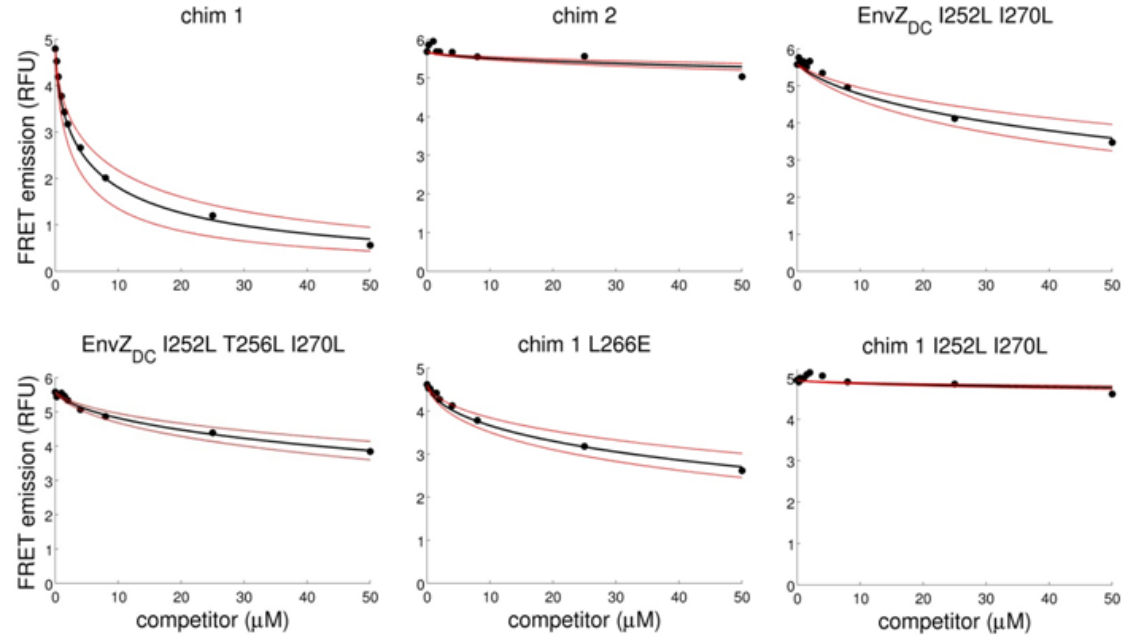


Figure 2.3. Fitting heterodimer dissociation constants.
As described in the methods, K_d values for heterodimerization between EnvZ$_{HDC}$ and mutant kinases were fit in MATLAB (black line) using the experimental FRET competition data (black circles). To determine a heterodimer K_d, K_d values for the two respective homodimers were first determined separately (see Fig. 2.2). Then the decrease in corrected FRET emission with increasing competitor was fit by simulating the equilibria, given the two previously measured homodimer K_d values and the unknown heterodimer K_d. The lower and upper limits of K_d values that fit the data with $0.95R^2_{max}$ are shown in red. For each kinase, a 0.5 μM equimolar mixture of CFP-kinase A and YFP-kinase A was mixed with unlabeled kinase-B competitor.

Table 2.1. Dissociation constants for wild-type kinases, chimeras, and cluster 1 and 2 mutants (μM).[a]

	K_d (homodimer)	K_d (heterodimer with $EnvZ_{HDC}$)
wild-type kinases and chimeras		
$EnvZ_{HDC}$	0.4	
$RstB_{HDC}$	0.3	>50
$EnvZ_{DC}$	0.1	0.1
$RstB_{DC}$	270	
chim 1	0.1	0.3
chim 2	0.3	20.5
cluster 1 mutants		
$EnvZ_{DC}$ T256L I270L	1.7	1.7
$EnvZ_{DC}$ I252L I270L	23.0	22.1
$EnvZ_{DC}$ I252L T256L I270L	4.7	16.6
chim 1 I252L I270L	0.1	19.6
cluster 2 mutants		
$EnvZ_{DC}$ A255R L266E	>50	
$EnvZ_{DC}$ L266E	>50	
chim 1 L266E	0.9	4.4

[a] K_d values shown were averaged across duplicates. Limits in sensitivity of the FRET competition assay prevent measurement of homodimer K_d values tighter than 0.1 μM or heterodimer K_d values weaker than 50 μM.

We also assayed dimerization of $EnvZ_{HDC}$ and $RstB_{HDC}$ using a pull-down assay. $EnvZ_{HDC}$ or $RstB_{HDC}$, N-terminally labeled with a FLAG epitope, was mixed with MBP-tagged $EnvZ_{HDC}$ or $RstB_{HDC}$. The protein complexes formed were isolated using anti-FLAG beads and examined by SDS-PAGE (Fig. 2.1C). Homospecific kinase complexes were clearly identified at a level above that due to non-specific bead binding (compare lanes 2 vs. 10 and 4 vs. 12) but heteromeric complexes were much weaker (compare lanes 6 vs. 12 and 8 vs. 10). These data corroborated our FRET data indicating that $EnvZ_{HDC}$ and $RstB_{HDC}$ specifically homodimerize.

Table 2.2. Dissociation constants for wild-type kinases, chimeras, and cluster 1 and 2 mutants (μM).[a]

	K_d (homodimer)	K_d (homodimer)	K_d (heterodimer with $EnvZ_{HDC}$)	K_d (heterodimer with $EnvZ_{HDC}$)
wild-type kinases and chimeras				
$EnvZ_{HDC}$	0.5 (0.1, 0.8)	0.3 (0.1, 1.8)		
$RstB_{HDC}$	0.4 (0.1, 2.1)	0.1 (0.1, 0.6)	>50	>50
$EnvZ_{DC}$	0.1 (0.1, 1.0)	0.1 (0.1, 0.9)	0.1 (0.1, 0.12)	0.1 (0.1, 0.12)
$RstB_{DC}$	154 (60, 318)	385 (213, 658)		
chim 1	0.1 (0.1, 0.7)	0.1 (0.1, 1.1)	0.3 (0.2, 0.4)	0.3 (0.2, 0.4)
chim 2	0.3 (0.1, 2.0)	0.3 (0.1, 2.1)	17.9 (14.6, 23.2)	23.0 (18.8, 29.8)
cluster 1 mutants				
$EnvZ_{DC}$ T256L I270L	2.4 (0.6, 5.9)	1.0 (0.1, 3.5)	1.9 (1.5, 2.5)	1.5 (1.2, 1.9)
$EnvZ_{DC}$ I252L I270L	19.7 (11.7, 32.6)	26.3 (17.6, 40.6)	18.9 (15.2, 24.5)	25.2 (20.2, 32.8)
$EnvZ_{DC}$ I252L T256L I270L	4.2 (1.5, 8.9)	5.2 (2.5, 9.6)	11.9 (9.9, 14.7)	21.3 (17.2, 27.5)
chim 1 I252L I270L	0.1 (0.1, 1.4)	0.1 (0.1, 1.4)	19.6 (16.0, 25.3)	>50
cluster 2 mutants				
$EnvZ_{DC}$ A255R L266E	>50	>50		
$EnvZ_{DC}$ L266E	>50	>50		
chim 1 L266E	0.8 (0.1, 3.1)	1.0 (0.1, 3.3)	3.7 (3.1, 4.7)	5.1 (4.1, 6.5)

[a] K_d values shown were measured in duplicate. In parentheses are listed the lower and upper limits of K_d values that fit the data with $0.95R^2_{max}$.

A recent crystal structure of a HAMP-DHp fusion protein shows the HAMP and DHp domains make up the cytoplasmic dimerization interface (PDB id: 3zrx). Furthermore, in EnvZ the CA domain is a monomer and is not required for dimerization (Park et al., 1998). To experimentally test contributions of the HAMP and DHp domains to dimerization specificity, we made constructs in which most of the HAMP domain was removed. The constructs we made had the same N-terminus as the EnvZ construct for which an NMR structure was solved, such that only 10 residues from the HAMP domain remained; in the NMR structure this region is unfolded and non-interacting (Tomomori et al., 1999). Interestingly, removing the HAMP domain significantly destabilized RstB but not EnvZ. We estimated a K_d for the wild-type $EnvZ_{DC}$ homodimer of 0.1 μM and a K_d for the wild-type $RstB_{DC}$ homodimer of ~300 μM (Table 2.1). To explore the

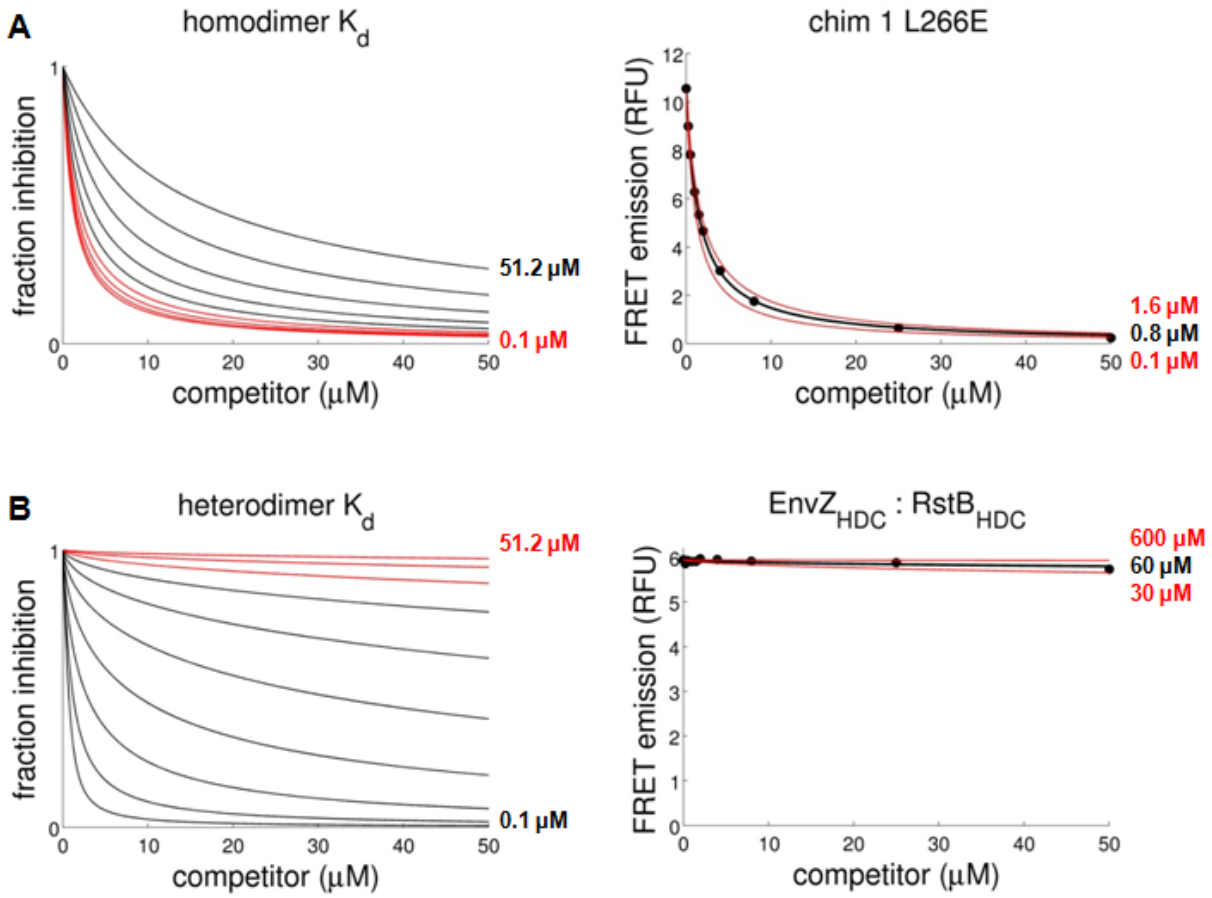


Figure 2.4. Limits in FRET assay sensitivity.
(A) To look at limits in measuring homodimerization affinity, FRET competition assays were simulated in MATLAB across a range of homodimer affinities. In the left panel, homodimer K_d values were varied from 0.1 μM to 51.2 μM, by factors of 2. Competition curves are poorly-separated for homodimer K_d's from 0.1 μM to 1.6 μM (red lines) and well-separated for homodimer K_d's weaker than 1.6 μM (black lines). As an example of a difficult-to-fit homodimer, in the right panel the best fit for chim 1 L266E homodimer is shown (black line) along with two curves that fit the data nearly as well (red lines). (B) FRET competition assays were simulated in MATLAB across a range of heterodimer affinities. In the left panel, heterodimer K_d values were varied from 0.1 μM to 51.2 μM, by factors of 2. Competition curves are well-separated for heterodimer K_d's from 0.1 μM to 12.8 μM (black lines) and poorly-separated for heterodimer K_d's weaker than 12.8 μM (red lines). In the right panel, the broad range of heterodimer K_d values that fit the $EnvZ_{HDC}$-$RstB_{HDC}$ heterodimer is shown. In both homodimer and heterodimer simulations, a 0.5 μM equimolar mixture of CFP-kinase A and YFP- kinase A was mixed with unlabeled kinase competitor. For the heterodimer simulation, the homodimer K_d's of kinase A and the competitor were both set to 0.5 μM.

dimerization specificity of these constructs, FRET competition experiments were performed at higher concentrations (20 μM) where both $EnvZ_{DC}$ and $RstB_{DC}$ dimerize. These higher concentrations allowed measurement of a range of weaker K_d's. Under these conditions, $EnvZ_{DC}$

homodimerization was inhibited by unlabeled $EnvZ_{DC}$, but not by $RstB_{DC}$, even at high concentrations (Fig. 2.5). Similarly, $RstB_{DC}$ homodimerization was inhibited by unlabeled $RstB_{DC}$, but not by $EnvZ_{DC}$ (Fig. 2.5). We estimated the $EnvZ_{DC}$-$RstB_{DC}$ heterodimer K_d as >100 μM; any K_d >100 μM fits the data equally well (Fig. 2.4). We conclude that $EnvZ_{DC}$, like $EnvZ_{HDC}$, retains a high degree of homodimerization specificity. Whether $RstB_{DC}$ specifically homodimerizes is less clear because we cannot determine the relative stabilities of $RstB_{DC}$ homodimer and $EnvZ_{DC}$-$RstB_{DC}$ heterodimer.

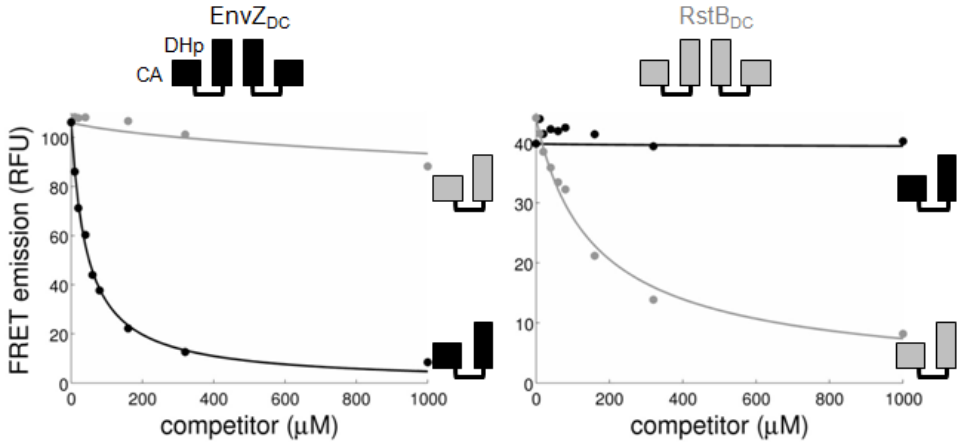


Figure 2.5. The DHp domains of EnvZ and RstB homodimerize.
FRET competition assay for homo-association and hetero-association of $EnvZ_{DC}$ (left panel) and $RstB_{DC}$ (right panel). CFP and YFP fusion protein concentrations were 20 μM.

Identification of putative specificity determining residues within the DHp domain

The specific homodimerization of $EnvZ_{DC}$ motivated us to look for specificity-determining residues within the DHp domain. To guide this search, we performed a computational analysis of amino-acid covariation in histidine kinases. The assumption underlying this approach is that homodimerization specificity-determining residues must coevolve to maintain the self-association of a kinase while disfavoring competing heterodimer states. We built a multiple sequence alignment of 4272 histidine-kinase sequences that each contain a HAMP domain

immediately N-terminal to the DHp and CA domains. We then measured covariation between all possible pairs of positions within the alignment using MIp, a scoring metric based on mutual information (MI) that includes a correction for covariation that arises from phylogenetic relationships and random noise (Dunn et al., 2008). Within the 25 most highly covarying pairs, 20 pairs were within the DHp domain and 5 pairs were within the HAMP domain. None of the HAMP residue pairs were in physical contact based on the Af1503 HAMP NMR structure (see Methods and Fig. 2.6). In contrast, 15 DHp pairs had atoms within 5.5 Å of one another and 7 of these pairs involved residues on different chains based on the EnvZ NMR structure. Six of the seven interchain DHp contacts identified by covariation mapped to the region of the DHp

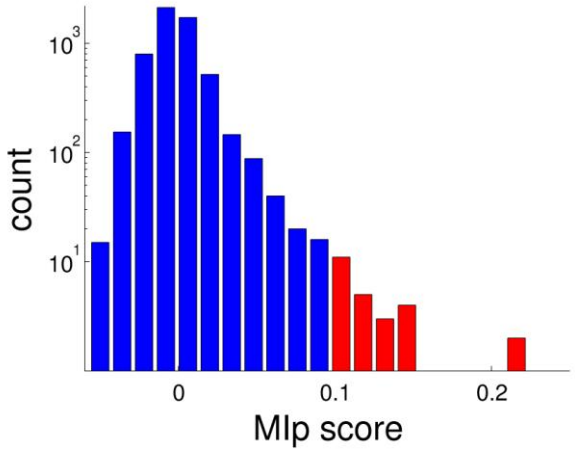


Figure 2.6. Histogram of amino-acid covariation scores.
Covariation scores (MIp) within the HAMP and DHp domains in an alignment of sequences with HAMP-DHp-CA domains was measured using mutual information with a correction for phylogenetic relationships and random noise (Dunn et al., 2008). The red bars represent the top 25 covarying pairs, with MIp score > 0.1. Above this cutoff, covarying pairs were filtered for interchain contacts and mapped onto the structure of the EnvZ DHp domain.

domain distal from the HAMP and near the loops of the four-helix bundle (Fig. 2.7A). We refer to this region, the lower half of the helix bundle, as the "base" of the bundle, and we refer to the upper half as the membrane-proximal region (Fig. 2.7A). Interestingly, recent crystal structures of the histidine kinase DesK indicate that the membrane-proximal region of the DHp domain

exhibits significant structural plasticity, consistent with the notion that residues in the more static base region may be important for maintaining dimerization and perhaps encoding specificity (Albanesi et al., 2009).

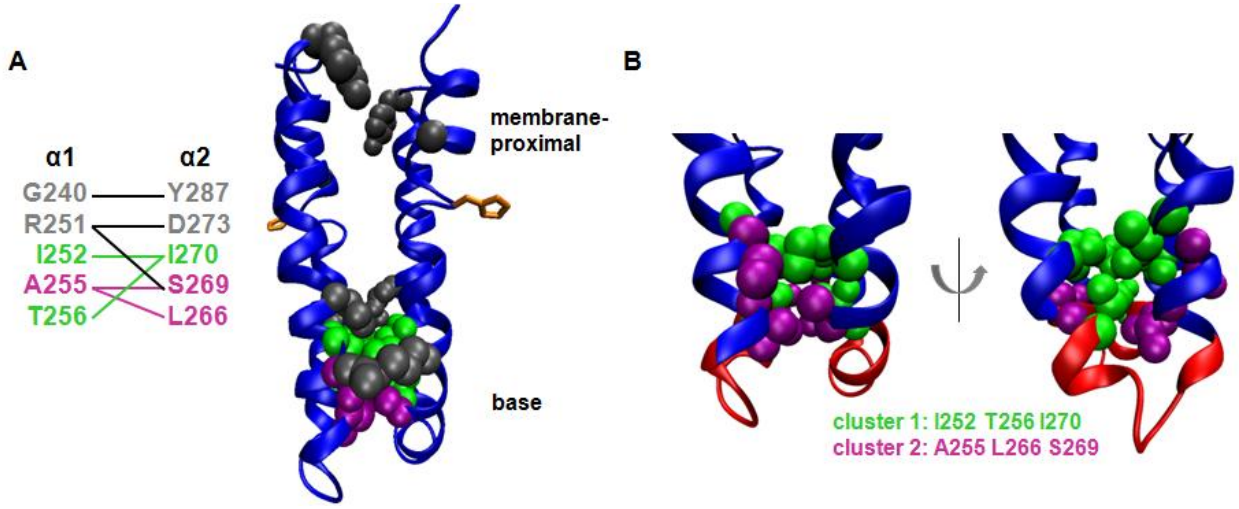


Figure 2.7. Covarying pairs in the EnvZ dimerization interface.
(A) Residues in EnvZ that highly covary and are within 5.5 Å across the dimerization interface are listed (at left) and shown on the NMR structure of EnvZ (at right) (Tomomori et al., 1999). Covarying residues in helices α1 and α2 are connected by lines. Cluster 1 residues are shown in green, cluster 2 residues are shown in purple and all other covarying, interchain residues are shown in grey. The conserved histidine is shown in orange. (B) Close-up views of clusters 1 and 2. The chimera-1-region backbone (residues 256-265) is in red; clusters 1 and 2 are outside that region.

To test the role of residues at the base of the DHp domain in dimerization specificity, we used a series of chimeric proteins in which portions of EnvZ were replaced with the corresponding residues from RstB. These same chimeric proteins were previously studied in the context of histidine kinase-response regulator interaction specificity (J. Skerker, M.T. Laub, unpublished data) (Skerker et al., 2008). In each of the four chimeras, progressively more of the base of the EnvZ helical bundle was replaced with the corresponding residues from RstB. The proteins were designed such that equivalent portions of α helices 1 and 2 were replaced in each case (Fig. 2.8A, B). Each of the four chimeric DHp domains was fused to the EnvZ CA domain.

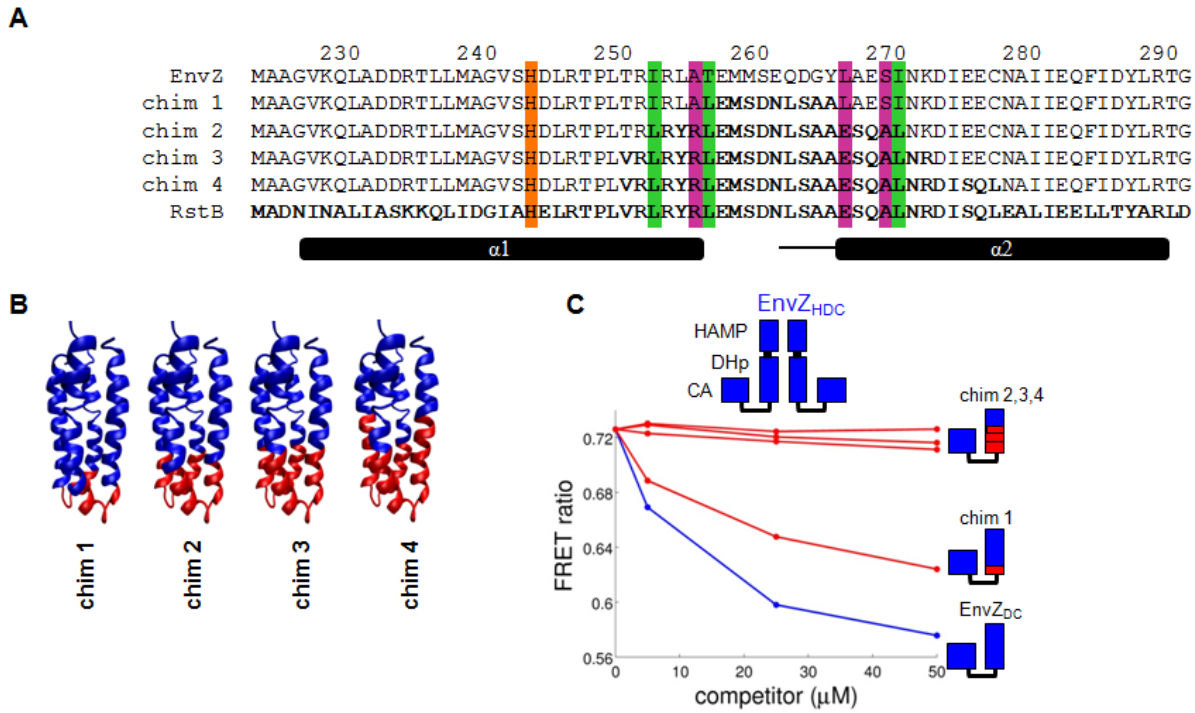


Figure 2.8. EnvZ-RstB chimeric proteins isolate dimerization specificity to the DHp domain base.
(A) Sequence alignment of the DHp domains of EnvZ, RstB and chimeras 1-4. Amino acids mutated, relative to EnvZ, to make the chimeras are shown in bold. The columns highlighted in color correspond to covarying residues in cluster 1 (I252, T256, I270) and cluster 2 (A255, L266, S269). The highlighted column in orange corresponds to the highly conserved histidine. Locations of the two helices in the DHp domain are indicated. (B) Models of chimeras 1-4 showing residues from EnvZ in blue and residues from RstB in red. The EnvZ NMR structure is used as a template. The chimeras are fused to the EnvZ CA domain, which is omitted in the models. (C) Interactions between $EnvZ_{DC}$ or chimeras 1-4 with $EnvZ_{HDC}$ (5 μM each of the CFP and YFP fusions).

We tested the chimeric proteins for homodimerization using FRET competition, as above, and observed that each was able to homodimerize (Table 2.1, Fig. 2.2, 2.9). We then assessed whether the chimeras lost the ability to interact with $EnvZ_{HDC}$. Interestingly, chimera 1 bound $EnvZ_{HDC}$ with a K_d of 0.3 μM, whereas chimeras 2, 3, and 4 did not interact with $EnvZ_{HDC}$ under the conditions tested (Fig. 2.8C, Table 2.1). Chimeras 2-4, but not chimera 1, share nine residues from RstB that replace the corresponding residues of EnvZ (Fig. 2.8A). Our results thus implicated some or all of these residues in establishing homodimerization specificity.

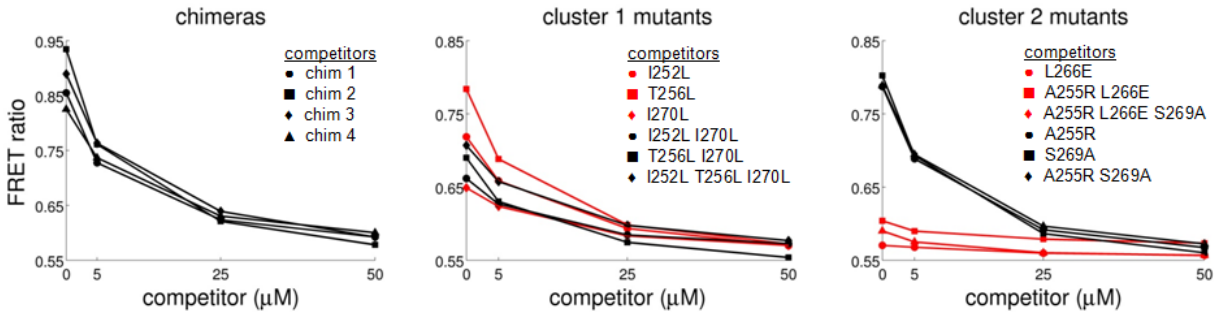


Figure 2.9. Homo-association of chimeras 1-4 and cluster 1 and 2 mutants.
Measuring homo-associations of chimeras 1-4 (left panel), $EnvZ_{DC}$ cluster 1 mutants (middle panel), and $EnvZ_{DC}$ cluster 2 mutants (right panel). CFP and YFP fusion protein concentrations were 5 μM.

Table 2.3. Dissociation constants for cluster 1 single mutants (μM).

	K_d (homodimer)	K_d (homodimer)	K_d (heterodimer with $EnvZ_{HDC}$)	K_d (heterodimer with $EnvZ_{HDC}$)
$EnvZ_{DC}$ I252L	4.7 (1.8, 9.7)	3.2 (0.9, 7.5)	1.2 (0.9, 1.6)	1.2 (0.9, 1.7)
$EnvZ_{DC}$ T256L	0.1 (0.1, 1.1)	0.1 (0.1, 0.7)	0.15 (0.11, 0.2)	0.14 (0.1, 0.19)
$EnvZ_{DC}$ I270L	12.0 (6.5, 20.8)	10.7 (6.0, 18.0)	2.7 (2.0, 3.8)	2.8 (2.0, 3.9)

Table 2.4. Dissociation constants for cluster 2 mutants (μM).

	K_d (homodimer)	K_d (homodimer)	K_d (heterodimer with $EnvZ_{HDC}$)	K_d (heterodimer with $EnvZ_{HDC}$)
$EnvZ_{DC}$ A255R	0.1 (0.1, 1.1)	0.1 (0.1, 1.4)	0.12 (0.1, 0.16)	0.1 (0.1, 0.13)
$EnvZ_{DC}$ S269A	0.1 (0.1, 0.5)	0.1 (0.1, 0.6)	0.1 (0.1, 0.13)	0.1 (0.1, 0.12)
$EnvZ_{DC}$ A255R S269A	0.1 (0.1, 0.4)	0.1 (0.1, 0.9)	0.3 (0.2, 0.4)	0.4 (0.3, 0.5)

Table 2.5. Dissociation constants for destabilized mutants (μM).

	K_d (homodimer)
$EnvZ_{DC}$ L266E	>50
$EnvZ_{DC}$ A255R L266E	>50
$EnvZ_{DC}$ A255R L266E S269A	>50
$EnvZ_{DC}$ M259S	>50
$EnvZ_{DC}$ M259S L266E	>50

Specificity determinants can be further localized within the chimera 1 and 2 regions

Chimera 2 showed dimerization behavior similar to RstB, in that it did not hetero-associate with $EnvZ_{HDC}$ but maintained an ability to homo-associate. To further localize specificity determinants we again used the covariation analysis described above. Strongly covarying residue pairs in the chimeria-2 region included 252-270 and 256-270, making up a 3-residue cluster (cluster 1, green in Fig. 2.7A, B), as well as 255-266 and 255-269, making another 3-residue cluster (cluster 2, purple in Fig. 2.7A, B). These six residues mediate interchain contacts in the NMR structure of EnvZ. Cluster 1 is located in the hydrophobic core of the DHp helical bundle whereas cluster 2 is located adjacent to the chimera 1 loop. To examine the contribution covarying residue pairs make to homodimerization specificity, we made mutations in $EnvZ_{DC}$ that substituted EnvZ residues with the corresponding ones from RstB. For each three-residue cluster we made each single mutant as well as two double mutants in which the covarying residues were both substituted with the corresponding residues from RstB.

Using the FRET competition assay, we tested each cluster-1 mutant for homodimerization and for interaction with $EnvZ_{HDC}$. Unlike chimera 2, none of the double mutants showed high homospecificity (Table 2.1, 2.2, 2.3, Fig. 2.9, 2.10). In fact, cluster 1 mutations, particularly the I252L/I270L double mutant, destabilized homodimer formation. We hypothesized that the effect of these mutations may be context dependent and require concomitant changes in the loop residues that comprise chimera 1. Indeed, introducing the double mutation I252L/I270L into the context of chimera 1 gave a highly homospecific protein with high homodimer stability (K_d = 0.1 μM) that did not heterodimerize significantly with $EnvZ_{HDC}$ (K_d = 20 μM) (Table 2.1, Fig. 2.2, 2.3). This analysis thus localized determinants of homospecificity to twelve residues at the base

of the EnvZ DHp bundle: the ten residues in chimera 1 and the two residues at positions 252 and 270. Interestingly, a significant part of the chimera-1 context effect can be explained by the role of residue T256. Changing this residue to Leu (as in RstB), along with the mutations I252L and I270L, produced a homospecific mutant, I252L/T256L/I270L. These three residues, which strongly covaried with each other, thus exert a strong influence on homodimerization specificity in EnvZ (Table 2.1, Fig. 2.2, 2.3).

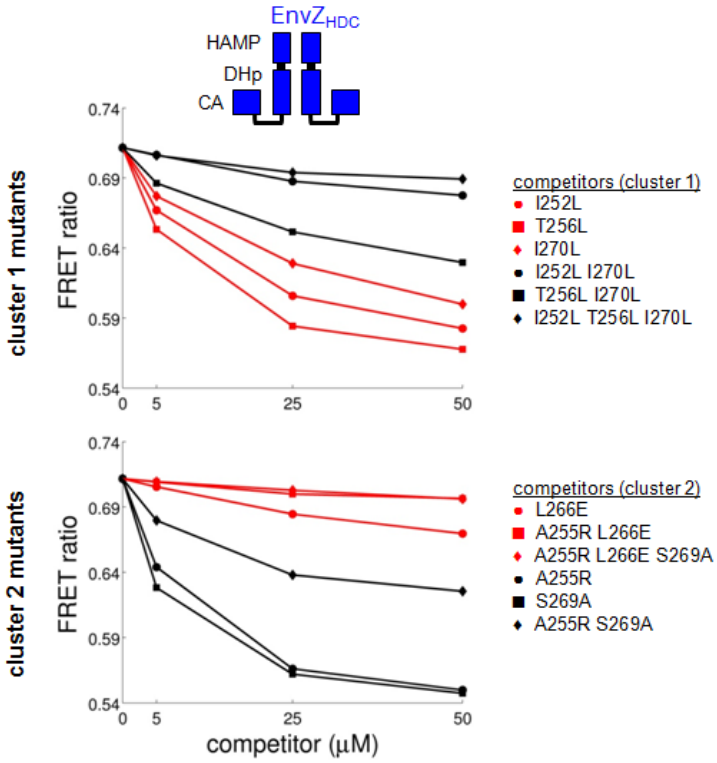


Figure 2.10. Interactions of cluster 1 and 2 mutants with EnvZ.
Measuring interactions between $EnvZ_{HDC}$ and either $EnvZ_{DC}$ cluster 1 (first row) or $EnvZ_{DC}$ cluster 2 mutants (second row). CFP and YFP fusion protein concentrations were 5 μM.

Similarly, we found examples of mutations in cluster 2 that were destabilizing when made individually but were accommodated in the context of chimera 1. The double mutant A255R/S269A had no effect on homo or hetero-associations, but A255R/L266E was significantly destabilized as a homodimer, an effect that was attributable to L266E alone (Table

2.1, 2.4, 2.5). Because position 266 covaries strongly with position 259 which is within the chimera 1 loop, we tested whether mutating M259 in EnvZ to serine, as found in RstB, would rescue homodimerization stability, but it did not (Table 2.5). However, the L266E substitution made in the context of chimera 1 did form a stable homodimer (K_d = 0.9 μM) and a slightly less stable heterodimer with $EnvZ_{HDC}$ (K_d = 4.4 μM), yielding modest homodimerization specificity (Table 2.1, Fig. 2.2, 2.3). Our results thus suggest that the loops present in chimera 1 are either very stabilizing, or provide an important context for the residues in the chimera-2 region.

Dimerization specificity switch mutants do not affect phosphotransfer specificity

We had created three homospecific kinases, I252L/T256L/I270L, chimera 1 + L266E, and chimera 1 + I252L/I270L. Although each kinase was capable of homodimerizing, we wanted to test whether each mutant protein could still catalyze autophosphorylation and phosphotransfer to response regulators. Previous studies have shown that wild-type EnvZ and RstB exhibit a kinetic preference for phosphotransfer *in vitro* to their cognate regulator substrates OmpR and RstA, respectively (Skerker et al., 2005). Further, chimera 1 was previously shown to have EnvZ-like substrate specificity while chimera 2 was shown to have altered substrate specificity such that it phosphorylated RstA but not OmpR (Fig. 2.11A). We tested our three most homospecific kinases for autophosphorylation and for phosphotransfer to OmpR and RstA (Fig. 2.11B). Each mutant was competent for autophosphorylation, providing further evidence that the mutant kinases homodimerize, because EnvZ, and likely RstB, autophosphorylate exclusively as dimers (Cai and Inouye, 2003; Yang and Inouye, 1991). Each mutant also maintained the phosphotransfer specificity of EnvZ, and phosphorylated OmpR but not RstA. Thus, the mutations characterized here affect dimerization specificity without disrupting phosphotransfer specificity.

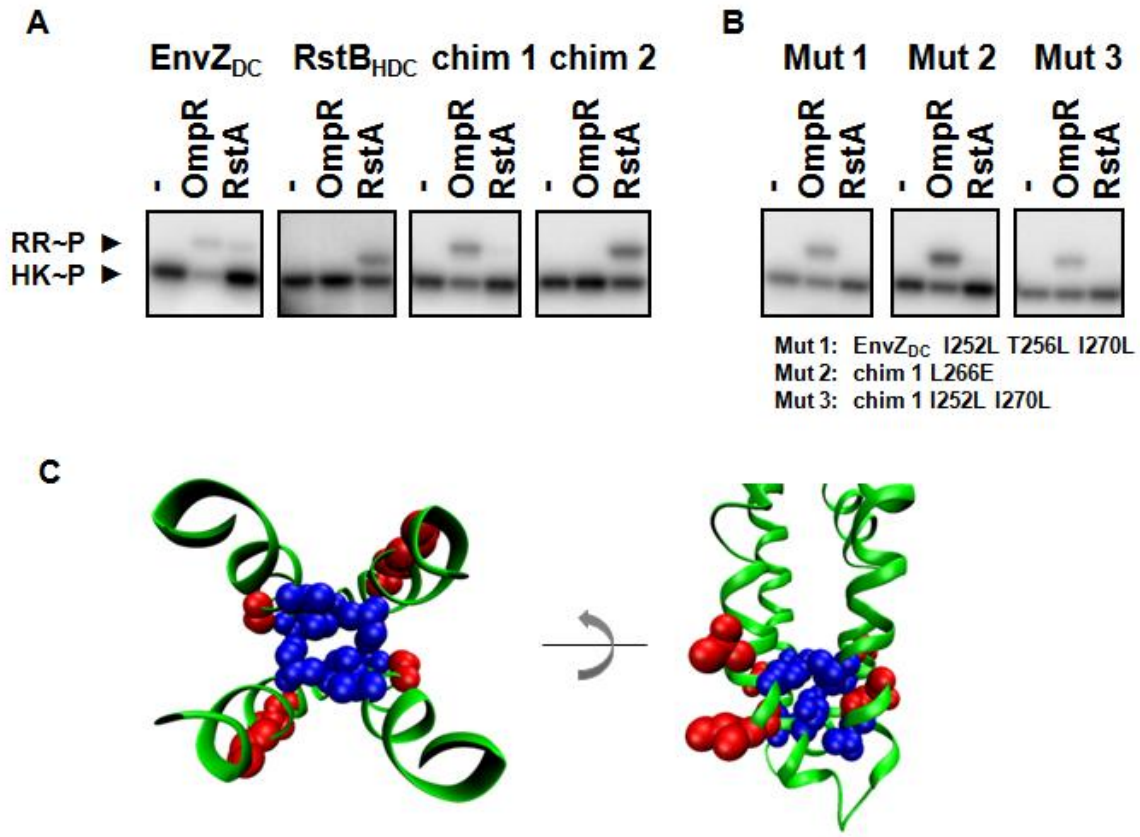


Figure 2.11. Phosphotransfer specificity for dimerization specificity switch mutants.
Each kinase (HK) was autophosphorylated with radiolabeled ATP and either incubated alone (-) or assayed for phosphotransfer to a response regulator (RR). (A) $EnvZ_{DC}$ and $RstB_{HDC}$ phosphorylated their cognate response regulators, OmpR and RstA, respectively. Chimera 2 showed a switch in phosphotransfer specificity relative to chimera 1. (B) Phosphotransfer specificity of $EnvZ_{DC}$ dimerization specificity switch mutants Mut 1-3, defined as in the figure. (C) View from the base of the four-helix bundle in EnvZ, with positions affecting interaction specificity shown in space-filling form. Cluster 1 positions 252, 256, and 270, and cluster 2 position 266 (blue) are buried in the dimerization interface. Positions 250, 254, 255, and 269, previously identified to affect histidine kinase-response regulator phosphotransfer specificity (red) are solvent-exposed. The residues are mapped onto the EnvZ DHp domain.

Discussion

In this study we investigated the dimerization specificity of a pair of histidine-kinase paralogs from *E. coli*, EnvZ and RstB. Using FRET to measure interactions, we showed that the EnvZ and RstB cytoplasmic domains preferentially self-associate, and we identified a set of residues in EnvZ that can be replaced with the corresponding residues from RstB to enable orthogonal homodimerization. Below, we discuss the possible roles of different parts of histidine-kinase structures in determining dimerization specificity and how our results illustrate a plausible evolutionary pathway for establishing histidine-kinase dimer homospecificity following gene duplication events.

Structural and functional modularity of histidine kinases

The strong preference of $EnvZ_{HDC}$ and $RstB_{HDC}$ for homodimerization *in vitro* is consistent with a mechanism in which molecular recognition establishes pathway isolation in two-component signaling. The HAMP and DHp domains likely comprise the dimerization interface and thereby determine the stability and specificity of histidine kinases. To localize the structural determinants of homodimerization specificity, we first made constructs in which most of the HAMP domain was removed. The K_D estimated for $EnvZ_{DC}$ was similar to that of $EnvZ_{HDC}$, consistent with previous thermal denaturation studies (Kishii et al., 2007). Either the EnvZ HAMP domain does not contribute significantly to dimer stability or it does not fold properly *in vitro* (Kishii et al., 2007). Nevertheless $EnvZ_{DC}$ preferentially homodimerized, with no evidence of a heterodimer forming with $RstB_{DC}$, even using high concentrations of the latter. These findings suggest that, at least for EnvZ, the DHp domain plays a critical role in determining homodimerization

specificity.

Other evidence supports an important role for the DHp domain in establishing dimerization specificity for histidine kinases more broadly. For example, this notion is consistent with the observation that many histidine kinases do not have HAMP domains. In fact, only ~24% of histidine kinases have a HAMP domain N-terminal to the DHp domain (Szurmant et al., 2007). Some kinases have a DHp domain immediately adjacent to the last transmembrane domain, with no additional domains, whereas others have a PAS or GAF domain in this region. Some kinases have multiple domains preceding the DHp domain, for example, in *B. subtilis* KinA there are three PAS domains (Lee et al., 2008). Although HAMP, PAS, and GAF domains can homodimerize and hence could contribute to specificity, these domains are often involved in transmitting signals from an extracellular or periplasmic domain to the kinase domain. A recent NMR study of full-length *E. coli* histidine kinase DcuS showed that increased disorder and flexibility in the N-terminal helix of the cytoplasmic PAS domain increased kinase activity (Etzkorn et al., 2008). Similarly, there is evidence that HAMP domains may fluctuate between two different packing states, which may correlate with different signaling states (Airola et al., 2010; Ferris et al., 2011; Hulko et al., 2006; Zhou et al., 2009). The DHp domain thus appears to be the most static part of the kinase structure and, consequently, well suited to enforcing dimerization specificity.

Covariation analysis also supports a primary role for DHp domains in establishing homodimer specificity. Of the 100 most strongly covarying residue pairs in our histidine-kinase sequence alignment, the number between residues within the HAMP domain was close to that expected at random (28 vs. 22) and the number between the HAMP and DHp domains was significantly less

than expected at random (4 vs. 50, p-value < 10^{-24}). By contrast, the number of highly covarying pairs within the DHp domain was significantly more than expected (68 vs. 28, $p < 10^{-16}$). We also assessed whether the strongly covarying residues within the DHp and HAMP domains were positioned to make contacts across the dimer interface. Among the 100 most strongly covarying pairs, the number of interchain pairs within contact distance in the DHp domain was greatly enriched relative to that expected by chance (7 vs. 0.7, $p < 10^{-5}$) whereas the number of interchain contact pairs within the HAMP domain was not as significantly enriched (4 vs. 1.1, $p \approx 0.02$). The enrichment of covarying residues mediating interchain contacts within the DHp domain suggests that this domain might be particularly important for dimerization specificity.

Results from co-variation analysis are based on thousands of kinase sequences and are in no way specific to EnvZ. For a residue pair to show significant co-variation, a very large number of proteins must contribute to the signal. Thus, although we have experimentally verified the importance of the DHp domain for dimerization specificity in only one kinase, EnvZ, the above analysis suggests this finding could generalize to many other kinases. Amino-acid covariation within the HAMP domains was also previously examined, and three clusters of residues identified were proposed to affect signaling and transitions between conformational states (Dunin-Horkawicz and Lupas, 2010).

Modularity within the DHp domain

The DHp four-helix bundle can be divided into the membrane proximal region and the base (Fig. 2.7A). Covariation analysis identified more interchain residue pairs in the base, arguing that this region might be more important for specificity. The importance of this region is also supported by recent structural analyses showing that there is more variability in the membrane proximal

part of the DHp structure than in the base. For example, in *T. maritima* TM0853, the first three turns of helix 1 in the DHp are unfolded when the kinase is bound to its cognate response regulator, but folded into a 2-helix coiled coil when the kinase is in the unbound state (Casino et al., 2009). For the *B. subtilis* histidine kinase DesK, conformational states likely corresponding to the kinase, phosphatase, and phosphotransfer states of the kinase were crystallized, and only in the putative phosphatase state was the N-terminus of the DHp folded into a coiled coil (Albanesi et al., 2009). The smaller amount of conformational diversity at the base of the DHp domain makes it more likely to mediate and specify selective dimerization.

Within the base of the DHp domain, the structure of EnvZ also appears to maintain a degree of modularity. It is notable that all four EnvZ/RstB chimeras examined were able to form stable homodimers, indicating that significant portions of the RstB bundle could be transplanted into an EnvZ context without disrupting the domain. However, different layers of the helix bundle are not completely insulated from one another, as changes in the chimera 1 loop significantly influence mutations in the chimera-2 region. The base of EnvZ also contains another level of modularity with respect to protein-protein interactions. This region is important for both homodimerization and for phosphotransfer to cognate response regulators (Skerker et al., 2008). However, residues we identified as important for establishing the dimerization specificity of EnvZ (252, 256, 270, 266) do not overlap with residues identified as important for phosphotransfer specificity (250, 254, 255, 269). These two sets of residues are positioned close together in the base of the DHp four-helix bundle, with the more solvent exposed sites governing regulator specificity and the more buried sites influencing dimerization specificity (Fig. 2.11C).

Localizing specificity determinants in the DHp domain using covariation analysis

To identify individual residues in the base of the EnvZ DHp domain that are sufficient to alter dimerization specificity, we used amino-acid covariation analysis. The sites identified with the greatest effects on specificity were positions 252, 256, 270, and 266 in EnvZ, although we do not rule out that additional specificity residues may be localized to other parts of the structure (Fig. 2.7B). Making substitutions in EnvZ with the corresponding residues from RstB at these covarying sites destabilized both homodimerization and heterodimerization with $EnvZ_{HDC}$. However, introducing the same mutations in the context of chimera 1 (chimera 1 + I252L/I270L and chimera 1 + L266E), which includes the loops at the base of the DHp domain, produced kinases that maintained stable self-association and only weakly hetero-associated with $EnvZ_{HDC}$ (Table 2.1). Replacing only the loop of EnvZ, as in chimera 1, produced a kinase that stably homodimerized but continued to heterodimerize with EnvZ. We thus speculate that the chimera 1 mutations may alter the structure near the base of the bundle, better accommodating substitutions I252L, I270L and L266E, and thereby stabilizing the homodimer and destabilizing the heterodimer with EnvZ. An alternative possibility is that instead of being structurally coupled, the loop region in chimera 1 and substitutions I252L, I270L, and L266E do not influence one another. Replacing the loop in EnvZ with that in RstB could instead simply stabilize the dimer, allowing it to accommodate destabilization from chimera 2 substitutions.

Although chimera 1 introduced 10 amino-acid changes into EnvZ, it may only be a subset of these residues that influence substitutions introduced in chimera 2. For example, two positions within chimera 1, 256 and 259, covaried strongly with positions 252, 270 and 266 within chimera 2. The substitution T256L along with I252L and I270L led to more homospecificity than

I252L and I270L alone. With only three mutations, this kinase achieved a notable fraction of the full homospecificity found in chimera 1 + I252L/I270L.

Our results, and those reported previously, demonstrate the power of using amino-acid coevolution analyses to guide the identification of critical specificity-determining residues (Skerker et al., 2008; Thattai et al., 2007). Here, covariation analyses highlighted a small number of interchain contacts, among the dozens of possible contacts in the DHp interface, many of which significantly influenced dimerization specificity. Identifying interchain contacts, however, relied on the EnvZ NMR structure. Also, not all of the covarying positions affected specificity when mutations were made at those sites. The covariation analysis also may have missed important residues within the loop at the base of the DHp domain. As noted previously, these loop regions do not align well, making their analysis by covariation methods difficult (Skerker et al., 2008). Finally, it is clear that the sequence context in which mutations are made can matter. For instance, adding the double mutant I252L/I270L into EnvZ versus chimera 1 resulted in homodimers with very different stabilities. These observations underscore the importance of combining analyses of coevolution with structural analysis and experimental studies.

Evolutionary implications

Paralogous kinases generated through gene duplication can initially physically associate. Establishing a new and distinct homodimerizing kinase requires a series of mutations in one or both kinases that prevent heterodimerization but retain function. The mutants of EnvZ created here, which homodimerize but no longer interact with EnvZ, represent possible evolutionary intermediates. In our most homospecific mutants, I252L/T256L/I270L, chimera 1 + I252L/I270L, and chimera 1 + L266E, we had to change both residues in the loop and residues

unique to chimera 2. One interesting and plausible evolutionary pathway for kinase dimerization specificity is that neutral sequence drift in the loop preceded the more destabilizing changes in the helix bundle base (positions 252, 270, and 266). This scenario would be consistent with our observations that mutations I252L/I270L and L266E alone destabilized both the homodimers and the heterodimers with EnvZ, but were accommodated in homodimers after introducing either T256L or the entire chimera 1 loop region.

The loop at the base of the four-helix bundle is the least conserved region of the DHp domain, and changes in the loop could place residues in the helix bundle base in a new structural context while being neutral with respect to interaction specificity. Chimera 1, which represents a loop swap between EnvZ and RstB, is neutral with respect to kinase dimerization specificity (Table 2.1), and with respect to phosphotransfer specificity (Fig. 2.11A). The notion that neutral mutations must accumulate first to provide a context for additional, selectively advantageous mutations has been suggested for a number of systems (Ortlund et al., 2007; Soskine and Tawfik, 2010; Wang et al., 2002).

The structural and functional modularity of EnvZ also informs models of how histidine kinases evolve. The ability to modulate dimerization specificity and kinase-regulator interaction specificity independently, each with a relatively small number of mutations, increases the evolvability of this large and important gene family. Coupled with mechanisms that add or replace sensory and signaling domains, facile routes to achieving the broad diversity of histidine kinases are not difficult to imagine.

Methods

Constructing pENTR clones and expression clones for histidine kinases

The Gateway recombinational cloning system (Invitrogen) was used to generate pENTR clones for all constructs. EnvZ and RstB constructs containing the HAMP, DHp, and CA domains ($EnvZ_{HDC}$ = amino acids 179-450, $RstB_{HDC}$ = amino acids 157-433) or the DHp and CA domains ($EnvZ_{DC}$ = amino acids 223-450, $RstB_{DC}$ = amino acids 201-433), as well as full-length response regulators OmpR and RstA were cloned previously (Skerker et al., 2005). Single-mutant EnvZ constructs were generated using QuikChange site-directed mutagenesis (Stratagene). Double and triple mutant EnvZ clones were generated using PCR-based site-directed mutagenesis, with DpnI digestion preceding blunt end ligation (Fisher and Pei, 1997). pENTR clones of chimeras 1-4 were constructed previously (Skerker et al., 2008).

The $EnvZ_{HDC}$ mutant in which the HAMP domain was replaced with the RstB HAMP was generated using splicing-by-overlap-extension PCR. In the first round the RstB HAMP and $EnvZ_{DC}$ were individually amplified and gel purified. The products were then mixed in a 1:1 molar ratio and amplified in a second round of PCR with outer primers and gel purified. The final product was cloned into pENTR/D-TOPO vector following the manufacturer's protocol (Invitrogen).

For constructs used as competitors in the FRET assays and for pull-down assays, a tandem FLAG tag (DYKDDDDKDYKDDDDKGSG) was introduced at the N-terminus of the appropriate kinase using PCR with a forward primer containing the tandem tag (Schneider et al., 2001). PCR products were then cloned into the pENTR/D-TOPO vector.

The Gateway recombinational cloning system was used to move genes from pENTR clones into IPTG inducible expression vectors with suitable purification tags. pENTR clones were recombined into either pHIS-MBP-DEST, pTRX-HIS-DEST, or pHIS-DEST destination vectors using the Gateway LR clonase reaction (Skerker et al., 2005).

Plasmids encoding N-terminal ECFP (pRG31) and N-terminal monomeric mYFP (pRG88) along with a His_6 tag were a generous gift from the laboratory of A.M. Stock (Gao et al., 2008a). EnvZ and RstB pENTR clones were amplified with forward and reverse primers containing NheI and NotI restriction sites and ligated into pRG31 and pRG88 vectors digested with NheI and NotI. All constructs were sequence verified.

Protein expression and purification

Proteins were expressed and purified as previously described (Skerker et al., 2005). Briefly, expression vectors were transformed into *E. coli* BL21-Tuner cells. Colonies were picked and grown in 1 L LB at 37 °C. After reaching OD_{600} ~0.6, cells were either induced with 0.3 mM IPTG for 4 hrs at 30 °C (for strains expressing proteins with His_6-MBP, TRX-His_6, or His_6 tags) or with 0.5 mM IPTG for 12-16 hrs at 18 °C (for strains expressing proteins tagged with CFP or YFP). Induced cells were then pelleted and stored at -80 °C. His_6-tagged proteins were purified using Ni-NTA agarose beads and final protein aliquots were stored in storage buffer (10 mM HEPES-KOH [pH 8.0], 50 mM KCl, 10% glycerol, 0.1 mM EDTA, 1 mM DTT) at -80 °C. Protein concentrations were determined using the Edelhoc method (1x PBS, 7 M GuHCl, pH 7.4) with absorbance measured at 280 nm (Edelhoch, 1967), except for fluorescent proteins for which concentration was measured by absorbance at 433 nm for CFP (ε 32500 M^{-1} cm^{-1}) or at

514 nm for YFP (ε 83400 M^{-1} cm^{-1}) (Shaner et al., 2005).

Pull-down assay

For pull-down assays, 2.5 μM $FLAG_2$-histidine kinase was mixed with 12.5 μM MBP-histidine kinase in HEPES buffer (10 mM HEPES-KOH, 50 mM KCl, 0.1 mM EDTA, pH 8.0). Non-specific bead-binding controls included only the MBP-tagged histidine kinase. Proteins were equilibrated for two hours at room temperature on an end-over-end rotator to allow time for subunit exchange before adding 40 μL of anti-FLAG M2 affinity gel (50% slurry, Sigma) that had been washed once in HEPES buffer. The protein mixture and beads were incubated for 30 minutes at 4 ^{o}C and then washed 3 times with HEPES buffer. Protein was competitively eluted using five column volumes of 3X-FLAG peptide (100 μg/mL, Sigma). Eluant was concentrated using StrataClean binding resin (Stratagene), run on 10% Tris-HCl SDS-PAGE gel (Bio-Rad), and visualized with Coomassie staining.

FRET competition binding assay

Equimolar mixtures of CFP-histidine kinase and YFP-histidine kinase at 0.5, 5, or 20 μM, and $FLAG_2$-histidine kinase were placed in 96-well plates (Corning), covered with a foil seal, and incubated for eight hours at 30 °C. Plates were incubated for eight hours because in a kinetic study of 5 μM CFP-$RstB_{HDC}$ mixed with 5 μM YFP-$RstB_{HDC}$, the FRET ratio required five hours to reach equilibrium (data not shown), although most other mixtures examined reached equilibrium within an hour. Fluorescence was measured using a Varioskan plate reader at 30 °C with three channels monitored: donor channel (excite 433 nm, emit 475 nm), acceptor channel (excite 488 nm, emit 527 nm), and FRET channel (excite 433 nm, emit 527 nm). For each well,

30 measurements were made and then averaged. The FRET ratio was calculated as the ratio of FRET channel signal to donor channel signal. To correct for crosstalk and bleed-through, a corrected FRET ratio was calculated as $(F_m-F_a(A_m/A_a)-F_d(D_m/D_d)) / D_m$ where D=donor channel, A=acceptor channel, F=FRET channel, d=donor sample alone, a=acceptor sample alone, and m=mix of donor and acceptor samples and unlabeled kinase (Gordon et al., 1998). The FRET ratio was measured when making a qualitative assessment of an interaction. The numerator of the corrected FRET ratio, $F_m-F_a(A_m/A_a)-F_d(D_m/D_d)$, referred to as the corrected FRET emission signal, was the signal used in fitting K_d values. K_d values were measured in duplicate.

Fitting equilibrium dissociation constants for homodimers and heterodimers

The experimental setup of mixing unlabeled $FLAG_2$-kinase (U) with an equimolar mixture of CFP-kinase (C) and YFP-kinase (Y) can be described by the following six equilibrium reactions: $2C\leftrightarrow C_2$, $2Y\leftrightarrow Y_2$, $2U\leftrightarrow U_2$, $C+Y\leftrightarrow CY$, $C+U\leftrightarrow CU$, and $Y+U\leftrightarrow YU$. To measure homodimer K_d values, the FRET competition was performed with kinases C, Y, and U having the same kinase sequences. For example, we mixed unlabeled $EnvZ_{HDC}$ with CFP-$EnvZ_{HDC}$ and YFP-$EnvZ_{HDC}$. We made the simplifying assumption that C_2, Y_2, and U_2 have the same equilibrium dissociation constant, K_d, and that CY, CU, and YU have the same dissociation constant, $K_d/2$ (heterodimer twice as likely as homodimer). The decrease in corrected FRET emission signal as the FRET complex CY is competed off by U is directly proportional to the concentration of FRET complex, and the concentration of the FRET complex is determined by the single fitting parameter K_d. A simulation of the FRET experiment was performed in MATLAB, where the system of ordinary differential equations describing the six reactions was integrated until they reached equilibrium. The initial conditions were given by the concentrations of C, Y, and U and

the single fitting parameter was the K_d. The concentrations of C and Y were 0.5 μM, except in the case of $RstB_{DC}$, where it was 20 μM. Starting with an initial guess for the K_d, the R^2 between the simulated data and the experimental data was calculated. The K_d was varied in 0.1 μM increments to maximize the R^2, referred to as R^2_{max} To estimate the range of K_d values that fit the data equally well, we calculated the lower and upper limits of K_d values that fit the data with $0.95R^2_{max}$.

To measure a heterodimer K_d, the FRET experiment was performed with C and Y as the same kinase and U a different kinase. As a result, three dissociation constants determine the equilibrium: the two homodimer dissociation constants for C_2 (same as for Y_2) and U_2, and the heterodimer dissociation constant for CU (same for YU). Because the homodimer K_d values are measured in separate experiments, the heterodimer K_d is the only fitting parameter. The fitting was done in a manner analogous to the fitting of homodimer K_d values.

To estimate the homodimer K_d for $RstB_{DC}$, we increased the concentration of CFP-$RstB_{DC}$ and YFP-$RstB_{DC}$ to 20 μM. We found that even in the presence of 50-fold more unlabeled $RstB_{DC}$, there was a significant amount of corrected FRET emission signal remaining. This was attributed to direct binding between the ECFP and mYFP at these higher concentrations. The corrected FRET emission signal for the mix of purified ECFP and mYFP, each at 20 μM, was comparable to the signal in the $RstB_{DC}$ experiment at 50-fold more unlabeled $RstB_{DC}$. We corrected for this background level of binding between CFP and YFP by subtracting the CFP and YFP binding acceptor emission from the $RstB_{DC}$ experiment acceptor emissions. After this background correction, fitting the homodimer K_d was carried out as described above. Such a background

correction was unnecessary when fluorescent protein concentrations were 0.5 μM as the magnitude of CFP and YFP binding signal was only ~1-5% of the signal in the binding experiments involving kinases.

Phosphotransfer assay

Phosphotransfer assays were performed as previously described (Skerker et al., 2005). Briefly, purified $FLAG_2$-tagged histidine kinases and thioredoxin tagged response regulators were diluted to 5 μM in storage buffer plus 5 mM $MgCl_2$. Autophosphorylation was initiated with the addition of 500 μM ATP and 5 μCi [γ^{32}P]ATP (Amersham Biosciences, 6,000 Ci/mmol) to the kinase, and reactions were incubated for one hour at 30 °C. To initiate phosphotransfer, response regulator was added to the kinase in a 2.5 μM:2.5 μM ratio at room temperature and the reaction was stopped after 10 seconds by adding sample buffer. Samples were then run on a 10% Tris-HCl SDS-PAGE gel (Bio-Rad). The gel was exposed to a phosphor screen for two hours at room temperature and then scanned with a Storm 860 imaging system (Amersham Biosciences) at 50 μm resolution.

Histidine-kinase sequence alignments

Sequence alignments were built using hidden Markov models obtained from the Pfam database (Eddy, 1998; Finn et al., 2009). A sequence alignment of kinases with the domain architecture HAMP-DHp-CA was constructed by taking histidine-kinase sequences containing a DHp domain (PF00512) and identifying, using HMMER (http://hmmer.org), the subset of those sequences containing a HAMP domain (PF00672) followed by a DHp domain followed by a CA domain (PF02518). The sequence alignment was subjected to a 90% sequence identity cutoff (no pair of

sequences could share more than 90% sequence identity) and a 10% gap cutoff (columns with >10% gaps were removed). The final alignment of sequences contained 4,272 sequences.

To calculate whether highly covarying positions in the multiple sequence alignment preferentially occurred in HAMP or DHp domains, position pairs were first divided into three classes: both residues within the HAMP domain (total number of pairs, m = 1225), both residues within the DHp domain (m = 1596), or one residue from each domain (m = 2850). The process of choosing the observed number of pairs from one class, k, is described by the hypergeometric distribution; i.e., the probability that k pairs out of n pairs are from a single class, when selected randomly without replacement from a set of N pairs containing m pairs of that class, is given by $P(X=k)=\binom{m}{k}\binom{N-m}{n-k}\Big/\binom{N}{n}$. In our case N = 5671 pairs, and n = 100 pairs. Of these 100 pairs, 28 pairs are between HAMP residues, 68 pairs are between DHp residues, and 4 pairs are between HAMP and DHp residues (these are the values of k for the 3 classes). The expected number of covarying pairs from a class is given by the mean of this distribution, nm/N. This null model, which assumes that covarying pairs are distributed randomly over domains, acts to correct for domain size because longer domains are expected to contain more covarying pairs. The probability that k or more covarying pairs from the same class were chosen from a distribution characterized by the above null model is $P(X \geq k)$.

The expected number of interchain contacts within the HAMP (PDB id: 2asw) or DHp (PDB id: 1joy) domains and the significance of seeing the observed number of contacts were modeled in the same manner as above. The null hypothesis was that interchain contacts were randomly distributed across a rank-ordered list of covarying pairs. When counting the observed number of

interchain contacts, as before the top 100 covarying pairs out of 5671 total pairs were considered. Covarying pairs were divided into three classes: HAMP domain interchain contacts (k = 4, m = 63), DHp domain interchain contacts (k = 7, m = 39), or none of the above.

Covariation and structure

For the HAMP-DHp-CA multiple sequence alignment, mutual information with the average product correction, MIp, was measured for every pair of positions within the HAMP and DHp domains (Dunn et al., 2008). When analyzing covariation, a MIp score threshold of 0.1 was chosen (Fig. 2.6). The 25 highest scoring pairs were mapped onto the structure of either the Af1503 HAMP domain or the EnvZ DHp domain (Hulko et al., 2006; Tomomori et al., 1999). Distances were measured between heavy atoms, and the distance between a pair of residues was measured in either an intrachain context or an interchain context. In an intrachain context, residues are on the same monomer in the dimer whereas in an interchain context, residues are on opposite monomers in the dimer. A pair was classified as a contact if the minimum distance between the residues was less than 5.5 Å.

Acknowledgements

We thank the laboratory of A.M. Stock for providing plasmid reagents. We thank members of the Keating laboratory (especially T.S. Chen, C. Negron, L. Reich, A. W. Reinke, and V. Potapov) and the Laub laboratory (especially E.J. Capra, B.S. Perchuk, and C.G. Tsokos) for helpful discussions. We thank the BioMicro Center of the Massachusetts Institute of Technology for use of the Varioskan plate reader. We thank the HHMI-MIT Summer Research Program in Chemical Biology for support of K. Rozen-Gagnon. This work was funded by National Institutes of Health award GM067681 to A.E. Keating, a National Science Foundation CAREER Grant to M.T. Laub, and the National Science Foundation GRFP fellowship to O. Ashenberg. M.T.L. is an Early Career Scientist of the Howard Hughes Medical Institute. We used computer resources provided by National Science Foundation award 0821391.

References

Airola, M. V., Watts, K. J., Bilwes, A. M., and Crane, B. R. (2010). Structure of concatenated HAMP domains provides a mechanism for signal transduction. Structure **18,** 436-48.

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. Proc Natl Acad Sci U S A **106,** 16185-90.

Bick, M. J., Lamour, V., Rajashankar, K. R., Gordiyenko, Y., Robinson, C. V., and Darst, S. A. (2009). How to switch off a histidine kinase: crystal structure of *Geobacillus stearothermophilus* KinB with the inhibitor Sda. J Mol Biol **386,** 163-77.

Bilwes, A. M., Alex, L. A., Crane, B. R., and Simon, M. I. (1999). Structure of CheA, a signal-transducing histidine kinase. Cell **96,** 131-41.

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). J Mol Biol **329,** 495-503.

Capra, E. J., Perchuk, B. S., Lubin, E. A., Ashenberg, O., Skerker, J. M., and Laub, M. T. (2010). Systematic dissection and trajectory-scanning mutagenesis of the molecular interface that ensures specificity of two-component signaling pathways. PLoS Genet **6,** e1001220.

Casino, P., Rubio, V., and Marina, A. (2009). Structural insight into partner specificity and phosphoryl transfer in two-component signal transduction. Cell **139,** 325-36.

Chen, L., Willis, S. N., Wei, A., Smith, B. J., Fletcher, J. I., Hinds, M. G., Colman, P. M., Day, C. L., Adams, J. M., and Huang, D. C. (2005). Differential targeting of prosurvival Bcl-2 proteins by their BH3-only ligands allows complementary apoptotic function. Mol Cell **17,** 393-403.

Cheung, J., and Hendrickson, W. A. (2010). Sensor domains of two-component regulatory systems. Curr Opin Microbiol **13,** 116-23.

Dunin-Horkawicz, S., and Lupas, A. N. (2010). Comprehensive analysis of HAMP domains: implications for transmembrane signal transduction. J Mol Biol **397,** 1156-74.

Dunn, S. D., Wahl, L. M., and Gloor, G. B. (2008). Mutual information without the influence of phylogeny or entropy dramatically improves residue contact prediction. Bioinformatics **24,** 333-40.

Eddy, S. R. (1998). Profile hidden Markov models. Bioinformatics **14,** 755-63.

Edelhoch, H. (1967). Spectroscopic determination of tryptophan and tyrosine in proteins. Biochemistry **6,** 1948-54.

Etzkorn, M., Kneuper, H., Dunnwald, P., Vijayan, V., Kramer, J., Griesinger, C., Becker, S., Unden, G., and Baldus, M. (2008). Plasticity of the PAS domain and a potential role for signal transduction in the histidine kinase DcuS. Nat Struct Mol Biol **15,** 1031-9.

Ferris, H. U., Dunin-Horkawicz, S., Mondejar, L. G., Hulko, M., Hantke, K., Martin, J., Schultz, J. E., Zeth, K., Lupas, A. N., and Coles, M. (2011). The Mechanisms of HAMP-Mediated Signaling in Transmembrane Receptors. Structure **19,** 378-85.

Finn, R. D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J. E., Gavin, O. L., Gunasekaran, P., Ceric, G., Forslund, K., Holm, L., Sonnhammer, E. L., Eddy, S. R., and Bateman, A. (2009). The Pfam protein families database. Nucleic Acids Res **38,** D211-22.

Fisher, C. L., and Pei, G. K. (1997). Modification of a PCR-based site-directed mutagenesis method. Biotechniques **23,** 570-1, 574.

Forst, S., Delgado, J., and Inouye, M. (1989). Phosphorylation of OmpR by the osmosensor EnvZ modulates expression of the ompF and ompC genes in *Escherichia coli*. Proc Natl Acad Sci U S A **86,** 6052-6.

Gao, R., Tao, Y., and Stock, A. M. (2008a). System-level mapping of *Escherichia coli* response regulator dimerization with FRET hybrids. Mol Microbiol **69,** 1358-72.

Gao, Z., Wen, C. K., Binder, B. M., Chen, Y. F., Chang, J., Chiang, Y. H., Kerris, R. J., 3rd, Chang, C., and Schaller, G. E. (2008b). Heteromeric interactions among ethylene receptors mediate signaling in *Arabidopsis*. J Biol Chem **283,** 23801-10.

Goodman, A. L., Merighi, M., Hyodo, M., Ventre, I., Filloux, A., and Lory, S. (2009). Direct interaction between sensor kinase proteins mediates acute and chronic disease phenotypes in a bacterial pathogen. Genes Dev **23,** 249-59.

Gordon, G. W., Berry, G., Liang, X. H., Levine, B., and Herman, B. (1998). Quantitative fluorescence resonance energy transfer measurements using fluorescence microscopy. Biophys J **74,** 2702-13.

Grefen, C., Stadele, K., Ruzicka, K., Obrdlik, P., Harter, K., and Horak, J. (2008). Subcellular localization and in vivo interactions of the *Arabidopsis thaliana* ethylene receptor family members. Mol Plant **1,** 308-20.

Hidaka, Y., Park, H., and Inouye, M. (1997). Demonstration of dimer formation of the cytoplasmic domain of a transmembrane osmosensor protein, EnvZ, of *Escherichia coli* using Ni-histidine tag affinity chromatography. FEBS Lett **400,** 238-42.

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,** 929-40.

Kishii, R., Falzon, L., Yoshida, T., Kobayashi, H., and Inouye, M. (2007). Structural and functional studies of the HAMP domain of EnvZ, an osmosensing transmembrane histidine kinase in *Escherichia coli*. J Biol Chem **282,** 26401-8.

Laub, M. T., and Goulian, M. (2007). Specificity in two-component signal transduction pathways. Annu Rev Genet **41,** 121-45.

Lee, J., Tomchick, D. R., Brautigam, C. A., Machius, M., Kort, R., Hellingwerf, K. J., and Gardner, K. H. (2008). Changes at the KinA PAS-A dimerization interface influence histidine kinase function. Biochemistry **47,** 4051-64.

Marina, A., Waldburger, C. D., and Hendrickson, W. A. (2005). Structure of the entire cytoplasmic portion of a sensor histidine-kinase protein. Embo J **24,** 4247-59.

Newman, J. R., and Keating, A. E. (2003). Comprehensive identification of human bZIP interactions with coiled-coil arrays. Science **300,** 2097-101.

Ninfa, E. G., Atkinson, M. R., Kamberov, E. S., and Ninfa, A. J. (1993). Mechanism of autophosphorylation of *Escherichia coli* nitrogen regulator II (NRII or NtrB): trans-phosphorylation between subunits. J Bacteriol **175,** 7024-32.

Ortlund, E. A., Bridgham, J. T., Redinbo, M. R., and Thornton, J. W. (2007). Crystal structure of an ancient protein: evolution by conformational epistasis. Science **317,** 1544-8.

Park, H., Saha, S. K., and Inouye, M. (1998). Two-domain reconstitution of a functional protein histidine kinase. Proc Natl Acad Sci U S A **95,** 6728-32.

Reinke, A. W., Grigoryan, G., and Keating, A. E. (2010). Identification of bZIP interaction partners of viral proteins HBZ, MEQ, BZLF1, and K-bZIP using coiled-coil arrays. Biochemistry **49,** 1985-97.

Scheu, P. D., Liao, Y. F., Bauer, J., Kneuper, H., Basche, T., Unden, G., and Erker, W. (2010).

Oligomeric sensor kinase DcuS in the membrane of *Escherichia coli* and in proteoliposomes: chemical cross-linking and FRET spectroscopy. J Bacteriol **192,** 3474-83.

Schneider, F., Hammarstrom, P., and Kelly, J. W. (2001). Transthyretin slowly exchanges subunits under physiological conditions: A convenient chromatographic method to study subunit exchange in oligomeric proteins. Protein Sci **10,** 1606-13.

Shaner, N. C., Steinbach, P. A., and Tsien, R. Y. (2005). A guide to choosing fluorescent proteins. Nat Methods **2,** 905-9.

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,** 1043-54.

Skerker, J. M., Prasol, M. S., Perchuk, B. S., Biondi, E. G., and Laub, M. T. (2005). Two-component signal transduction pathways regulating growth and cell cycle progression in a bacterium: a system-level analysis. PLoS Biol **3,** e334.

Soskine, M., and Tawfik, D. S. (2010). Mutational effects and the evolution of new protein functions. Nat Rev Genet **11,** 572-82.

Stiffler, M. A., Chen, J. R., Grantcharova, V. P., Lei, Y., Fuchs, D., Allen, J. E., Zaslavskaia, L. A., and MacBeath, G. (2007). PDZ domain binding selectivity is optimized across the mouse proteome. Science **317,** 364-9.

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

Szurmant, H., White, R. A., and Hoch, J. A. (2007). Sensor complexes regulating two-component signal transduction. Curr Opin Struct Biol **17,** 706-15.

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

Thattai, M., Burak, Y., and Shraiman, B. I. (2007). The origins of specificity in polyketide synthase protein interactions. PLoS Comput Biol **3,** 1827-35.

Tomomori, C., Tanaka, T., Dutta, R., Park, H., Saha, S. K., Zhu, Y., Ishima, R., Liu, D., Tong, K. I., Kurokawa, H., Qian, H., Inouye, M., and Ikura, M. (1999). Solution structure of the homodimeric core domain of *Escherichia coli* histidine kinase EnvZ. Nat Struct Biol **6,** 729-34.

Wang, X., Minasov, G., and Shoichet, B. K. (2002). Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. J Mol Biol **320,** 85-95.

Yang, Y., and Inouye, M. (1991). Intermolecular complementation between two defective mutant signal-transducing receptors of *Escherichia coli*. Proc Natl Acad Sci U S A **88,** 11057-61.

Zhou, Q., Ames, P., and Parkinson, J. S. (2009). Mutational analyses of HAMP helices suggest a dynamic bundle model of input-output signalling in chemoreceptors. Mol Microbiol **73,** 801-14.

Chapter 3

Helix bundle loops determine whether histidine kinases autophosphorylate *in cis* or *in trans*

This work is a currently being prepared as a manuscript with Amy E. Keating and Michael T. Laub.

Abstract

Bacteria frequently use two-component signal transduction pathways to sense and respond to environmental and intracellular stimuli. Upon receipt of a stimulus, a homodimeric sensor histidine kinase autophosphorylates and then transfers its phosphoryl group to a cognate response regulator. Although kinase autophosphorylation has been reported to occur both *in cis* and *in trans*, the molecular determinants dictating which mechanism is employed are unknown. Based on structural considerations, one model posits that the handedness of a loop at the base of the helical dimerization domain plays a critical role. Here, we directly tested this model by replacing the loop from *E. coli* EnvZ, which autophosphorylates *in trans*, with the loop from three PhoR orthologs that autophosphorylate *in cis*. These chimeric kinases autophosphorylated in *cis*, suggesting that this small loop is sufficient to determine autophosphorylation mechanism. Further, we report that the mechanism of autophosphorylation is conserved in orthologous sets of histidine kinases, despite highly dissimilar loop sequences. We found that four EnvZ orthologs and nine PhoR orthologs taken from widely divergent organisms are each autophosphorylated *in trans* or *in cis*, respectively. These findings suggest that histidine kinases are under selective pressure to maintain their mode of autophosphorylation, but can do so with a wide range of sequences.

Introduction

Organisms must sense and respond to their environments to survive. In many cases, organisms use membrane-bound protein kinases to directly sense environmental stimuli and, in response, phosphorylate substrates that can effect cellular changes. These phosphorylation events are typically reversible, allowing for temporal control of protein function. In bacteria, the predominant phosphorylation-based systems are two-component signal transduction pathways, which typically involve a sensor histidine kinase and its cognate substrate, a response regulator (Stock et al., 2000). Following activation by an input signal, a histidine kinase homodimer autophosphorylates on a conserved histidine. The phosphoryl group is then transferred to a cognate response regulator, which can often trigger a transcriptional response inside cells. Many histidine kinases are bifunctional and exhibit phosphatase activity toward their cognate regulators, allowing a system to reset once the input signal is removed. Using these systems, prokaryotes are able to respond to a large and diverse array of stimuli including light, carbon sources, quorum signals, antibiotics, and others (Mascher et al., 2006). Most bacteria encode dozens of paralogous histidine kinases and response regulators, with some encoding hundreds (Gooderham and Hancock, 2009).

The typical histidine kinase is an integral membrane homodimer in which an extracytoplasmic sensory domain is linked to a cytoplasmic transmitter domain through a transmembrane helix. Sensory domains are often members of the PAS or GAF family, but they show low sequence similarity owing to the diversity of inputs they recognize (Moglich et al., 2009). In contrast, the transmitter region of the kinase is highly conserved and always consists of a DHp (*d*imerization and *h*istidine *p*hosphotransfer) domain linked to a CA (*c*atalytic and *A*TP binding) domain (Fig.

3.1A). The DHp domain includes two α-helices that mediate homodimerization through the formation of a four-helix bundle (Tomomori et al., 1999). The CA domain forms an α/β-sandwich that binds ATP and catalyzes autophosphorylation of an exposed histidine residue in the DHp domain (Tanaka et al., 1998).

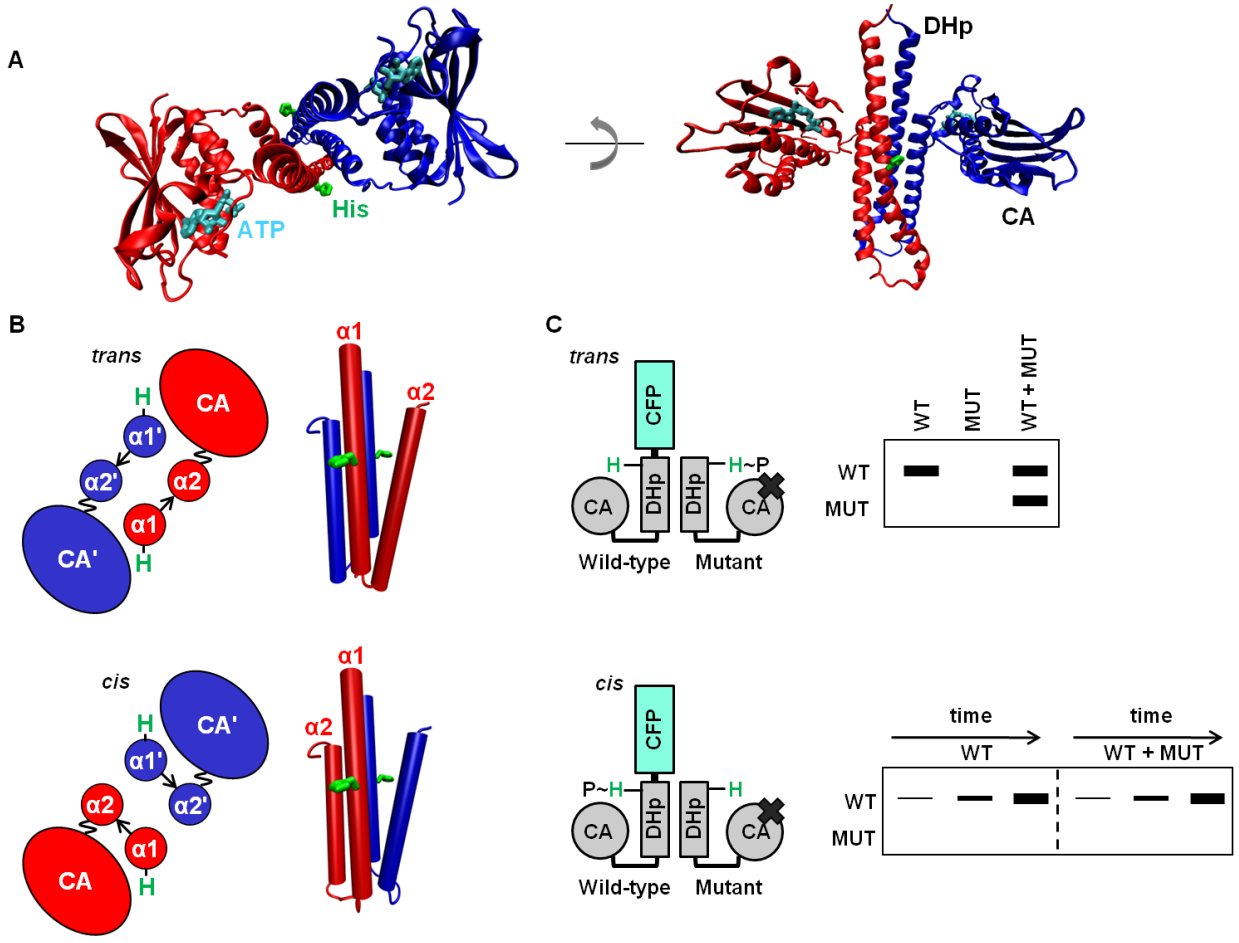


Figure 3.1. Histidine kinases autophosphorylate *in cis* or *in trans*.
(A) The cytoplasmic region of a histidine kinase (PDB ID 2c2a) consists of conserved DHp and CA domains. The histidine site of phosphorylation on the DHp domain, and the ATP analog bound by the CA domain, are shown in stick form in green and cyan, respectively. (B) Cartoon (left) looking down the four-helix bundle (right) of the DHp domain dimer. α1 and α2 helices in the DHp domain are labeled, with the ' symbol denoting the opposite chain (PDB ID 3ZRX). The loop at the base of the DHp domain is depicted by an arrow, and the linker between the DHp and CA domains is depicted as a wavy line to reflect the mobility of the CA domain. Depending on loop handedness in the DHp domain, the CA domain is either closer to the histidine on the same chain (*cis*) or the histidine on the opposite chain (*trans*). (C) Schematic of the assay to test *in cis* vs. *in trans* autophosphorylation. A wild-type (WT) histidine kinase homodimer is mixed with excess mutant (MUT) histidine-kinase homodimer unable to bind ATP. Autophosphorylation within the heterodimer is initiated by the addition of radiolabeled ATP, and the chains in the dimer are then separated by SDS-PAGE. Either the WT or MUT chain is labeled depending on whether the kinase

autophosphorylates *in cis* or *in trans*, respectively. When identifying *in cis* autophosphorylation, this assay is performed as a time course where autophosphorylation rates of WT and WT plus excess MUT are compared.

Early studies of the model histidine kinase EnvZ from *E. coli* demonstrated it autophosphorylates *in trans*, as the CA domain from one subunit of a homodimer phosphorylates the DHp domain of the other subunit (Fig. 3.1B) (Yang and Inouye, 1991). The *E. coli* kinases NtrB, AtoS, and CheA, and *S. aureus* AgrC, also autophosphorylate in *trans* (Filippou et al., 2008; George Cisar et al., 2009; Ninfa et al., 1993; Swanson et al., 1993; Trajtenberg et al., 2010), and all histidine kinases were initially assumed to function similarly. However, the kinases *T. maritima* HK853, *S. aureus* PhoR, and *E. coli* ArcB were recently shown to autophosphorylate exclusively *in cis* (Fig. 3.1B), *i.e.* the CA domain from one subunit autophosphorylates the DHp domain of the same subunit (although these kinases still homodimerize) (Casino et al., 2009; Pena-Sandoval and Georgellis, 2010). Whether these two mechanisms of autophosphorylation have functional consequences for signaling, and what determines whether a kinase autophosphorylates *in cis* or *in trans*, are currently unknown.

One model suggests that the autophosphorylation mechanism is determined by the loop at the base of the four-helix bundle formed through homodimerization of the DHp domain (Casino et al., 2009). This loop connects two α-helices (Fig. 3.1B) and is right-handed in *E. coli* EnvZ, which autophosphorylates *in trans*, but left-handed in *T. maritima* HK853, which autophosphorylates *in cis* (Casino et al., 2009; Tomomori et al., 1999). Differences in loop handedness likely do not significantly affect the structure of the dimeric helix bundle (Fig. 3.1A), but would change whether the ATP bound by a CA domain of one subunit is closer to the autophosphorylation site of the same subunit or the opposite subunit (Fig. 3.1B) (Casino et al., 2010). Although the CA domain moves significantly to autophosphorylate the DHp domain, a

relatively short linker between these two domains likely restricts motion of the CA domain such that it cannot autophosphorylate both DHp domains in a dimer. Multiple sequence alignments for DHp domains indicate that the sequence of the DHp domain loop is highly variable, even in orthologous histidine kinases (Fig. 3.2A). Such diverse loops may differ in their handedness, and therefore orthologs may differ with respect to whether they autophosphorylate *in cis* or *in trans*.

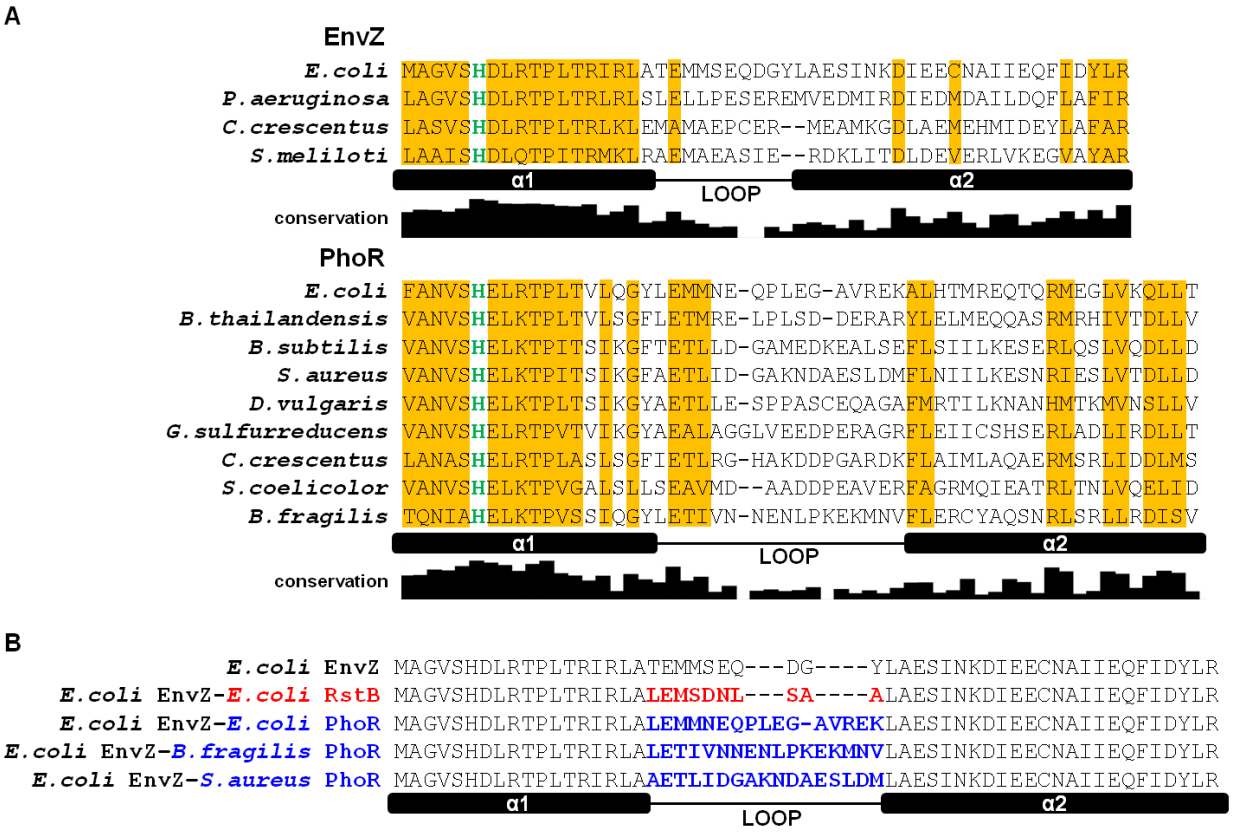


Figure 3.2. Alignments of EnvZ and PhoR orthologs and loop chimeras.
(A) Sequence alignment of DHp domains from EnvZ and PhoR orthologs characterized in this study. Locations of the two helices in the DHp domain, α1 and α2, and the loop connecting them are indicated. The histidine site of phosphorylation is colored in green. Columns where one residue is conserved in more than half of all available EnvZ or PhoR orthologs are shaded in orange (see Methods). The conservation graph below each alignment measures sequence information for each DHp position across all orthologs. (B) Alignment of DHp domains of loop chimeras. Sequence changes relative to *E. coli* EnvZ are highlighted in either red (RstB) or blue (PhoR).

Here, we tested whether autophosphorylation mechanism is conserved among orthologs of the kinases EnvZ and PhoR (Casino et al., 2009; Yang and Inouye, 1991). Surprisingly, we found that the mode of autophosphorylation is highly conserved, as diverse orthologs of EnvZ and PhoR autophosphorylated *in trans* and *in cis*, respectively. To directly probe the contribution of the DHp loop to autophosphorylation mechanism, we replaced the DHp loop in EnvZ with the loops from kinases that autophosphorylate *in cis*. Resulting chimeras autophosphorylated *in cis,* indicating that the loops are sufficient to dictate the autophosphorylation mechanism. These findings, along with the conservation of mechanism in orthologous kinases, indicate that the DHp loop is a functionally important determinant of autophosphorylation mechanism in histidine kinases, supplanting the notion that these loops are simple linkers between helices.

Results

Methodology for determining whether autophosphorylation occurs *in cis* or *in trans*

To determine whether a kinase autophosphorylated *in cis* or *in trans*, we characterized the *in vitro* phosphorylation of heterodimers comprising one wild-type subunit (WT) and one mutant subunit (MUT) that cannot bind ATP because it harbors an alanine in place of a conserved glycine in the G2 box of the CA domain (Fig. 3.1C) (Ninfa et al., 1993; Zhu and Inouye, 2002). If a kinase autophosphorylates exclusively *in cis*, then only the wild-type subunit in the heterodimer will be phosphorylated. If a kinase autophosphorylates *in trans*, then only the mutant subunit in the heterodimer will be phosphorylated (Fig. 3.1C). Thus, distinguishing between autophosphorylation *in trans* and *in cis* can be determined by assessing whether the mutant or wild-type subunits, respectively, are autophosphorylated. To distinguish the subunits, we fused CFP to the N-terminus of the wild-type subunit and a smaller FLAG epitope to the N-

terminus of the mutant subunit. Autophosphorylation was detected through addition of [γ-^{32}P]ATP followed by resolution of WT and MUT subunits with SDS-PAGE.

To form the heterodimer, homodimeric wild-type and mutant kinases were mixed together. As such a mixture equilibrates, three different dimers can form: the two possible homodimers and a heterodimer. We added 10-fold excess of the mutant kinase to favor formation of mutant homodimer and the heterodimer. The mutant homodimer cannot bind ATP, so any autophosphorylation activity observed should stem primarily from the heterodimer. For kinases that operate *in trans*, the mutant kinase chain in the heterodimer will be phosphorylated. For kinases that operate *in cis*, the wild-type kinase chain in the heterodimer will be phosphorylated. In the latter case, autophosphorylation of the wild-type kinase chain could also result from the small pool of remaining wild-type homodimer. Hence, to confirm cases of autophosphorylation *in cis*, we compared time courses for the wild-type kinase alone or in the presence of excess mutant kinase (Fig. 3.1C). If the time courses are similar, it indicates that the heterodimer was, in fact, autophosphorylating *in cis*. We assumed that the *in cis* autophosphorylation activity of the wild-type kinase would be independent of whether it was operating as a WT homodimer or a WT-MUT heterodimer.

Heterodimer formation in our assay results from subunit exchange following mixing of the wild-type and mutant kinases, and we measured the kinetics of this exchange using a previously developed FRET assay (Ashenberg et al., 2011). Briefly, we mixed CFP- and YFP-tagged versions of a wild-type kinase in the absence or presence of mutant kinase, and measured the change in FRET signal over time (for details, see Methods) (Fig. 3.3A). This assay confirmed that the wild-type and mutant kinases were competent to interact and indicated the appropriate

equilibration time necessary to form heterodimers.

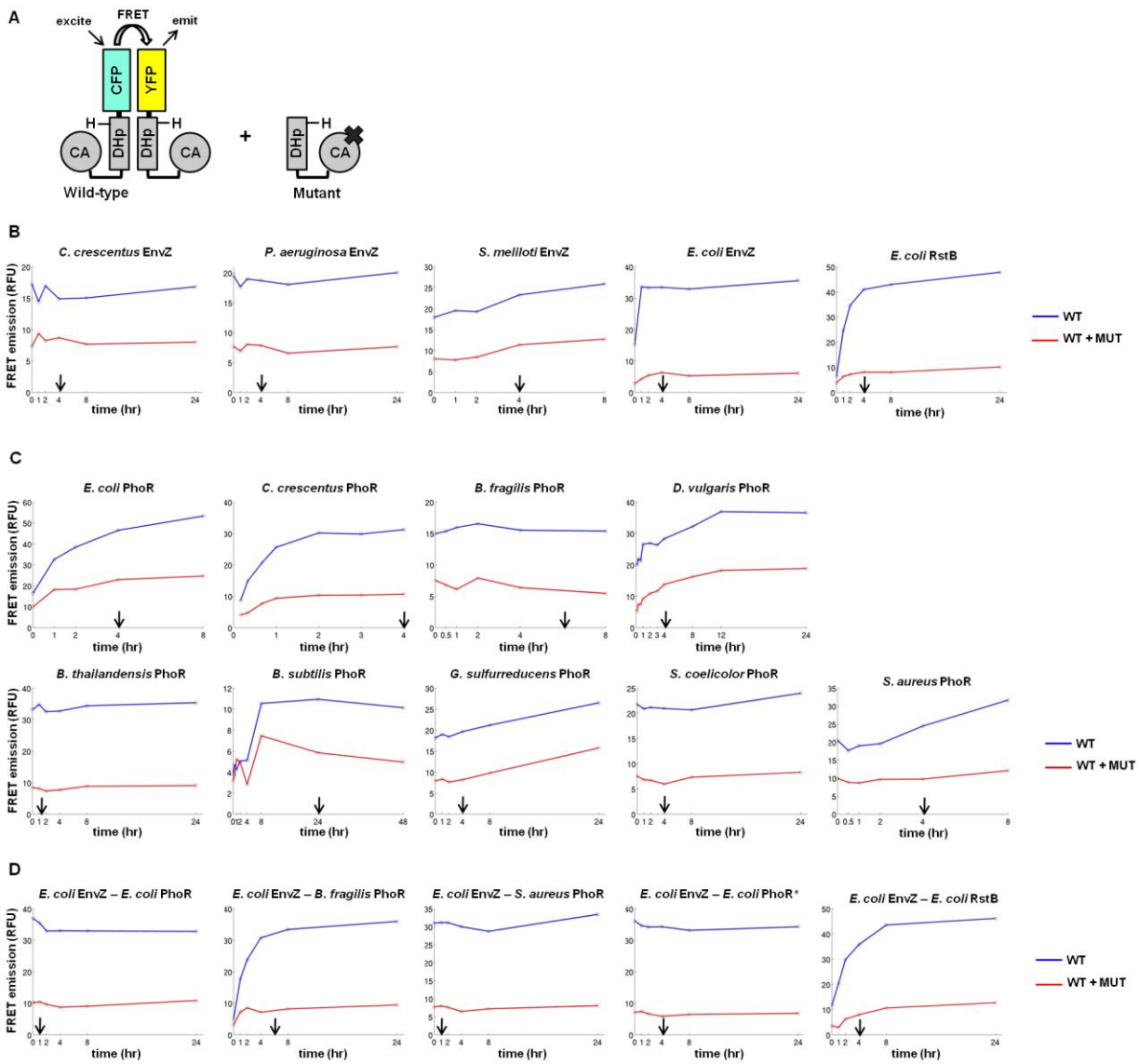


Figure 3.3. Determining equilibration times.
(A) CFP- and YFP-labeled wild-type kinase constructs (WT) were mixed together at time t = 0 and FRET signal from the resulting complex was monitored over time. FRET signal was reduced when excess mutant kinase (MUT) was included in the incubation. CFP and YFP fusion protein concentrations were 5 μM, with the indicated amount of mutant kinase (blue: mutant kinase absent; red: 50 μM mutant kinase). The experiments were carried out at 30 °C. FRET subunit exchange controls for EnvZ orthologs (B), for PhoR orthologs (C), and for loop chimeras (D). The equilibration time used for each kinase in the autophosphorylation mechanism determination assays is marked by an arrow.

Autophosphorylation mechanism is conserved among EnvZ and PhoR orthologs

E. coli EnvZ autophosphorylates *in trans* and *S. aureus* PhoR autophosphorylates *in cis* (Casino et al., 2009; Yang and Inouye, 1991). To assess whether the mechanisms are conserved, we examined autophosphorylation of four EnvZ orthologs and nine PhoR orthologs taken from phylogenetically diverse organisms (Table 3.1).

Table 3.1. Characterized EnvZ orthologs, PhoR orthologs, and loop chimeras.

PhoR ortholog	Species	Equilibration time (hrs)[a]	Class[b]	Region[c]
b0400	*Escherichia coli*	4	γ	N166-D431
CC0289	*Caulobacter crescentus*	4	α	L57-A471
BF1589	*Bacteroides fragilis*	6	Bacteroidia	D375-R610
DVU0013	*Desulfovibrio vulgaris*	4	δ	D357-V602
BTH2767	*Burkholderia thailandensis*	1	β	D196-A436
BSU29100	*Bacillus subtilis*	24	Bacillales	M231-A579
GSU1101	*Geobacter sulfurreducens*	4	δ	D302-T537
SCO4229	*Streptomyces coelicolor*	4	Actinobacteridae	D145-P426
NWMN_1585	*Staphylococcus aureus*	4	Bacillales	M316-E554
EnvZ ortholog				
b3404	*Escherichia coli*	4	γ	L223-G450
PA5199	*Pseudomonas aeruginosa*	4	γ	L212-A439
CC2932	*Caulobacter crescentus*	4	α	R237-G445
SM_b20218	*Sinorhizobium meliloti*	4	α	V181-L454
Loop chimera				
E. coli EnvZ–*E. coli* RstB		4		
E. coli EnvZ–*E. coli* PhoR		1		
E. coli EnvZ–*B. fragilis* PhoR		6		
E. coli EnvZ–*S. aureus* PhoR		1		
E. coli EnvZ–*E. coli* PhoR*		4		

[a] Equilibration times measured using FRET assay in Figure 3.3.
[b] The class of bacterial species a given ortholog belongs to. The Greek letters correspond to subdivisons of Proteobacteria.
[c] The sequence boundary (first and last amino-acid residue from full-length ortholog) used in constructing a given ortholog.

An alignment of the DHp domains from these orthologs highlights the diversity of the loops relative to the conservation in the two α-helices, α1 and α2 (Fig. 3.2A). Even for orthologs from the same proteobacterial subdivision, such as *E. coli* EnvZ and *P. aeruginosa* EnvZ (γ-proteobacteria) or *D. vulgaris* PhoR and *G. sulfurreducens* PhoR (δ-proteobacteria), there is often significant diversity in the loop sequences.

For each EnvZ and PhoR ortholog, we purified a wild-type construct and an ATP-binding deficient construct, each harboring the DHp and CA domains. We used the FRET assay described above to determine the time necessary to form sufficient heterodimer after mixing the two homodimers. For each kinase, the FRET signal formed by CFP- and YFP-tagged wild-type kinase was significantly reduced by addition of untagged mutant kinase (Fig. 3.3). Some kinases, like *P. aeruginosa* EnvZ, underwent subunit exchange within minutes whereas others, like *E. coli* PhoR, underwent subunit exchange over several hours (Fig. 3.3). This broad range of subunit exchange times for different kinases is consistent with previous reports (Cai and Inouye, 2003; Ninfa et al., 1993; Park et al., 2004).

Using the equilibration times indicated by the FRET assay, we tested the autophosphorylation mechanism for each EnvZ ortholog. We observed autophosphorylation for each wild-type EnvZ ortholog but not any of the mutated variants, as expected (Fig. 3.4A). When the wild-type and mutant kinases were mixed together, the mutant subunit was strongly phosphorylated in each case, clearly indicating that phosphorylation in the heterodimer occurred *in trans*. This indicates that the autophosphorylation mechanism for EnvZ is conserved, at least across α- and γ-proteobacteria.

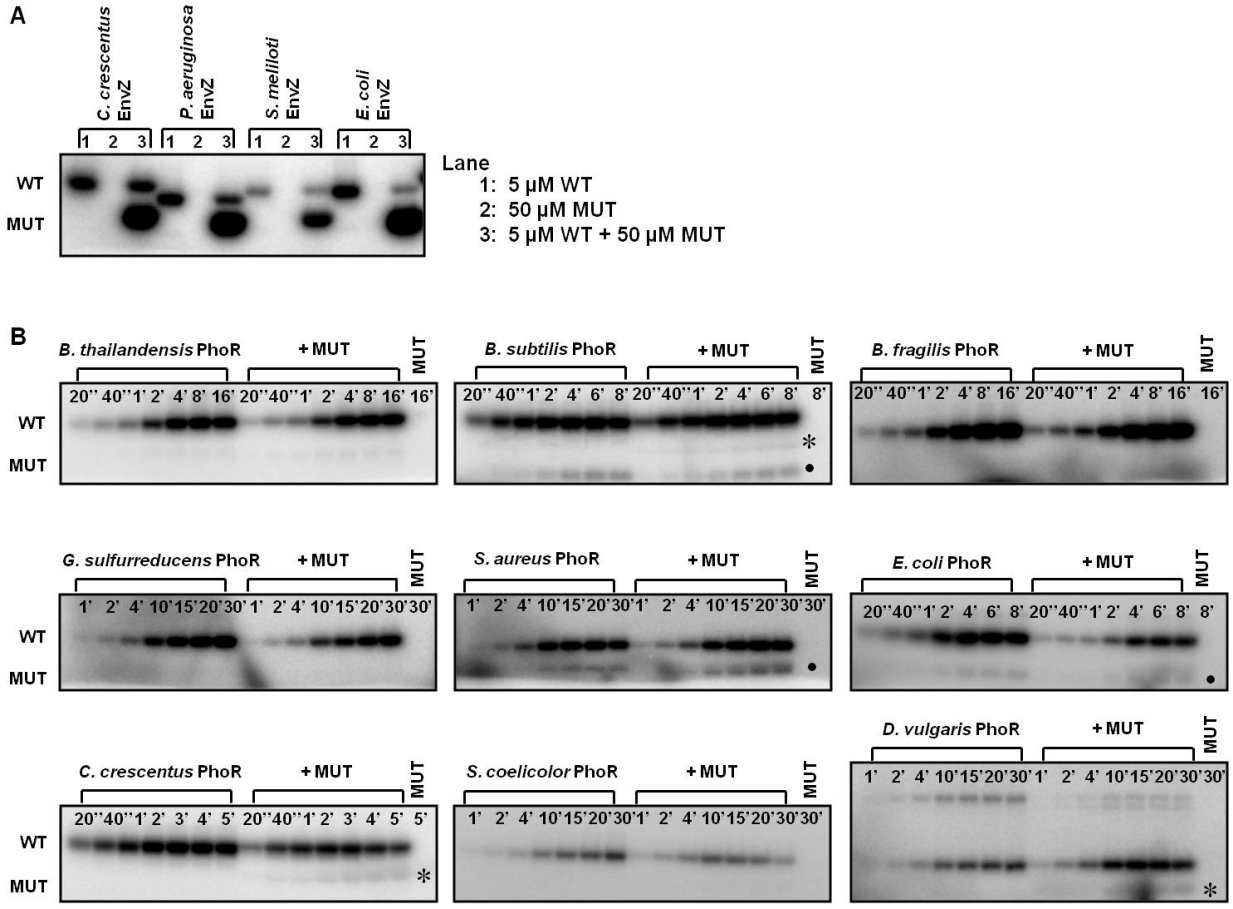


Figure 3.4. Autophosphorylation mechanism within EnvZ and PhoR orthologs is conserved.
(A) Characterization of autophosphorylation mechanism for EnvZ orthologs. In each ortholog series, the first lane contained 5 µM wild-type (WT) kinase, the second lane contained 50 µM mutant (MUT) kinase, and the third lane contained a mixture of the two kinases. The mixture was incubated four hours at 30 °C before initiating autophosphorylation. (B) Characterization of autophosphorylation mechanism for PhoR orthologs. The gel for each ortholog included two autophosphorylation time courses: the first for 5 µM wild-type kinase alone and the second for 5 µM wild-type kinase plus 50 µM mutant kinase. The final lane contained 50 µM mutant kinase. The autophosphorylation time is marked in each lane. Solid circles mark a breakdown product from the wild-type kinase, and asterisks indicate possible weak *in trans* autophosphorylation. Gel bands for each time course point are quantified and plotted in Fig. 3.5.

We carried out the same experiments with PhoR orthologs. As for the EnvZ orthologs, the wild-type PhoR constructs underwent autophosphorylation whereas the mutant constructs did not (Fig. 3.4B). For each ortholog, when the wild-type and mutant constructs were mixed together to form heterodimers, only the wild-type subunit was significantly phosphorylated, suggesting that each PhoR ortholog autophosphorylated *in cis*. In three cases (*B. subtilis*, *S. aureus*, and *E. coli*) we

observed an additional, smaller band. However, this band also appeared with WT homodimers, so it must represent a breakdown product from the WT kinase construct, not autophosphorylated heterodimer.

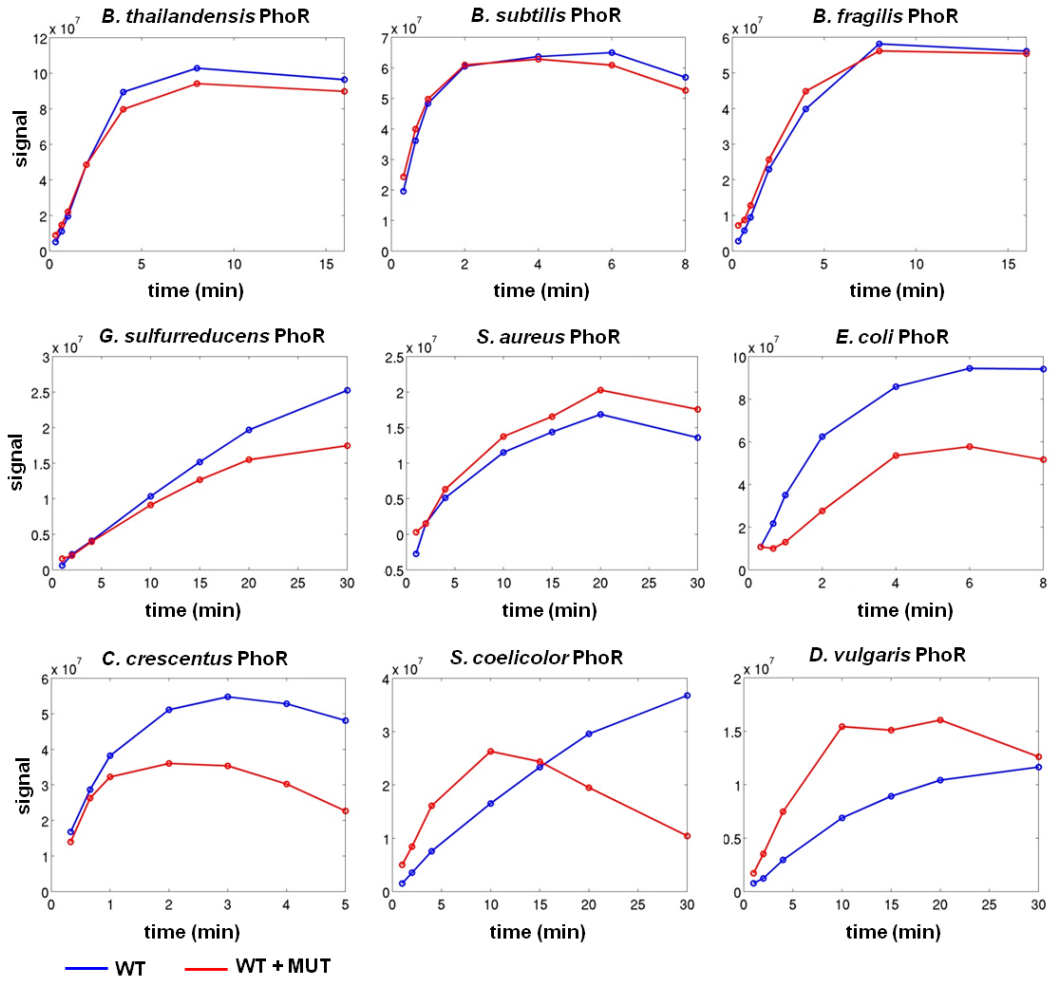


Figure 3.5. Quantification of autophosphorylation in PhoR orthologs.
Autophosphorylation signal of wild-type (WT) PhoR orthologs in the absence (blue) and presence (red) of the corresponding mutant (MUT) orthologs was compared. For each gel from Figure 3.4B, the autophosphorylation signal of the wild-type PhoR ortholog in each of the two time courses was quantified and plotted.

To confirm the *in cis* autophosphorylation results, we performed time-course experiments to compare the rate of autophosphorylation of the heterodimer with the wild-type homodimer. For five of the nine orthologs (*B. thailandensis*, *B. subtilis*, *B. fragilis*, *G. sulfurreducens*, *S. aureus*),

the rates were similar, strongly supporting the conclusion that these kinases autophosphorylated *in cis* (Fig. 3.4B, 3.5). For three PhoR orthologs (*E. coli*, *C. crescentus*, *S. coelicolor*), the rate of heterodimer autophosphorylation was moderately reduced relative to the homodimer. This reduction in phosphorylation could indicate that the WT-MUT heterodimer is less active than the wild-type homodimer, or that mutant kinase present in these mixtures can inhibit phosphorylation of the heterodimer. To test the latter possibility, we mixed wild-type *E. coli* and *C. crescentus* PhoR with their corresponding MUT variants, but omitted the equilibration step such that very little heterodimer should have formed (Fig. 3.3). A similar decrease in autophosphorylation occurred (Fig. 3.6), indicating that the wild-type PhoR constructs were partially inhibited simply by the presence of mutant protein, perhaps through dimer-dimer interactions. Finally, for one PhoR construct (*D. vulgaris*), autophosphorylation of the heterodimer was moderately higher than that of the homodimer. However, the wild-type construct appeared to form multimers by SDS-PAGE, possibly due to the cysteine residue in its DHp loop.

In sum, for all nine PhoR orthologs, our data indicate that autophosphorylation occurs *in cis*. In three cases (*B. subtilis*, *C. crescentus*, *D. vulgaris*), we observed a faint band corresponding to phosphorylation of the mutant chain (Fig. 3.4B). This band may indicate low rates of autophosphorylation *in trans*. Alternatively, it could result from dimer-dimer interactions and the phosphorylation of a mutant chain in one dimer by a wild-type chain from another dimer. Whatever the case, the mutant band was consistently much less intense than the wild-type band, supporting the conclusion that PhoR orthologs autophosphorylate primarily *in cis*.

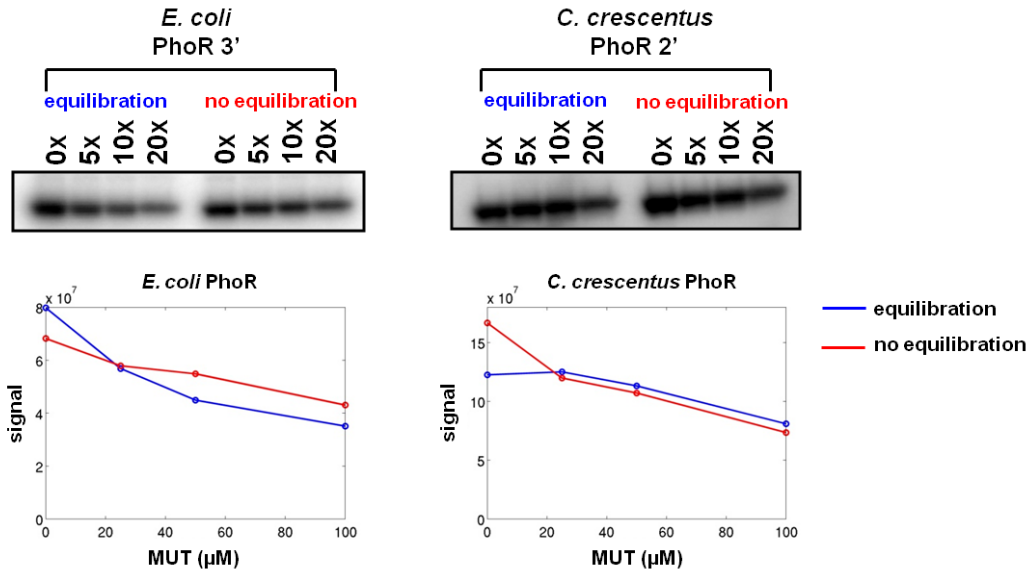


Figure 3.6. *E. coli* PhoR and *C. crescentus* PhoR.
Titration of mutant (MUT) PhoR ortholog into wild-type (WT) PhoR ortholog. Each lane contained 5 μM wild-type PhoR and the indicated amount of mutant PhoR (0x, 5x, 10x, 20x). For each PhoR ortholog, two titrations were performed. For the titration on the left, subunit equilibration in the mixture proceeded for 4 hours whereas for the titration on the right, the subunit equilibration step was omitted. In both titrations, autophosphorylation proceeded for the time listed above the respective gel. The plot below each gel quantifies the autophosphorylation level of wild-type PhoR ortholog as mutant PhoR ortholog is added. The experiments with and without equilibration are overlaid.

Autophosphorylation mechanism of loop chimeras

The DHp domains of *E. coli* EnvZ, which autophosphorylates *in trans*, and *T. maritima* TM853, which autophosphorylates *in cis*, both adopt four-helix bundle structures, but the loop connecting α–helices 1 and 2 in the two structures have different handedness. To test whether this difference in loops is related to the difference in autophosphorylation mechanism, we constructed chimeras in which we replaced the loop from *E. coli* EnvZ (Fig. 3.7A, B), with three diverse loops from *E. coli* PhoR, *B. fragilis* PhoR, or *S. aureus* PhoR, each of which autophosphorylates *in cis* (Fig. 3.2B, 3.4B). We also replaced the loop from *E. coli* EnvZ with that from *E. coli* RstB, another kinase that autophosphorylates *in trans* (Fig. 3.2B, 3.7A). Because there are no structures of the PhoR orthologs or of RstB, we do not know the exact boundary between helix and loop, so our

chimeras may have introduced a small number of helical residues flanking the DHp loop.

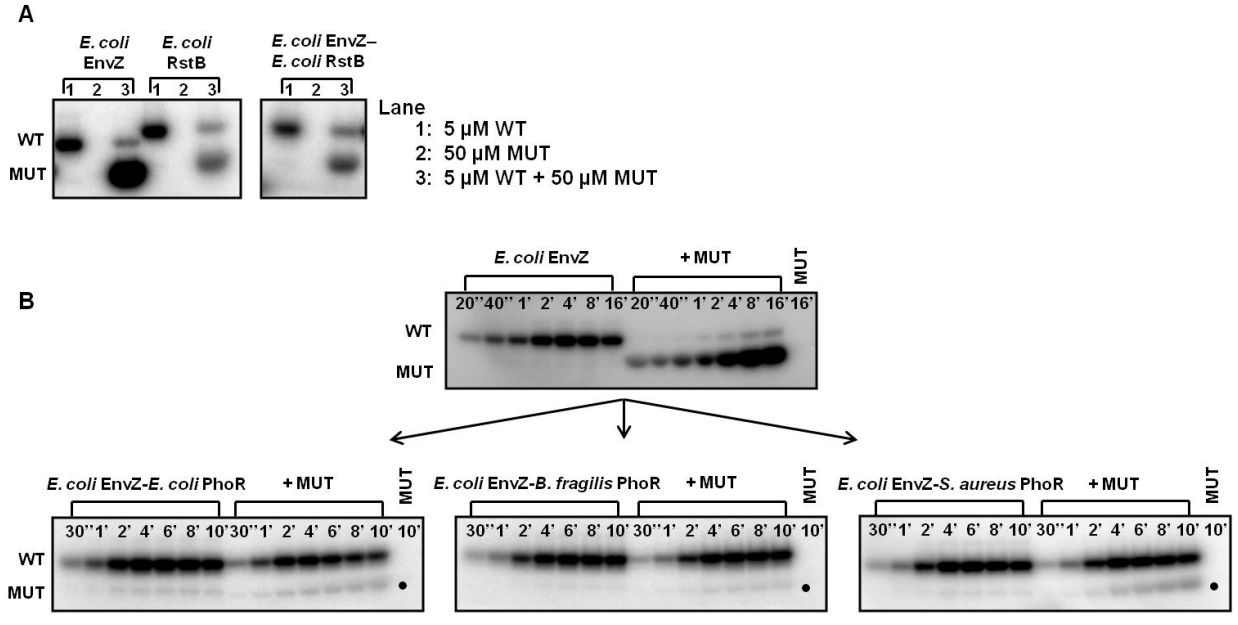


Figure 3.7. Changing DHp loop sequence changes autophosphorylation mechanism.
(A) Characterization of autophosphorylation mechanism in *E. coli* EnvZ, *E. coli* RstB, and the EnvZ-RstB chimera, as in Fig. 3.4A. (B) Autophosphorylation time courses for *E. coli* EnvZ and the EnvZ-PhoR loop chimeras, as in Fig. 3.4B. Gel bands for each time course point are quantified and plotted in Fig. 3.8.

The three EnvZ-PhoR chimeras each had autophosphorylation rates comparable to *E. coli* EnvZ, and reached their maximum level of autophosphorylation after ~5 minutes, indicating that the loop swaps did not significantly affect overall autokinase activity (Fig. 3.7B, 3.8). To test whether these chimeras autophosphorylated *in cis* or *in trans*, we mixed each chimera with its corresponding ATP-binding deficient mutant, as above. In each case, we continued to see high rates of autophosphorylation, with a single band corresponding to the kinase-active chain (Fig. 3.7B), indicating that autophosphorylation occurred *in cis*. For the chimera containing the loop from *E. coli* PhoR, the level of autophosphorylation observed after mixing with the corresponding mutant was similar to that of the chimera alone after 2 minutes, but was ~50% lower after more extended incubation times (Fig. 3.7B). However, we generated a new chimera

(*E. coli* EnvZ-*E. coli* PhoR*) in which we transplanted additional, loop-proximal residues from *E. coli* PhoR sequence into EnvZ (Fig. 3.9). This chimeric kinase also autophosphorylated *in cis*, but phosphorylation no longer significantly changed when mixed with its corresponding mutant.

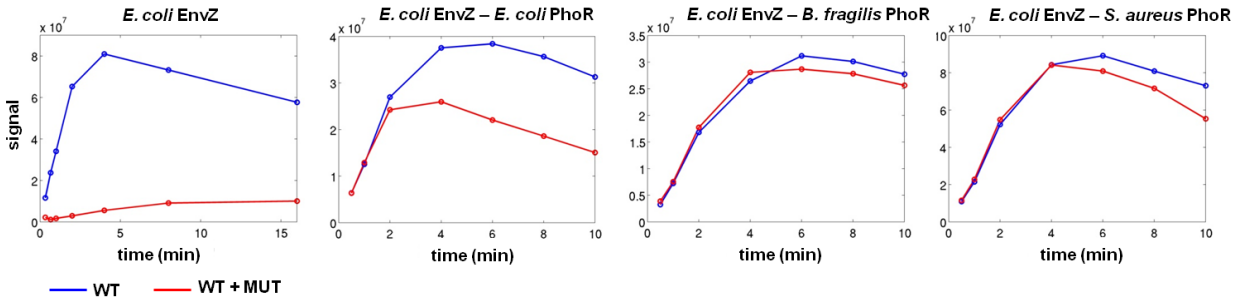


Figure 3.8. Quantification of autophosphorylation in loop chimeras.
Autophosphorylation signal of *E. coli* EnvZ or loop chimeras (WT) in the absence (blue) and presence (red) of the corresponding mutant (MUT) orthologs was compared. For each gel from Figure 3.7B, the autophosphorylation signal over time in each of the two time courses was quantified and plotted.

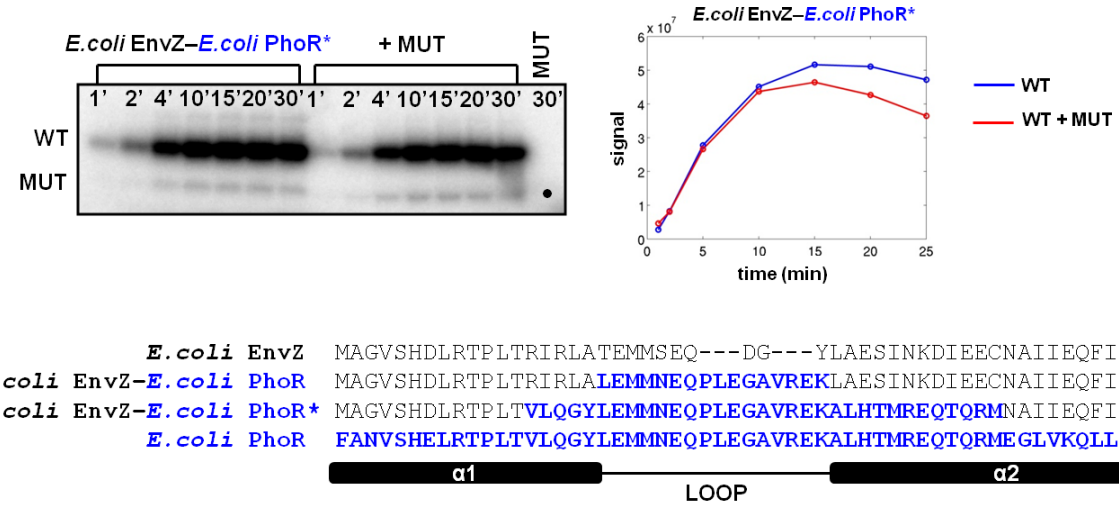


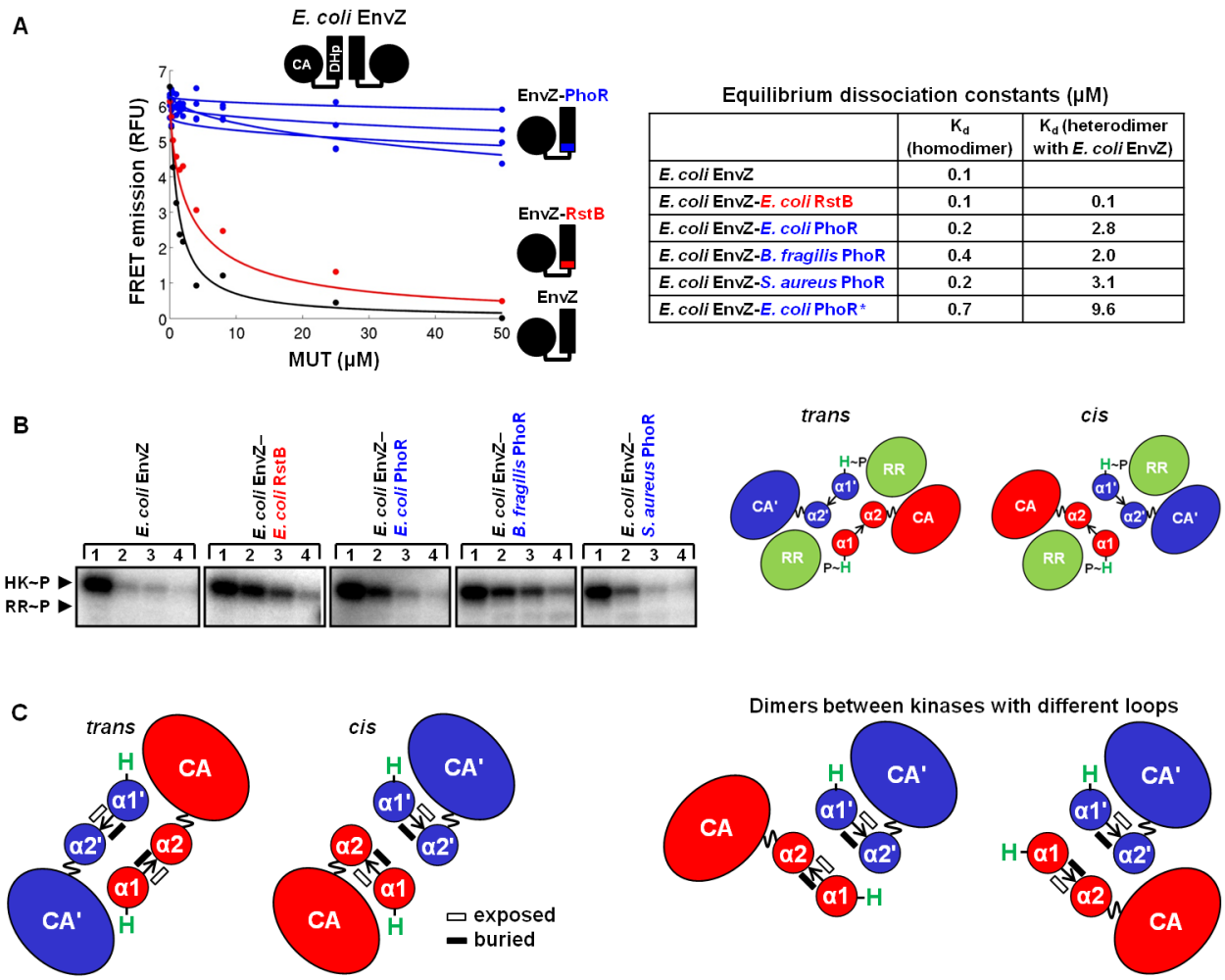
Figure 3.9. Autophosphorylation mechanism of *E. coli* EnvZ-*E. coli* PhoR* chimera.
Characterization of autophosphorylation mechanism in the *E. coli* EnvZ-*E. coli* PhoR* chimera. The chimera has additional *E. coli* PhoR sequence relative to the *E. coli* EnvZ – *E. coli* PhoR chimera. The sequence alignment highlights in blue the extent of *E. coli* PhoR sequence placed into *E. coli* EnvZ. The gel included two autophosphorylation time courses: the first contained 5 μM wild-type chimera (WT) and the second contained 5 μM wild-type chimera plus 50 μM mutant (MUT) chimera. The final lane contained 50 μM mutant chimera. The solid circle indicates a breakdown product from the wild-type kinase. The autophosphorylation signal from wild-type and chimera in each of the two time courses was quantified and plotted.

Finally, we tested the EnvZ-RstB chimera and found that it autophosphorylated *in trans*, like the wild-type EnvZ and RstB kinases from which it was derived (Fig. 3.7A). Taken together, our results demonstrate that the loop in the DHp domain of a histidine kinase is sufficient to change its autophosphorylation mechanism from operating *in trans* to *in cis*.

Loop swaps change dimerization specificity but not phosphotransfer specificity

We next assessed whether the EnvZ-RstB and EnvZ-PhoR loop chimeras showed any changes in kinase dimerization specificity or interaction with OmpR, the cognate response regulator for EnvZ. Previous studies demonstrated that histidine kinases specifically form homodimers and specifically phosphotransfer to their cognate response regulators *in vitro* (Ashenberg et al., 2011; Skerker et al., 2008).

We tested whether the chimeras could interact with *E. coli* EnvZ, using a FRET assay that allows fitting of equilibrium dissociation constants for both homodimers and heterodimers (Ashenberg et al., 2011). This assay was performed in a competitive-inhibition format, where the FRET signal from a mixture of CFP- and YFP- labeled *E. coli* EnvZ was inhibited by increasing concentrations of unlabeled, mutant kinase (Fig. 3.10A). We found that the chimera homodimers were similar in stability to the EnvZ homodimer, with dissociation constants ranging from ~0.1 μM to 0.5 μM (Fig. 3.10A, 3.11, Table 3.2). However, the heterodimers formed between EnvZ and the chimeras showed marked differences. The EnvZ-RstB chimera, which autophosphorylates *in trans* like EnvZ, interacted tightly with EnvZ. But the EnvZ-PhoR chimeras, which autophosphorylate *in cis*, interacted with EnvZ with dissociation constants approximately one order of magnitude weaker.



	K_d (homodimer)	K_d (heterodimer with *E. coli* EnvZ)
E. coli EnvZ	0.1	
E. coli EnvZ-*E. coli* RstB	0.1	0.1
E. coli EnvZ-*E. coli* PhoR	0.2	2.8
E. coli EnvZ-*B. fragilis* PhoR	0.4	2.0
E. coli EnvZ-*S. aureus* PhoR	0.2	3.1
E. coli EnvZ-*E. coli* PhoR*	0.7	9.6

Figure 3.10. Phosphotransfer specificity and dimerization specificity of loop chimeras.
(A) Each kinase was assayed for its ability to interact with *E. coli* EnvZ using a FRET competition assay. FRET signal from a complex of CFP-EnvZ and YFP-EnvZ was inhibited by interaction with unlabeled mutant (MUT) kinase (*E. coli* EnvZ, EnvZ-RstB chimera, EnvZ-PhoR loop chimeras, or *E. coli* EnvZ-*E. coli* PhoR*). From this FRET data, equilibrium dissociation constants (K_d) were fit for each kinase homodimer and for each kinase heterodimer with EnvZ. .(B) *E. coli* EnvZ and the EnvZ loop chimeras were assayed for their ability to phosphotransfer to the response regulator (RR) *E. coli* OmpR. Autophosphorylated histidine kinase (HK) was either incubated alone (first lane), or for 10" (second lane), 1' (third lane), or 5' (fourth lane) with OmpR. Each kinase was able to phosphotransfer to OmpR as seen by decreased levels of autophosphorylated kinase upon incubation with OmpR. On the right are cartoon depictions of a response regulator interacting with either the DHp domain dimer from a kinase that phosphorylates *in trans* or a from a kinase that phosphorylates *in cis*. The mode of interaction in the *cis* kinase is based on a crystal structure (PDB ID 3DGE), and the mode of interaction in the *trans* kinase is a model. (C) A DHp dimer formed between chains with different loop handedness cannot form the stereotypical DHp interface. In such a dimer, either the usual buried dimer interface will not be formed, or the arrangement of the α1 and α2 helices will be altered from what is observed in known structures.

We also tested the phosphotransfer activity of the chimeric kinases by incubating each kinase with *E. coli* OmpR and [γ^{32}P]ATP. Wild-type *E. coli* EnvZ phosphorylates its cognate response regulator OmpR *in vitro*, as manifested by a significant and rapid decrease in phosphorylated EnvZ upon addition of OmpR (Fig. 3.10B). Note that a strong band corresponding to OmpR is not expected, because EnvZ also drives the dephosphorylation of OmpR~P after phosphotransfer. Each chimera demonstrated robust phosphotransfer to OmpR indicating that the change in loop and consequent change in autophosphorylation mechanism did not affect interaction between the kinase and its cognate response regulator.

Table 3.2. Equilibrium dissociation constants for EnvZ and EnvZ chimeras (μM).[a]

	K_d (homodimer)	K_d (homodimer)	K_d (heterodimer with EnvZ)	K_d (heterodimer with EnvZ)
E. coli EnvZ	0.1 (0.1, 0.5)	0.1 (0.1, 0.8)		
E. coli EnvZ–*E. coli* RstB	0.1 (0.1, 0.9)	0.1 (0.1, 0.8)	0.1 (0.1, 0.19)	0.1 (0.1, 0.16)
E. coli EnvZ–*E. coli* PhoR	0.2 (0.1, 1.8)	0.1 (0.1, 0.8)	3.3 (2.8, 4.2)	2.3 (2.0, 2.9)
E. coli EnvZ–*B. fragilis* PhoR	0.6 (0.1, 2.7)	0.1 (0.1, 1.4)	2.0 (1.8, 2.2)	2.0 (1.7, 2.4)
E. coli EnvZ–*S. aureus* PhoR	0.1 (0.1, 1.1)	0.2 (0.1, 1.6)	2.8 (2.3, 3.6)	3.4 (2.8, 4.3)
E. coli EnvZ–*E. coli* PhoR*	0.3 (0.1, 1.9)	1.0 (0.1, 3.3)	14.7 (11.6, 19.9)	4.5 (4.0, 5.0)

[a] The lower and upper limits of K_d values that fit the data with $0.95R^2_{max}$ are listed in parentheses. Limits in sensitivity of the FRET competition assay prevent measurement of homodimer K_d values tighter than 0.1 μM.

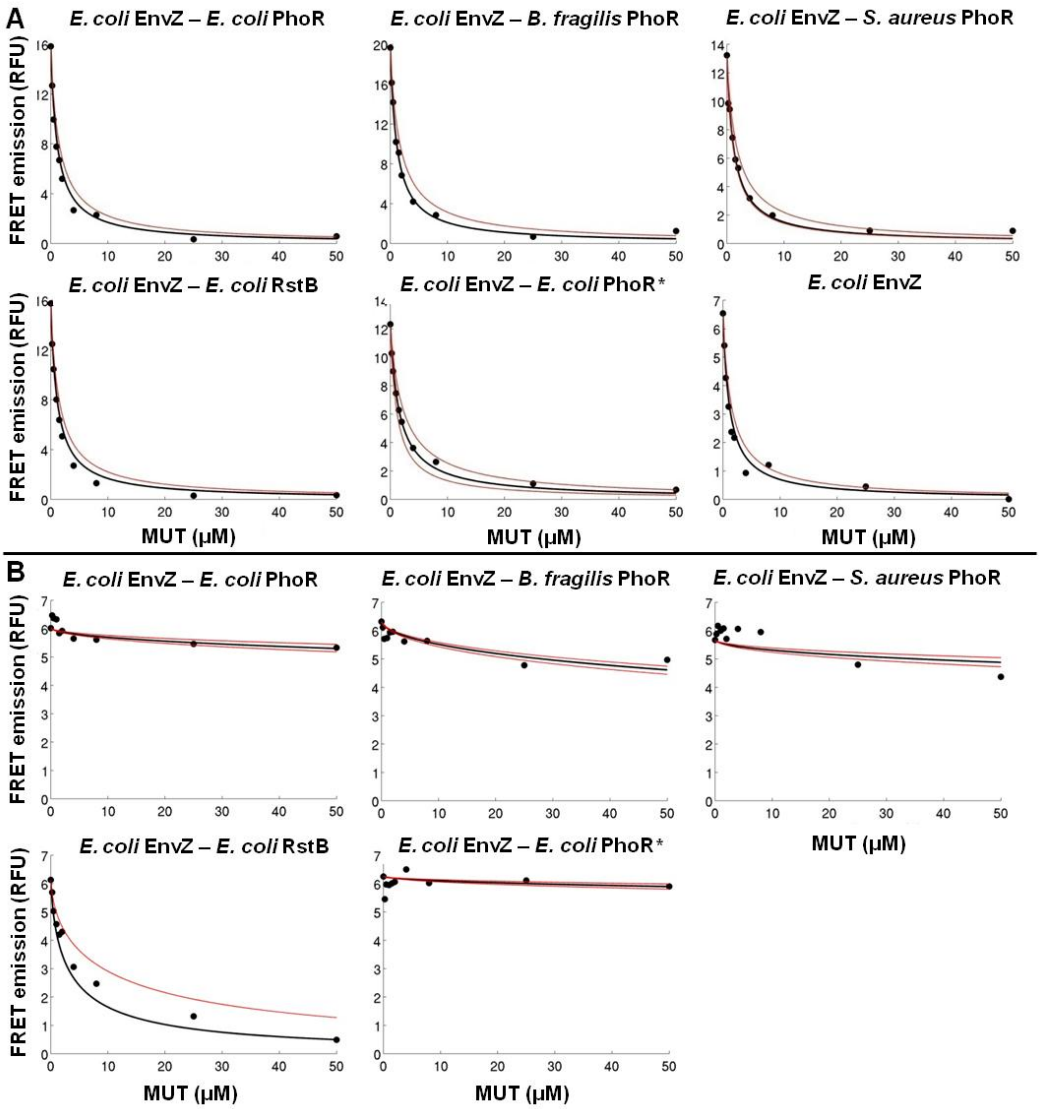


Figure 3.11. Fitting homodimer and heterodimer dissociation constants.
Equilibrium dissociation constants were fit for each kinase homodimer (A), or each kinase heterodimer with *E. coli* EnvZ (B) as described in the methods. FRET emission signal from a complex of CFP-kinase and YFP-kinase (black circles) was inhibited by interaction with unlabeled (MUT) mutant competitor kinase. The decrease in FRET emission signal with increasing mutant competitor was fit in MATLAB (black line) by simulating the equilibrium reactions between the three kinases. The lower and upper limits of K_d values that gave R^2 values at least 95% as good as that of the best fit are shown in red. (A) Kinase homodimers were measured by mixing a 0.5 μM equimolar mixture of CFP-kinase A and YFP-kinase A with unlabeled mutant kinase-A. (B) Kinase heterodimers were measured by mixing a 0.5 μM equimolar mixture of CFP-EnvZ and YFP-EnvZ with unlabeled mutant kinase-A.

Discussion

By investigating the autophosphorylation mechanisms of various histidine kinases and kinase chimeras, we found that the loop in the DHp domain is a key molecular determinant of autophosphorylation mechanism. Given this observation, it is striking that orthologous kinases with widely divergent loop sequences have conserved their autophosphorylation mechanism. Below we discuss both how the kinase loops might specify mechanism and possible functional pressures influencing *in cis* vs. *in trans* autophosporylation.

Kinase loops as determinants of autophosphorylation mechanism

Introducing loops from PhoR orthologs that autophosphorylate *in cis* into an EnvZ ortholog that autophosphorylates *in trans* caused a switch in mechanism. We favor a model in which the directionality of the loop determines the arrangement of helices in the DHp four-helix bundle, and is thus central to establishing autophosphorylation *in cis* vs. *in trans* (Fig. 3.1B). Based on this model, we expect the EnvZ orthologs, *E. coli* RstB, and the EnvZ-RstB chimera to have right-handed loops, and we expect the PhoR orthologs and the EnvZ-PhoR chimeras to have left-handed loops.

Studies of loops connecting α-helices demonstrate that geometric parameters of the flanking helices often correlate with different loop structure motifs (Engel and DeGrado, 2005; Lahr et al., 2005). Thus, in DHp domains, changes in loop handedness might be accompanied by small adjustments in the helix bundle geometry. A crystal structure of *T. maritima* HK853, a kinase that has a left-handed loop and autophosphorylates *in cis*, shows a helix bundle that is less parallel than the helix bundle in *E. coli* EnvZ (Marina et al., 2005; Tomomori et al., 1999). Interestingly,

the DHp domains of PhoR orthologs characterized here are more similar in sequence to HK853 (31-46% identity) than to *E. coli* EnvZ (23-37% identity), and may have more HK853-like helix bundles. In our studies of EnvZ-PhoR loop chimeras, the loop from one protein was accommodated in a different helix bundle, and presumably with a handedness different from the native loop. Whether small accompanying changes occurred in the helix bundle is not known, but could be resolved through additional structural studies. In addition, characterizing the autophosphorylation mechanism for kinases with existing structures, like KinB or ThkA, could further support the loop-handedness model and provide additional information about whether loop changes are accompanied by changes in the helix-bundle structure (Bick et al., 2009; Yamada et al., 2009).

Although we demonstrated a key role for the DHp loop in determining autophosphorylation mechanism, we cannot rule out that other kinase structural features may also contribute. In particular, the crystal structure of *B. subtilis* DesK has a left-handed loop, similar to that seen in HK853. Cross-linking studies suggested that DesK may autophosphorylate *in trans*, but did not conclusively rule out autophosphorylation occurring *in cis* (Trajtenberg et al., 2010). If DesK does autophosphorylate *in trans*, then regions other than the DHp domain loop must be contributing to autophosphorylation mechanism specificity, at least for DesK, which is a member of a relatively small subfamily of histidine kinases (Stock et al., 2000). One additional structural element that may contribute to autophosphorylation is the linker between the DHp and CA domains, which is 15 residues shorter in DesK relative to HK853(Casino et al., 2010). The length and structure of this linker may influence which DHp domain a given CA domain can autophosphorylate.

Autophosphorylation mechanism conservation

The EnvZ and PhoR orthologs we characterized come from bacterial species that, in some cases, diverged from one another hundreds of millions of years ago. However, we found that their autophosphorylation mechanisms are conserved. This conservation across long periods of time suggests that the mechanism of histidine kinase autophosphorylation may be under selection. Why different kinases would be under pressure to retain different mechanisms is unclear, because in either *in cis* or *in trans* autophosphorylation the same product is made, a phosphorylated histidine available for phosphotransfer to a cognate response regulator. A possible explanation is that other kinase functions linked to DHp loop handedness, separate from autophosphorylation mechanism, are under selection.

One possibility is that loop handedness may contribute to specificity in histidine-kinase interactions, thereby promoting the isolation of signaling pathways (Capra and Laub, 2012). For example, EnvZ and its paralog RstB were previously shown to specifically form homodimers; presumably, a heterodimer between these two kinases would lead to detrimental cross-talk between their respective pathways (Ashenberg et al., 2011). An *E. coli* EnvZ-RstB loop chimera, identical to the one in this study, maintained an interaction with EnvZ. This suggested that loop swaps do not affect kinase dimerization. However, loop changes in the three EnvZ-PhoR chimeras characterized here did lead to differences in dimerization. Interactions of EnvZ with the EnvZ-PhoR chimeras were destabilized relative to the interaction of EnvZ with the EnvZ-RstB chimera (Fig. 3.10A). The loop model supported by this work suggests that RstB has a right-handed loop, like EnvZ, whereas PhoR and the EnvZ-PhoR chimeras have left-handed loops. It appears that mixed dimers between kinases with different loop handedness are destabilized.

Possible reasons for this are illustrated in Fig. 3.10C, which shows that such kinases would be unable to form the stereotypical dimer. Either solvent-exposed residues would be buried in the helix-bundle core, or the α1 and α2 helices would not be correctly positioned relative to one another in the helix bundle. Avoiding these highly atypical arrangements could require that the loop adopt a higher energy conformation. Destabilization of mixed dimers via this mechanism could contribute to kinase pathway isolation. After gene duplication, a new histidine kinase could change its DHp loop handedness as one step towards evolving dimerization specificity against its new paralog. Conserving the changed loop handedness, and the resulting autophosphorylation mechanism, could be important for maintaining dimerization specificity.

Loop handedness may also play a role in phosphotransfer specificity. Loop changes in the chimeras were neutral with respect to phosphotransfer specificity, both here and in earlier studies (Fig. 3.10B) (Skerker et al., 2008). However, in earlier work that aimed to change the phosphotransfer specificity of kinases by making mutations at the base of their DHp helix bundles, it was observed that accompanying changes in the loops were sometimes required (Skerker et al., 2008). For example, rewiring EnvZ to phosphorylate PhoB, the cognate response regulator of PhoR, required introduction of the PhoR loop into EnvZ (Skerker et al., 2008). We suggest a model in which binding mode of the response regulator can be somewhat different for kinases with different loop handedness. Based on the structure of *cis* kinase HK853 in complex with its cognate response regulator, the EnvZ-PhoR chimeras likely contact the regulator using the α1 and α2 helices from the same chain in their DHp domain dimers (Fig. 3.10B) (Casino et al., 2009). Interestingly, a few positions in the DHp loop from HK853 (S279, E282, L283, and T287) also contact the regulator, adding to the kinase-regulator interface. In the EnvZ-PhoR

chimera, similar specific contacts made by the PhoR loop may have been required in switching EnvZ phosphotransfer specificity to PhoB. In contrast, a *trans* kinase like EnvZ likely interacts with its response regulator using the α1 helix of one chain and the α2 helix of the partner chain in its DHp domain dimer (Fig. 3.10B). In this arrangement, the DHp loop is not close to the interface with the regulator and not important for determining binding specificity. This model is supported by the observation that rewiring EnvZ to phosphorylate RstA, the cognate regulator of RstB, did not require changing the loop (Skerker et al., 2008). It is also consistent with our observation here, that EnvZ-PhoR chimeras could phosphorylate OmpR, because the interaction between *E. coli* EnvZ and OmpR is not expected to require a loop contact. These arguments suggest that loop handedness could help insulate pathways by destabilizing certain kinase-response regulator interactions that lack key loop interactions.

In closing, we emphasize that although our study supports conservation of mechanism, the autophosphorylation mechanisms of most kinases are not known. Including our work here, 20 histidine kinases now have a characterized autophosphorylation mechanism. But there are tens of thousands of kinases. A predictive rule for autophosphorylation mechanism, which could possibly be established by studying determinants of loop handedness, would help establish this more broadly across kinase families.

Methods

Cloning and protein purification of kinase orthologs and chimeras

PhoR orthologs (341 sequences) were previously identified using a reciprocal best BLAST hit procedure, where full-length sequences were used as the queries. The same approach was used to identify EnvZ orthologs (762 sequences) (Capra et al., 2012). Ortholog sequences were aligned using ClustalW (Larkin et al., 2007). When calculating sequence conservation and information content for each position in the DHp domain, a sequence alignment of all available orthologs was first filtered with a 90% sequence identity cutoff such that no pair of sequences was greater than 90% identical.

For each identified kinase ortholog, pENTR clones for use in the Gateway recombinational cloning system were generated as previously described (Ashenberg et al., 2011). The amino acid boundaries for each cloned ortholog are listed in Table 3.1, and orthologs were cloned from genomic DNA. Each clone contained the kinase DHp and CA domains. Some EnvZ or PhoR orthologs have a cytoplasmic HAMP or PAS domain, respectively, N-terminal to their DHp domain. The HAMP domain was included in the *S. meliloti* EnvZ construct and the PAS domain was included in the *E. coli*, *C. crescentus*, and *B. subtilis* PhoR constructs. To construct EnvZ-PhoR loop chimeras, the loop region of *E. coli* EnvZ was replaced by PhoR loop sequences using PCR-based site-directed mutagenesis as previously described (Ashenberg et al., 2011). Each ortholog or chimera was then cloned into His_6-expression vectors encoding either N-terminal ECFP (pRG31) or N-terminal monomeric YFP (pRG88) (Ashenberg et al., 2011; Gao et al., 2008). The pRG31 clones are referred to as wild-type (WT) constructs.

To disrupt ATP binding by the kinase, the third glycine in the G2 box (GxGxG) of the CA domain was mutated to alanine using QuikChange site-directed mutagenesis (Stratagene). The mutant construct (MUT), used in the autophosphorylation and FRET-based assays, was created by taking this resulting construct and adding a tandem FLAG tag (DYKDDDDKDYKDDDDKGSG) to the N-terminus using PCR. These pENTR clones were recombined into a His_6-expression vector, pHIS-DEST, using the Gateway LR clonase reaction (Skerker et al., 2005). Full-length response regulator *E. coli* OmpR was previously cloned into a His_6-expression vector, pTRX-HIS-DEST (Skerker et al., 2005). Protein expression and protein purification were performed as previously described (Ashenberg et al., 2011).

FRET equilibration assay

To determine appropriate equilibration times for the autophosphorylation assays, interaction between CFP-histidine kinase (WT) and $FLAG_2$-mutant histidine kinase (MUT) was monitored by measuring changes in FRET (fluorescence resonance energy transfer) signal over time. $FLAG_2$-mutant histidine kinase at 50 μM was added to an equimolar mixture of CFP-histidine kinase and YFP-histidine kinase at 5 μM. These mixtures were prepared in 96-well plates (Corning), covered with a foil seal, and incubated at 30 °C. Fluorescence in each well was measured over time using a Varioskan plate reader at 30 °C. To measure FRET signal, three channels were monitored: donor channel (excite 433 nm, emit 475 nm), acceptor channel (excite 488 nm, emit 527 nm), and FRET channel (excite 433 nm, emit 527 nm). In each well at the indicated time, 5 fluorescence measurements were made in each channel and then averaged. A corrected FRET emission signal was calculated as previously described (Ashenberg et al., 2011). The equilibration times chosen for the autophosphorylation assay are listed in Table 3.1 and

marked in Figure 3.3. Each FRET equilibration assay was performed at least twice, and results were similar across duplicates.

Determining autophosphorylation mechanism

To demonstrate a kinase ortholog autophosphorylated *in trans*, 5 μM wild-type CFP-kinase (WT) was mixed in the absence or presence of 50 μM $FLAG_2$-mutant kinase (MUT) in HEPES storage buffer (10 mM HEPES-KOH (pH 8.0), 50 mM KCl, 10% glycerol, 0.1 mM EDTA, 1 mM DTT) plus 5 mM $MgCl_2$. The mixtures were equilibrated at 30 °C according to times established by the FRET equilibration assay (Table 3.1). As a negative control, 50 μM mutant kinase was also equilibrated. After equilibration, the mixtures were autophosphorylated with 500 μM ATP and 5 μCi [γ^{32}P]ATP (Amersham Biosciences, 6000 Ci/mmol) for 30 min at 30 °C. Autophosphorylation was stopped by adding 4X sample buffer (500 mM Tris (pH 6.8), 8% SDS, 40% glycerol, 400 mM β-mercaptoethanol), and mixtures were placed on ice before being separated on a 10% Tris-HCl SDS-PAGE gel (Bio-Rad). The gel was exposed to a phosphor screen for 2 h at room temperature, the resulting screen was scanned using a Storm 860 imaging system (Amersham Biosciences) at 50 μm resolution, and the images were quantified with ImageQuant 5.2.

To demonstrate a kinase ortholog autophosphorylated in *cis*, the same procedure as in the previous paragraph was performed; the only difference was that the autophosphorylation reaction was performed as a time-course. At each time point, an aliquot of the autophosphorylation mixture was removed, and the reaction was stopped by adding sample buffer. The autophosphorylation time-courses of 5 μM wild-type kinase in the absence and presence of 50 μM mutant kinase were quantified and compared. Each autophosphorylation assay was

performed at least twice and results were similar across duplicates.

Measuring autophosphorylation without equilibration

For *E. coli* PhoR and *C. crescentus* PhoR, 5 µM wild-type CFP-kinase was mixed with increasing amounts of $FLAG_2$-mutant kinase (0, 20, 50, 100 µM) in HEPES storage buffer plus 5 mM $MgCl_2$. After equilibration according to times established by the FRET equilibration assay (Table 3.1), each mixture was autophosphorylated as in the mechanism determination assays, and autophosphorylation reactions were run for 3 min (*E. coli* PhoR) or 2 min (*C. crescentus* PhoR). Samples were separated on a 10% Tris-HCl SDS-PAGE gel and exposed to a phosphor screen as in the mechanism determination assays. In parallel, the same set of experiments was performed, but the equilibration step was omitted. Instead, $FLAG_2$-mutant kinase was added to the wild-type CFP-kinase, and autophosphorylation was then initiated within 20 s.

Phosphotransfer assay

Kinase chimeras were tested for their ability to phosphotransfer to *E. coli* response regulator OmpR as previously described (Skerker et al., 2005). CFP-histidine kinase and thioredoxin-tagged OmpR, both at 5 µM, were prepared in storage buffer plus 5 mM $MgCl_2$. Kinases were autophosphorylated as in the mechanism determination assays, and autophosphorylation reactions were incubated for 20 min at 30 °C. Phosphotransfer was initiated by adding response regulator to the phosphorylated kinase in a 2.5 µM: 2.5 µM ratio, and reactions were stopped after 10 s, 1 min, or 4 min with sample buffer. Samples were heated for 16 min at 65 °C, before being separated on a 10% Tris-HCl SDS-PAGE gel and exposed to a phosphor screen as in the mechanism determination assays.

FRET assay to measure equilibrium dissociation constants

A previously described FRET assay was used to measure equilibrium dissociation constants, K_d's, of chimera kinase homodimers, and of the heterodimers between chimera kinases and *E. coli* EnvZ (Ashenberg et al., 2011). Increasing amounts of $FLAG_2$-mutant histidine kinase were added to an equimolar mixture of CFP-histidine kinase and YFP-histidine kinase at 0.5 μM. Mixtures were equilibrated in sealed 96-well plates for 8 h at 30 °C. As previously described, fluorescence was then measured and the resulting corrected FRET emission signal was used to fit homodimer or heterodimer K_d values (Ashenberg et al., 2011). Each K_d value was measured in duplicate.

When measuring a homodimer K_d, the FRET assay was performed with CFP-histidine kinase, YFP-histidine kinase, and mutant histidine kinase all sharing the same kinase sequence. We assumed that the homodimers formed by these three kinases shared the same K_d value. When measuring a heterodimer K_d, CFP-EnvZ and YFP-EnvZ were mixed with a mutant histidine kinase that had a sequence different from EnvZ. Fitting of K_d values was performed in MATLAB by modeling a system of ordinary differential equations that described the possible equilibrium reactions between the CFP-histidine kinase, the YFP-histidine kinase, and the mutant histidine kinase (Ashenberg et al., 2011). Each K_d was fit so as to maximize the R^2 between the experimental data and the MATLAB-simulated data, and we also estimated the range of K_d values that could fit the data with similar accuracy to this best fit. The lower and upper limits of K_d values that gave R^2 values at least 95% as good as that of the best fit, $R^2_{\max}$, was calculated.

Acknowledgements

We thank T.A. Washington for providing *B. subtilis* A-168 genome DNA, and E.A. Lubin and E.J. Capra for providing purified response regulators. We thank members of the Keating laboratory (especially T. S. Chen, J. B. Kaplan, C. Negron, and A. W. Reinke) and the Laub laboratory (especially E. J. Capra, A. I. Podgornaia, and C. G. Tsokos) for insightful discussions. We used the Varioskan plate reader from the BioMicro Center of the Massachusetts Institute of Technology. This work was funded by National Institutes of Health award GM067681 to A.E. Keating, a National Science Foundation CAREER Grant to M.T. Laub, and the National Science Foundation GRFP fellowship to O. Ashenberg. M.T.L. is an Early Career Scientist of the Howard Hughes Medical Institute. We used computer resources provided by National Science Foundation award 0821391.

References

Ashenberg, O., Rozen-Gagnon, K., Laub, M. T., and Keating, A. E. (2011). Determinants of homodimerization specificity in histidine kinases. J Mol Biol **413,** 222-35.

Bick, M. J., Lamour, V., Rajashankar, K. R., Gordiyenko, Y., Robinson, C. V., and Darst, S. A. (2009). How to switch off a histidine kinase: crystal structure of *Geobacillus stearothermophilus* KinB with the inhibitor Sda. J Mol Biol **386,** 163-77.

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). J Mol Biol **329,** 495-503.

Capra, E. J., and Laub, M. T. (2012). Evolution of Two-Component Signal Transduction Systems. Annu Rev Microbiol.

Capra, E. J., Perchuk, B. S., Skerker, J. M., and Laub, M. T. (2012). Adaptive Mutations that Prevent Crosstalk Enable the Expansion of Paralogous Signaling Protein Families. Cell **150,** 222-32.

Casino, P., Rubio, V., and Marina, A. (2009). Structural insight into partner specificity and phosphoryl transfer in two-component signal transduction. Cell **139,** 325-36.

Casino, P., Rubio, V., and Marina, A. (2010). The mechanism of signal transduction by two-component systems. Curr Opin Struct Biol **20,** 763-71.

Engel, D. E., and DeGrado, W. F. (2005). Alpha-alpha linking motifs and interhelical orientations. Proteins **61,** 325-37.

Filippou, P. S., Kasemian, L. D., Panagiotidis, C. A., and Kyriakidis, D. A. (2008). Functional characterization of the histidine kinase of the *E. coli* two-component signal transduction system AtoS-AtoC. Biochim Biophys Acta **1780,** 1023-31.

Gao, R., Tao, Y., and Stock, A. M. (2008). System-level mapping of *Escherichia coli* response regulator dimerization with FRET hybrids. Mol Microbiol **69,** 1358-72.

George Cisar, E. A., Geisinger, E., Muir, T. W., and Novick, R. P. (2009). Symmetric signalling within asymmetric dimers of the *Staphylococcus aureus* receptor histidine kinase AgrC. Mol Microbiol **74,** 44-57.

Gooderham, W. J., and Hancock, R. E. (2009). Regulation of virulence and antibiotic resistance by two-component regulatory systems in *Pseudomonas aeruginosa*. FEMS Microbiol Rev **33,** 279-94.

Lahr, S. J., Engel, D. E., Stayrook, S. E., Maglio, O., North, B., Geremia, S., Lombardi, A., and DeGrado, W. F. (2005). Analysis and design of turns in alpha-helical hairpins. J Mol Biol **346,** 1441-54.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. Bioinformatics **23,** 2947-8.

Marina, A., Waldburger, C. D., and Hendrickson, W. A. (2005). Structure of the entire cytoplasmic portion of a sensor histidine-kinase protein. Embo J **24,** 4247-59.

Mascher, T., Helmann, J. D., and Unden, G. (2006). Stimulus perception in bacterial signal-transducing histidine kinases. Microbiol Mol Biol Rev **70,** 910-38.

Moglich, A., Ayers, R. A., and Moffat, K. (2009). Structure and signaling mechanism of Per-ARNT-Sim domains. Structure **17,** 1282-94.

Ninfa, E. G., Atkinson, M. R., Kamberov, E. S., and Ninfa, A. J. (1993). Mechanism of

autophosphorylation of *Escherichia coli* nitrogen regulator II (NRII or NtrB): trans-phosphorylation between subunits. J Bacteriol **175,** 7024-32.

Park, S. Y., Quezada, C. M., Bilwes, A. M., and Crane, B. R. (2004). Subunit exchange by CheA histidine kinases from the mesophile *Escherichia coli* and the thermophile *Thermotoga maritima*. Biochemistry **43,** 2228-40.

Pena-Sandoval, G. R., and Georgellis, D. (2010). The ArcB sensor kinase of *Escherichia coli* autophosphorylates by an intramolecular reaction. J Bacteriol **192,** 1735-9.

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,** 1043-54.

Skerker, J. M., Prasol, M. S., Perchuk, B. S., Biondi, E. G., and Laub, M. T. (2005). Two-component signal transduction pathways regulating growth and cell cycle progression in a bacterium: a system-level analysis. PLoS Biol **3,** e334.

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

Swanson, R. V., Bourret, R. B., and Simon, M. I. (1993). Intermolecular complementation of the kinase activity of CheA. Mol Microbiol **8,** 435-41.

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

Tomomori, C., Tanaka, T., Dutta, R., Park, H., Saha, S. K., Zhu, Y., Ishima, R., Liu, D., Tong, K. I., Kurokawa, H., Qian, H., Inouye, M., and Ikura, M. (1999). Solution structure of the homodimeric core domain of *Escherichia coli* histidine kinase EnvZ. Nat Struct Biol **6,** 729-34.

Trajtenberg, F., Grana, M., Ruetalo, N., Botti, H., and Buschiazzo, A. (2010). Structural and enzymatic insights into the ATP binding and autophosphorylation mechanism of a sensor histidine kinase. J Biol Chem **285,** 24892-903.

Yamada, S., Sugimoto, H., Kobayashi, M., Ohno, A., Nakamura, H., and Shiro, Y. (2009). Structure of PAS-linked histidine kinase and the response regulator complex. Structure **17,** 1333-44.

Yang, Y., and Inouye, M. (1991). Intermolecular complementation between two defective mutant signal-transducing receptors of *Escherichia coli*. Proc Natl Acad Sci U S A **88,** 11057-61.

Zhu, Y., and Inouye, M. (2002). The role of the G2 box, a conserved motif in the histidine kinase superfamily, in modulating the function of EnvZ. Mol Microbiol **45,** 653-63.

Chapter 4

Conclusions and Future Directions

In this thesis, I have taken biochemical and evolutionary approaches to dissect two molecular functions in histidine kinases: dimerization specificity and autophosphorylation. The identified functional determinants extend our understanding of histidine kinase signaling, and give us feedback as to how effective our approaches were.

Homodimerization specificity future directions

The work in chapter 2 established that histidine kinase dimerization specificity can be encoded in the sequence, in an analogous manner to kinase-regulator phosphotransfer and response regulator dimerization specificity. The transmitter domains of histidine kinase paralogs EnvZ and RstB specifically formed homodimers. Amino acid coevolution analysis of the kinases pointed to dimerization specificity determinants at the base of the four-helix bundle in the DHp domain, and these were verified by characterizing chimeras between EnvZ and RstB. Interestingly, the residues identified as important for dimerization specificity (252, 256, 266, and 270) and for phosphotransfer specificity are all closely located together at the base of the four-helix bundle. This region of the kinase governs two different modes of protein interaction specificity.

Although the amino acid coevolution analyses successfully identified specificity determinants, the sequence context in which the mutations were made was important. In particular, replacing the loop from EnvZ with the loop from RstB (chimera 1) did not change dimerization specificity. However introducing two mutations (I252L/I270L) into that context resulted in a kinase that homodimerized, but no longer interacted with EnvZ. These two mutations when introduced into EnvZ destabilized both the homodimer, and the heterodimer with EnvZ.

Ideally, we would like a method that can select an appropriate sequence context for introducing

mutations. Given a set of EnvZ orthologs, some similar to *E. coli* EnvZ and some similar to chimera 1, we want this method to choose the latter group of orthologs as the best sequence context for introducing I252L/I270L. The best sequence context would be one in which the mutations do not produce a protein with properties very different from the extant sequences. In dimerization specificity, an appropriate context would be a sequence in which the mutations did not destabilize the homodimer.

One intuitive approach for choosing an appropriate sequence context is that mutations will more likely be tolerated if the resulting mutant protein is similar to an extant protein. This idea can be extended to covariation. The set of highly covarying pairs in a protein form a network of coupled residues. Whether mutations at covarying sites can be accommodated depends on how well they fit into the context of this network. After mutations are made, how similar is the resulting network of covarying pairs compared to the covarying networks from existing proteins? In essence, we would like to score the likelihood of a network of covarying pairs given information from that protein family's sequence alignment. Several new covariation approaches address this problem by treating covarying pairs as edges linked together in a probabilistic graph model (Thomas et al., 2009; Weigt et al., 2009). Given a sequence alignment, these methods can choose an informative set of covarying pairs, and assign a likelihood for observing an entire set of residues covarying in a network. To choose an appropriate sequence context for a set of mutations like I252L/I270L, we can introduce the mutations into each extant EnvZ ortholog and score that resulting sequence's likelihood. We would introduce mutations into the EnvZ ortholog with the maximum likelihood. Although I propose this approach as a way to choose the most acceptable sequence context for a given set of mutations, it could also be developed in the

opposite direction, where the most acceptable set of mutations at some sites were chosen given a sequence context.

Autophosphorylation mechanism future directions

In chapter 3, I identified the loop in the DHp domain as a determinant of histidine kinase autophosphorylation mechanism. Replacing the loop from *trans* kinase *E. coli* EnvZ with the loop from *cis* PhoR orthologs resulted in chimeric kinases that autophosphorylated in *cis*. In addition, the observed conservation of autophosphorylation mechanism within orthologs suggested DHp loop handedness may be under purifying selection. Whether the mechanism occurs in *cis* or in *trans* may be important to signaling fitness; alternatively, loop handedness may be conserved because it links to a different kinase function, for example dimerization specificity. This latter view, that loop handedness may be a determinant of dimerization specificity, is supported by measurements showing a weakened interaction between EnvZ and the EnvZ-PhoR loop chimeras. To further study the conservation of autophosphorylation mechanism, below I discuss methods to predict mechanism for a given kinase.

Predicting autophosphorylation mechanism

A predictive rule for which autophosphorylation mechanism a kinase adopts would be a useful tool. By predicting autophosphorylation mechanism for histidine kinases across all species, we could learn if there was any preference for *cis* or *trans* mechanisms. Currently, only ~20 histidine kinases out of more than 50,000 have characterized mechanisms. We could also predict on a large scale whether mechanism is conserved across orthologs, or whether mechanism changes after gene duplication. Prediction accuracy could be measured by characterizing a test set of

kinases using the experiments developed in chapter 3.

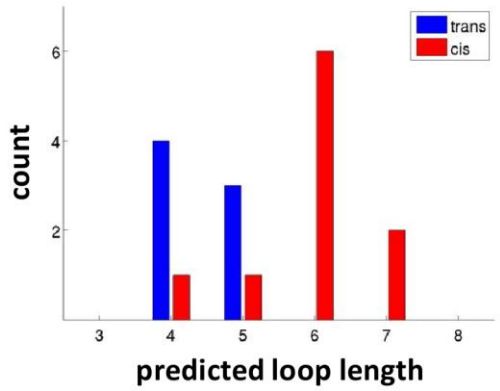


Figure 4.1 Predicting loop handedness using loop length.
For each histidine kinase with a characterized autophosphorylation mechanism (*trans* or *cis*), the length of the loop in the DHp domain was predicted using PSIPRED. Kinases with the same autophosphorylation mechanism were binned by predicted loop length, and the number of kinases with each loop length was shown in a histogram.

Given the likely link between autophosphorylation mechanism and DHp loop handedness, predicting loop handedness may be sufficient to predict autophosphorylation mechanism. Both sequence-based and structure-based predictions of loop handedness should be tried. The ideal rule would be sequence-based, as this would likely be simpler and faster to compute. To collect a set of loop sequences, loop boundaries must be reliably identified, and this may be accomplished with PSIPRED, the most accurate secondary structure prediction program (Bryson et al., 2005). The significant divergence within loop sequences lowers the prospects for classifying them using sequence conservation. Still, there may be some structural features predictable from sequence. Different kinases may have different loop lengths, as seen in comparing the structures of EnvZ and HK853 (Marina et al., 2005; Tomomori et al., 1999). In particular, I have predicted loop lengths for the previously characterized kinases: seven kinases are *in trans* (four EnvZ orthologs, *E. coli* AtoS, *E. coli* NtrB, *E. coli* RstB), and ten kinases are *in cis* (nine PhoR orthologs, *T.*

maritima HK853) (Fig 4.1). In this small set of kinases, loops from *in trans* kinases tend to be predicted as four to five residues long, and loops from *in cis* kinases tend to be predicted as six to seven residues long. This trend can be experimentally tested by characterizing kinases with long and short predicted loop lengths. Besides loop length, the presence of predicted helix-capping motifs may also distinguish loops of different handedness (Aurora and Rose, 1998).

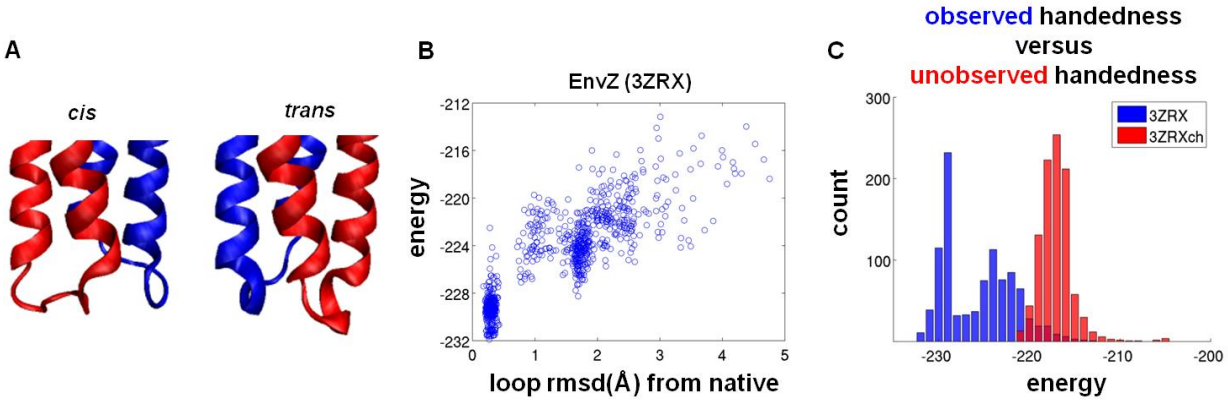


Figure 4.2 Predicting loop handedness using Rosetta.
(A) The loop in the DHp domain of EnvZ (PDB ID 3ZRX) was modeled with either left (*cis*) or right (*trans*) handedness using the macromolecular modeling program Rosetta. (B) 1000 models of the EnvZ loop in its observed right-handedness were generated using the kinematic closure algorithm, and the energy and rmsd of these modeled loops were compared to the loop in the crystal structure. Each blue circle corresponds to one model. The most stable loop models (most negative in energy) were also those most similar in structure to the loop in the crystal structure. (C) The models of the EnvZ loop in its observed handedness (3ZRX) were compared to models of the EnvZ loop generated with the unobserved handedness (3ZRXch). Loop models generated with the observed handedness were on average more stable than those generated with the unobserved handedness.

Structure-based predictions of loop handedness would require building homology models for DHp domains, based off existing DHp domain structures. This could be done in the macromolecular modeling program Rosetta, which contains one of the most accurate loop reconstruction algorithms, kinematic closure (KIC) (Mandell et al., 2009). Loop handedness can be predicted by modeling the kinase loops in either handedness, using KIC, and then choosing

the more energetically favorable structure (Fig 4.2A).

As a first step, we have tested whether Rosetta can predict that *E. coli* EnvZ has a right-handed loop. Using the crystal structure of EnvZ (PDB ID 3ZRX) as a template, we modeled EnvZ's loop with its observed handedness. In this test, no structural information from EnvZ's native loop was used as the KIC algorithm removed native loop coordinates in the first step of modeling. Rosetta generated 1000 possible loop models and we compared the models to the loop observed in the crystal structure. The more similar in structure to the native loop (lower root mean square deviation), the more favorably Rosetta scored the loop model (Fig 4.2B), and this demonstrated accurate structure prediction. The lowest energy loop model was only 0.3 Å rmsd away from the native loop.

We next modeled EnvZ's loop in the unobserved left-handed conformation, and compared the energies between left-handed and right-handed models (Fig 4.2C). Structures with the native right-handedness were energetically favored, and this difference stemmed from poor packing between the loops with the unobserved handedness. Slight movements in the helix backbones may better accommodate the loop in its unobserved handedness so incorporating backbone flexibility into the template would be an important follow-up step. Still, the energetic difference between EnvZ loop models indicates predicting autophosphorylation mechanism using structure prediction may succeed.

The next step is to predict loop handedness for the EnvZ orthologs and the EnvZ-PhoR loop chimeras, which presumably have helix backbones close to the native *E. coli* EnvZ template. Because these chimeras autophosphorylate *in cis*, they should also be modeled on the *T.*

maritima HK853 template. It would be especially promising if the EnvZ-PhoR loop chimeras were predicted to have left-handed loops, compared to the right-handed loop in *E. coli* EnvZ. To further test ability to predict a *cis* autophosphorylation mechanism, we would also predict loop handedness in PhoR orthologs. Because no PhoR structure exists, extensive homology modeling would be required, making this the most challenging type of prediction. PhoR orthologs would need to be modeled on several templates and backbone flexibility would need to be considered.

Linking new sensory domains to histidine kinases

The mechanisms by which a recently-duplicated two-component signaling pathway can acquire a new function are becoming increasingly understood. Significant effort has been spent on learning how a new pathway becomes insulated from existing pathways, at the level of phosphotransfer specificity and kinase and regulator dimerization specificity (Ashenberg et al., 2011; Capra et al., 2012; Gao et al., 2008). But how histidine kinases have evolved to sense new inputs remains an open question. The conserved transmitter domains of kinases are linked to an extraordinary number of diverse sensory domains, sensing hundreds of different signals. Large amounts of domain shuffling, and many adaptive mutations, have occurred across the sensory domains (Alm et al., 2006). Insight into how such diversity evolved could be gained by determining how sensory domains and transmitter domains are linked together in existing kinases. This understanding could also be used to create kinases with new signal sensing capabilities, allowing for different ways to modulate existing or engineered cell signaling networks.

Many different sensory domains are found covalently linked to the kinase transmitter domain, and conversely, these same sensory domains are also found linked to transmitter domains from signaling proteins other than histidine kinases (Moglich et al., 2009b). The

interchangeability between these diverse domains suggests there are common structural mechanisms for signaling between the sensory and transmitter domains. The two most common sensory domains, PAS and GAF, are found N-terminal to the transmitter domain in thousands of kinases. Although both domains have very low sequence homology, they have roughly similar folds and both their C-termini consist of α helices (Moglich et al., 2009b).

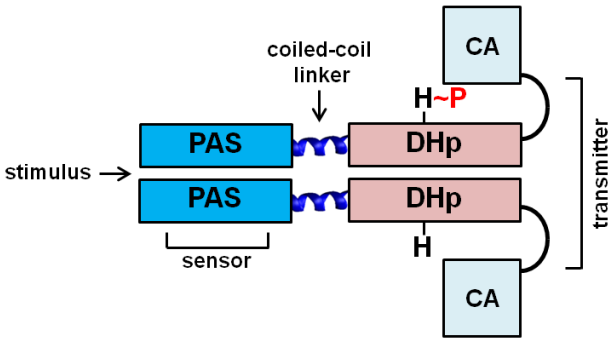


Figure 4.3 Histidine kinase domain organization
A sensor domain in the histidine kinase receives a stimulus and transduces that signal through a coiled-coil linker to the kinase transmitter region, leading to changes in kinase activity. Several sensor domains are used by histidine kinases, including PAS and GAF domains.

The presence of the C-terminal α helix is important as the linker connecting the sensory domain to the DHp domain in the transmitter region likely forms a helix, in particular a coiled coil (Anantharaman et al., 2006; Moglich et al., 2009a; Stewart and Chen, 2010) (Fig 4.3). Because α-helices behave as rigid rods below their persistence lengths, a contiguous coiled coil linking the sensory and transmitter domains could transmit a signal a significant distance. Work on the light-regulated FixL kinase, introduced in chapter 1, further suggests that a coiled-coil linker transduces signal through a rotation mechanism (Moglich et al., 2009a). The linker connects the light-sensing PAS domain to the FixL transmitter region. A seven-residue (heptad) insertion in the linker, which in a coiled coil corresponds to two turns preserving the coiled-coil register, did

not change the light-regulated activity. In contrast, insertions of two, three, five, or six residues in the middle of the coiled coil, which would produce incomplete turns and disrupt the coiled-coil core, resulted in complete inactivity. Given a rotation mechanism, the coiled-coil linker likely allows at least two different conformational states, possibly analogous to the strong and weak helix packing states seen in HAMP domains (Airola et al., 2010). Indeed, light-regulated FixL loses ~5% of helical structure during light absorption, which may correspond to partial unfolding of the coiled coil linker (Moglich et al., 2009a).

Experiments on the signaling helix (S-helix), a five-heptad coiled-coil linker found in 10% of histidine kinases, further suggest that the linker may have multiple states. The S-helix commonly occurs between the DHp domain and a HAMP, PAS, or GAF domain, and structural models suggested it forms a parallel dimer (Anantharaman et al., 2006). Deletion of the S-helix in osmosensing yeast histidine kinase Sln1p removed kinase activity *in vitro* (Tao et al., 2002). However, introducing the coiled coil from the bZIP C/EBP in a manner that preserved coiled-coil register restored kinase activity. *In vivo* though, kinase activity deviated from wild type and showed hyperactive osmosensitivity. This indicated that the S-helix coiled coil had specific physical properties that differentiated it from the C/EBP coiled coil in signaling, which is expected to form a static dimer. Below I discuss bioinformatic and experimental approaches to study the nature of this coiled-coil linker.

Bioinformatic studies of coiled-coil linkers

What are the sequence features that define the coiled coils linking the sensory and transmitter domains? First, linkers from PAS-containing and GAF-containing kinases can be collected. The boundaries of these linkers can be identified using Hidden Markov Models from the Pfam

database for PAS, GAF, and DHp domains (Punta et al., 2012). Whether these linkers are helical and form coiled coils can be predicted using PSIPRED and LEARNCOIL (Bryson et al., 2005; Singh et al., 1998). Possible transmembrane helices could be filtered out using TMHMM (Krogh et al., 2001). Sequence conservation and the presence of sequence motifs could then be assessed.

In addition, the linkers can be grouped by length. This has already been done for PAS domains upstream of DHp domains, and the linkers were either 7n or 7n+2 residues long, where n=3, 4, 5, or 6 (Moglich et al., 2009a). The significant linker length variability may be functionally relevant. A heptad deletion scan in the signaling helix of kinase NarX showed impaired signaling in some deletions and wild-type signaling in other deletions (Stewart and Chen, 2010). Looking at linker length variation across kinase orthologs and paralogs could be particularly informative. Only ~30% of recent kinase paralogs have the same input domain architecture as their closest paralog (Alm et al., 2006). This percentage is likely higher for kinase orthologs. Do orthologs with the same sensory domain conserve their linker length, or are there heptad insertion/deletion events? If length conservation is relaxed, this would suggest the signaling output of a given sensory domain is not specific to linker length. Likewise, do paralogs with different sensory domains share the same linker or does the linker get shuffled together with the sensory domain? If very similar linkers are found in the context of different sensory domains, this would suggest interchangeability between linkers.

The PAS linkers and GAF linkers may have specifically coevolved with their respective sensory domains. Alternatively, these linkers may be interchangeable across sensory domains. One test for this is to bin all linker sequences by their lengths and then run all-against-all BLAST searches within these bins (Altschul et al., 1990). Sequences can then be clustered by E-value, and if the

PAS linkers and GAF linkers fall into two separate clusters, the choice of coiled-coil linker sequence may be optimized and specific for each sensory domain. A different test is to measure amino acid coevolution between the C-termini of the PAS or GAF domains and their respective linkers. If different sites covary in these two types of linkers, exchanging linkers between PAS and GAF domains would likely disrupt signaling. Similarly, coevolution within the linker, and coevolution between the linker and the N-termini of the DHp domain, should be measured. These coevolution analyses, combined with identified linker sequence motifs, may identify important sequence features in the junctions between the linker and its abutting domains.

Experimental studies of coiled-coil linkers

The sequence-based studies above can guide the rational design of new signal-sensing kinases, where sensory domains and their linkers are fused to non-cognate transmitter domains. These designs would also test to what extent a sensory domain and its linker can function in the context of a different transmitter domain. To begin, a set of PAS, GAF, and kinase transmitter domains can be chosen to be linked together. The PAS and GAF domains should be previously characterized and have a known input to which they respond. For transmitter domains, EnvZ or FixL are appropriate choices as they have been used in past sensory domain chimeras (Moglich et al., 2009a; Utsumi et al., 1989). In creating light-regulated FixL, four different fusion points between the sensory domain and FixL transmitter domain were tried (Moglich et al., 2009a). These points were chosen as the linker positions with matching residues in an alignment of the sensory domain and its linker region to FixL. Two of the fusion points resulted in light-regulation, but why those two points worked and the other two did not is unknown. The above considerations of sequence motifs and coevolution in linkers of existing kinases would

potentially be a systematic and accurate way to choose a fusion point. A fusion point should be chosen so that the new kinase has sequence motifs and networks of covarying residues similar to existing kinases. The likelihood that residues from the linker and transmitter domain are compatible in a covarying network could be measured with the tools discussed above.

To characterize signal-responsiveness in an engineered kinase, a transcriptional reporter assay may be used with *LacZ* placed under control of the promoter of the appropriate response regulator (OmpR or FixJ). Changes in β-galactosidase activity can then be measured in the presence or absence of input stimulus. Because FixL is a cytoplasmic kinase, its full-length version may also be characterized *in vitro*. If several kinases were successfully made signal-responsive, this would suggest a good understanding of how to bring different signaling modules together, whereas if they were not responsive, this would suggest an incorrect choice of fusion points. If the appropriate fusion point had been missed, it could be identified by characterizing a series of kinases with all possible fusion points within the linker. If none of these kinases were signal-responsive, this would suggest other sequence changes are needed after fusion to make a functional signaling unit. The necessary changes would likely be difficult to predict, and this would also indicate that the sensory and transmitter domains are not completely swappable modules.

New kinases could also be engineered by swapping linkers between existing kinases. To what extent can a coiled-coil linker maintain function in the context of a different sensory domain? This is a refined version of the C/EBP experiment above, where the coiled-coil linker that is being introduced is now presumably capable of multiple conformational states. Experiments may be performed using the newly engineered kinases above, or if those did not function, extant

histidine kinases with characterized signaling. The linker for a given PAS domain could be swapped with the linker from another PAS domain. Also if the bioinformatic analysis above showed linkers did not fall into separate sequence clusters depending on sensory domain, linker swaps could be made between PAS and GAF domains. Whether there is a directionality to signaling encoded within the linker could also be tested by swapping linkers between a kinase whose activity is turned on in the presence of signal (kinase activity) and a kinase whose activity is turned off in the presence of signal (phosphatase activity). Switches in kinase and phosphatase activity may suggest there are two classes of linker, with each class sufficient to determine the type of signaling activity. No activity switch would indicate the type of signaling activity is determined by the sensory and transmitter domains.

Finally, the coiled-coil linkers may be better understood through their biophysical characterization. As discussed above, the coiled coil from bZIP C/EBP cannot fully complement the deletion of the coiled coil in Sln1p (Tao et al., 2002). This may be because the coiled-coil linkers have other stable conformational states not encoded in C/EBP, or because the coiled-coil linkers partially unfold during signaling whereas C/EBP is too stable to do so. I am interested in the latter hypothesis: is C/EBP significantly more stable than kinase coiled-coil linkers? The equilibrium dissociation constant for C/EBP homodimer is ~10 nM (experiment performed by Aaron Reinke). Coiled-coil linkers from Sln1p and the kinases characterized above can be purified and tested for dimerization using size exclusion chromatography. Either thermal melts monitored by circular dichroism, or a FRET-based assay, can report on dimerization stability. If there were significant stability differences between the coiled-coil linkers and C/EBP, mutations that tune affinity may produce interesting phenotypes. Destabilizing mutations could be

introduced into C/EBP, and stabilizing mutations could be introduced into the coiled-coil linkers, based on qualitative rules for coiled-coil stability (Vinson et al., 2006). Correlation in these variants between coiled-coil dimerization affinity and signal responsiveness would implicate dimer stability as an important feature of kinase signaling.

The goal of the above project is to understand how sensory and transmitter domains are linked in existing kinases, and then to test that understanding through kinase engineering. In a similar approach to the kinase dimerization studies, this work would take advantage of the huge number of sequences available for existing sensory domains and kinases. Careful sequence analyses would ideally prioritize a small set of engineering experiments most likely to succeed, and those experiments would give feedback as to how effective the sequence studies were. Establishing the important sequence properties of coiled-coil linkers could also allow us to engineer new sensory kinases with similar confidence to our ability to rewire kinase-regulator phosphotransfer specificity (Skerker et al., 2008).

References

Airola, M. V., Watts, K. J., Bilwes, A. M., and Crane, B. R. (2010). Structure of concatenated HAMP domains provides a mechanism for signal transduction. Structure **18,** 436-48.

Alm, E., Huang, K., and Arkin, A. (2006). The evolution of two-component systems in bacteria reveals different strategies for niche adaptation. PLoS Comput Biol **2,** e143.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. J Mol Biol **215,** 403-10.

Anantharaman, V., Balaji, S., and Aravind, L. (2006). The signaling helix: a common functional theme in diverse signaling proteins. Biol Direct **1,** 25.

Ashenberg, O., Rozen-Gagnon, K., Laub, M. T., and Keating, A. E. (2011). Determinants of homodimerization specificity in histidine kinases. J Mol Biol **413,** 222-35.

Aurora, R., and Rose, G. D. (1998). Helix capping. Protein Sci **7,** 21-38.

Bryson, K., McGuffin, L. J., Marsden, R. L., Ward, J. J., Sodhi, J. S., and Jones, D. T. (2005). Protein structure prediction servers at University College London. Nucleic Acids Res **33,** W36-8.

Capra, E. J., Perchuk, B. S., Skerker, J. M., and Laub, M. T. (2012). Adaptive Mutations that Prevent Crosstalk Enable the Expansion of Paralogous Signaling Protein Families. Cell **150,** 222-32.

Gao, R., Tao, Y., and Stock, A. M. (2008). System-level mapping of *Escherichia coli* response regulator dimerization with FRET hybrids. Mol Microbiol **69,** 1358-72.

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol **305,** 567-80.

Mandell, D. J., Coutsias, E. A., and Kortemme, T. (2009). Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. Nat Methods **6,** 551-2.

Marina, A., Waldburger, C. D., and Hendrickson, W. A. (2005). Structure of the entire cytoplasmic portion of a sensor histidine-kinase protein. Embo J **24,** 4247-59.

Moglich, A., Ayers, R. A., and Moffat, K. (2009a). Design and signaling mechanism of light-regulated histidine kinases. J Mol Biol **385,** 1433-44.

Moglich, A., Ayers, R. A., and Moffat, K. (2009b). Structure and signaling mechanism of Per-ARNT-Sim domains. Structure **17,** 1282-94.

Punta, M., Coggill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., Heger, A., Holm, L., Sonnhammer, E. L., Eddy, S. R., Bateman, A., and Finn, R. D. (2012). The Pfam protein families database. Nucleic Acids Res **40,** D290-301.

Singh, M., Berger, B., Kim, P. S., Berger, J. M., and Cochran, A. G. (1998). Computational learning reveals coiled coil-like motifs in histidine kinase linker domains. Proc Natl Acad Sci U S A **95,** 2738-43.

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,** 1043-54.

Stewart, V., and Chen, L. L. (2010). The S helix mediates signal transmission as a HAMP domain coiled-coil extension in the NarX nitrate sensor from *Escherichia coli K-12*. J Bacteriol **192,** 734-45.

Tao, W., Malone, C. L., Ault, A. D., Deschenes, R. J., and Fassler, J. S. (2002). A cytoplasmic coiled-coil domain is required for histidine kinase activity of the yeast osmosensor, SLN1. Mol Microbiol **43,** 459-73.

Thomas, J., Ramakrishnan, N., and Bailey-Kellogg, C. (2009). Graphical models of protein-protein interaction specificity from correlated mutations and interaction data. Proteins **76,** 911-29.

Tomomori, C., Tanaka, T., Dutta, R., Park, H., Saha, S. K., Zhu, Y., Ishima, R., Liu, D., Tong, K. I., Kurokawa, H., Qian, H., Inouye, M., and Ikura, M. (1999). Solution structure of the homodimeric core domain of *Escherichia coli* histidine kinase EnvZ. Nat Struct Biol **6,** 729-34.

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 **245,** 1246-9.

Vinson, C., Acharya, A., and Taparowsky, E. J. (2006). Deciphering B-ZIP transcription factor interactions in vitro and in vivo. Biochim Biophys Acta **1759,** 4-12.

Weigt, M., White, R. A., Szurmant, H., Hoch, J. A., and Hwa, T. (2009). Identification of direct residue contacts in protein-protein interaction by message passing. Proc Natl Acad Sci U S A **106,** 67-72.

Appendix A

Assays to Measure Histidine Kinase Dimerization

This appendix is an overview of four experimental techniques for measuring histidine kinase dimerization interactions. These techniques were explored before my development of the competition FRET assay discussed in Chapter 2. The ability to make quantitative measurements of binding at equilibrium is the main advantage of the FRET assay over these other assays. Two of the assays I discuss, pull-downs and yeast-two-hybrid, are common tools for measuring protein-protein interactions, whereas the other two assays I discuss, separation of complexes via anion exchange and competitive inhibition of autophosphorylation, are based around specific properties of histidine kinases.

Pull-down of histidine-kinase dimers

In this assay, a histidine kinase (HK) is prepared with an N-terminal $FLAG_2$-tag and mixed with potential kinase interaction partners. Kinases interacting with $FLAG_2$-labeled kinase are pulled-down using anti-FLAG beads. I tested the assay with both HDC (HAMP+DHp+CA) and DC (DHp+CA) constructs of EnvZ and RstB. Each assayed interaction consisted of a mix of a $FLAG_2$-tagged kinase and a MBP- or Trx-labeled kinase. Pull-downs were done in HEPES buffer (10 mM HEPES-KOH, 50 mM KCl, 0.1 mM EDTA, pH 8.0). The concentration of $FLAG_2$-HK versus MBP-HK (or Trx-HK) was typically 2.5 μM:12.5 μM. The mixture of $FLAG_2$-HK and MBP-HK was incubated for 2 hours at room temperature on a rotisserie to allow sufficient time for subunit exchange. Anti-FLAG beads were equilibrated in HEPES buffer and added to the reaction mixture. The protein mixture and beads were incubated 30 minutes at 4 °C to allow for complexes to bind the beads. After letting protein bind the beads, beads were washed 3x with HEPES buffer, and protein complexes were eluted with 5 column volumes of 3X FLAG peptide at 100 μg/mL. Elution fractions were analyzed with SDS-PAGE and Coommassie

staining.

The first set of experiments examined interactions between HDC constructs for EnvZ, RstB, and CpxA. Each kinase was tagged with either His_6-$FLAG_2$ or His_6-MBP. The three homodimers were preferentially pulled-down over the six possible heterodimers, and the amount of non-specific bead binding was minimal (Fig. A.1A). Approximately 5-fold less heterodimer was pulled down than homodimer. However this did not necessarily indicate that the homodimers were more stable than the heterodimers, as the pull-down assay was not an equilibrium measurement. During the washing steps, the pulled-down complexes could dissociate, and off-rates across different homodimers and heterodimers may not be equal. If a heterodimer had both a high off-rate and a high on-rate, it may have a similar K_d to a homodimer, but never be pulled down due to the complex quickly dissociating during washes.

A pull-down similar to the one above was performed, but the His_6-MBP construct was replaced with a Trx-His_6 construct. The purpose of this experiment was to evaluate whether Trx-tagged kinases could be used along with MBP tagged kinases in a competition experiment to bind $FLAG_2$-kinase. $EnvZ_{HDC}$ homodimer was significantly pulled-down, whereas $RstB_{HDC}$ homodimer was not pulled-down significantly compared to the heterodimer mixes (Fig. A.1B). This result was reproducible, suggesting that the Trx tag behaves differently from the MBP tag, and that the Trx tag would not be suitable for a competitive pull-down with these kinases.

To evaluate the importance of the HAMP domain in maintaining interaction specificity, I also tried pulling down DHp+CA (DC) constructs. $FLAG_2$-$EnvZ_{HDC}$ or $FLAG_2$-$RstB_{HDC}$ was used to pull-down either MBP-$EnvZ_{DC}$ or MBP-$RstB_{DC}$ (Fig. A.1C). The EnvZ homodimer was significantly pulled-down, but the RstB homodimer control was not pulled-down.

Therefore, the heterodimer mixes could not be interpreted because MBP-$RstB_{DC}$ did not interact as expected. $RstB_{DC}$ could have a weaker affinity and higher off-rate due to the loss of interactions made by the HAMP. This interpretation is consistent with the destabilized dissociation constant for $RstB_{DC}$ measured using the FRET competition assay.

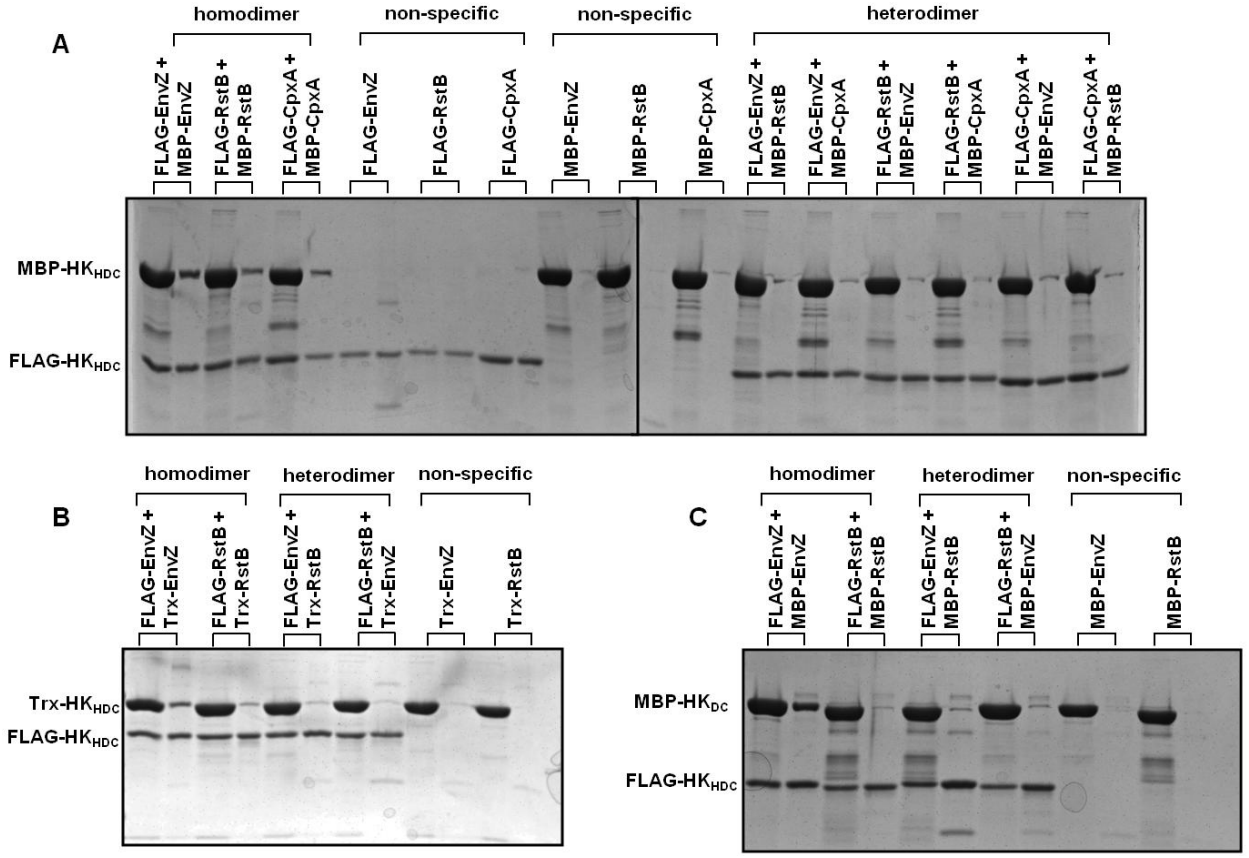


Figure A.1. Pull-down of histidine kinase dimers.
Pull-down assay with purified histidine kinases (HK) EnvZ, RstB, and CpxA. FLAG-labeled kinase was mixed with MBP-labeled kinase or Trx-labeled kinase, and then incubated 2 hrs at room temperature. Complexes were then isolated using anti-FLAG beads, and washed 3 times. Odd lanes show sample before being applied to anti-FLAG beads, and even lanes show elutions. (A) Pull-downs of EnvZ, RstB, and CpxA homodimer and heterodimer interactions. $FLAG_2$-HK_{HDC} and MBP-HK_{HDC}. (B) Pull-downs of EnvZ and RstB homodimer and heterodimer interactions. $FLAG_2$-HK_{HDC} and Trx- HK_{HDC}. (C) Pull-downs of EnvZ and RstB homodimer and heterodimer interactions. $FLAG_2$-HK_{HDC} and MBP-HK_{DC}.

Separation of histidine-kinase dimers by anion exchange

Another way to detect an interaction between $FLAG_2$-HK A and HK B is through separation on an anion exchange column. An equilibrated mix of HK A and HK B can form up to 3

dimer complexes, separable by the different number of $FLAG_2$ tags in A:A (2 $FLAG_2$), A:B (1 $FLAG_2$), and B:B (0 $FLAG_2$). This assay was previously used to study the subunit exchange of transthyretin tetramers (Schneider et al., 2001). Importantly transthyretin exchanged subunits slowly relative to the timescale of separation on the column, and there was significant separation in elution positions of the non-$FLAG_2$ tetramer and the completely $FLAG_2$-tagged tetramer. Satisfying these two conditions is necessary for distinguishing A:B from A:A and B:B.

As with the pull-down assay, I first focused on identifying homodimer interactions. Reaction mixtures of MBP-$FLAG_2$-EnvZ and Trx-EnvZ at different concentrations were incubated at room temperature for at least 2 hours. Mixtures were then loaded with a 500 μL injection loop onto a Mono Q HR 5/5 column (1 mL column volume, Baker Lab) in an FPLC system. After loading the sample, the column was washed with at least 5 column volumes Buffer A (25 mM Tris-HCl, pH 8.0). The mixture was then eluted with a linear gradient going from 0% to 100% Buffer B (25 mM Tris-HCl, 1 M NaCl, pH 8.0) over 40 minutes. Protein peaks were monitored by measuring A_{280}, and aliquots from eluted peaks were run on SDS-PAGE to verify protein identity.

Mixing MBP-$FLAG_2$-EnvZ at 3.4 μM and Trx-EnvZ at 5.5 μM did not produce a third peak, possibly corresponding to a dimer between the two species (Fig. A.2A). MBP-$FLAG_2$-EnvZ eluted at 34% B and Trx-EnvZ eluted at 42% B. In the mixture, two overlapping peaks eluted with maxima roughly corresponding to the elution positions of the proteins alone.

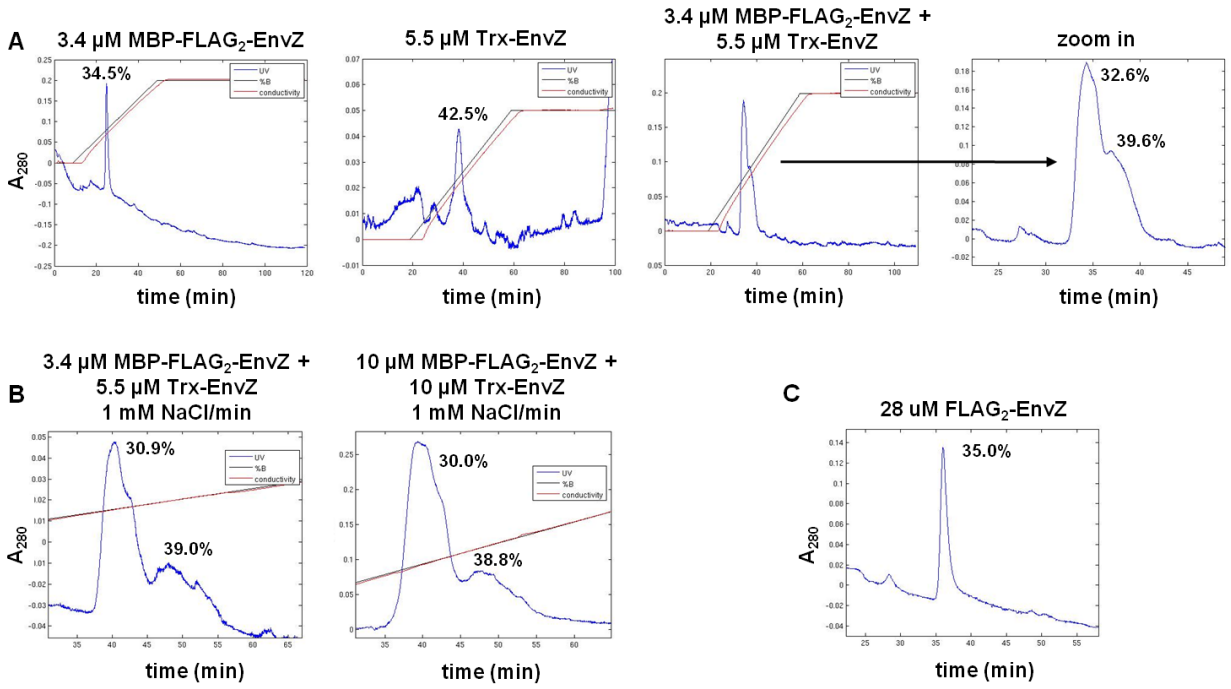


Figure A.2. Anion exchange to separate differentially-charged histidine-kinase dimers
Anion exchange assays were used to separate three dimer complexes: $(MBP\text{-}FLAG_2\text{-}EnvZ)_2$, MBP-$FLAG_2$-EnvZ:Trx-EnvZ, and $(Trx\text{-}EnvZ)_2$. MBP-$FLAG_2$-EnvZ and Trx-EnvZ were equilibrated, and then complexes were separated with a Mono Q column using a linear salt gradient. All EnvZ constructs had the HAMP, DHp, and CA domains. (A) MBP-$FLAG_2$-EnvZ alone, Trx-EnvZ alone, and a mixture of the two species, separated at 2.5 mM NaCl/min. MBP-$FLAG_2$-EnvZ and Trx-EnvZ eluted too closely to distinguish a middle peak in their mixture that could correspond to a heterodimer between the two proteins. (B) Mixtures of MBP-$FLAG_2$-EnvZ and Trx-EnvZ at different concentrations, separated at 1 mM NaCl/min to enhance peak resolution. Although peaks were more resolved, no third peak was detected. (C) To test a different $FLAG_2$-HK construct, MBP-$FLAG_2$-EnvZ was cleaved using TEV protease to create $FLAG_2$-EnvZ, which was then purified with the Mono Q column. $FLAG_2$-EnvZ showed similar elution behavior to its full-length parent MBP-$FLAG_2$-EnvZ.

To increase peak resolution, the salt gradient was decreased from 2.5 mM NaCl/min to 1 mM NaCl/min. This experiment was also attempted at increased protein concentrations of 10 µM. In both experiments, the two peaks were further resolved in time, but no third peak appeared (Fig. A.2B). Overall these two protein constructs were probably still not separated enough to see a possible third peak. Furthermore, the MBP and Trx tags were determining the elution positions of the proteins rather than the presence or absence of the $FLAG_2$ tags, as suggested by the fact that Trx-EnvZ was eluting later than MBP-$FLAG_2$-EnvZ. To follow up on this, I removed the

MBP tag from MBP-$FLAG_2$-EnvZ using TEV cleavage. The TEV-cleaved MBP-$FLAG_2$-EnvZ eluted at 35% B, which was close to the 30-34% B elution position of MBP-$FLAG_2$-EnvZ (Fig. A.2C). Therefore, this construct was not suitable for the desired experiment. Although conditions may be found where anion exchange separation works, it is unlikely to work for all kinases as the subunit exchange rate in some kinases is significantly faster than the timescale of anion exchange separation.

Using yeast-two-hybrid to measure histidine-kinase dimers

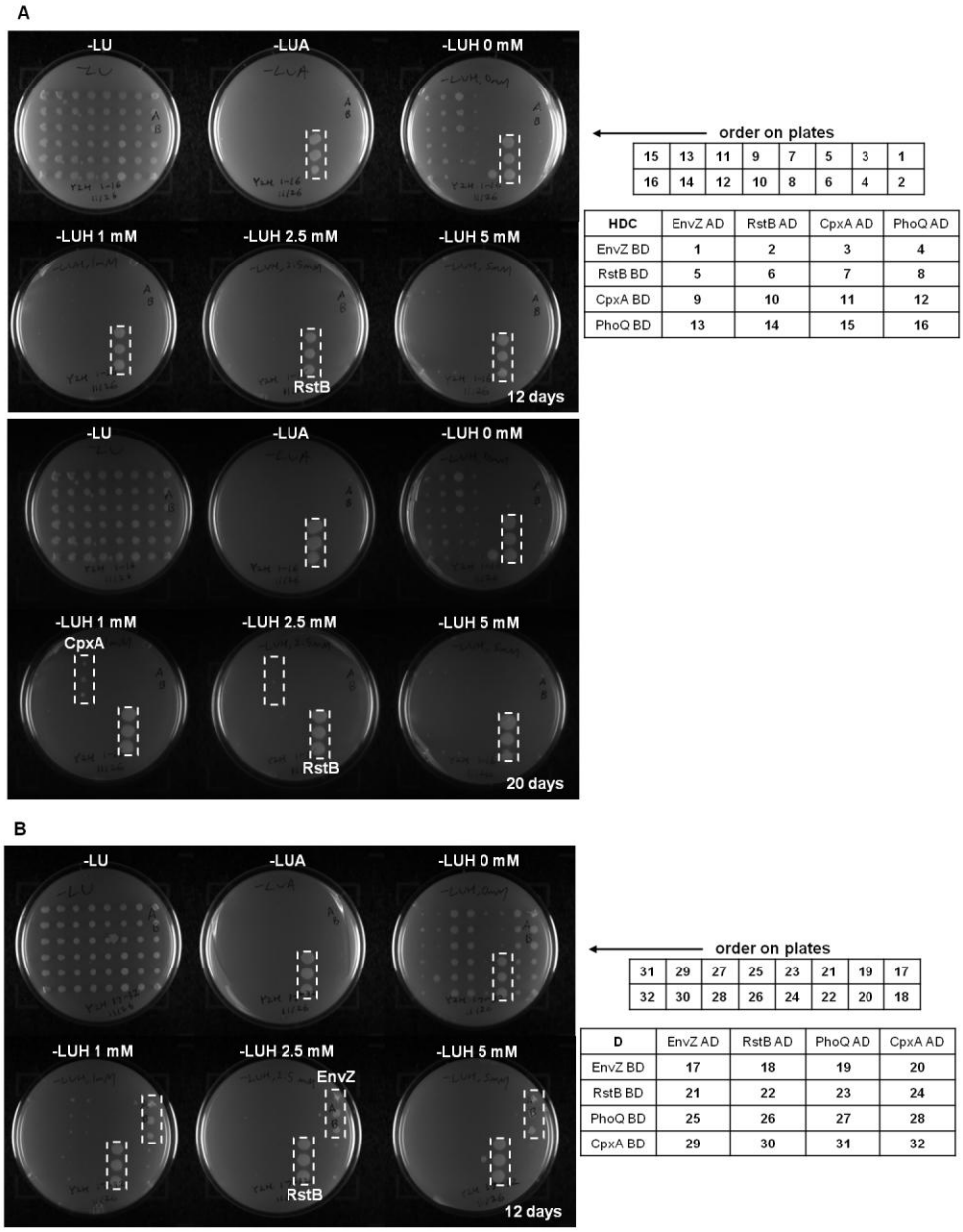
As an alternative to *in vitro* biochemical assays, I also tried yeast-two-hybrid (Y2H), which is based on the yeast GAL4 transcription factor. Prey activation domain (AD/pGAD-C1) and bait DNA binding domain (BD/pGBDU-C3) vectors were prepared as destination vectors compatible with the Gateway system by Christos Tsokos (James et al., 1996). The LR reaction was carried out to make AD and BD destination clones. In these clones, Gal4 AD and Gal4 BD were N-terminal fusions. For each interaction to be assayed, the corresponding AD and BD clones were co-transformed into the yeast strain PJ69-4a, which is a TRP, LEU, URA, ADE, HIS auxotroph. An interaction in this system could drive expression of 3 different reporter genes: HIS3 (H), ADE2 (A), and lacZ, each under control of a different Gal promoter.

For each co-transformation, three independent colonies were picked and re-suspended in wells of 96-well plates. Using a pin-stamp, colonies were stamped onto plates lacking leucine and uracil (-LU), which selects respectively for the AD and BD clones (Fig A.3A). Colonies were also concurrently stamped onto selection media –LUA, and onto selection media –LUH supplemented with 0 mM, 1 mM, 2.5 mM, or 5 mM 3-AT (Fig A.3B). 3-AT is a commonly used drug that inhibits leaky activity of the HIS3 gene product. To test for autoactivation of the

BD clones, yeast transformants with only the BD clones were stamped onto -U, -UA, and -UH+3-AT plates. Growth across all plates was logged every 3 days for 20 days. An interaction was scored as positive if all 3 colonies grew under selection.

Interactions between kinases EnvZ, RstB, PhoQ, and CpxA were assayed under different combinations of HAMP, DHp, and CA domains. Only when a pair of proteins showed up as homodimers growing under the same selection did I interpret the presence or absence of their heterodimer. As positive controls, kinase-regulator pairs from *C. crescentus*, DivJ-DivK (done reciprocally) and PhoR AD-PhoB BD, were tested. Autoactivation controls were done for all 10 kinase BD constructs, in addition to 4 BD constructs from *C. crescentus.* As a negative control, the kinase-regulator pair DivJ AD–PhoB BD was tested. As additional negative controls, each of the 10 HK BD constructs was co-transformed with the regulator DivK AD and each of the 10 HK AD constructs was co-transformed with the regulator PhoB BD. Kinases known not to interact with my kinases of interest would be more appropriate negative controls but that data did not exist at the time of these experiments.

The positive controls (DivJ-DivK and PhoR-PhoB) grew within 3 days on –LUA, and on -LUH at all concentrations of 3-AT (data not shown). No negative or autoactivation controls grew on –LUA, but almost all grew at –LUH 0 mM 3-AT after 20 days. However on –LUH 1 mM 3-AT, only two negative controls and two autoactivation controls grew. The autoactivation controls showed that the two negative controls growing at 1 mM 3-AT were autoactivators. For –LUH 2.5 mM 3-AT and 5 mM 3-AT, no autoactivation or negative controls grew. From these controls we saw –LUA was the most stringent selection condition, consistent with the literature, and that at least 1 mM 3-AT was needed to interpret growth on –LUH plates.



order on plates

15	13	11	9	7	5	3	1
16	14	12	10	8	6	4	2

HDC	EnvZ AD	RstB AD	CpxA AD	PhoQ AD
EnvZ BD	1	2	3	4
RstB BD	5	6	7	8
CpxA BD	9	10	11	12
PhoQ BD	13	14	15	16

order on plates

31	29	27	25	23	21	19	17
32	30	28	26	24	22	20	18

D	EnvZ AD	RstB AD	PhoQ AD	CpxA AD
EnvZ BD	17	18	19	20
RstB BD	21	22	23	24
PhoQ BD	25	26	27	28
CpxA BD	29	30	31	32

Figure A.3. Yeast-two-hybrid to detect histidine-kinase dimers

Colony growth on selection plates using yeast-two-hybrid. Single colonies were spotted on plates using pin-stamping. The upper left plate was the positive control selecting only for presence of activation domain (AD) and DNA-binding domain (BD) hybrids. The other five plates were selection for either adenine, or for histidine under increasing concentrations of inhibitory drug 3-AT. In each column, 2 different interactions were tested with 3 independent colonies picked for each. Growth was shown either after 12 days or 20 days. (A) Growth after 12 days for HDC constructs. $RstB_{HDC}$ homodimer detected. (B) Growth after 20 days for HDC constructs. $RstB_{HDC}$ homodimer detected and $CpxA_{HDC}$ homodimer detected at 1 mM and 2.5 mM 3-AT. (C) Growth after 12 days for D constructs. $RstB_D$ homodimer and $EnvZ_D$ homodimer detected. At 1 mM, 2.5 mM, and 5 mM 3-AT, no EnvZ-RstB heterodimer growth was detected.

Under –LUA selection, RstB homodimer grew within 3 days in both the HDC (HAMP+DHp+CA) construct and the D (DHp) construct (Fig A.3A, B). This was also the only interaction showing up under –LUA (Fig A.3A). Under –LUH 1 mM, 2.5 mM, 5 mM 3-AT, RstB homodimer (HDC and D) and EnvZ homodimer grew (D) within 3 days (Fig A.3A, B). For D constructs, no EnvZ-RstB heterodimer grew, and this lack of interaction was reciprocal. After 15 days, CpxA homodimer (HDC) grew at –LUH 1 mM 3-AT, and after 20 days it grew at 2.5 mM 3-AT (Fig. A.3A). No RstB-CpxA heterodimers grew (Fig. A.3A).

The yeast-two-hybrid experiments gave 0 false positives, but several false negatives, such as never seeing a PhoQ homodimer. Overall 2/4 homodimers grew as HDC constructs, 2/4 homodimers grew as D constructs. No heterodimers grew for either HDC or D constructs.

Competitive inhibition of kinase autophosphorylation

The competitive inhibition of kinase autophosphorylation assay was based around histidine kinases undergoing autophosphorylation *in trans.* When this assay was developed, *in cis* autophosphorylation in kinases had not yet been detected (Casino et al., 2009). However in this assay, I used two kinases, *E. coli* EnvZ and *E. coli* RstB, that autophosphorylated *in trans*. First, single and double mutants were constructed. Mutating the phosphorylation-site histidine in the DHp domain to an alanine removes a kinase's ability to autophosphorylate, and mutating a conserved glycine in the G2 box of CA domain to an alanine removes a kinase's ability to bind ATP (Fig A.4A, B). A wild-type kinase that dimerized with this double mutant would be unable to get autophosphorylated (Fig. A.5A). For example, titrating in EnvZ double mutant into wild-type EnvZ would reduce autophosphorylation signal, as wild-type EnvZ would get bound in non-functional complexes (Fig A.5A). The level of inhibition in this homodimer mix can be

compared to the level of inhibition found when titrating RstB double mutant into wild-type EnvZ.

All characterized EnvZ and RstB constructs (wild-type, single and double mutants) had N-terminal MBP tags, and were in the HDC form (HAMP+DHp+CA), except for the EnvZ double mutant, which was in the DC (DHp+CA) form.

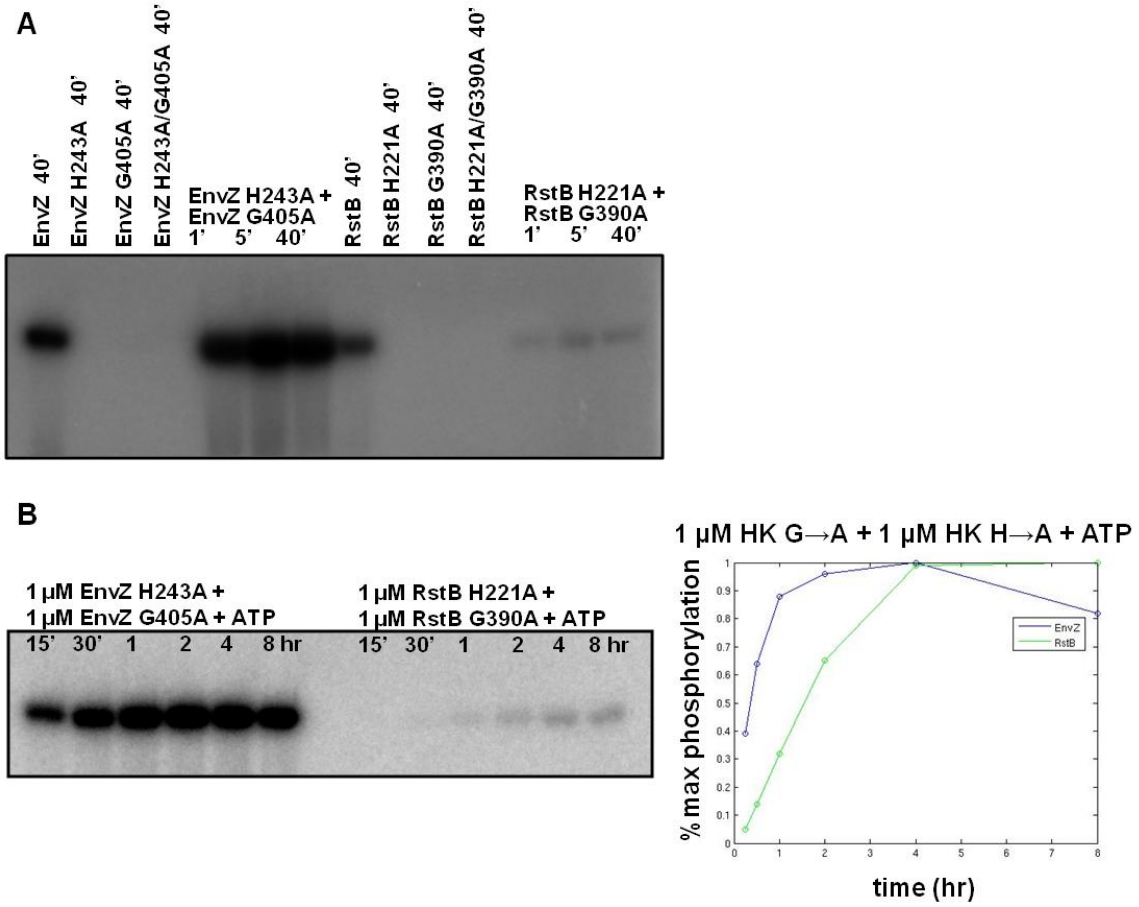


Figure A.4. Subunit exchange between histidine-kinase mutants
Control autophosphorylation reactions for $EnvZ_{HDC}$ and $RstB_{HDC}$. Either the conserved histidine phosphorylation site was mutated to alanine, or the third glycine in the G2 box of the CA domain was mutated to alanine. (A) All proteins were equilibrated 2 hrs at room temperature, and then autophosphorylation reactions were run for the indicated times. Both EnvZ and RstB autophosphorylated *in trans*. (B) To assess the rate of subunit exchange in $EnvZ_{HDC}$ and $RstB_{HDC}$, the H→A and G→A mutants of each kinase were mixed together with radiolabeled ATP, and aliquots of this reaction were collected at the indicated times. EnvZ exchanged more quickly than RstB.

I first ran autophosphorylation controls to show that wild-type kinases were functional, that single and double mutant kinases were not functional, and that autophosphorylation signal could be restored *in trans* by mixing together the single mutant kinases (H→A + G→A). Autophosphorylation was initiated by adding [γ-^{32}P]ATP, and reactions were stopped through the addition of sample buffer containing SDS. After 40 minute autophosphorylation reactions, wild-type $EnvZ_{HDC}$ and $RstB_{HDC}$ autophosphorylated, but none of the mutants autophosphorylated (Fig A.4A). Autophosphorylation in both EnvZ and RstB was restored when mixing the single mutants, with RstB giving significantly weaker signal (Fig A.4A, B).

The kinetics of the reactions, subunit exchange and phosphorylation, involved in the competitive-inhibition experiment were considered next. To measure subunit exchange, I mixed 1 μM HK G→A, 1 μM HK H→A, and ATP, and performed a time course (Fig A.4B). Assuming similar rates of autophosphorylation, EnvZ underwent exchange faster than RstB, as EnvZ reached maximum autophosphorylation signal near 1 hr whereas RstB reached maximum signal near 4 hrs. When performing the competitive inhibition assay below, all equilibration steps proceeded for 4 hrs.

To look at phosphorylation time scales, I equilibrated $EnvZ_{HDC}$ or $RstB_{HDC}$ in the presence or absence of different double mutant HKs, and then ran autophosphorylation time courses (Fig A.5B). EnvZ and RstB alone both reached maximum autophosphorylation signal after 5 min, a faster timescale than that for subunit exchange. For EnvZ, greater levels of autophosphorylation inhibition were observed when adding double mutant $EnvZ_{DC}$ than when adding double mutant $RstB_{HDC}$ (Fig A.5B). Similarly for RstB, greater levels of autophosphorylation inhibition were observed when adding double mutant $RstB_{HDC}$ than when adding double mutant $EnvZ_{DC}$ (Fig

9b). The increase in RstB autophosphorylation signal upon addition of double mutant EnvZ was no longer seen when the experiment was done in Eppendorf tubes designed to have lower protein binding. Based on these results, I ran later autophosphorylation reactions for 5 min, as this gave sufficient range in signal, and was shorter than the timescale of subunit exchange.

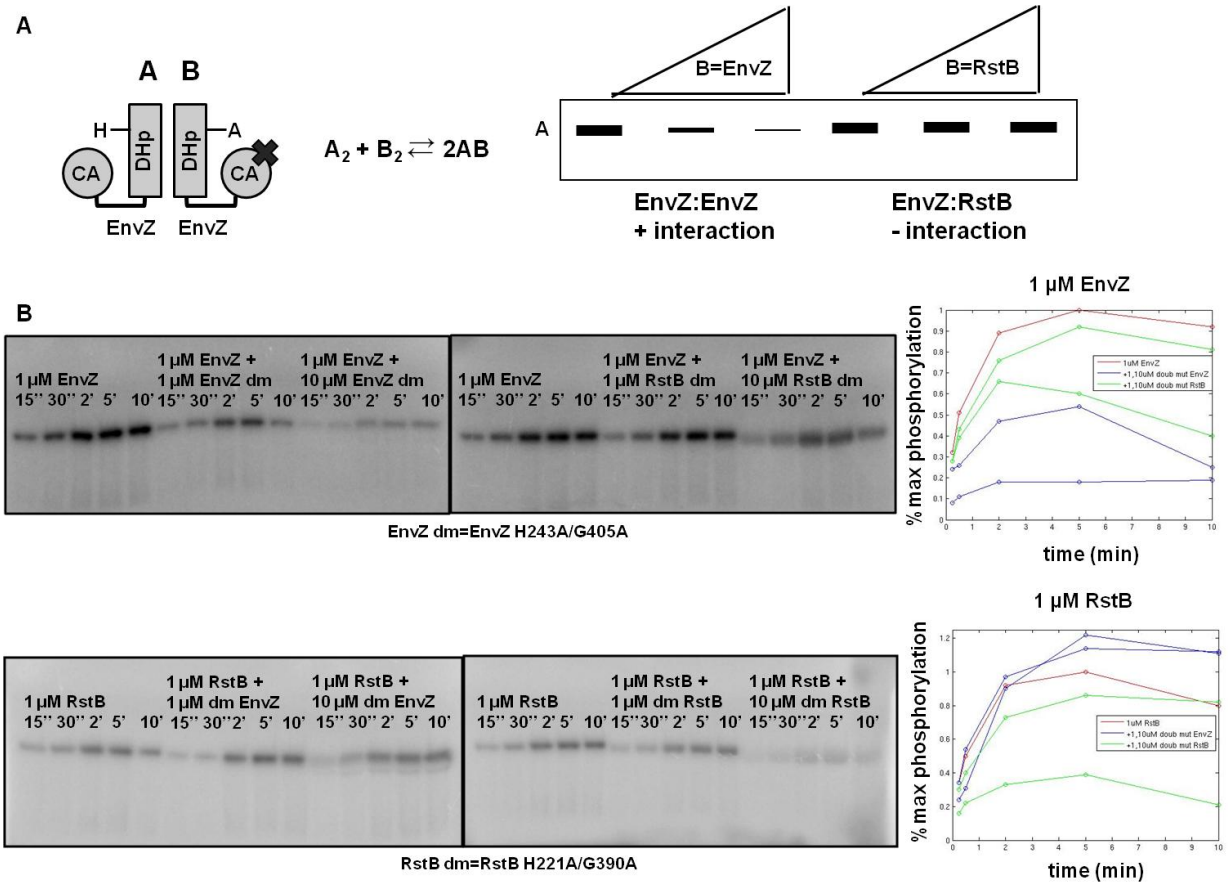


Figure A.5. Competitive inhibition of autophosphorylation time-course.
(A) Cartoon of competitive inhibition assay. Wild-type kinase (A) was mixed with double mutant kinase (B), which had both the H→A and G→A mutants. These mutations remove the kinase autophosphorylate site, and disrupted ATP-binding, respectively. If wild-type kinase A autophosphorylation was inhibited upon adding double mutant kinase B, this indicated an interaction between A and B. Importantly, this assay assumed kinase A autophosphorylated *in trans*. (B) Autophosphorylation time courses for $EnvZ_{HDC}$ and $RstB_{HDC}$ in presence or absence of $EnvZ_{DC}$ and $RstB_{HDC}$ double mutants. Proteins were equilibrated 4 hrs before starting autophosphorylation reactions.

With the two parameters of subunit equilibration time and autophosphorylation reaction time set, I performed competitive inhibition assays where (0, 1, 2, 5, 10, 20)x $EnvZ_{DC}$ or $RstB_{HDC}$ double mutant kinase was titrated into 1 μM $EnvZ_{HDC}$ or 1 μM $RstB_{HDC}$ (Fig. A.6). EnvZ autophosphorylation was inhibited by EnvZ double mutant, and RstB autophosphorylation was inhibited by RstB double mutant. Addition of double mutant EnvZ to RstB, or double mutant RstB to EnvZ, resulted in weaker inhibition compared to the corresponding homodimer mixes.

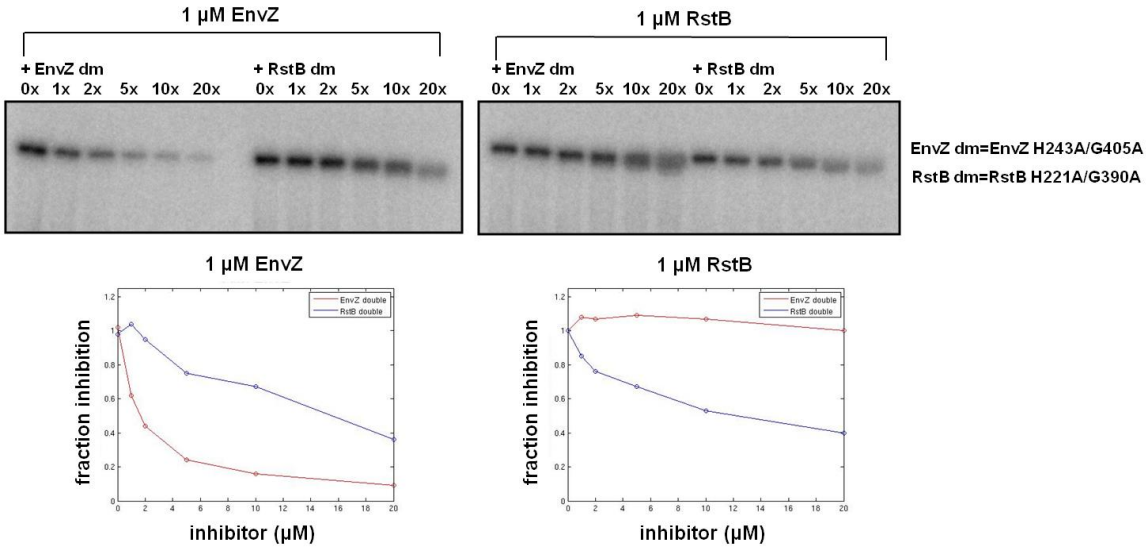


Figure A.6. Competitive inhibition of autophosphorylation titration.
Competitive inhibition titration curves for $EnvZ_{HDC}$ and $RstB_{HDC}$. Increasing amounts of $EnvZ_{DC}$ double mutant or $RstB_{HDC}$ double mutant were added to wild-type kinase, and mixtures were equilibrated 4 hrs. Autophosphorylation reactions proceeded for 5 min. Inhibition of autophosphorylation signal was evidence of an interaction between the wild-type and double-mutant kinases.

References

Casino, P., Rubio, V., and Marina, A. (2009). Structural insight into partner specificity and phosphoryl transfer in two-component signal transduction. Cell **139,** 325-36.

James, P., Halladay, J., and Craig, E. A. (1996). Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast. Genetics **144,** 1425-36.

Schneider, F., Hammarstrom, P., and Kelly, J. W. (2001). Transthyretin slowly exchanges subunits under physiological conditions: A convenient chromatographic method to study subunit exchange in oligomeric proteins. Protein Sci **10,** 1606-13.