ELUCIDATING THE ENERGETICS OF BACTERIAL SIGNAL TRANSDUCTION: INSIGHTS FROM PHOQ

Kathleen S. Molnar

A DISSERTATION

in

Biochemistry and Molecular Biophysics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2015

Supervisor of Dissertation

_____
William F. DeGrado, Ph.D.
Professor of Pharmaceutical Chemistry

Co-Supervisor of Dissertation

_____
S. Walter Englander, Ph.D.
Professor of Biochemistry and Biophysics

Graduate Group Chairperson

_____
Kim A. Sharp, Ph.D.
Associate Professor of Biochemistry and Biophysics

Dissertation Committee

Mark Goulian, Ph.D.
Edmund J. and Louise W. Kahn Endowed Term Professor of Biology, Physics and Astronomy

Mark A. Lemmon, Ph.D.
George W. Raiziss Professor of Biochemistry and Biophysics

Bohdana M. Discher, Ph.D.
Research Assistant Professor of Biochemistry and Biophysics

James Shorter, Ph.D.
Associate Professor of Biochemistry and Biophysics

Yoshitomo Hamuro, Ph.D.
Adjunct Associate Professor of Biochemistry and Biophysics

ELUCIDATING THE ENERGETICS OF BACTERIAL SIGNAL TRANSDUCTION:

INSIGHTS FROM PHOQ

This work is dedicated to my parents,

Arpad and Susana Molnar,

who are the source of inspiration and courage to dream big and work hard.

Acknowledgements

This work was completed within the walls of two great institutions, with the collaboration of many scientists and the support of even more friends. I have limited space to express my gratitude to this deserving group.

I thank my thesis advisor, Bill DeGrado, for all that he provided, namely, his time, his lab, and his grilled salmon. I especially appreciate how supportive Bill has been through the many challenges I faced scientifically as well as circumstantially. I admire how he balances scientific passion with a passion for enjoying life. This balance extends also to the members of DeGrado lab who were thoughtful critics and editors as well as fantastic friends. I must thank several past and present lab members by name – Yoshitomo Hamuro, Shalom Goldberg, Graham Clinthorne, Manasi Bhate, Bruk Mensa – because their suggestions and contributions were crucial to the completion of this work.

I also thank my committee – Mark Goulian, Walter Englander, Mark Lemmon, and Jim Shorter – not only for their scientific critiques and compelling suggestions, but for their patience and understanding. My committee chair, Mark, has also been a great collaborator and I thank him and his lab for being such a superb resource. I thank my co-advisor, Walter, for his support that was felt even across the country. Finally, I thank Yoshitomo Hamuro and Bohdana Discher for agreeing to join the committee as additional examiners; your efforts are greatly appreciated.

I thank the Sali lab for their collaboration and I greatly appreciate their quick email responses.

My heartfelt gratitude goes to my extremely supportive family; my parents, to whom I dedicate this thesis, and my sister, Elizabeth Molnar, who helped me in countless ways over the years.

Finally, I thank Matthew Puster for his exuberant encouragement, his unwavering support, and especially his delicious breakfast sandwiches.

ABSTRACT

ELUCIDATING THE ENERGETICS OF BACTERIAL SIGNAL TRANSDUCTION:

INSIGHTS FROM PHOQ

Kathleen S. Molnar

Dr. William F. DeGrado

Dr. S. Walter Englander

Bacteria transduce signals across the membrane using two-component systems, consisting of a membrane-spanning sensor histidine kinase and a cytoplasmic response regulator. The histidine kinase, PhoQ, serves as a master regulator of virulence response in *S. typhimurium* and *E. coli*. It also is inhibited by divalent cations, particularly Mg^{2+}. While the periplasmic sensor domain of this protein has a unique function, the cytoplasmic portion of this modular protein is made of structurally conserved domains found in many other bacterial sensor kinases. Signal transduction through these conserved domains is thought to be universal; however, the structural and energetic rearrangements that occur during signaling have generated numerous models.

Through Bayesian inference we constructed a two-state model based on cysteine crosslinking data and homologous crystal structures. These two signaling states differ in membrane depth of the periplasmic acidic patch as well as the reciprocal displacement of diagonal helices along the dimer interface. Comparative studies of multiple histidine kinases suggest that diagonal displacement of helices is a common mode of signal transduction. A similar scissor-like model was previously ruled out in CheA-linked chemoreceptors; therefore, this new evidence suggests that sensor His-kinase and CheA-linked receptors possess different signaling mechanisms.

To unify the various signaling mechanisms that exist for the different protein domains, we built a thermodynamic model based on Linked Equilibrating Domains (LED). We used this model to quantitatively interpret functional data of single-point Ala, Phe and Cys mutants throughout the signal transducing regions of PhoQ. Data from 35 mutants, including both activating and deactivating phenotypes, were globally fit using LED, and gross features such as V_{max} and K_d were related to more nuanced population distributions and thermodynamic coupling. LED analysis highlights the principles by which individual signaling domains can be connected to create a functional signal transducer. These principles allow us to quantitatively explain signaling in histidine kinases and are likely to be broadly applicable to many other signal transduction proteins.

Table of Contents

List of Tables

List of Figures

CHAPTER ONE

1.1 Introduction

Bacteria utilize sensor histidine kinases to sense and interact with their environment (Stock et al., 2000). Signal transduction through these proteins controls critical processes from biofilm formation (Lv et al., 2012) to chemotaxis (Hazelbauer, 2012). Important in human health, histidine kinases (HKs) also regulate virulence (Tobe, 2008) and contribute to antibiotic resistance (Perry et al., 2011). Besides being attractive drug targets (Bem et al., 2015), they are also attractive model proteins for studying transmembrane signal transduction (Hall et al., 2012; Lemmin et al., 2013; Lowe et al., 2012; Molnar et al., 2014).

This thesis focuses on signal transduction through one particular HK, PhoQ, which is required for virulence in *Salmonella typhimurium* (Miller and Mekalanos, 1990). In this chapter, I introduce the features of PhoQ that make this protein a tractable and canonical system for studying reversible protein phosphorylation. I will also provide background on the techniques used to probe PhoQ signaling later in the thesis. Finally, I present a review article, in press at *Structure*, that summarizes the current state of the HK signal transduction field in terms of structural models and thinking. My role in this review article was to contribute text and figures to a collaborative document in conjunction with a postdoc, Manasi Bhate. She contributed the structural analysis of the numerous crystal structures and is therefore listed as first author.

1.2 A model signal transduction system, PhoP/Q

PhoP/Q is a bacterial two-component system (TCS) with canonical architecture. The sensor protein, PhoQ, is membrane associated and controls the phosphorylation state of a cognate response regulator (RR), PhoP. Environmental deprivation of Mg^{2+} stimulates PhoP phosphorylation and the expression of genes required for bacterial virulence. Additionally, this system can respond to other divalent cations (e.g. Ca2+), low pH (Gao and Lynn, 2005), and antimicrobial peptides (AMPs) (Kindrachuk et al., 2007). Finally, this system has two single-pass

transmembrane effector proteins, SafA (Eguchi et al., 2012) and MgrB (Lippa and Goulian, 2012), which have opposing effects on PhoQ signaling.

While much of the biology of these bacterial systems are known, the question that remains is: how do these proteins transmit a signal across the membrane? In the case of PhoQ, sensor domain is located in between two transmembrane (TM) regions. Sensor HKs are dimeric proteins and many form a helical bundle core that spans the length of the protein (Goldberg et al., 2008). The PhoQ dimer forms an antiparallel four-helix bundle in the TM that is connected to a parallel four helix bundle and that is connected to the last antiparallel bundle. This structural core and packing is not unique; instead, all proteins in this class have high structural homology at the domain level and you will find many examples of functional chimeric proteins. Chimeras between proteins have been used to aid in crystallization of flexible domains (Ferris et al., 2012), and chimeras between species have shown conserved functionality (Lesley and Waldburger, 2001).

Table 1.1 – Functional overview of PhoQ domains

	Function	Abbr.	Notes
1	Sense [Mg2+], low pH, and AMPs	PAS	Largest class of PAS fold proteins found among TCS sensor kinases (Moglich et al., 2009)
2	Localize to membrane	TM	Few crystal structures
3	Transduce signal	HAMP	Relay signal from membrane bound components to cytoplasmic domains (Schultz et al., 2015)
4	Location of catalytic His and RR docking	DHp	Crystalized in active and inactive conformations
5	Bind ATP and perform autokinase activates	CA	Highly conserved domain among HKs

Much is known about the functionality of these different domains. Briefly, PhoQ has five domains and the function of these domains are listed in a table and labeled in a corresponding figure. The detailed structures, abbreviations, and specifics on all of the domains can be found in the last section of this chapter. This modularity leads to the conclusion that there must be universal signaling mechanism; however, the field is full of competing models of signal transduction (some are highlighted in the review). This counterintuitive result is likely an artifact of working within one system or within one particular domain.

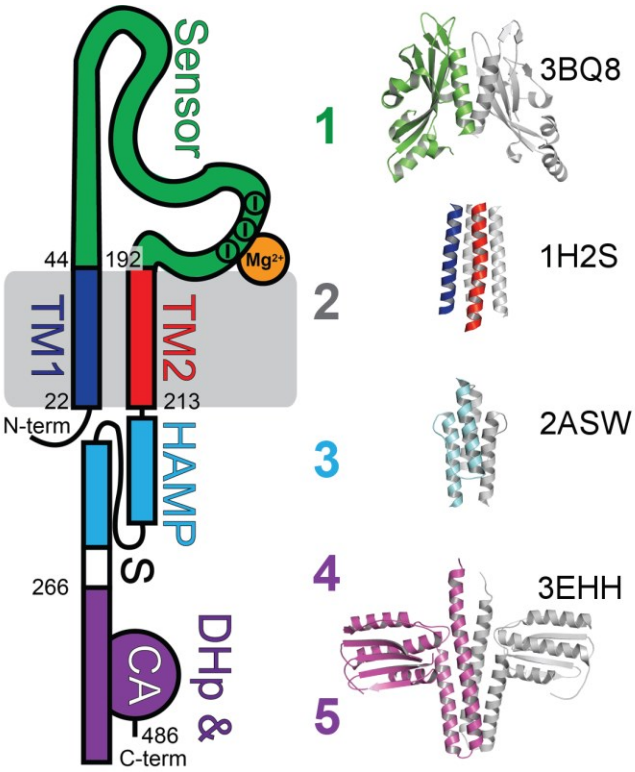


Figure 1.1 – Two structural representations of PhoQ. The numbers in the middle of the figure correspond to the numbering in Table 1.1.
(Left) Topological schematic of a PhoQ monomer. Residue numbers for *E. coli* PhoQ (UniProt: P23837) are included for scaling.
(Right) Crystallographic view of a PhoQ dimer composed from homologous protein structures used to model of each domain of PhoQ. The corresponding PDB ID is listed. One monomer is color-coded and the other monomer is shown in grey.

There are several experimental considerations that make the PhoP/Q system amenable to biochemical study. First, unlike some of the redundant pathways in bacteria, PhoP/Q signaling is relatively insulated from crosstalk. In E. coli, the PhoQ knockout strain is viable and the background phosphorylation of PhoP is quite low as judged by vector plasmid transformants. Second, PhoQ kinase and phosphatase activity on PhoP can be modulated by several ligands, the most straightforward being magnesium. Bacteria grow without defect in a large range of magnesium, which allows for consistent measurements of protein activity unhindered by the bacterial growth. Finally, it is possible to probe the signal transduction pathway of PhoQ as a full-length protein in a native membrane. The major techniques in this thesis rely either on bacterial membrane fractions containing PhoQ (membrane prep) or on PhoQ expressed from a plasmid in an in vivo reporter assay.

1.3 Overview of techniques used to probe signal transduction

Both of the techniques we used to probe the signal transduction pathway rely on mutagenesis. In each we look at a large group of mutants to give us an overall picture of the protein landscape.

Disulfide scanning. Mutational analysis using single cysteine mutations has been shown to effectively probe conformational changes in the periplasm of bacterial TCS (Careaga and Falke, 1992; Falke and Koshland, 1987). This technique is used to determine the distance and orientation of two cysteine mutants along a dimer interface, thus giving restraints to the conformational state various functional conditions. The chemotaxis protein, Tar, has been extensively studied with this technique (Butler and Falke, 1998; Chervitz and Falke, 1995; Swain et al., 2009). Falke and coworkers showed that the dimeric interface of Tar remained static with and without aspartate (Chervitz et al., 1995). They also used disulfide scanning to later show aspartate dependent motion within one monomer of Tar (helix 4) (Chervitz and Falke, 1996). The major limitation of these studies is that the models generated naturally become a two state model (with and without stimulus). They presupposed that the coupling between signaling domains will be a rigid and these models cannot account for the possible contributions from signaling intermediates.

LacZ reporter assay. In this study we focused on probing PhoQ by varying the concentration of Mg(II). This ligand does not have a distinct binding site, but instead is thought to bridge a cluster of acidic residues and the bacterial inner membrane. This system is the reverse of a ligand binding and causing activation. For PhoQ, the ligand binding activates the phosphatase activity of the protein towards phosphorylated PhoP, but even at high Mg(II) concentrations PhoQ remains autophosphorylated (Castelli et al., 2000). The next step along the signal cascade is for phosphate group to transfer from the catalytic Histidine to a conserved aspartate on PhoP. A dimer of PhoP-P will then bind DNA and act as a transcription factor for its targeted genes. Here is where reporter assays are inserted into the genome to measure the activity of these proteins *in vivo*. Phosphorylated PhoP and binds known promoter sequences (*mgtA*, *mgrB*, etc.) and the desired reporter protein is placed downstream. Reporter proteins have utilized enzymatic activity (beta-galactosidase) as well as fluorescent proteins (e.g. GFP/YFP) for

fluorescent microscopy (Ram and Goulian, 2013) and flow cytometry (Podgornaia and Laub, 2015).

Other Techniques. There are also several important techniques used to characterize *in vitro* preparations of PhoQ. Here, I will mention: 1) radioactive phosphoimaging and 2) hydrogen/deuterium exchange measured by mass spectrometry (HX-MS). By far the oldest technique, determining PhoQ activity using radiolabeled ATP, is still a required method as validation of protein activity in conjunction with *in vitro* assays (Podgornaia and Laub, 2015). The key reagent, hot-ATP, has the ^{32}P at the gamma position, which can then be transferred only to active protein (autophosphorylation) and further tracked in subsequent steps (phospo-transfer). Proteins are then separated by electrophoresis and the gels are exposed to a phosporimage plate for detection and quantification via a scanner such as a Typhoon.

A much younger technique, HX-MS measures protein stability and dynamics by monitoring the rates of exchange of the amide protons on the protein backbone. Here, the protein is mixed with deuterated buffers and allowed to exchange for varying time intervals. Deuterium exchange is then quenched at low temperatures (0°C) and acidic pH (2.3). The protein is then proteolyzed into small peptides by an acid stable protease, pepsin (Hamuro et al., 2008), and sent directly to the mass spectrometer. The mass shift is determined and compared to a theoretical mass shift to determine the degree of protection for that peptide (Kan et al., 2011; Skinner et al., 2012a, b). These measurements are amenable to a variety of experimental conditions and has been used to study membrane associated proteins (Chetty et al., 2013), large protein complexes (Walters et al., 2013), as well as bacterial HKs (Koshy et al., 2013; Wang et al., 2012).

1.4 Signal Transduction in Histidine Kinases: Insights from New Structures

In press manuscript at Structure*, a Cell Press journal, reproduced here under fair use guidelines*

Histidine kinases (HKs) are major players in bacterial signaling. There has been an explosion of new HK crystal structures in the last five years. We globally analyze the structures of HKs to yield insights into the mechanisms by which signals are transmitted to and across protein structures in

this family. We interpret known enzymological data in the context of new structural data to show how asymmetry across the dimer interface is a key feature of signal transduction in HKs, and discuss how different HK domains undergo asymmetric-to-symmetric transitions during signal transduction and catalysis. A thermodynamic framework for signaling that encompasses these various properties is presented and the consequences of weak thermodynamic coupling are discussed. The synthesis of observations from enzymology, structural biology, protein engineering and thermodynamics paves the way for a deeper molecular understanding of histidine kinase signal transduction.

1.5 Review Introduction

Two-component systems represent a critical stimulus-response mechanism abundant in most bacteria. Stimulus is sensed by a histidine kinase (HK) and transmitted to a response regulator (RR), which in most cases binds to DNA and mediates a cellular response. A given bacterium can have tens to hundreds of different two-component systems (Ulrich and Zhulin, 2010) that allow it to sense and adapt to a variety of environmental signals such as osmotic changes (EnvZ/OmpR), temperature (DesK/C), small-molecules (NarX/Q, CitA/B) and antimicrobials (PhoQ/P) (Aguilar et al., 2001; Albanesi et al., 2009; Cheung and Hendrickson, 2010; Groisman, 2001; Kaspar et al., 1999; Miller et al., 1989; Rabin and Stewart, 1992; Russo and Silhavy, 1991; Tanaka et al., 1998). A molecular description of the structures and conformational dynamics of HKs is therefore fundamental to understanding bacterial signal transduction.

HKs catalyze the net transfer of a phosphoryl group from ATP to a histidine residue, and then to an aspartate of a response regulator resulting in signaling. Intriguingly, many HKs also function as phosphatases, which catalyze the hydrolytic cleavage of the phosphoryl group from the Asp residue of the response regulator. Sensory input in the form of ligand-binding or other environmental cues regulates the balance between these two opposing catalytic states.

The last five years have witnessed an explosion in high-resolution crystal structures of various domains of histidine kinases in various states (Albanesi et al., 2009; Diensthuber et al., 2013a; Ferris et al., 2014; Ferris et al., 2012; Marina et al., 2005; Mechaly et al., 2014; Wang et al.,

2013). Although the structure of a full-length membrane-spanning HK has not yet been reported, low-resolution models based on disulfide cross-linking have been built (Molnar et al., 2014). Structural studies of intracellular and extracellular sensors have elucidated how these domains bind their ligands (Cheung and Hendrickson, 2010). The mechanisms by which HKs recognize their response regulators have been revealed through bioinformatics, protein engineering, and structural studies (Casino et al., 2010; Laub and Goulian, 2007; Podgornaia et al., 2013), and recently two structures of the ATP-bound Michaelis complex primed for autophosphorylation were reported (Casino et al., 2014; Mechaly et al., 2014). Despite this recent progress, many questions remain about the structural mechanism by which signals pass from the extracellular sensors to the cytoplasmic catalytic domain.

Here we compare and analyze the recent high-resolution structures of HK domains to glean insight into the structural ensembles and perturbations associated with signal transduction. Cryogenic crystal structures cannot tell us about the kinetics or molecular dynamics of signal transduction and catalysis, but they likely represent limiting states of conformational trajectories that can be used to inform various mechanistic models. Signaling can be understood in terms of thermodynamic coupling between distinct conformational states of individual protein domains. Transitions between symmetric and asymmetric states appear to be important in some domains. We explore how transitions within individual domains affect neighboring domains as signals are transmitted across the protein.

1.6 Domain architecture of histidine kinases

Histidine kinases are multi-domain proteins with considerable architectural variety across the family. The class of HKs discussed here are constitutive homodimers with extracellular sensors (Capra and Laub, 2012; Gao and Stock, 2009; Jung et al., 2012; Stock et al., 2000). Other HKs have transmembrane or intracellular sensors. CheA-linked chemo-attractant receptors are proteins that share some domains with HKs, but have distinct higher order structures yielding different biochemical properties (Hazelbauer et al., 2008).

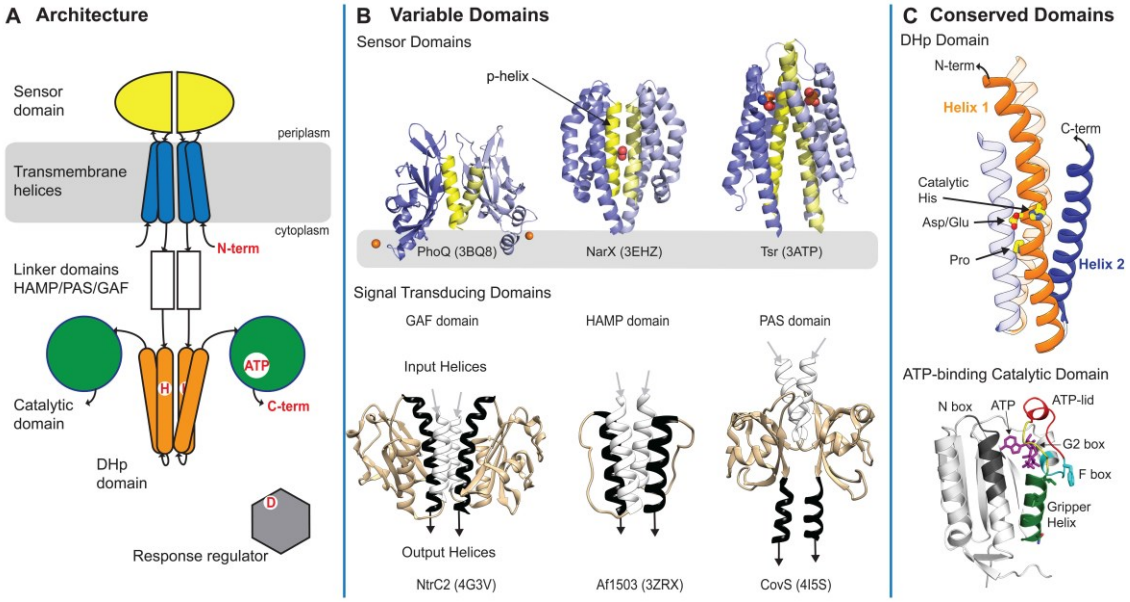


Figure 1.2 – *Two-component system architecture.* **(A)** Schematic representation of a canonical TCS. Many of these domains are repeated in other protein classes. The DHp is the catalytic core of Histidine Kinases. **(B)** Variable domain structures from three sensor and signal transducing domains colored to highlight key features. The yellow helix (top) shows the dimeric interface. The white and black helices (bottom) show input and output helices. **(C)** Structures of conserved domains, DHp and catalytic, colored to highlight various features. DHp dimeric structure (top) has the N-terminal helix colored in orange and the C-terminal helix in blue. Critical regions for function are labeled and color-coded in the CA structure (bottom).

Each subunit of a canonical HK (**Fig. 1.2.A**) has an antiparallel pair of transmembrane (TM) helices near the N-terminus with the two helices separated by a periplasmic sensor domain. We use the term periplasmic (instead of “extracytoplasmic”) to refer to both Gram negative and Gram positive bacteria (Matias and Beveridge, 2006). A variety of different signal transducing elements connect the sensor/TM regions to the catalytic regions near the C-terminus of the protein.

The periplasmic sensor domain is comprised of one or more diverse folds that include α/β PAS domains such as in CitA (Sevvana et al., 2008) and PhoQ (Cheung et al., 2008), or all α-helical structures like NarX (Cheung and Hendrickson, 2009) and TorS (Moore and Hendrickson, 2012) (Fig. 1B, top). The modes by which HK sensors respond to stimuli vary widely, as do the nature of the stimuli. Generally, stimulus sensing is transmitted through a key helix, which we define here as a periplasmic helix (p-helix) that connects the sensor to the transmembrane

domain. Some HKs which respond to bilayer stress have only a small extracellular loop (Mascher et al., 2006).

The transmembrane domain in many HKs (PhoQ, CitA, LuxQ, EnvZ etc.) forms a 4-helical bundle in the membrane, with two TM helices from each monomer. In some cases like DesK and FixL each monomer has multiple helices, forming large transmembrane bundles. Intracellular signal transducing domain(s) include a variety of domains like HAMP[1], PAS[2], GAF[3] (**Fig. 1.2.B**, bottom) and coiled-coils that reside in the cytoplasm just below the transmembrane domain. Transducing domains often exist as combinations or tandem-repeats and are generic signal transducers found in many other bacterial and mammalian proteins.

The dimerization and histidine phosphotransfer (DHp) domain is the site of the three catalytic reactions: Histidine phosphorylation, phosphotransfer to the response regulator, and, for bifunctional HKs, a phosphatase reaction (not a simple reversal of the phosphotransfer reaction). The DHp forms a homodimeric antiparallel 4-helical bundle with two helices connected by a hairpin loop (**Fig. 1.2.C,** top). The catalytic histidine is located a few turns down the bundle on the solvent exposed side of helix 1. It begins a seven amino acid stretch of conserved sequence (**Fig. 1.3.A**)

Residue #:	**1**	**2**	**3**	**4**	**5**	**6**	**7**
AA sequence:	**H**	**-(D/E)-**	**(L/I)-**	**(K/R)-**	**(T/N)-**	**P**	**- L**

The sidechain of the Asp or Glu (residue 2) is positioned to serve as a hydrogen bond acceptor or general acid/base while the Lys or Arg interact with acidic phosphoryl groups during phosphotransfer. The conserved (T/N)-P dipeptide (residues 5,6) is a locus for helix bending that allows the N-terminal end of helix 1 to adopt multiple conformations during catalysis (**Fig. 1.3.B**). The bottom of the DHp bundle hosts the binding interface for the cognate response regulator as discussed previously (Capra and Laub, 2012; Podgornaia et al., 2013). The DHp is connected via a flexible loop to the catalytic domain.

---

[1] HAMP domains are named for being present in Histidine kinases, Adenylate cyclases, Methyltransferases and Phosphodiesterases.
[2] PAS domains are named for being first discovered in Periodic circardian proteins, Aryl hydrocarbon nuclear translocator protein and Single-minded protein.
[3] GAF domains are named after being present in cGMP-specific phosphodiesterases, Adenylyl cylases and Formate hydrogenlases.

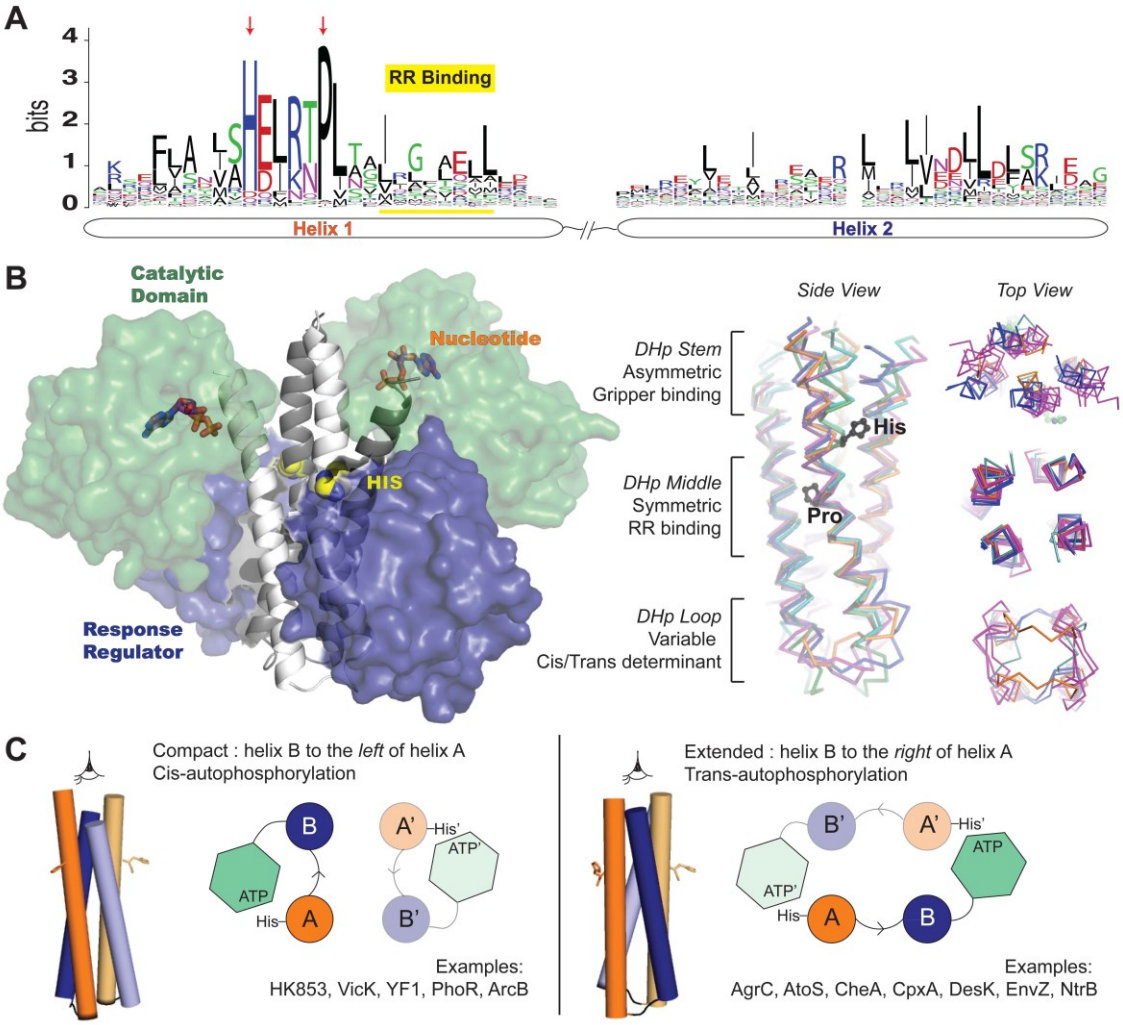


Figure 1.3 – *Catalytic core of Histidine Kinases.* **A)** Surface representation of the catalytic domain (green) and the response regulator (blue) are shown on their respective docking sites on the DHp bundle (left). Structural alignment of DHp bundles from various HKs (right) show that the core of the bundle is highly symmetric and conserved while the ends of the bundle vary. **B)** A sequence logo for the helices of DHp domains (based on Pfam 00512) with conserved positions. An acidic residue always follows the catalytic histidine. There is a loop of variable length between the two helices. **C)** A schematic showing how the handedness of the loop between the two DHp helices determines auto-phosphorylation geometry in the DHp bundle.

Examination of DHp domains from distantly related proteins reveal regions of high conformational conservation as well as high variability (**Fig. 1.3.B**). These regions are also mirrored in comparisons of different catalytic states in a single HK. The top third of the bundle, above the conserved proline, is the docking site for the catalytic domain during autophosphorylation. It switches between symmetric and asymmetric conformations, which

appear to correlate with phosphatase and kinase states of the enzyme, respectively. In contrast, the central core of the bundle is highly symmetric and conformationally homogeneous over a large number of HKs and catalytic states. Finally, the lowest part of the bundle is highly variable in length and handedness of the interhelical loop, fitting its role of providing sequence-specific interactions for cognate response regulator recognition.

The *cis* vs. *trans* autophosphorylation propensity of different histidine kinases stems from the intrinsic handedness of the hairpin loop between the two DHp helices (Ashenberg et al., 2013). If this loop turns right (**Fig. 1.3.C**, right), then the catalytic domain of one monomer is closer to the catalytic histidine of the dimeric partner, and autophosphorylation proceeds in trans. In the alternate case (**Fig. 1.3.C**, left), the loop turns left and the catalytic domain is closer to its own histidine and autophosphorylation occurs in cis. Altering the handedness of the loop between the helices can interconvert cis- and trans-kinases (Ashenberg et al., 2013; Casino et al., 2014). The catalytic domain (CA), also known as the ATP binding domain, is a highly conserved α/β-sandwich with three alpha helices packed against five anti-parallel beta strands (**Fig. 1.2.C**, bottom). The nucleotide binds between two helices, and is held by a loop known as the ATP lid. Conserved nucleotide binding sequences that comprise the binding site are called the N, G1, G2 and F boxes named for critical residues within these sites (Kim and Forst, 2001).

We also define a “Gripper” helix in the CA domain, which works in conjunction with a Phe in the F-box to bind to the DHp domains in different manners as the protein transitions from one catalytic activity to the next (**Fig. 1.5**). It associates with the DHp using surface-exposed sidechains arranged along one face of the helix (at positions i, i+4, i+8, i+11), which differ from one kinase to another, but tend to be primarily hydrophobic. These sidechains are arranged like four sticky fingers, while the Phe serves as thumb. The length and orientation of the Gripper helix adjusts in response to the bound nucleotide.

The response regulator (RR) is typically a separate soluble protein with mixed α/β structure and a conserved aspartate that is phosphorylated during catalysis (Bourret, 2010). In some cases, a receiver domain can be fused directly to the HK.

1.7 Catalytic properties and enzyme mechanistic investigations of HKs

Autophosphorylation involves transferring the terminal phosphate from ATP to the Nε atom of the catalytic histidine to form a high-energy N-P bond. *In vitro* biochemical assays show that the K_M for ATP for the autokinase reaction is ~10-200 μM across various HKs. This reaction is slow, with reported k_{cat} values in the regime of 0.1-5 min^{-1} (Gutu et al., 2010; Jiang et al., 2000; Trajtenberg et al., 2010; Yeo et al., 2012). The binding constant (K_D) for ATP is estimated to be tight, ~10 μM, from binding-studies of the non-hydrolyzable analogue ANP-PNP (Casino et al., 2014). The K_D for ADP binding is weaker; ~80-100 μM based on isothermal calorimetry and photo-affinity labeling (Yeo et al., 2012). For reference, a typical bacterial cell has an ambient ATP concentration of ~1 mM (Buckstein et al., 2008); the ADP concentration is generally constant at about one tenth this value. Given these nucleotide affinities, it follows that at least one binding site of a typical HK is saturated with ATP/ADP at any given time in a cell.

Several HKs exhibit an intrinsic asymmetry across the dimer in the autokinase reaction. Early enzymological studies of NRII showed that in the presence of physiological concentrations of ATP and ADP, the protein is predominantly phosphorylated on only one of its two identical subunits. Both subunits are phosphorylated equally only when ADP is continuously recycled back to ATP (Jiang et al., 2000). Similar half-of-sites reactivity was reported for the kinase HK853, which showed fast accumulation of the hemi-phosphorylated form when treated with excess ATP (Casino et al., 2014). These observations propose that ADP inhibits the autokinase reaction by competing with ATP.

Further characterization of nucleotide-binding kinetics is required to understand how different nucleotides influence the kinase and phosphatase states. It has been known for a long time that HK phosphatase activity is stimulated by ADP as well as by ATP and non-hydrolyzable analogs (Igo et al., 1989). Recently, Groisman and coworkers have proposed a kinetic mechanism by which ADP inhibits kinase activity in PhoQ, analogous to end-product inhibition in enzymes. While they also propose that asymmetry is important, they predict that the release of ADP from the kinase is remarkably slow (t½ ~ 30 min) and limits turnover (Yeo et al., 2012).

Indeed their model predicts that a single turnover occurs on a time scale slower than that of the doubling time for *E. coli.* These studies, however, were not performed under a constant ADP/ATP ratio, nor corrected for the build-up of ADP during the course of the reaction. Thus, it would be of interest to directly measure the off-rate of ADP following ATP hydrolysis for PhoQ.

His-to-RR phosphotransfer involves transfer of the phosphoryl group from phospho-His to an aspartate on the bound response regulator. The affinity (K_D) between a HK and its cognate RR is of the order of 1 μM. Phosphotransfer is faster than the autophosphorylation, with reported k_{cat} values on the order of 20-100 min^{-1} (Fisher et al., 1996; Yeo et al., 2012). Comparative studies show that non-cognate RRs have both a lowered affinity and slower phosphotransfer kinetics, which inhibits cross-talk between different two-component circuits *in vivo* (Fisher et al., 1996; Laub and Goulian, 2007). Unlike the autokinase reaction, enzymatic studies of phosphotransfer do not report any asymmetry across the dimer. RRs are typically present in greater abundance their cognate HKs (Li et al., 2014). For example it was estimated that in *E. coli* the steady state concentration of the kinase EnvZ is ~0.1 μM whereas its RR, OmpR, exists at 3.5 μM (Russo and Silhavy, 1991). Corresponding numbers for the kinase PhoQ and its response regulator PhoP are ~0.5 μM and 2.7 μM under non-inducing conditions (Yeo et al., 2012). Thus, since a given RR is not in large molar excess over its cognate HK, only a few turnovers are required to change the fractional population of phosphorylated response regulators, a fact consistent with relatively slow kinetic rates for *in vitro* autophosphorylation. Furthermore since many HKs exhibit both phosphatase and kinase activity, two hurdles are crossed with one leap and a single turnover affects both the production and the consumption of the phosphorylated response regulator.

The lifetime of the phosphorylated response regulator is typically on the order of seconds to days (Bourret et al., 2010) and is controlled by both auto-dephosphorylation and intrinsic phosphatase activity from the HK. The phosphatase reaction is mechanistically distinct from the phosphotransfer step, and in some cases does not require the catalytic histidine. In cases where the catalytic histidine is required, it likely acts as a base to assist the attack of water or hydroxide on the phosphoryl-aspartyl group.

1.8 Is half-of-sites reactivity reflected in structures?

Individual domains of HKs adopt both symmetric and asymmetric conformations in different catalytic states, and there is now good reason to believe that some of these symmetry-asymmetry transitions are important for function.

Structural Analysis of Recent Structures of DHp and CA domains

The structures and symmetry of DHp and CA domains depends markedly on the catalytic act in which they have been crystallized. There are over 20 different structures of linked DHp-CA domains deposited in the PDB. In addition, there are several structures of DHp domains without the catalytic domains.

Structures with response regulator bound are believed to be in either the phosphotransfer or the phosphatase states. In both states, two RRs bind symmetrically to each dimeric DHp-CA unit and occupy the bottom half of the DHp bundle. The CA domain occupies the top half and also forms sparse contacts with the bound RR (Casino et al., 2009) (**Fig. 1.3.A**). The phospho-accepting Asp group of the RR is bound to a structurally invariant and sequentially conserved region of the DHp near the catalytic His (Casino et al., 2010) and ADP nucleotides are bound to both catalytic subunits, consistent with the observation that ADP stimulates phosphatase activity. Overall, the structures with bound RRs are highly symmetric.

DHp-CA constructs crystallized in the absence of RR domains are much less symmetric (Casino et al., 2014; Diensthuber et al., 2013b; Mechaly et al., 2014; Wang et al., 2013). We structurally aligned all the known structures using the middle of the DHp bundle as an anchor because this yielded the best overall alignment (see supplement for details). We then defined a coordinate system in which the dimeric axis of the DHp was coincident with the z-axis and the Cα atoms of the conserved proline were on the x-axis. Each structure was transformed into this coordinate system (see supplement for details).

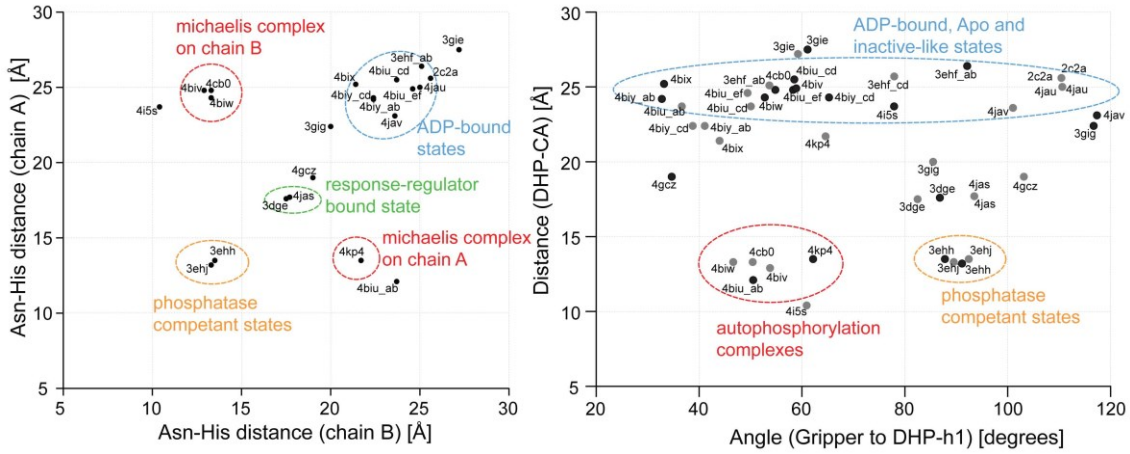


Figure 1.4 – *Asymmetry in CA-DHp distances.* **A)** The distances between the catalytic His (on the DHp) and an ATP-binding Asn (on the CA) on each side of the dimer are correlated for all known HK structures. Structures representing the autophosphorylation Michaelis complex show significant asymmetry **B)** The DHp-CA distance is correlated with the angle between the Gripper helix and the DHp helix. The Gripper helix shows a very specific angular preference in in the autokinase Michaelis structures.

Asymmetry in CA-DHp distances

We noticed marked asymmetry in our aligned ensemble. To distil a parameter that defines the structural state of each monomer, we measured the distance between the catalytic His and a conderved ATP-binding Asn in the CA domain, and constructed a correlation plot of this distance in chain A versus B (**Fig. 1.4.A**). The autophosphorylation Michaelis complex is defined by a close approach between the CA and DHp domains; about 10-15 Å in the Michaelis complex versus 20-30 Å in other states, including the doubly ADP-bound states. The approach is asymmetric in that it occurs on only one side of the homodimer in all known structures. The other, inactive subunit shows an interaction more typical of the ADP-bound state. In contrast, structures representative of the phosphatase or phosphotransfer to the response regulator are much more symmetric.

Gripper Helix mediates CA-DHp interactions

The Gripper helix, which is at the edge of the ATP lid near the G2 box (**Fig. 1.5**), adopts different orientations relative to the DHp in different catalytic states. The local structure of Gripper also varies with the bound nucleotide; when ATP is bound, Gripper is shortened by a turn, which

allows the G2 box to form a P-loop like motif to bind the beta phosphate of ATP. Frequently, the hydrophobic residues on Gripper and/or the Phe from the F-box insert between DHp helices 1 and 2 (or 2', depending on the bundle topology) wedging them apart.

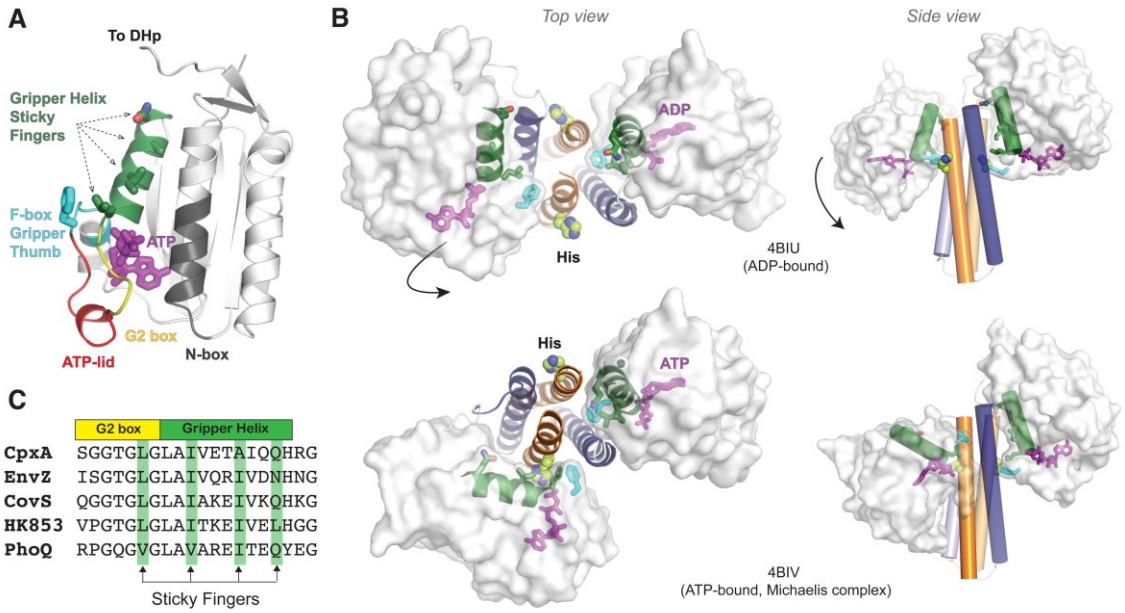


Figure 1.5 – *Nucleotide dependent placement of the Gripper helix.* Structural representations of the DHp and CA domains are shown in ADP- (left) and ATP- (right) bound states show different associations with the Gripper. On the inactive side, Gripper residues interact with the DHp stem to form a 3 helix bundle, with the Phe (cyan) wedged into the dimeric interface. This conformation is also seen on both monomers in the ADP-bound state. On the active side, Gripper is released to facilitate autokinase activity. This conformation is only seen in one monomer of the ATP-bound state. DHp helices are colored as before. Gripper residues are conserved across several HKs.

We correlated the distance between the CA and DHp domains with the angle of the Gripper helix relative to the first helix of the DHp. For catalytically competent states the Gripper helix is close to the DHp helix (less than 15 Å) and oriented about 60 degrees from the DHp helix (**Fig. 1.4.B**, red circle). This tight cluster is in direct contrast with the broad range of orientations adopted by the Gripper in the apo- and inactive states (**Fig. 1.4.B**, blue circle).

Asymmetry of the DHp bundle

We see a large variation in the plasticity of the DHp bundle: the central core is highly invariant whereas the top and the bottom of the bundle vary significantly depending on the catalytic state.

To quantify the variation we extracted Cα coordinates at four inward-facing positions along the bundle (**Fig. 1.6.A**). Slice 1 corresponds to positions that had the closest distance between Cβ atoms across the dimer. At each slice, we computed the inter-monomer distance, and the distance to the bundle axis (z). In general, helix 1 is closer to its dimeric partner (8-16 Å) than

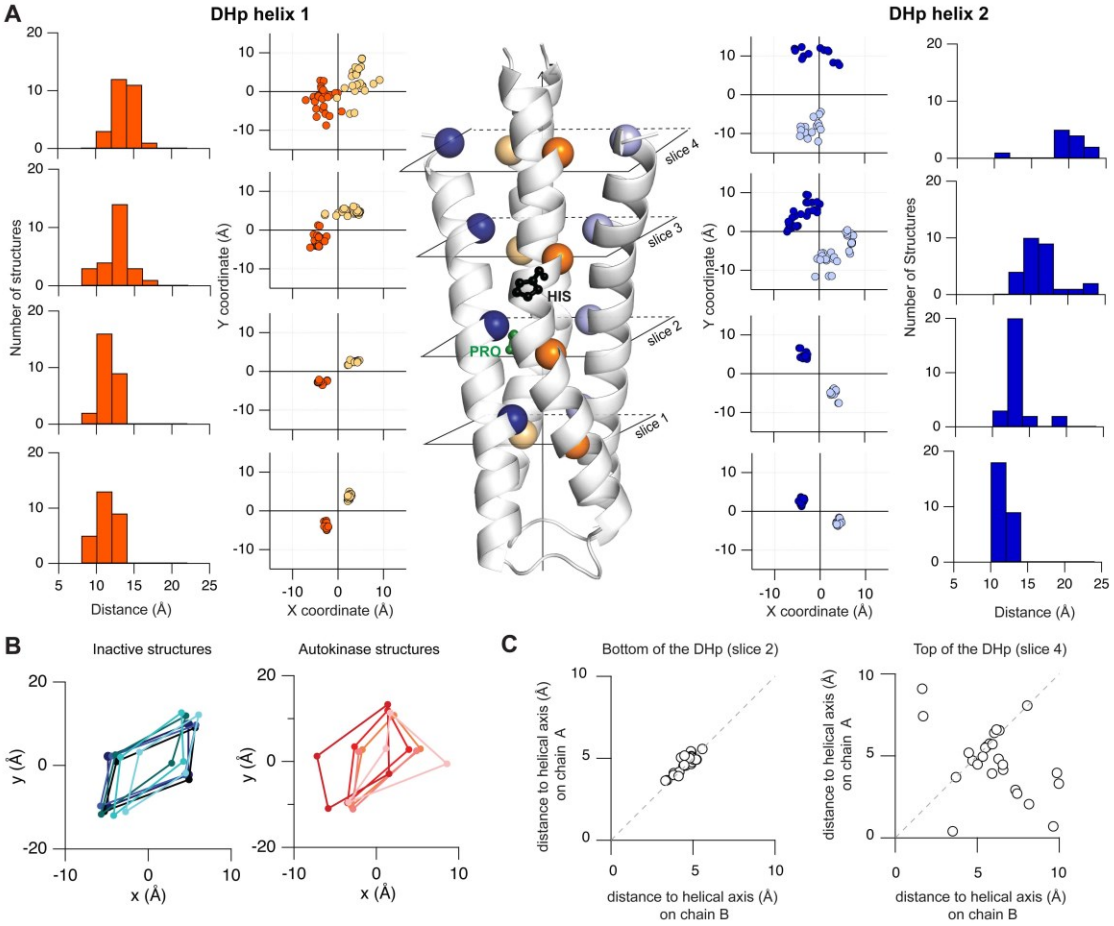


Figure 1.6 – *Asymmetry in the DHp bundle.* **A)** 23 structures of DHp domains were structurally aligned such that the z-axis coincided with the dimer axis (see supplement). X,Y coordinates extracted at 4 different slices across the bundle show that the core of the bundle is highly invariant whereas the top of the bundle varies significantly. Inter-monomer Cα-Cα distance distributions are also shown. **B)** Structures that represent ADP-bound, Apo and inactive states show a symmetric diamond like geometry (blue) in the XY coordinates at slice 4, whereas ATP-bound Michaelis complex structures show a distorted kite-like geometry (red). **C)** A third correlation plot of distances from the central helical axis to chain A vs. chain B shows that slice 2 is highly symmetric whereas slice 4 is asymmetric.

helix 2 (10-25 Å). The variance of the distance distributions increased from 1.1 Å and 0.4 Å in the conserved core to 4.6 Å and 12.9 Å near the top of the bundle for helices 1 and 2 respectively. This is a remarkable change in variance across the bundle.

The large variance at the top of the bundle is presumably because of conformational transitions associated with docking the CA domain in different catalytic states. Projections in the x and y plane of coordinates at the top of the DHp (slice 4) reveal that in apo and ADP-bound states, the bundle forms a symmetric diamond-like shape whereas in ATP- and ATP-analogue bound structures, the bundle is more distorted (**Fig. 1.6.B**)**.** In some extreme cases like DesK, the helical crossing angles at the top of the DHp switch from being left-handed (in kinase-competent states) to right-handed (in phosphatase-competent states).

We and others (Mechaly et al., 2014; Wang et al., 2013) propose that the signal triggers catalysis by inducing an asymmetric bend in the DHp stem that promotes strong interactions between the Gripper and the DHp bundle on the inactive side of the dimer while concomitantly releasing the Gripper from the active side of the dimer. In general, the consequence of ligand-induced activation involves remodeling the energetic landscape of the DHp stem so that it preferentially populates an asymmetric structure. Thus structural asymmetry is directly mirrored in functional asymmetry (half-of-sites reactivity discussed above).

1.9 Signal transducing elements: HAMP, PAS and polar linkers

Stimulus-dependent remodeling of the DHp bundle relies on the extracellular signal reaching the top of the DHp bundle. Commonly the signal must pass through a HAMP[1], PAS[2] or GAF[3] domain, and exit C-terminal helices that encode the output signal.

Structural topology of a HAMP domain

HAMP domains are parallel 4-helical bundles with two helices from each monomer. A highly conserved glycine marks the end of the first helix and the beginning of a loop that wraps around the structure (**Fig. 1.7.A**). A conserved glutamate marks the beginning of the second HAMP helix. When viewed from the top, the bundle has two interfaces: a close inter-subunit interface (**Fig. 1.7.E**, Cα-Cα distance 4.0-5.5 Å) and a far intra-subunit interface (Cα-Cα distance 7.0-8.5 Å)**.**

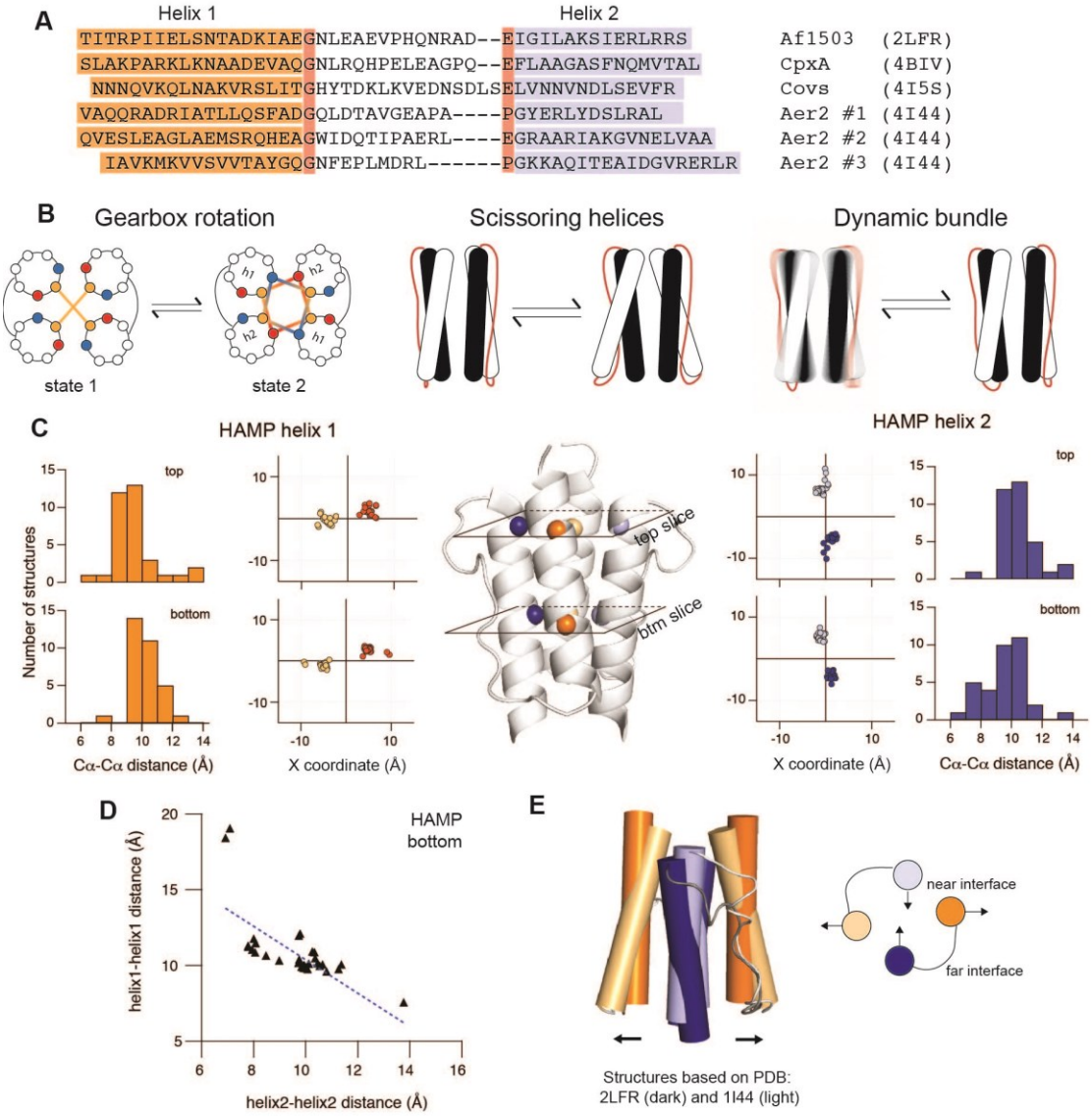


Figure 1.7 – *HAMP Domains*. **A)** Sequence alignment of six HAMP domains of known structure show a conserved Gly and a semi-conserved Glu. **B)** Three different mechanisms proposed for HAMP signaling are described in cartoon form. **C)** An aligned ensemble of 26 HAMP domains was used to extract XY coordinates and distance distributions at 2 locations as in Fig. 5. The distribution for helix 1 (orange) is skewed outwards at the bottom of the bundle whereas helix 2 is skewed inwards (blue). **D)** The negative slope of the correlation between intermonomer Cα-Cα distances show that displacements in helix 1 and 2 are anti-correlated. **E)** The two overlaid structures, from Af1503 (2LFR,dark) and Aer2 (1I44, light) show a diagonal displacement of helices.

Proposed models for signal transduction via HAMPs

Different mechanisms have been proposed to describe HAMP mediated signaling; these are not always mutually exclusive (**Fig. 1.7.B**).

The *gear-box mechanism* suggests that the HAMP is a bi-stable structure with two nearly iso-energetic states that interconvert via helical rotations. This model was proposed because the original structure of the Af1503 HAMP showed knobs-to-knobs coiled-coil packing that differed from what is typically seen in four-fold symmetric homotetrameric coiled-coils, so the authors postulated that a second state with a more canonical knobs-to-holes packing must also exist (Hulko et al., 2006). More recent studies provided some support for this hypothesis, but also showed that the structures deviate significantly from ideal coiled-coils (as reflected in deviations in Crick angles). Thus, the precise formalism of coiled coil registry change envisioned in the gearbox mechanism is restrictive, and might be expanded to include additional changes in helical packing and dynamics.

The *diagonal scissoring mechanism* proposes that the HAMP alternates between one state in which the helices are tightly packed, and a second in which they splay out at the C-terminal and form a loose bundle. These states have been crystallized in structures of tandem HAMP repeats (**Fig. 1.7.E**) (Airola et al., 2010). Further evidence comes from disulfide cross-linking studies of Tar, which show that interactions across the close interface are critical for function. Stapling these helices together yielded the kinase-on or kinase-off state of the protein depending on the position of the staple (Swain and Falke, 2007; Swain et al., 2009).

The *dynamic bundle model* proposes that one signaling state of the HAMP has significantly more disorder than the other (Parkinson, 2010; Stewart, 2014). In its most extreme form, controlled local unfolding of the HAMP has been proposed as the final signal transduction event (Schultz and Natarajan, 2013). Specifically, a folded and tightly packed HAMP domain restrains the downstream domain in an inactive state, whereas an unfolded, loosely packed HAMP allows the subsequent domain to access all the necessary signaling states. The rationale for this model comes from HAMP-adenylate cyclase proteins, which show localized unfolding of a few helical turns between the HAMP and the enzyme in structures of the active state (Tews et al., 2005). There is no direct evidence for HAMPs unfolding in HKs, although a number of mutational studies on NarX (Appleman et al., 2003), Af1503 (Ferris et al., 2012; Ferris et al., 2011), Tsr

(Ames et al., 2008) and Tar (Swain and Falke, 2007) reveal that adding large sidechains to the core of the HAMP has a striking effect on kinase activity.

Recent structures of CpxA suggest that the bottom of the HAMP might also be involved in a tertiary interaction with the catalytic domain on the inactive side of the asymmetric autokinase state. The generality of this interaction remains to be seen. Consensus in the field is that HAMPs are highly malleable structures that occupy a relatively shallow energy landscape allowing them to shuttle between different conformations very easily. This marginal stability is a key feature of signal transduction.

Symmetric, anti-correlated helical displacement in HAMPs

Although the HAMP often connects directly to the highly asymmetric stem of the DHp bundle, all known structures of HAMP domains are relatively symmetric homodimers: the Cα RMSD between monomeric chains is < 1Å on average. In order to quantify the symmetry, we generated an aligned ensemble of 26 reported HAMP structures in a manner analogous to the DHp and extracted the X and Y coordinates for inward facing Cα atoms at the top and the bottom of the bundle (see supplement for technical details).

Unlike the DHp (**Fig. 1.6.A**), the HAMP (**Fig. 1.7.C**) does not have distinct regions of structural variability; instead the coordinates suggest smaller rigid body shifts that are concerted over the whole bundle. Earlier analysis showed that helical rotation is a small component of the observed variance across structures (Molnar et al., 2014). Remarkably, displacements in the two HAMP helices are anti-correlated: in structures where helix 1 tilts outwards, helix 2 moves inwards (**Fig. 1.7.D**) suggesting that helical tilts are a large component of the input-output signaling through HAMPs. Distance distributions show a variance of ~2 Å at both the top of helix 1 and the bottom of helix 2, consistent with these regions being important for signal input and output.

The variance in Cα coordinates from the aligned HAMPs is much smaller than the variance at the DHp stem, so some sort of signal amplification must occur between the HAMP

and the DHp. Furthermore since the HAMP conformations are all symmetric, whereas the autokinase states of the DHp and CA are decidedly asymmetric, a symmetry-asymmetry transition must also occur between the HAMP and the DHp.

Structure and signal transduction through PAS domains

The canonical PAS domain is a mixed α/β structure containing a central antiparallel 5-stranded beta sheet (**Fig. 1.2.C**) surrounded by several helices. This central beta sheet is curved like a baseball glove and the two faces make contacts critical in signaling. The inner surface, or palm of the glove, is responsible for binding cofactors or ligands (e.g. heme, tricarboxylic acids) while the outer surface makes contacts between dimers or with flanking output helices (via a conserved Asp-Ile-Thr (DIT) motif). The PAS architecture allows ligand-binding to alter the packing and dynamics of the flanking alpha helices that transmit the signal. PAS domains are found as linkers as well as periplasmic sensors; they are the dominant sensor domains for HKs in *B.subtilis* (Chang et al., 2010; Cheung and Hendrickson, 2010).

Structural alignments of all known PAS domains, and details of their signaling mechanisms have been reviewed extensively elsewhere (Henry and Crosson, 2011; Möglich et al., 2009b), so we highlight only a few results from the laboratory of Möglich and Moffat. These coworkers have shown that the length of the helical linker between the PAS domain and the catalytic histidine is tightly controlled and varies by multiples of seven across the family, indicating that a heptad register is involved in transmitting signals downstream of the PAS (Moglich et al., 2009). Using these principles, they engineered a blue light sensitive HK by adding a light sensing PAS domain above the DHp of FixL (Diensthuber et al., 2013). Recently, the structure of the CovS histidine kinase has provided the first structural view of a natural HK connected to a PAS domain (Wang et al., 2013). The structure is remarkably asymmetric, especially in the PAS output helices that connect it to the DHp.

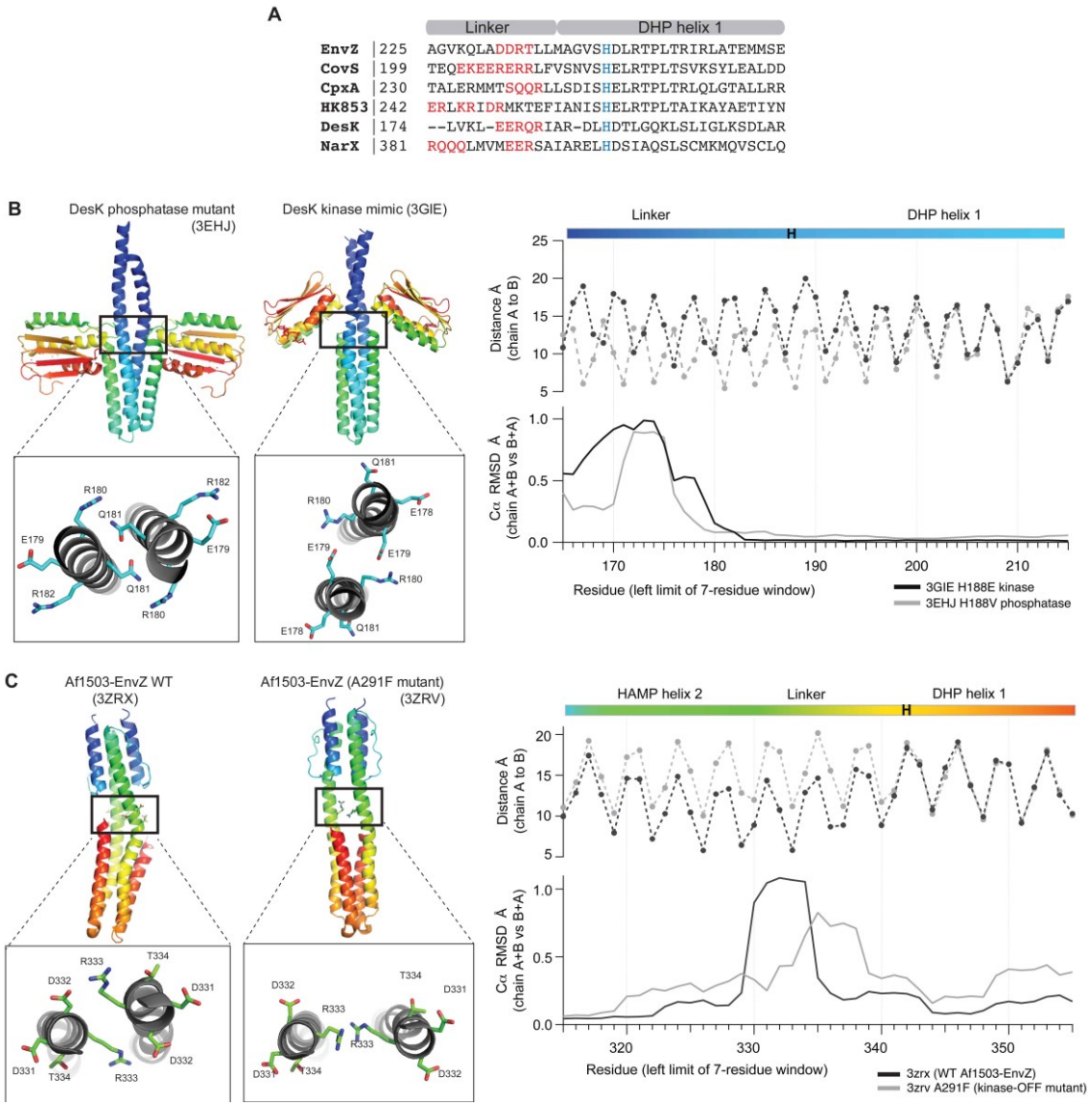


Figure 1.8 – *Symmetry-asymmetry transitions through polar linkers.* **A)** Polar clusters located a few residues above the catalytic histidine are present in several HKs**.** Alternate packing arrangements of the polar clusters result in asymmetric bending of the linker as seen in structures of the temperature-sensitive DesK **(B, left)** and Af1503-EnvZ chimera **(C, left)**. The asymmetry profiles (**B,C right**) plot the Cα RMSD of computed over a 7-residue moving window for chains A and B superimposed on chains B and A respectively, and the Cα-Cα distance for each residue across the dimer. There is striking asymmetry in the linker in many HK structures.

Symmetry-Asymmetry transitions through polar linkers

The short helical linkers between the HAMP/PAS and the DHp have special properties that facilitate asymmetric signaling. The HAMP and DHp domains are typical helical bundles with hydrophobic interiors and polar exteriors yielding a regular pattern of polar and apolar residues in

their sequences. This pattern is interrupted in the short linker, often by insertion of a single residue such that an ideal coiled-coil cannot smoothly connect the two domains. Biochemical studies on several HKs show that systematically manipulating the helical registry in these linkers can completely alter signaling (Moglich et al., 2009; Stewart and Chen, 2010; Winkler et al., 2012). Furthermore, these linkers are often rich in polar residues (**Fig. 1.8.A**), which must be buried (at the thermodynamic cost of dehydration) at the dimer interface. The net result is conformational strain such that no single symmetric state presents a deep energy well, and the linker helices are free to bend asymmetrically as seen in structures of DesK (**Fig. 1.8.B**) and the AF1503-EnvZ chimera (**Fig. 1.8.C**). Thus the symmetric conformational of the HAMP is converted into an asymmetric conformation in the DHp.

1.10 Sensory elements: Periplasmic and Transmembrane domains

Sensor domains are the most diverse part of histidine kinases. They can be extracellular, transmembrane or intracellular and they differ in their structures and ligand binding modes (Cheung and Hendrickson, 2010; Mascher et al., 2006).

The structurally elusive transmembrane domains:

There is no crystal structure for any TM in this family. A set of NMR structures for the monomeric transmembrane units from ArcB, QseC and KdpD (Maslennikov et al., 2010) have been solved in detergent micelles at high temperature, however these structures have limited utility because the dimer interface is not obvious. Instead, the structure of a phototaxis sensory rhodopsin II-transducer complex (HtrII) is often used as a model for the HK transmembrane four-helical bundle (Moukhametzianov et al., 2006).

Site-directed mutagenesis, disulfide cross-linking and molecular dynamics simulations have been used to characterize structure and helical periodicity in the TM domain of PhoQ (Goldberg et al., 2010; Lemmin et al., 2013). These studies show that a polar residue, which forms a small water pocket in the core of the TM domain, is critical for signal transduction. Remarkably, this residue can be moved up or down a turn or even moved to a neighboring TM helix without significant perturbations to the function. Replacements with other polar residues

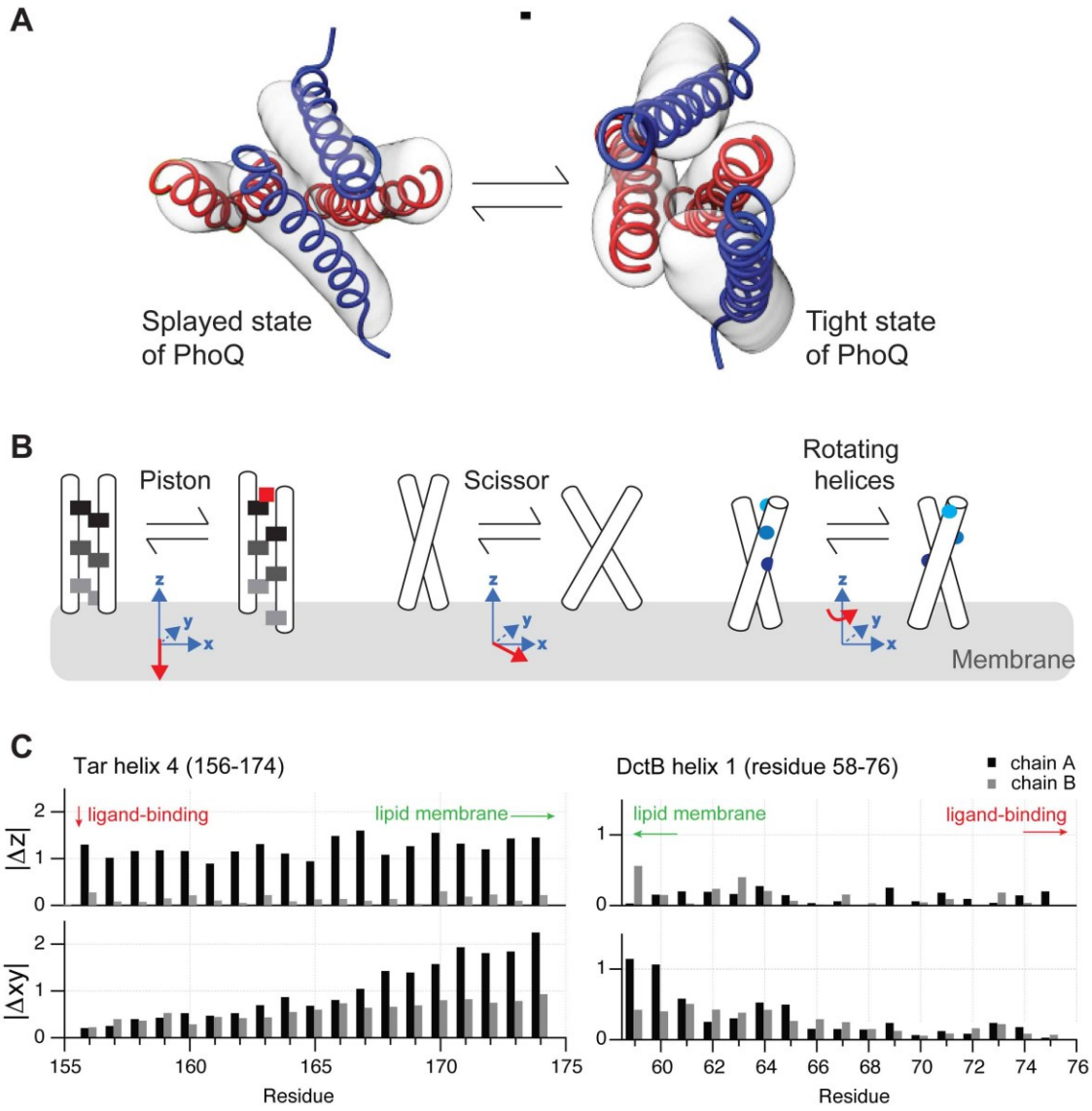


Figure 1.9 – *The TM and periplasmic sensory elements*. **A)** Two structural states of the transmembrane bundle of PhoQ inferred using Bayesian modeling of disulfide crosslinking data. The states differ in their helical packing by a diagonal or reciprocal displacement of the helices. **B)** A simplified cartoon of three proposed models for conformational changes in the sensor domain. **C)** The magnitude of the displacement in the z and xy coordinates of the p-helices of Tar (left) and DctB (right) upon ligand binding. A small z-shift (pistoning) and a larger xy-shift (scissoring) are seen in both cases as the helix enters the membrane.

(e.g. Asn, Gln, Arg, His) at that position maintain the phenotype, while a hydrophobic mutation switches the protein to a phosphatase state. Thus, the helical bend induced at the water pocket allows the TM helices to switch between left- and right-handed crossing angles in the two signaling states.

Recent data on PhoQ in which ligand-dependent disulfide cross-linking experiments were interpreted using a Bayesian framework, showed that two different structural states were required to adequately fit the data (Molnar et al., 2014). The main difference between these states is a diagonal displacement of the TM helices (**Fig. 1.9.A**).

Diversity in Extracellular Sensory mechanisms:

Sensory mechanisms are diverse across the HK family. Many sensors bind to a ligand directly, but the binding modes vary. For example nitrate binds in a cavity at the dimer interface of NarX (Cheung and Hendrickson, 2009); magnesium ions bind to an acidic patch at the periphery between the PhoQ sensor and the lipid bilayer (Cheung et al., 2008), and in CitA and DctB, the ligands (citrate and malonate) bind inside the PAS domain (**Fig. 1.2.B**, top) (Sevvana et al., 2008; Zhou et al., 2008)**.** Other HK sensors sense the state of a secondary protein that binds to the ligand; for example the sensor domain of LuxQ interacts with LuxP, which binds directly to the ligand AI-2 (Neiditch et al., 2006).

Despite this diversity, all sensors couple ligand-binding/stimulus-sensing to a conformational change that propagates across the receptor. The p-helix (defined earlier) is often found at the dimer interface of the sensor. Many of the known ligand binding sites are proximal to the p-helix and trigger structural rearrangements in it that propagate to the TM domains.

Asymmetry in sensor domains.

Asymmetry is often a critical feature of HK stimulus sensing, but the details of the asymmetry vary. LuxQ, for example, forms an asymmetric 2:2 complex upon binding to its cognate periplasmic protein, LuxP (Neiditch et al., 2006). TorS begins in an asymmetric 2:2 complex with its periplasmic partner, TorT, in the apo-state, and the activated complex becomes symmetric upon TMAO binding (Moore and Hendrickson, 2012). The homodimeric aspartate receptor, closely related to sensor HKs, binds two Asp ligands with negative cooperativity, and signals from the asymmetric, singly-bound state (Chervitz and Falke, 1996; Yu and Koshland, 2001). Even in cryogenic crystal structures of dimeric sensor domains, large deviations from 2-fold rotational

symmetry exist. For example, the PhoQ sensor domain forms a number of asymmetric interactions across the dimer interface, which suggests they are important for signaling and cannot be maintained in a two-fold symmetrical structure (Cheung et al., 2008).

Proposed models for stimulus-structure coupling

Given the diversity in stimulus-sensing mechanisms across different HKs, a number of models have been proposed for ligand-induced structural perturbations.

The piston model (**Fig. 1.9.B**, left), which was based on the chemoreceptor Tar (Chervitz and Falke, 1996), proposes that ligand binding generates an ~1 Å piston-like downward displacement in the p-helix that propagates into the TM bundle along an axis normal to the bilayer plane. Pistoning has generally been discussed within a rigid-coupling framework, with many papers suggesting that the 1Å displacement propagates to the TM helices in an all-or-nothing fashion (Chervitz and Falke, 1996; Falke and Erbse, 2009). The scissor-blade model (**Fig. 1.9.B**, middle), was also initially proposed for Tar based on crystallography (Milburn et al., 1991), but fell out of favor after disulfide crosslinking supported pistoning. Here, the ligand/signal induces a scissor-blade like diagonal displacement of the p-helices relative to the membrane normal, such that they splay out as they enter the membrane. Recently there has been renewed interest in scissoring since the sensor domains of the heparin binding hybrid kinase BT4663 (Lowe et al., 2012) and the magnesium sensitive PhoQ (Molnar et al., 2014) were both proposed to scissor. A third model suggests that helical rotations are the primary conformational motion required for signaling (**Fig. 1.9.B**, right). This is based on the LuxP-LuxQ complex in which the asymmetric ligand-bound state requires rotation of the transmembrane helices (Neiditch et al., 2006).

Pistoning and Scissoring in Tar and DctB

We analyzed the sensors of the aspartate receptor, Tar and the dicarboxylate receptor DctB which have physiologically relevant dimeric structures in both apo- and ligand-bound states. We aligned all structures to a z-axis, normal to the membrane, and measured changes in the x,y and z coordinates of Cα atoms to evaluate ligand-induced pistoning and scissoring (details in

supplement). Structural alignments between two distinct conformations of other sensors have been analyzed in detail elsewhere (Cheung and Hendrickson, 2009; Molnar et al., 2014).

Two general results emerged. First, diagonal displacement explains a large component of the variance between apo- and ligand-bound structures. While Tar shows a greater vertical displacement compared to DctB, consistent with the long-standing idea that ligand-binding induces a piston-like vertical displacement in Tar, for both DctB and Tar the lateral xy displacement is much more pronounced than the vertical displacement and increases as the p-helix approaches the membrane (**Fig. 1.9.C**). There is no appreciable change in helical phase or rotation. Pistoning and scissoring cannot be completely decoupled in a simple x-y-z coordinate system, but this simple analysis qualitatively recapitulates results from a full parameterization using an orthogonal basis set (Molnar et al., 2014). Additionally, the conformational changes are highly asymmetric across the dimer interface; in both Tar and DctB, the magnitude of the shifts was generally larger on chain A of the dimer, than on chain B (**Fig. 1.9.C**).

Many more structures of sensors are required to assess if there is any generality in sensory mechanisms across the HK family. It is especially useful to think of the sensor-TM as a thermodynamically coupled equilibrium because it is the energetic effect of the ligand-induced structural change rather than the mechanical details that dictates how the signal is transduced.

1.11 Statistical-Thermodynamic Model and Coupling of Conformational Transitions

Given the multi-domain architecture of histidine kinases, and the relatively independent nature of each domain, it is helpful to think of signal transduction as a series of equilibria between thermodynamically linked protein domains. These equilibria could be tightly coupled leading to a switchable cooperative two-state system, or they can be weakly coupled, leading to the population of many intermediates.

To illustrate this conceptually, consider an isolated catalytic domain that can sample two states, an active state (**X**) or an inactive state (**0**) (**Fig. 1.10.A**). The fractional population of the

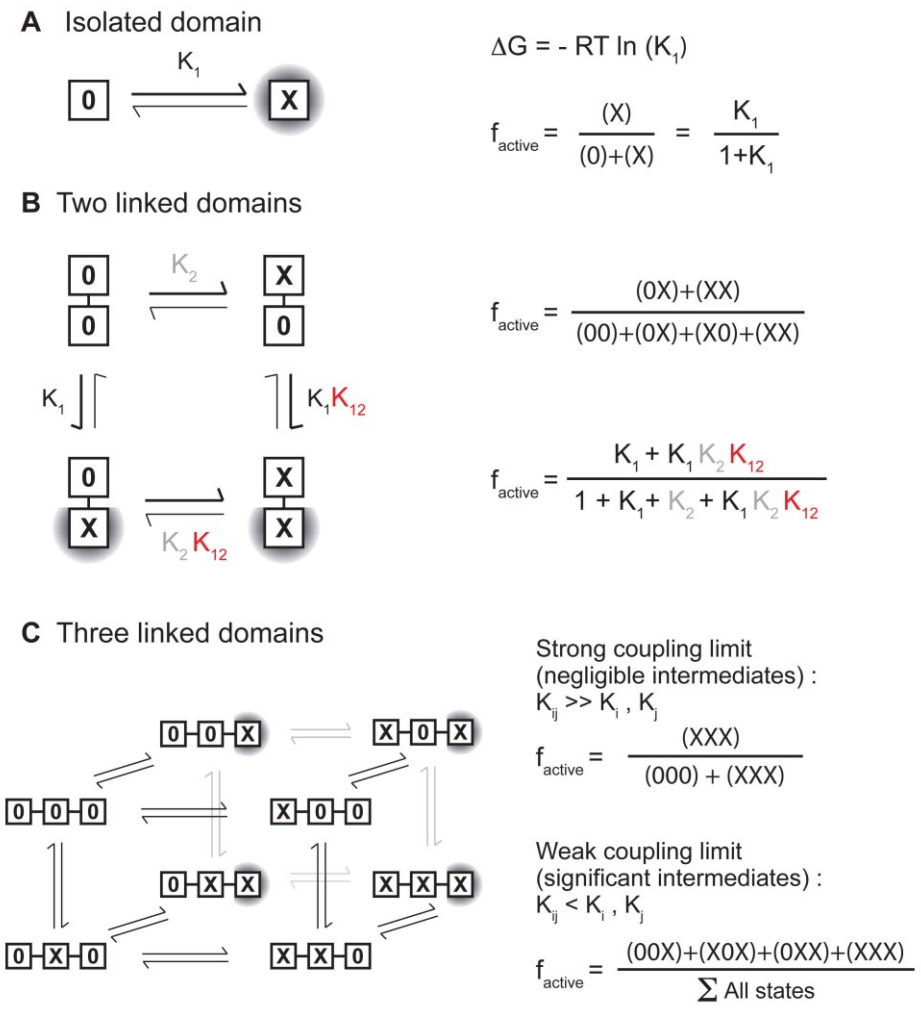


Figure 1.10 – *A thermodynamic framework for signal transduction.* **A)** shows a single domain that has two states, an active state (X) or an inactive state (0). The population of the active state is directly related to the energy difference between these two states i.e. the equilibrium constant. **B)** shows two thermodynamically coupled domains. The fraction of the active state depends on two equilibrium constants and a coupling energy. **C)** extends this formalism to three linked domains. In the limit of rigid, all-or-nothing coupling between the domains, the system reduces to a 2 state system but in the limit of weak coupling, multiple intermediates contribute to the active population.

active state dictates the maximal response of the protein to a stimulus and is dependent on the free energy difference (ΔG_1) between the two states and temperature. If this catalytic domain is linked to a transmembrane domain, which can also be either active (**X**) or inactive (**0**) (**Fig. 9B**), the system has 2^2 = 4 states and the population of the active catalytic unit is governed not only by it's own energy gap (ΔG_1), but also by the energy gap of the linked domain (ΔG_2) and the strength of the coupling (ΔG_{12}). Three such domains yield 2^3 = 8 states and so on. Formally, the

fractional population of a given state is the statistical weight of that state divided by the sum of the statistical weights for all the states, or the partition function. Each additional linked domain contributes an intrinsic energy gap, and a coupling energy to the partition function. These additional energetic terms can alter the basal population of the active kinase, thereby modulating the inducible response (Grigoryan et al., 2011).

Why is a thermodynamic coupling framework useful for HKs?

Thermodynamic linkage is particularly appropriate for multi-domain proteins like HKs because it allows each domain to access independent conformational states and dynamics. The entire protein need not utilize the same mode of motion in order to transmit a signal. As long as each domain has an energetic input and output, it will sense the state of a neighboring domain and populate conformational states accordingly.

This energetic effect of one domain on its neighbors is called a coupling energy. The molecular origin of coupling can be enthalpic (via the formation of new bonds or the removal of steric hindrance) or entropic (via increased dynamics or release of solvent). Indeed, it is not always necessary for a large $\Delta G_{coupling}$ to yield a large structural change or vice versa. In the strong coupling limit, the coupling interaction is larger than the intrinsic energy gap of a given domain, intermediates are sparsely populated and a multi-state system is effectively reduced to a 2-state (on-off) system (**Fig. 1.10.C**). Conversely, in the non-rigid or weak coupling limit, the effect of turning “on” one domain is to alter the equilibrium constant between the different states of the coupled domain. Weak thermodynamic coupling allows for the accrual of multiple domains capable of binding different ligands and populating intermediates along the way. For example VirA binds sugars via the periplasmic-binding protein, ChvE, as well as phenolics. Moreover, both VirA (Lin et al., 2014) and ChvE (Hu et al., 2013) are regulated by pH. PhoQ responds to divalent cations, protons and antimicrobials. In both cases each ligand can alter the signaling response (Bader et al., 2005; García Véscovi et al., 1996).

The same framework also explains why domains from different HKs can be combined to form functional chimeric proteins when the energetic input and output for each domain is

preserved (Utsumi et al., 1989). It can easily be extended to include additional domains, asymmetric conformations of individual domains, and ligand binding (Grigoryan et al., 2011). The challenge is to molecularly understand how proteins achieve weak-thermodynamic coupling and whether allosteric networks and mechanisms are conserved across the family (Kuriyan and Eisenberg, 2007).

1.12 Universal Themes in HK Signal Transduction

Recent structures of HKs yielded many general insights into bacterial signal transduction. Symmetric to asymmetric transitions within the homodimer, appear to be ubiquitous in many HK domains and also in their enzymological function. A similar connection between half-of-sites reactivity and asymmetry is seen in homodimeric receptor tyrosine kinases like EGFR (Leahy, 2010) and in chemoreceptors (Li and Hazelbauer, 2014), suggesting that this might be a general strategy for signal transduction. In HKs, asymmetric transitions are often encoded into structures via the fundamental properties of helical bundles such as their hydrophobic registry and side-chain packing propensities. Diagonal displacement of 4-helix bundles where pairs of helices slip in opposing directions is a general mode of helical-bundle malleability. This allows HKs to have shallow but a finely tuned energy landscapes such that small changes induced by stimuli, can propagate via thermodynamic coupling to yield a significant physiological response. The controlled reshuffling of the populations of different conformers as a consequence of thermodynamic coupling is at the heart of transmembrane signal transduction.

1.13 Acknowledgements

We thank Bruk Mensa and Nathan Schmidt for helpful discussions, and Thomas Lemmin for carefully reading the manuscript and providing feedback. The DeGrado lab acknowledges support from NIH GM54616 and AI074866, as well as support from the MRSEC program of NSF (DMR-1120901). M.P.B. was supported by the Jane Coffin Childs foundation.

1.14 Supplementary methods

Alignment and analysis of DHp and CA domain structures

Structures of DHp-CA domains were obtained from the PDB: [2C2A, 3EHF, 3EHH, 3EHJ, 3GIE, 3DGE, 3GIG, 4BIU, 4BIV, 4BIW, 4BIX, 4BIY, 4BIY, 4CB0, 4JAS, 4JAU, 4JAV, 4KP4, 4I5S, 4GCZ]. Crystal structures of HAMP-DHp constructs: [3ZCC, 3ZRX, 3ZRV] were also considered. In cases where the asymmetric unit contained more than one dimer (3EHF, 4BIU and 4BIY), each dimer was considered separately, yielding 25 different structures. The structures were compared by computing the pairwise Cα RMSD between all pairs of structures over a dimeric 7-residue moving window to find the region of highest structural similarity (lowest overall Cα RMSD). For each pair, chainA-A, chainB-B and chainA-B, chainB-A alignments were considered and the one that yielded the lower RMSD was used. Cα RMSDs were assessed using the pair_fit function in pymol. A coordinate system was defined based on the symmetric structure of HK853 (PDB: 2c2a). The biological unit of 2C2A was generated using the C2 crystallographic axis of symmetry that was also the dimer axis (x-axis in the molecular frame). Coordinates were transformed such that the dimer axis was along Z, and the Cα atoms of inward facing hydrophobic residue (corresponding to L 266 in 2C2A) were defined to be on the x-axis. All ensemble of all DHp structures with the best overall structural alignment was transformed into this coordinate system for further analysis. X,Y coordinates were extracted for all structures in the ensemble at positions that were multiples of 7 residues relative to L266. This ensured that all atoms considered had the same relative phase.

Alignment and analysis of HAMP structures

Structures HAMP domains were obtained from the PDB: [2L7H, 2L7I, 2LFR, 2LFS, 3Y0Q, 2Y20, 2Y21, 3ZRV, 3ZRW, 3ZRX, 4BIU, 4BIV, 4I3M, 4I44, 4CTI, 4GN0, 3ZCC, 3ZX6, 4I3M]. In cases where the asymmetric unit contained more than one dimer, each dimer was considered separately. Each HAMP in tandem repeats was also considered separately yielding 26 dimeric

HAMP structures. The first HAMP domain from the tandem repeat structure 4I44 was used as an anchor. This structure was transformed such that the dimer axis was coincident with the z-axis and the x-axis passed through the Cα atoms of the core position, I17. Each of the other HAMP domains was structurally aligned to 4I44 using the Cα-Cα pair_fit algorithm in pymol. Inter-dimer distances (Cα, chain A to Cα-chain B) were extracted at the top of the HAMP (7 residues upstream of I17) and the bottom of the HAMP (7 residues downstream of I17) and used for further analysis.

Alignment and analysis of sensor domain structures for Tar and DctB

Two pairs of structures: 3E4P, 3E4Q for DctB and 1LIH, 2LIG for Tar were structurally aligned using the pymol "super" algorithm. The structures were transformed so that the z-axis was normal to the membrane and bisecting the 2 p-helices of interest. For DctB the z=0 plane was set to be in the plane containing the CA atom of residue 75 and residues 58-75 on helix 1 analyzed. For Tar and the z=0 plane was set to contain the CA atom of residue 165 and residues 160-175 on helix 4 were analyzed. We also analyzed residues 42-55 on helix 1 of Tar, but did not include it in the figure because extensive work by Falke et al. showed that the pistoning is more significant in helix 4. Ligand-induced differences in coordinates of the aforementioned residues were analyzed; changes in the magnitude of the z-coordinate $\Delta z=\sqrt{(z_{apo}-z_{bound})^2}$ were measured to evaluate pistoning; and in the xy plane $\Delta xy=\sqrt{(x_{apo}-x_{bound})^2+\left(y_{apo}-y_{bound}\right)^2}$, to evaluate scissoring. A structural alignment for these two proteins suggested no change in helical phase.

1.15 References

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

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*, 436-448.

Albanesi, D., Martín, 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*, 16185-16190.

Ames, P., Zhou, Q., and Parkinson, J.S. (2008). Mutational analysis of the connector segment in the HAMP domain of Tsr, the Escherichia coli serine chemoreceptor. J Bacteriol *190*, 6676-6685.

Appleman, J.A., Chen, L.L., and Stewart, V. (2003). Probing conservation of HAMP linker structure and signal transduction mechanism through analysis of hybrid sensor kinases. J Bacteriol *185*, 4872-4882.

Ashenberg, O., Keating, A.E., and Laub, M.T. (2013). Helix bundle loops determine whether histidine kinases autophosphorylate in cis or in trans. J Mol Biol *425*, 1198-1209.

Bader, M.W., Sanowar, S., Daley, M.E., Schneider, A.R., Cho, U., Xu, W., Klevit, R.E., Le Moual, H., and Miller, S.I. (2005). Recognition of antimicrobial peptides by a bacterial sensor kinase. Cell *122*, 461-472.

Bem, A.E., Velikova, N., Pellicer, M.T., Baarlen, P., Marina, A., and Wells, J.M. (2015). Bacterial histidine kinases as novel antibacterial drug targets. ACS Chem Biol *10*, 213-224.

Bourret, R.B. (2010). Receiver domain structure and function in response regulator proteins. Curr Opin Microbiol *13*, 142-149.

Bourret, R.B., Thomas, S.A., Page, S.C., Creager-Allen, R.L., Moore, A.M., and Silversmith, R.E. (2010). Measurement of response regulator autodephosphorylation rates spanning six orders of magnitude. Methods Enzymol *471*, 89-114.

Buckstein, M.H., He, J., and Rubin, H. (2008). Characterization of nucleotide pools as a function of physiological state in Escherichia coli. J Bacteriol *190*, 718-726.

Butler, S.L., and Falke, J.J. (1998). Cysteine and Disulfide Scanning Reveals Two Amphiphilic Helices in the Linker Region of the Aspartate Chemoreceptor. Biochemistry *37*, 10746-10756.

Capra, E.J., and Laub, M.T. (2012). Evolution of two-component signal transduction systems. Annu Rev Microbiol *66*, 325-347.

Careaga, C.L., and Falke, J.J. (1992). Thermal motions of surface alpha-helices in the D-galactose chemosensory receptor. Detection by disulfide trapping. J Mol Biol *226*, 1219-1235.

Casino, P., Miguel-Romero, L., and Marina, A. (2014). Visualizing autophosphorylation in histidine kinases. Nature communications *5*, 3258.

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

Casino, P., Rubio, V., and Marina, A. (2010). The mechanism of signal transduction by two-component systems. Curr Opin Struct Biol *20*, 763-771.

Castelli, M.a.E., Véscovi, E.G.a., and Soncini, F.C. (2000). The Phosphatase Activity Is the Target for Mg2+Regulation of the Sensor Protein PhoQ in Salmonella. J Biol Chem *275*, 22948-22954.

Chervitz, S.A., and Falke, J.J. (1995). Lock on/off disulfides identify the transmembrane signaling helix of the aspartate receptor. J Biol Chem *270*, 24043-24053.

Chervitz, S.A., and Falke, J.J. (1996). Molecular mechanism of transmembrane signaling by the aspartate receptor: a model. Proc Natl Acad Sci U S A *93*, 2545-2550.

Chervitz, S.A., Lin, C.M., and Falke, J.J. (1995). Transmembrane signaling by the aspartate receptor: engineered disulfides reveal static regions of the subunit interface. Biochemistry *34*, 9722-9733.

Chetty, P.S., Nguyen, D., Nickel, M., Lund-Katz, S., Mayne, L., Englander, S.W., and Phillips, M.C. (2013). Comparison of apoA-I helical structure and stability in discoidal and spherical HDL particles by HX and mass spectrometry. J Lipid Res *54*, 1589-1597.

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

Cheung, J., and Hendrickson, W.A. (2009). Structural analysis of ligand stimulation of the histidine kinase NarX. Structure (London, England: 1993) *17*, 190-201.

Cheung, J., and Hendrickson, W.A. (2010). Sensor Domains of Two-Component Regulatory Systems. Curr Opin Microbiol *13*, 116-123.

Diensthuber, R.P., Bommer, M., Gleichmann, T., and Moglich, A. (2013a). Full-length structure of a sensor histidine kinase pinpoints coaxial coiled coils as signal transducers and modulators. Structure *21*, 1127-1136.

Diensthuber, R.P., Bommer, M., Gleichmann, T., and Möglich, A. (2013b). Full-Length Structure of a Sensor Histidine Kinase Pinpoints Coaxial Coiled Coils as Signal Transducers and Modulators. Structure *21*, 1127-1136.

Eguchi, Y., Ishii, E., Yamane, M., and Utsumi, R. (2012). The connector SafA interacts with the multi-sensing domain of PhoQ in Escherichia coli. Mol Microbiol *85*, 299-313.

Falke, J.J., and Erbse, A.H. (2009). The piston rises again. Structure (London, England: 1993) *17*, 1149-1151.

Falke, J.J., and Koshland, D.E. (1987). Global flexibility in a sensory receptor: a site-directed cross-linking approach. Science *237*, 1596-1600.

Ferris, H.U., Coles, M., Lupas, A.N., and Hartmann, M.D. (2014). Crystallographic snapshot of the Escherichia coli EnvZ histidine kinase in an active conformation. J Struct Biol *186*, 376-379.

Ferris, Hedda U., Dunin-Horkawicz, S., Hornig, N., Hulko, M., Martin, J., Schultz, Joachim E., Zeth, K., Lupas, Andrei N., and Coles, M. (2012). Mechanism of Regulation of Receptor Histidine Kinases. Structure *20*, 56-66.

Ferris, Hedda U., Dunin-Horkawicz, S., Mondéjar, Laura G., Hulko, M., Hantke, K., Martin, J., Schultz, Joachim E., Zeth, K., Lupas, Andrei N., and Coles, M. (2011). The Mechanisms of HAMP-Mediated Signaling in Transmembrane Receptors. Structure *19*, 378-385.

Fisher, S.L., Kim, S.K., Wanner, B.L., and Walsh, C.T. (1996). Kinetic comparison of the specificity of the vancomycin resistance VanSfor two response regulators, VanR and PhoB. Biochemistry *35*, 4732-4740.

Gao, R., and Lynn, D.G. (2005). Environmental pH sensing: resolving the VirA/VirG two-component system inputs for Agrobacterium pathogenesis. J Bacteriol *187*, 2182-2189.

Gao, R., and Stock, A.M. (2009). Biological insights from structures of two-component proteins. Annu Rev Microbiol *63*, 133-154.

García Véscovi, E., Soncini, F.C., and Groisman, E.A. (1996). Mg2+ as an extracellular signal: environmental regulation of Salmonella virulence. Cell *84*, 165-174.

Goldberg, S.D., Clinthorne, G.D., Goulian, M., and DeGrado, W.F. (2010). Transmembrane polar interactions are required for signaling in the Escherichia coli sensor kinase PhoQ. Proc Natl Acad Sci U S A *107*, 8141-8146.

Goldberg, S.D., Soto, C.S., Waldburger, C.D., and DeGrado, W.F. (2008). Determination of the Physiological Dimer Interface of the PhoQ Sensor Domain. J Mol Biol *379*, 656-665.

Grigoryan, G., Moore, D.T., and DeGrado, W.F. (2011). Transmembrane communication: general principles and lessons from the structure and function of the M2 proton channel, K(+) channels, and integrin receptors. Annu Rev Biochem *80*, 211-237.

Groisman, E.A. (2001). The pleiotropic two-component regulatory system PhoP-PhoQ, Vol 183 (Am Soc Microbiol).

Gutu, A.D., Wayne, K.J., Sham, L.T., and Winkler, M.E. (2010). Kinetic characterization of the WalRKSpn (VicRK) two-component system of Streptococcus pneumoniae: dependence of WalKSpn (VicK) phosphatase activity on its PAS domain. J Bacteriol *192*, 2346-2358.

Hall, B.A., Armitage, J.P., and Sansom, M.S. (2012). Mechanism of bacterial signal transduction revealed by molecular dynamics of Tsr dimers and trimers of dimers in lipid vesicles. PLoS Comput Biol *8*, e1002685.

Hamuro, Y., Coales, S.J., Molnar, K.S., Tuske, S.J., and Morrow, J.A. (2008). Specificity of immobilized porcine pepsin in H/D exchange compatible conditions. Rapid Commun Mass Spectrom *22*, 1041-1046.

Hazelbauer, G.L. (2012). Bacterial chemotaxis: the early years of molecular studies. Annu Rev Microbiol *66*, 285-303.

Hazelbauer, G.L., Falke, J.J., and Parkinson, J.S. (2008). Bacterial chemoreceptors: high-performance signaling in networked arrays. Trends Biochem Sci *33*, 9-19.

Hu, X., Zhao, J., DeGrado, W.F., and Binns, A.N. (2013). Agrobacterium tumefaciens recognizes its host environment using ChvE to bind diverse plant sugars as virulence signals. Proc Natl Acad Sci U S A *110*, 678-683.

Igo, M.M., Ninfa, A.J., Stock, J.B., and Silhavy, T.J. (1989). Phosphorylation and dephosphorylation of a bacterial transcriptional activator by a transmembrane receptor. Genes Dev *3*, 1725-1734.

Jiang, P., Peliska, J.A., and Ninfa, A.J. (2000). Asymmetry in the autophosphorylation of the two-component regulatory system transmitter protein nitrogen regulator II of Escherichia coli. Biochemistry *39*, 5057-5065.

Jung, K., Fried, L., Behr, S., and Heermann, R. (2012). Histidine kinases and response regulators in networks. Curr Opin Microbiol *15*, 118-124.

Kan, Z.Y., Mayne, L., Chetty, P.S., and Englander, S.W. (2011). ExMS: data analysis for HX-MS experiments. J Am Soc Mass Spectrom *22*, 1906-1915.

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

Kim, D., and Forst, S. (2001). Genomic analysis of the histidine kinase family in bacteria and archaea. Microbiology *147*, 1197-1212.

Kindrachuk, J., Paur, N., Reiman, C., Scruten, E., and Napper, S. (2007). The PhoQ-Activating Potential of Antimicrobial Peptides Contributes to Antimicrobial Efficacy and Is Predictive of the Induction of Bacterial Resistance? Antimicrob Agents Chemother *51*, 4374-4381.

Koshy, S.S., Eyles, S.J., Weis, R.M., and Thompson, L.K. (2013). Hydrogen exchange mass spectrometry of functional membrane-bound chemotaxis receptor complexes. Biochemistry *52*, 8833-8842.

Kuriyan, J., and Eisenberg, D. (2007). The origin of protein interactions and allostery in colocalization. Nature *450*, 983-990.

Laub, M.T., and Goulian, M. (2007). Specificity in two-component signal transduction pathways. Annu Rev Genet *41*, 121-145.

Leahy, D.J. (2010). The ins and outs of EGFR asymmetry. Cell *142*, 513-515.

Lemmin, T., Soto, C.S., Clinthorne, G., DeGrado, W.F., and Dal Peraro, M. (2013). Assembly of the Transmembrane Domain of E. coli PhoQ Histidine Kinase: Implications for Signal Transduction from Molecular Simulations. PLoS Comput Biol *9*, e1002878.

Lesley, J.A., and Waldburger, C.D. (2001). Comparison of the Pseudomonas aeruginosa and Escherichia coli PhoQ sensor domains: evidence for distinct mechanisms of signal detection. J Biol Chem *276*, 30827-30833.

Li, G.W., Burkhardt, D., Gross, C., and Weissman, J.S. (2014). Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. Cell *157*, 624-635.

Li, M., and Hazelbauer, G.L. (2014). Selective allosteric coupling in core chemotaxis signaling complexes. Proc Natl Acad Sci U S A *111*, 15940-15945.

Lin, Y.H., Pierce, B.D., Fang, F., Wise, A., Binns, A.N., and Lynn, D.G. (2014). Role of the VirA histidine autokinase of Agrobacterium tumefaciens in the initial steps of pathogenesis. Frontiers in plant science *5*, 195.

Lippa, A.M., and Goulian, M. (2012). Perturbation of the oxidizing environment of the periplasm stimulates the PhoQ/PhoP system in Escherichia coli. J Bacteriol *194*, 1457-1463.

Lowe, E.C., Basle, A., Czjzek, M., Firbank, S.J., and Bolam, D.N. (2012). A scissor blade-like closing mechanism implicated in transmembrane signaling in a Bacteroides hybrid two-component system. Proc Natl Acad Sci U S A *109*, 7298-7303.

Lv, Y., Xiao, J., Liu, Q., Wu, H., Zhang, Y., and Wang, Q. (2012). Systematic mutation analysis of two-component signal transduction systems reveals EsrA-EsrB and PhoP-PhoQ as the major virulence regulators in Edwardsiella tarda. Vet Microbiol *157*, 190-199.

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

Mascher, T., Helmann, J.D., and Unden, G. (2006). Stimulus Perception in Bacterial Signal-Transducing Histidine Kinases. Microbiol Mol Biol Rev *70*, 910-938.

Maslennikov, I., Klammt, C., Hwang, E., Kefala, G., Okamura, M., Esquivies, L., Mörs, K., Glaubitz, C., Kwiatkowski, W., Jeon, Y.H.*, et al.* (2010). Membrane domain structures of three classes of histidine kinase receptors by cell-free expression and rapid NMR analysis. Proc Natl Acad Sci U S A *107*, 10902-10907.

Matias, V.R., and Beveridge, T.J. (2006). Native cell wall organization shown by cryo-electron microscopy confirms the existence of a periplasmic space in Staphylococcus aureus. J Bacteriol *188*, 1011-1021.

Mechaly, A.E., Sassoon, N., Betton, J.M., and Alzari, P.M. (2014). Segmental helical motions and dynamical asymmetry modulate histidine kinase autophosphorylation. PLoS Biol *12*, e1001776.

Miller, S.I., Kukral, A.M., and Mekalanos, J.J. (1989). A two-component regulatory system (phoP phoQ) controls Salmonella typhimurium virulence. Proc Natl Acad Sci U S A *86*, 5054-5058.

Miller, S.I., and Mekalanos, J.J. (1990). Constitutive expression of the phoP regulon attenuates Salmonella virulence and survival within macrophages. J Bacteriol *172*, 2485-2490.

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

Molnar, K.S., Bonomi, M., Pellarin, R., Clinthorne, G.D., Gonzalez, G., Goldberg, S.D., Goulian, M., Sali, A., and DeGrado, W.F. (2014). Cys-scanning disulfide crosslinking and bayesian modeling probe the transmembrane signaling mechanism of the histidine kinase, PhoQ. Structure *22*, 1239-1251.

Moore, J.O., and Hendrickson, W.A. (2012). An Asymmetry-to-Symmetry Switch in Signal Transmission by the Histidine Kinase Receptor for TMAO. Structure *20*, 729-741.

Moukhametzianov, R., Klare, J.P., Efremov, R., Baeken, C., Goppner, A., Labahn, J., Engelhard, M., Buldt, G., and Gordeliy, V.I. (2006). Development of the signal in sensory rhodopsin and its transfer to the cognate transducer. Nature *440*, 115-119.

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*, 1095-1108.

Parkinson, J.S. (2010). Signaling mechanisms of HAMP domains in chemoreceptors and sensor kinases. Annu Rev Microbiol *64*, 101-122.

Perry, J., Koteva, K., and Wright, G. (2011). Receptor domains of two-component signal transduction systems. Molecular bioSystems *7*, 1388-1398.

Podgornaia, A.I., Casino, P., Marina, A., and Laub, M.T. (2013). Structural basis of a rationally rewired protein-protein interface critical to bacterial signaling. Structure *21*, 1636-1647.

Podgornaia, A.I., and Laub, M.T. (2015). Protein evolution. Pervasive degeneracy and epistasis in a protein-protein interface. Science *347*, 673-677.

Rabin, R.S., and Stewart, V. (1992). Either of two functionally redundant sensor proteins, NarX and NarQ, is sufficient for nitrate regulation in Escherichia coli K-12. Proc Natl Acad Sci U S A *89*, 8419-8423.

Ram, S., and Goulian, M. (2013). The architecture of a prototypical bacterial signaling circuit enables a single point mutation to confer novel network properties. PLoS Genet *9*, e1003706.

Russo, F.D., and Silhavy, T.J. (1991). EnvZ controls the concentration of phosphorylated OmpR to mediate osmoregulation of the porin genes. J Mol Biol *222*, 567-580.

Schultz, J.E., Kanchan, K., and Ziegler, M. (2015). Intraprotein signal transduction by HAMP domains: a balancing act. Int J Med Microbiol *305*, 243-251.

Schultz, J.E., and Natarajan, J. (2013). Regulated unfolding: a basic principle of intraprotein signaling in modular proteins. Trends Biochem Sci *38*, 538-545.

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

Skinner, J.J., Lim, W.K., Bedard, S., Black, B.E., and Englander, S.W. (2012a). Protein dynamics viewed by hydrogen exchange. Protein Sci *21*, 996-1005.

Skinner, J.J., Lim, W.K., Bedard, S., Black, B.E., and Englander, S.W. (2012b). Protein hydrogen exchange: testing current models. Protein Sci *21*, 987-995.

Stewart, V. (2014). The HAMP signal-conversion domain: static two-state or dynamic three-state? Mol Microbiol *91*, 853-857.

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*, 734-745.

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

Swain, K.E., and Falke, J.J. (2007). Structure of the conserved HAMP domain in an intact, membrane-bound chemoreceptor: a disulfide mapping study. Biochemistry *46*, 13684-13695.

Swain, K.E., Gonzalez, M.A., and Falke, J.J. (2009). Engineered socket study of signaling through a four-helix bundle: evidence for a yin-yang mechanism in the kinase control module of the aspartate receptor. Biochemistry *48*, 9266-9277.

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

Tews, I., Findeisen, F., Sinning, I., Schultz, A., Schultz, J.E., and Linder, J.U. (2005). The structure of a pH-sensing mycobacterial adenylyl cyclase holoenzyme. Science *308*, 1020-1023.

Tobe, T. (2008). The roles of two-component systems in virulence of pathogenic Escherichia coli and Shigella spp. Adv Exp Med Biol *631*, 189-199.

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*, 24892-24903.

Ulrich, L.E., and Zhulin, I.B. (2010). The MiST2 database: a comprehensive genomics resource on microbial signal transduction. Nucleic Acids Res *38*, D401-407.

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*, 1246-1249.

Walters, B.T., Mayne, L., Hinshaw, J.R., Sosnick, T.R., and Englander, S.W. (2013). Folding of a large protein at high structural resolution. Proc Natl Acad Sci U S A *110*, 18898-18903.

Wang, C., Sang, J., Wang, J., Su, M., Downey, J.S., Wu, Q., Wang, S., Cai, Y., Xu, X., and Wu, J. (2013). Mechanistic Insights Revealed by the Crystal Structure of a Histidine Kinase with Signal Transducer and Sensor Domains. PLoS Biol *11*, e1001493.

Wang, L.C., Morgan, L.K., Godakumbura, P., Kenney, L.J., and Anand, G.S. (2012). The inner membrane histidine kinase EnvZ senses osmolality via helix-coil transitions in the cytoplasm. EMBO J *31*, 2648-2659.

Winkler, K., Schultz, A., and Schultz, J.E. (2012). The S-helix determines the signal in a Tsr receptor/adenylyl cyclase reporter. J Biol Chem *287*, 15479-15488.

Yeo, W.-S., Zwir, I., Huang, H.V., Shin, D., Kato, A., and Groisman, E.A. (2012). Intrinsic negative feedback governs activation surge in two-component regulatory systems. Mol Cell *45*, 409-421.

Yu, E.W., and Koshland, D.E., Jr. (2001). Propagating conformational changes over long (and short) distances in proteins. Proc Natl Acad Sci U S A *98*, 9517-9520.

Chapter Two

2.1 Overview

Proposals concerning the mechanisms of bacterial protein signaling vary widely because of the multitude of systems studied as well as the methods employed to track movements during signaling. As mentioned in the introduction, previously proposed TCS signaling mechanisms such as piston shift or gearbox rotation are not necessarily mutually exclusive, and instead might represent one component of more complex structural transitions within a given domain. Also, one type of conformational change such as the pivoting of a helix might connect to a completely different structural change in an adjoining domain.

From all of the data available in several well-studied systems, we sought to determine whether a single mode of structural change most directly connected to signal propagation through the domains. To do this we took a two-pronged approach. First, we coupled cysteine-scanning data to Bayesian modeling in a canonical TCS, PhoQ, to model the various states of the protein. Next, we analyzed structural changes within existing crystal structures of HK domains believed to represent an on or off state. The results of the experimental data along with Bayesian modeling suggest that PhoQ moves between two states in a scissoring fashion, where the scissor motion is most prominent in the periplasmic domain, but less prominent in the TM and HAMP domains. The results of the PCA suggest that this scissoring motion is found to be prominent in all other TCS queried and magnitude of the scissoring displacement is more prominent than other displacement in the coordinate system.

This chapter describes a collaborative project, published in *Structure* (Molnar et al., 2014). My role was to carry out a quantitative comparison of structures of PhoQ and the chemoreceptor, Tar, as well as crystal structures that were published after the submission of our manuscript (requested from the reviewers of the paper). I also conducted the analysis of a number of discrepancies between the experimental data and our computational model for PhoQ's two conformational states. Cross-linking was observed in several regions that could not be

explained by either of the two resultant models. Several of the conflicts could be explained by my activity measurements, which showed that the mutations induced an inactive conformation (locked phosphatase) in the protein. A more functionally interesting series of conflicts was in the periplasmic dimer interface, residues 50-52, where we observed elevated levels of crosslinking at all three consecutive positions. Not only is this observation in conflict with the model, but also it is inconstant with the known helical structure (Cheung et al., 2008). As discussed in the paper below, we found that cross-linking depended markedly on the signaling state of the protein, changing in the presence and absence of Mg^{2+}. Finally, we extended this observation to evaluate the role of a negative regulator of PhoQ, MgrB. These very interesting observations are currently being pursued in the Goulian group, and ultimately will be analyzed using the Linked Equilibrating Domain model described in Chapter 3.

2.2 Cys-scanning Disulfide crosslinking and Bayesian modeling probe the transmembrane signaling mechanism of the histidine kinase, PhoQ

Published manuscript in Structure*, Cell Press journal, reproduced here under fair use guidelines*

Bacteria transduce signals across the membrane using two-component systems (TCSs), consisting of a membrane-spanning sensor histidine kinase and a cytoplasmic response regulator. In Gram negative bacteria, the PhoPQ TCS senses cations and antimicrobial peptides, yet little is known about the structural changes involved in transmembrane signaling. We construct a model of PhoQ signal transduction using Bayesian inference, based on disulfide crosslinking data and homologous crystal structures. The data are incompatible with a single conformation but are instead consistent with two interconverting structures. These states differ in membrane depth of the periplasmic acidic patch and the reciprocal displacement of diagonal helices along the dimer interface. Studies of multiple histidine kinases suggest this repacking might be a common mode of signal transduction in sensor His-kinase receptors. Since a similar scissors model has been ruled out in CheA-linked chemoreceptors, the new evidence suggests that sensor His-kinase and CheA-linked receptors possess different signaling mechanisms.

2.3 Introduction

The PhoQ sensor histidine kinase belongs to a family of two-component signal transduction systems, which dominate signaling across prokaryotic membranes (Stock et al., 2000). These systems respond to diverse environmental signals, such as pH (Gao and Lynn, 2005), small molecules (Kaspar et al., 1999; Lee et al., 1999), ions (García Véscovi et al., 1996), and peptides (Kindrachuk et al., 2007), and regulate critical responses, such as ion transport and virulence (Miller et al., 1989). A prototypical two-component system includes a transmembrane (TM) histidine kinase (HK) receptor and a cytoplasmic response regulator (Mascher et al., 2006). The periplasmic sensor responds to environmental signals by promoting

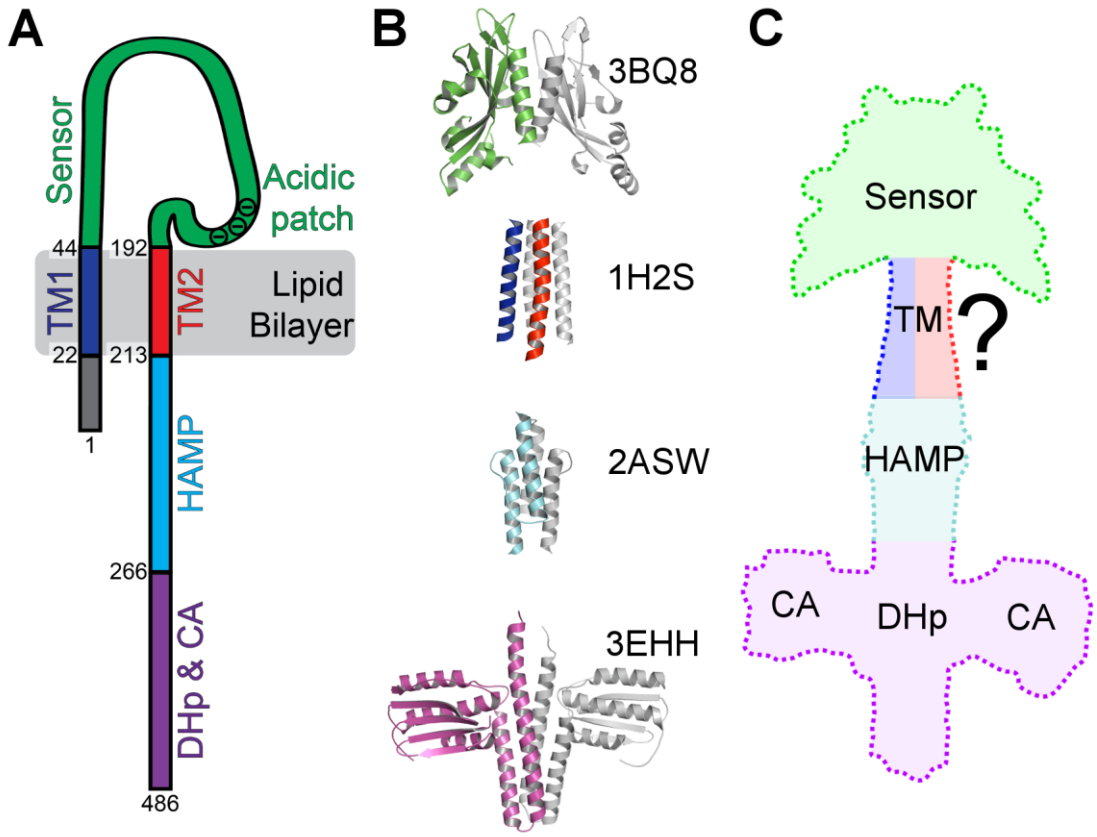


Figure 2.1 – Structural representations of PhoQ. (A) Schematic of the topology of a PhoQ monomer. The numbers indicate residue numbers for *E. coli* PhoQ (UniProt: P23837). (B) Crystal structures used for structural comparison of each domain of PhoQ. The corresponding PDB ID is listed next to the structure. One monomer is color-coded and the other monomer is in grey. (C) A model of the first three domains of PhoQ: sensor, transmembrane (TM), and HAMP domains. The dimerization and histidine phosphotransfer domain (DHp) and ATP-binding domain (ABD) are added for clarity but were not modeled.

autophosphorylation of a conserved histidine, followed by phosphotransfer to a conserved aspartate residue on its corresponding cytoplasmic response regulator. Phosphotransfer activates the response regulator, which in turn modulates genetic response (Robinson et al., 2000).

Although TCSs have been shown to be diverse (Dutta et al., 1999), the topology of a canonical HK receptor (**Fig. 2.1A**) consists of a periplasmic sensing domain flanked by two TM helices, followed by one or more small cytoplasmic domains. PhoQ has a HAMP domain (named for the domain's presence in *hi*stidine kinases, *a*denylyl cyclases, *m*ethyl-accepting chemotaxis proteins, and *p*hosphatases) (Galperin et al., 2001) followed by a kinase domain. The first part of this kinase domain is typically known as the *d*imerization and *h*istidine *p*hosphotransfer domain (DHp), which contains the substrate for autophosphorylation (a conserved histidine) and mediates phosphotransfer reactions. The second part is a catalytic, ATP-binding domain (CA) that is part of the family of highly conserved GHKL domains. A functional histidine kinase is homodimeric (**Fig. 2.1B**) with an extended dimer interface along the entire length of the molecule (Goldberg et al., 2008). Since TCSs frequently reuse these domains, mechanistic insights into PhoQ can inform general TCS signal transduction.

Structural efforts have attempted to elucidate the mechanistic details of signal transduction spanning several domains from the periplasmic sensors to the cytoplasmic DHp domain, and numerous structures have been reported. Crystal structures are now available for multiple domains of two-component and chemotaxis systems (Albanesi et al., 2009; Ferris et al., 2012; Lowe et al., 2012; Xie et al., 2010; Zhou et al., 2008), including a structure of the periplasmic sensor domain of PhoQ (Cheung et al., 2008). NMR and X-ray structures have been solved for HAMP domains as well as TM regions (Dunin-Horkawicz and Lupas, 2010; Royant et al., 2001). Many recent multi-domain crystal structures give us detailed view of the connections between cytoplasmic domains. A full length structure of an engineered, cytoplasmic two-component sensor (lacking a TM domain) was determined (Diensthuber et al., 2013), as was the structure of the cytoplasmic region of VicK, from *Streptococcus mutans* (Wang et al., 2013). Despite these advances, several competing proposals still remain for a mechanism of transmembrane signaling.

Early studies focused on the aspartate receptor, Tar, in *Salmonella typhimurium*. Although Tar is a member of the CheA-linked receptor class, and not a HK receptor, it shares several domains with TCS sensor kinases. Falke and coworkers demonstrated a swinging piston mechanism for signal transduction in this protein based on both disulfide-scanning and crystallographic studies (Chervitz and Falke, 1996). Multiple independent lines of evidence, many obtained using functional, full-length, membrane bound receptors in the working complex with CheA kinase, have supported the importance of the piston displacement of the signaling helix TM2 in transmembrane signal transduction and CheA kinase regulation (Chervitz and Falke, 1995; Chervitz et al., 1995; Falke and Hazelbauer, 2001; Hazelbauer, 2012; Hughson and Hazelbauer, 1996). These studies provide strong evidence that the subunit interface is largely static during on-off switching, thereby ruling out an early model (Milburn et al., 1991) proposing a scissors-type displacement of the two subunits in CheA-linked receptors. Other signal propagation models address the signaling role of the cytoplasmic HAMP domain that immediately adjoins the TM. The gearbox model is based on a series of NMR and crystallographic structures of the HAMP domain (Ferris et al., 2012; Hulko et al., 2006). Alternatively, a folding/unfolding transition has been proposed for signaling through the HAMP domain of CheA-linked receptors (Schultz and Natarajan, 2013; Zhou et al., 2009). In summary, the piston mechanism is well established for transmembrane signaling in CheA-linked receptors. However, current evidence does not rule out possible contributions of the helix tilt component of the swinging piston TM2 displacement (Chervitz and Falke, 1996) in signal transduction in this receptor class.

The hypothesis that a piston displacement could also play a role in HK receptors was supported by later structural comparisons of the TorS (Moore and Hendrickson, 2009, 2012) and NarX (Cheung and Hendrickson, 2009) HK signaling states. However, the measured displacements for TorS and NarX in the presence *versus* the absence of signaling ligands were small (< 1 Å) when compared to those seen in Tar. Other observed motions include inter-helical torqueing (Diensthuber et al., 2013), helix-bending (Wang et al., 2013), or DHp domain cracking (Dago et al., 2012). Another study posits a combination of these models (Casino et al., 2009).

Critical to a transmembrane signal transduction model is a structural model of the TM portions of a sensor HK. Three structures of monomeric HK TM domains were recently solved using NMR of isolated domains in micelles (Maslennikov et al., 2010). All three of the reported structures are limited in their utility for modeling a physiological dimeric interface, and without structural analyses from a HK TM domain, the structural starting point is not obvious. However, one crystal structure has been solved for the dimeric TM domain of a homologous protein, the HtrII sensory transducer (Gordeliy et al., 2002). A previous study utilized the HtrII X-ray structure as a model for the TM domain in HKs, and we have also reported similarities between the TM domains of HtrII and PhoQ (Goldberg et al., 2010). We demonstrated that the same pronounced water hemi-channel observed in HtrII plays an important mechanistic role within PhoQ.

Previously, we explored local changes in the TM domain by combining molecular dynamics simulations with disulfide crosslinking data (Lemmin et al., 2013). To elucidate larger scale changes across the membrane, we incorporate new crosslinking data in the HAMP and juxtamembrane regions of PhoQ with previous data and analyze it, using multi-state Bayesian modeling (Rieping et al., 2005). This approach provides the first investigation into the structures of the two signaling states of PhoQ, which interconvert through a motion in which opposing helices move towards or away from the bundle axis. Our subsequent quantitative structural analysis of additional receptor HK domains also divulge similar large and recurring motions of the helices relative to the central helical bundle axis. Scissoring motions account for a greater proportion of coordinate variation between HK receptor structures in distinct states than the swinging piston displacements observed in CheA-linked receptors (Chervitz and Falke, 1996), which signal in large hexagonal lattices (Liu et al., 2012). Thus it appears that sensor HKs and CheA-linked receptors may possess different signal transduction mechanisms.

2.4 Results

We probed the TM domain and the neighboring HAMP and periplasmic domains of PhoQ using disulfide-scanning mutagenesis. Building upon our previous analysis of the periplasmic helix at the dimer interface (Goldberg et al., 2008), new single cysteine residue mutations were

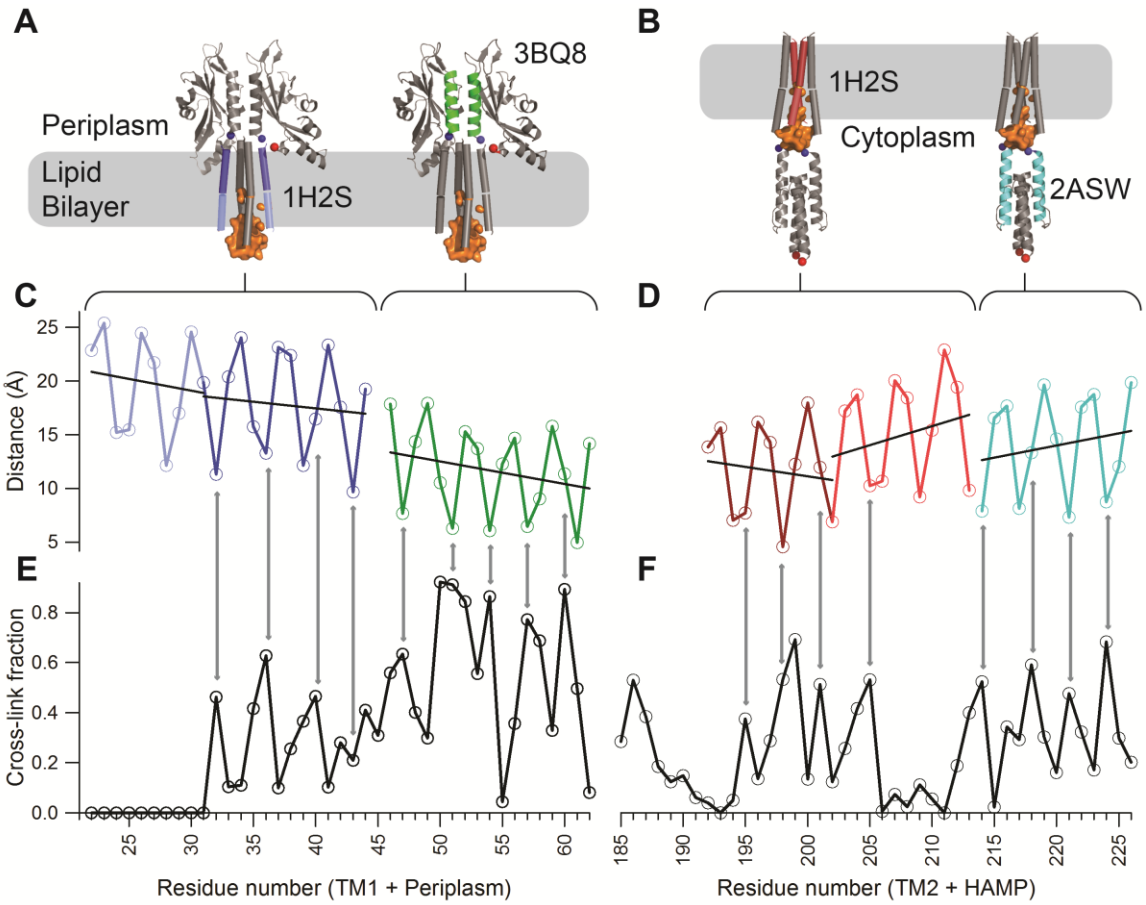


Figure 2.2 – Comparison of the crosslinking efficiency with structural models. (A) PhoQ TM1-periplasm-TM2 model in a lipid bilayer. The color-coded helical regions (blue-green-red, respectively) indicate where cysteine mutations were made. An orange envelope marks the water hemichannel. (B) PhoQ TM1-TM2-HAMP model in a lipid bilayer. Color-coding (blue-red-cyan, respectively) is applied to the regions probed by cysteine mutations. The water hemichannel is shown as in (A). (C) Inter-monomer distances between the dimeric structures of structural models for TM1-periplasm. The first TM helix is modeled from HtrII (PDB ID: 1H2S) and the periplasmic helix is from *E. coli* PhoQ (PDB ID: 3BQ8). The measured distances are between Cβ-Cβ' of corresponding residues (or Cα-Cα' for glycine). Black lines indicate linear fits to each helical segment. (D) Inter-monomer distances between dimeric structures of structural models for TM2-HAMP. The second TM helix is from HtrII, and the HAMP helix is from *Archaeoglobus fulgidus* (PDB ID: 2ASW). Distances and fits were done as in (C). (E) Crosslinking data from the full length PhoQ protein in a native membrane for cysteine mutants 22 through 61. (F) Crosslinking data from the full-length PhoQ protein in a native membrane for cysteine mutants 192 to 226. Western blot data from sections (E) and (F) are shown in **Fig. S2.1**.

introduced along the TM helices and at selected positions within the HAMP domain (**Fig. 2.2A** and **2.2B**). Without the oxidizing environment of the periplasm, measuring the extent of disulfide bond formation in these mutants required the presence of an oxidative catalyst, Cu(II)(1,10-phenanthroline)$_3$ (CuPhen). For each residue in the predicted TM domain, we calculated the

fraction of crosslinking from the measured intensities of covalent dimer and monomer bands on a Western blot.

Comparison of disulfide crosslinking efficiency to homologous crystal structures. The crosslinking efficiency depends on collisional dynamics of sulfhydryl groups, relative orientations of Cys side chains, and their accessibility to oxidants. This dependence leads to a roughly inverse relationship between crosslinking efficiency and the distance between the reacting thiol groups (Careaga and Falke, 1992; Hughson and Hazelbauer, 1996; Metcalf et al., 2009), so we compared the measured crosslinking efficiency for all three domains against their individual or homologous structures: periplasmic crosslinking data to the crystal structure of the PhoQ periplasmic sensor; TM crosslinking data to the TM structure of HtrII; and cytoplasmic crosslinking data on the HAMP structure of Af1503 from *Archaeoglobus fulgidus* (**Fig. 2.1B**). These comparisons test how faithfully these individual domains represent the full-length structure of PhoQ (**Fig. 2.1C**). Importantly, the crosslinking data also adds structural insight by spanning the intact juxtamembrane regions, and suggest how the single domain structures connect.

For the periplasmic linker that connects TM1 to the periplasmic helix (res. 42-49), the crosslinking efficiency maintains a sinusoidal variation with a consistent phase (**Fig. S2.1A** and **S2.1B**), suggesting an uninterrupted helix. The similarity in the phases can be seen qualitatively (**Fig. 2.2E**), but we also fit this data to a sinusoidal function (**Fig. S2.1A**). The fit deviates a bit near residue 43, which could reflect a kinking of the helix as it leaves the membrane.

The cytoplasmic linker that connects TM2 to the HAMP (res.207-214) is a region of reduced crosslinking and likely corresponds to a divergent portion of the bundle. A conserved Pro residue at position 208 can cause a bend in the helix in this region (Lemmin et al., 2013; Yohannan et al., 2004). The sinusoidal fit to the crosslinking data of TM2 are out of phase with the fit for the HAMP helix (**Fig. S2.1C** and **S2.1D**), suggesting this linker can access a distorted helical or non-helical geometry.

To check how faithful a model HtrII is for the PhoQ TM bundle, we compared the inter-residue distances to corresponding experimental crosslinking data (**Fig. 2.2**). The topology of the HtrII bundle is like other HKs, where the helices are anti-parallel and the N-terminus of the first TM is

towards the cytoplasm (**Fig. 2.1A**). The crosslinking fractions agree qualitatively with bundle being well packed near the periplasm, but diverges slightly near the cytoplasm. At the periplasmic side (**Fig. 2.2**, dark blue and dark red), we observe a periodic pattern of crosslinking efficiency, close to that of an ideal α-helix, with a period of 3.6 residues. Fitting a sinusoidal function to the data resulted in a period of 3.5 residues for TM2 and 3.7 residues for TM1 (**Table 2.1**). We also computed a phase offset to determine if we achieve the expected inverse relationship between crosslinking and distance variation for an alpha helix (180°).

Table 2.1. Least-squares fitting of a sinusoidal function to the crosslinking efficiency of PhoQ and the inter-residue distances of HtrII and Af1503 crystal structures.

Helix	Period[1]	Phase offset[2]
PhoQ TM1	3.67 ± 0.13	173°
HtrII TM1	3.69 ± 0.03	
PhoQ TM2	3.53 ± 0.30	168°
HTrII TM2	3.67 ± 0.08	
PhoQ HAMP	3.53 ± 0.20	153°
AF1503 HAMP	3.54 ± 0.02	

[1] Number of residues per repeat
[2] Differences in phase for the fitted sinusoidal waves between the experimental crosslinking data and the inter-monomer distance data (Cβ-Cβ' distance or Cα-Cα' for Gly) taken from corresponding crystal structure

To check how faithful a model HtrII is for the PhoQ TM bundle, we compared the inter-residue distances to corresponding experimental crosslinking data (**Fig. 2.2**). The topology of the HtrII bundle is like other HKs, where the helices are anti-parallel and the N-terminus of the first TM is towards the cytoplasm (**Fig. 2.1A**). The crosslinking fractions agree qualitatively with bundle being well packed near the periplasm, but diverges slightly near the cytoplasm. At the periplasmic side (**Fig. 2.2**, dark blue and dark red), we observe a periodic pattern of crosslinking efficiency, close to that of an ideal α-helix, with a period of 3.6 residues. Fitting a sinusoidal function to the data resulted in a period of 3.5 residues for TM2 and 3.7 residues for TM1 (**Table 2.1**). We also computed a phase offset to determine if we achieve the expected inverse relationship between crosslinking and distance variation for an alpha helix (180°).

At the cytoplasmic end of the TM bundle, there was little crosslinking observed (**Fig. 2.2**, light blue and light red). This low degree of crosslinking near the cytoplasmic side of the bundle agrees with the presence of a water hemichannel, shown as solvent accessible surface in orange in **Fig. 2.2**. However, the complete lack of crosslinking on the cytoplasmic side of PhoQ TM1 helices suggests a larger separation in the PhoQ hemi-channel compared to that in the HtrII structure. At

the periplasmic side of the TM bundle, the TM1 helices crosslink as strongly as the TM2 helices, despite the shorter helical distance in the HtrII TM2 helices. Taken together, these data indicate that HtrII is only an approximate model for the TM domain of PhoQ.

Multi-state Bayesian modeling. We collected data using the full-length PhoQ protein in a native membrane, which was free to structurally fluctuate between signal transduction states. Therefore, we do not assume that all crosslinking experiments necessarily probe a single structural state. For example, one structural state cannot explain both high TM1-TM1' (residues 32-45) as well as high TM2-TM2' crosslinking (residues 192-206) (**Fig. 2.2E** and **2.2F**) without introducing steric clashes. Consequently, we hypothesize the presence of multiple, distinct structural states in the sample. We used a multi-state Bayesian modeling of cysteine crosslinking data, which simultaneously models several structures based on experimental and prior information (such as available structural information), and infers additional parameters (*e.g.*, uncertainty in the data, σ_0, and population fractions, w_i). The Bayesian approach (Habeck et al., 2006; Rieping et al., 2005) estimates the probability that a single model or set of models explains the available information about a system (see Material and Methods).

We divided the PhoQ dimer into 6 rigid bodies for each monomer, for a total of 12 rigid bodies in the dimer (Materials and Methods). A coarse-grained representation of PhoQ was used, in which each residue is modeled as a bead centered on the Cα atom. The conformations of the dimer were explored without imposing any symmetry between the two chains, using a Gibbs sampling scheme relying on a Monte Carlo algorithm enhanced by Replica Exchange (Rieping et al., 2005). The sampled models were clustered based on the predicted cross-linked fractions. Thus, members of the same cluster predict similar data, although they might be structurally different, especially in regions that are not restrained by the data (**Fig. 2.3**).

We focused the structural analysis on the most populated cluster, which corresponds to the peak with the greatest probability in the posterior probability distribution of states, given the cross-linking data and domain models. Cluster representatives and predicted cross-linked fractions for all clusters with a population greater than 3% are reported in **Table S2.1**. To predict the minimal

number of states that best explain the crosslinking data, the Bayesian approach was applied independently for 1, 2, and 3 states.

1-state modeling. The cluster analysis of the sampled models (**Table S2.1**) revealed that the experimental data could not be fully explained by a single structure. The 1-state model was in good agreement with the predicted cross-linked fractions in the periplasmic side of TM1 (residues 13 to 45) and cytoplasmic side of TM2 (residue 205 to 215). However, the model does not match a large region of data with a high cross-linked fraction – the periplasmic side of TM2 (residues 195 to 205). A single structure cannot simultaneously reconcile high crosslinking on the periplasmic side of both TM1 *and* TM2. Instead, for the TM2 periplasmic region, the model-predicted cross-linked fractions equal to zero. Therefore, in the 1-state model, proximity between the periplasmic region of TM2 and TM2' is not observed due to steric exclusion by the TM1 and TM1' helices.

2-state modeling. The most populated cluster of two states found by 2-state modeling approach explained crosslinking data better than 1-state modeling, as shown by the lower likelihood score (**Table S2.1**) and the improved agreement between the model and the data for the periplasmic TM2 region and surrounding residues (185-205) (**Fig. 2.3**). By postulating the existence of a mixture of states, the 2-state model explains the presence of conflicting crosslinking data observed within the periplasmic side of the TM domain. The inferred population fractions of State-1 and State-2 were 40.5% and 59.5%, respectively. The two computed structures deviate from symmetry, although it is not possible to determine whether the deviations are significant given the precision of the ensemble of models. Overall, the two models differ at the dimeric interface in the arrangement of the helices from each domain. We assessed the robustness of the two-state modeling Bayesian approach by data jackknifing (**Fig. S2.2**).

For the periplasmic region, State-2 resembles the crystallographic structure of the PhoQ sensor domain, 3BQ8 (Cα RMSD = 3 Å), previously proposed to correspond to the activated state (Cheung et al., 2008). In contrast, in State-1 the periplasmic helices are closer to a parallel configuration. The periplasmic helices transition between a parallel (State-1) and a crossing configuration (State-2); this transition corresponds to a scissoring motion. A consequence of the

scissoring motion is a displacement of the periplasmic acidic patch (residues 145-154) from resting on the surface of the membrane in State-1 to a position deeper in the membrane in State-2. Also, the 2-state model suggests that the reorientation of the periplasmic domain due to the scissoring motion propagates to remodel the TM and HAMP helical bundles.

The scissoring motion of the periplasmic, interfacial dimer helices induces a different type of structural change to the TM domain. This motion is best seen on the periplasmic side of the TM bundle (top down view of helical bundle in **Fig. 2.3B** and **2.3E**), where the diagonal pairs of helices take turns displacing each other. State-1 predicts that the TM1 and TM1' helices (blue) pack close and displace the TM2 and TM2' helices (red), while in State-2, the TM2 and TM2' helices move towards the center of the bundle and displace the TM1-TM1' intersubunit helical contacts. More specifically, all the helices in the bundle undergo a collective rearrangement, because the movement towards the bundle center of one helix pair accompanies the outward displacement of the other pair.

The large changes seen in the TM domains are coupled with smaller changes in the HAMP domains. Specifically, the diagonal displacement seen in the TM domain is also observed for the HAMP helices. In State-1, the helix 1-helix 1' distance is shorter than the helix 2-helix 2' distance near the N-terminal end of the bundle; this relationship reverses in State-2 (**Fig. 2.3C** and **2.3F**). Presumably, this conformational change is coupled to additional, previously characterized changes in the catalytic and DHp domains (Albanesi et al., 2009; Ferris et al., 2012).

3-state modeling. Models in the most populated clusters in 3-state modeling do not improve the fit relative to the 2-state model, as indicated by the average and best likelihood scores for the clusters (**Table S2.1**). The previously identified models were not found here because we imposed a lower bound of 0.2 on the individual population fractions w_i (Materials and Methods).

Selecting the best model. The 2-state model fits the data significantly better than the 1-state model, while the 3-state model does not improve the fit. Therefore, the 2-state model is the most parsimonious explanation of the data.

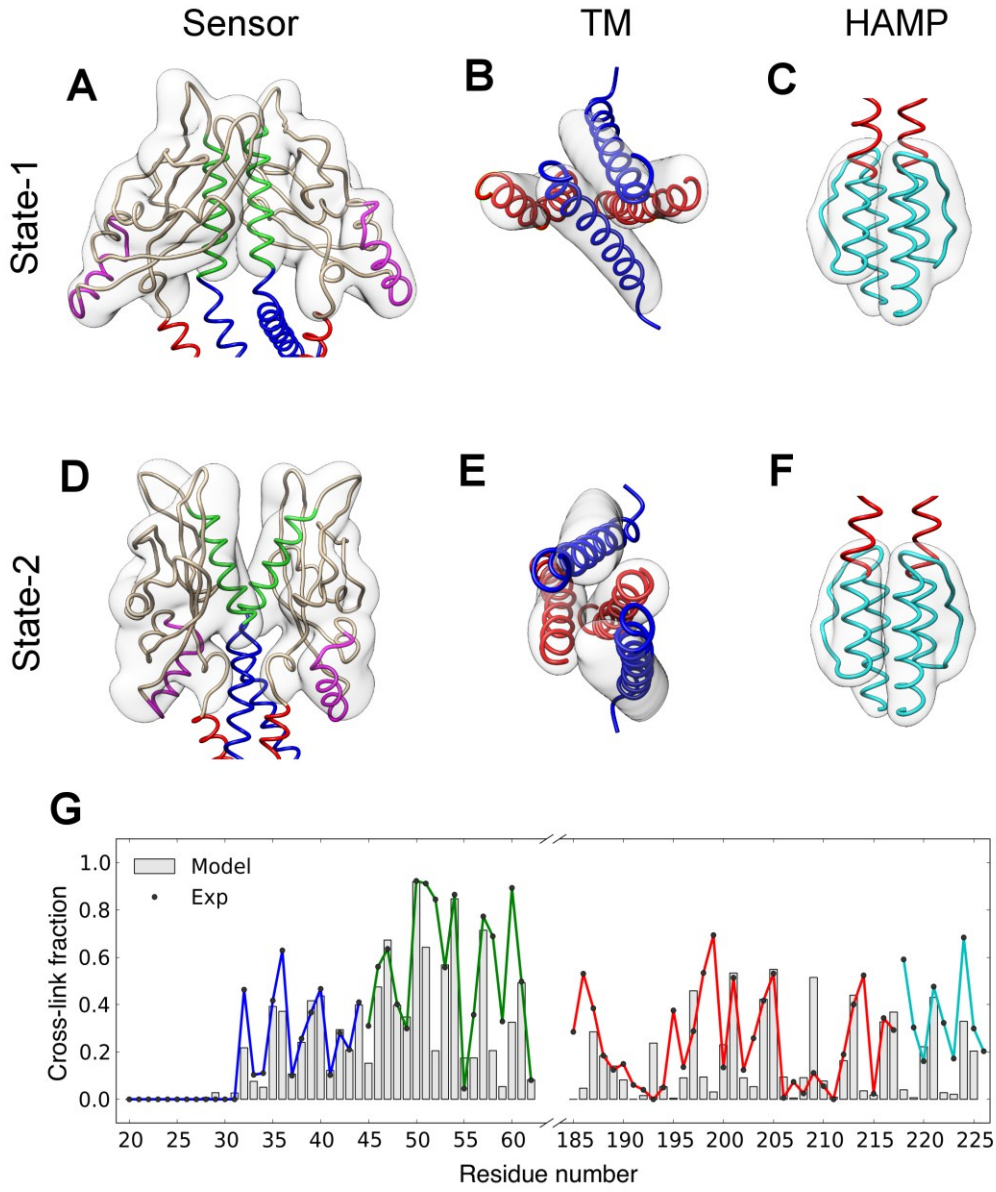


Figure 2.3 – Analysis of the most populated cluster found in 2-state modeling. Backbone ribbon representation of the cluster representative of: (A) sensor domain in State-1; (B) TM domain in State-1 as viewed from the sensor domain looking into the cytoplasm; (C) HAMP domain in State-1; (D) sensor domain in State-2; (E) TM domain in State-2, viewed from periplasm looking into the cytoplasm; (F) HAMP domain in State-2. The cluster structural variability is represented by the transparent density volumes calculated using the VMD VolMap tool (Humphrey et al., 1996) . The color-coding for (A-F) is as follows: periplasmic sensor helices (residues 45 to 61) are in green, the Mg^{2+}-binding, acidic patch (residues 137 to 150) in magenta; the TM1 and TM2 helices of the TM domain are in blue and red, respectively; and the HAMP domain in cyan. (G) Overlay of model data, predicted by the highest likelihood model of the cluster (grey bars), and experimental cross-linked fractions, color-coded by the domain definition above. Model robustness assessment summarized in **Fig. S2.2**.

Functional measurements of Cys mutants explain most of the deviations between crosslinking data and the 2-state model. While the 2-state model best fits the experimental observations, a few data points still could not be explained. In particular, isolated deviations were observed at residues 52, 195, 199, 208, 209, and 218 (**Fig. 2.3G**). These discrepancies can originate from computational inaccuracies of the Bayesian model (including the forward model, noise model, sampling, and the assumed number of states) or the representation of the system. Alternatively, the Cys mutations could induce a non-native structure with disruption of function. To discriminate between these possibilities, we investigated the phenotypes of the cysteine mutants, by measuring transcriptional activity at low and high Mg^{2+} concentration (**Fig. 2.4C** and **Fig. S2.3**), as described previously (Soncini et al., 1996).

Three of the outliers were associated with mutations that greatly decreased the function of PhoQ. The mutants P208C and L209C have low β-galactosidase activity at both high and low concentrations of Mg^{2+}. By contrast, the wild-type protein activity changes 2 to 5 fold between these Mg^{2+} concentrations. Interestingly, a kink in TM2 occurs between P208 and L209 in an MD model of the TM domain of PhoQ (Lemmin et al., 2013), suggesting this region is a fulcrum of movement. It is possible that these cysteine mutations abolish signal transduction by tampering with the helical kink. Similarly, mutant L218C, which lies at a connecting loop between the periplasmic and the TM domains, shows low activity at both low and high Mg^{2+} concentrations.

However, the deviation for the final outlier, F195C, could not be explained by a structural disruption that results in loss of function. This mutant resides in the TM2 helix where the helical period between experimental and model data is shifted, indicating a potential helical rotation or bending in that region (residues 195 to 199). TM2 was modeled as a rigid body extending from 194-205, but the discrepancy at F195C suggests that two rigid bodies or a flexible chain might be more appropriate representations for this helix.

Residue 52 has robust activity, but does not agree with the 2-state model; it is also the first of three consecutive residues that shows high crosslinking (**Fig. 2.4B**, green curve). High crosslinking at three consecutive residues is inconsistent with the known α-helical structure in this region of the protein (Cheung et al., 2008), as it would require residues on both sides of a helix to

crosslink to a neighbor efficiently. This discrepancy encouraged us to repeat the previously published crosslinking experiments for a portion of the periplasmic helix at the dimer interface. In this region, disulfide crosslinks occur spontaneously and do not require the aid of an oxidant like CuPhen, as is required for HAMP and TM domains. The previous periplasmic crosslinking experiments (Goldberg et al., 2008) used long, overnight incubations in LB media. However, when we incubated for shorter periods of time (to avoid spurious crosslinks) in minimal media (for precise control of Mg^{2+} concentration), we found that residue 50 does not crosslink nearly as much as residue 51, and the extent of crosslinking at 52 was dependent on the Mg^{2+} concentration (**Fig. 2.4B**). Activity measurements also showed that, although the mutants had

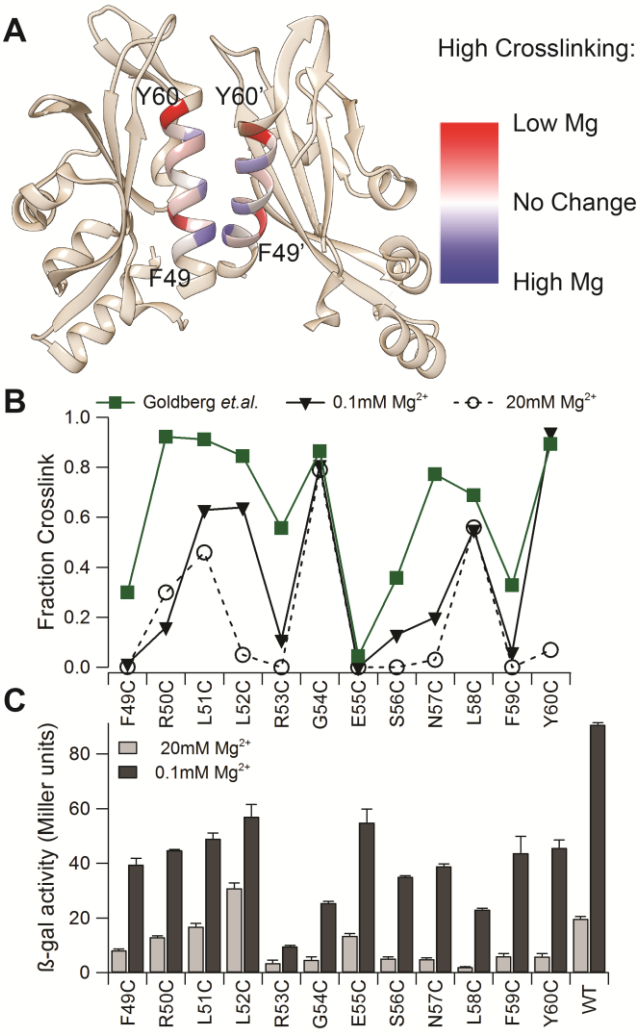


Figure 2.4 – Change in crosslink fraction for the periplasmic helix of PhoQ at low and high [Mg^{2+}].
(A) X-ray structure of the periplasmic domain of PhoQ (3BQ8) with the dimer interfacing helices (residues 49-60) colored according to change in crosslink fraction. Red represents higher crosslinking at low Mg^{2+} conditions, white indicates no Mg-dependent change, and blue represents higher crosslinking under high Mg^{2+} (key at right).
(B) Crosslinking data from the full-length PhoQ protein in a native membrane for cysteine mutants 49 to 60 under both high and low [Mg^{2+}] conditions at mid-log phase growth. The green curve is data reproduced from (Goldberg et al., 2008).
(C) Activity response of β-galactosidase reporter in whole cells for the mutants involved (data for extended region in **Fig. S2.3**).

somewhat lower transcriptional activity than WT, they were still responsive to changes in the Mg^{2+} concentration. The reduced crosslinking at position 50 improves agreement between the crosslinking data and an ideal helical period.

An unexpected finding of this crosslinking experiment is that the extent of crosslinking in the first periplasmic helix depended very markedly on the concentration of Mg^{2+}, particularly at residues 52 and 60 (**Fig. 2.4**). This finding is in contrast with the effect of activating ligands on crosslinking of Tar, in which there is very little change in crosslinking in the first TM and periplasmic helices (Chervitz and Falke, 1995; Chervitz et al., 1995; Pakula and Simon, 1992). Instead, Tar helix α1 and α1' interact to form a static homodimeric core that supports a piston-shift of periplasmic α4 and TM α2 during signaling.

In summary, Bayesian modeling helped us rationalize flaws originating from artifactual disulfide formation (residue 50), inactive constructs (residues 208, 209, 218) and a potential representation inaccuracy (residue 195). In traditional modeling, these points would be considered as outliers and removed from the data set. In the Bayesian framework, such a manual intervention is not necessary because an uncertainty parameter is associated to each data point, thus allowing those points that are not consistent with the bulk of the data to be properly down-weighted in the construction of the model. In contrast to traditional modeling, the 2-state model motivated additional functional experiments to explain the large differences between the observed and predicted data.

Structural variation between signaling states. The 2-state model proposes conformational changes between State-1 and State-2 in which the helices show larger displacements (relative to the helical bundle axis) than the anticipated motions from either the swinging piston or gearbox models. To test whether or not these diagonal bundle displacements are unique to PhoQ, we quantified the structural variability among known structures of other two-component HK domains. We selected several homodimeric 4-helix bundles where each dimer contributes two helices and required that the domains be crystallized as physiologically relevant dimers. These states represent distinct signaling states, either by virtue of having a bound signaling ligand or a mutation that modulated the degree of activation (**Table 2.2**). We examined the structures used to

propose the piston shift found in the aspartate sensor, Tar (Chervitz and Falke, 1996), and the gearbox model proposed based on HAMP structures including a HAMP(Af1503)/DHp(EnvZ) chimera (Ferris et al., 2012). We also quantified the structural variability seen in two crystallographically determined signaling states of the sensor domain of TorS (Moore and Hendrickson, 2009) as well as DesK DHp structures believed to represent different signaling states (Albanesi et al., 2009).

We describe the changes in helix orientation by relying on six degrees of freedom that define a convenient coordinate system, previously used to analyze four-helix bundles (Lombardi et al., 2000; Summa et al., 1999), as described in **Fig. S2.4**. Two degrees of freedom match previous signaling models; a translational motion parallel to the bundle axis ("height"=z) corresponds to the piston model, and a rotation about the helix axis ("helix phase"$=\psi$) corresponds to the gearbox model. The remaining four degrees of freedom are helix tilt towards the bundle axis ("towards tilt"=φ_1), helix tilt perpendicular to the "towards tilt" ("sideways tilt"=φ_2), radial displacement from the central bundle axis ("radius"=r), and global rotation of individual helices relative to their neighbors around the central bundle axis ("bundle phase"=θ) (**Fig. S2.4**). Of these degrees of freedom, radial displacements (r), "towards tilt" (φ_1), and "sideways tilt" (φ_2) define lateral displacements of individual helices, which might apply tension that alters the conformation/energy landscape of neighboring domains.

The position of each helix in the four-helix bundles was analyzed using this parameterization. For each helix in the four-helix bundle we then measured the contribution of the six parameters to the coordinate variation between two states believed to comprise different signaling states of the domains. The calculated values for the parameters are listed in **Table S2.3**, and the displacements along these modes are listed in **Table 2.2**.

The computed changes for the aspartate sensor domain agree with the analysis of Falke (Chervitz and Falke, 1996), who documented a downward shift (z) along with a "swinging" (φ_2) piston shift motion that was localized to a single helix. We also examined structural changes in TorS. Although an early report (Moore and Hendrickson, 2009) suggested that this protein signaled via a piston motion, subsequent crystallographic investigations of TorS complexed with

Table 2.2. Coordinate displacements between 4-helix bundle crystal structures

		PDB IDs	Chain	Helix		Displacements in Å					
						$\Delta\psi$ [a]	$\Delta\varphi_1$ [b]	$\Delta\varphi_2$ [b]	Δr	Δz	$\Delta\theta$ [c]
Sensor Domains	Tar	**2LIG-1LIH**	A	1	(42-57)	-0.18	0.37	0.48	0.24	-0.12	-0.20
		Asp bound	B	1	(42-57)	-0.09	-0.20	0.12	0.09	-0.25	-0.15
		to apo	A	2	(155-174)	0.09	0.06	1.11	0.30	1.06	0.14
			B	2	(155-174)	0.00	0.07	0.07	0.10	0.24	0.09
	TorS	**3O1H-3O1I**	A	1	(52-67)	-0.10	-0.04	0.38	-0.71	0.21	-0.71
		TorT/TMAO bound	B	1	(52-67)	0.38	0.04	-0.14	0.83	-0.14	-0.38
		to apo TorT	A	2	(300-317)	0.40	0.72	0.55	-0.19	0.05	0.28
			B	2	(300-317)	-0.22	0.22	0.74	-0.22	0.25	0.61
HAMP	AF1503	**3ZRX-3ZRV**	A	1	(280-297)	-0.30	-0.24	0.11	0.58	-0.19	0.07
		WT to A291F	B	1	(280-297)	-0.03	-0.49	-0.24	0.11	0.05	-0.53
			A	2	(310-328)	0.83	1.95	1.11	-1.73	-0.73	0.32
			B	2	(310-328)	-0.08	1.15	-0.04	-0.75	0.25	0.09
DHp Domains	EnvZ	**3ZRX-3ZRV**	A	1	(333-345)	1.39	-1.50	1.51	-1.60	-0.06	-0.90
		WT to A291F	B	1	(333-345)	0.92	-2.06	2.19	-1.17	0.11	-1.20
			A	2	(373-385)	-0.23	-0.79	1.13	0.90	-0.01	0.65
			B	2	(373-385)	0.63	-1.51	2.34	0.57	0.51	0.70
	DesK	**3EHF-3GIG**	A	1	(182-198)	-0.19	1.65	-1.06	0.51	-0.02	0.41
		WT to WT	B	1	(182-198)	-0.10	-0.18	-0.60	-0.74	-0.03	-0.06
		+ADP +AMPPCP	A	2	(224-238)	-0.18	1.36	1.09	-0.70	0.44	1.12
			B	2	(224-238)	0.36	2.47	1.81	-1.23	-0.14	1.67

Each helix was parameterized as described in Lombardi *et al*, 2000. For each domain, we analyzed pairs of helical bundles believed to correspond to different signaling states, and compare the differences in the computed parameters associated with helical phase, tilts, rotation, and displacements, and translations. Coordinate displacement associated with angular displacements was determined from the following equations:

[a] $4*\sin(\Delta\psi)$ [b] $1.5*(\text{"helix length"})/2*\sin(\Delta\varphi)$ [c] $(\text{"average radius"})*\sin(\Delta\theta)$

Cell color indicates the largest (**red** or **blue**) and 2nd largest (salmon or light blue) value in each row. Angular valves for $\Delta\psi$, $\Delta\varphi_1$, $\Delta\varphi_2$, and $\Delta\theta$ are listed in **Supplemental Table 3**.
(related to **Fig. S4**, and **Supplemental Table 2** and **3**)

its partner protein TorT in the presence and absence of the signaling ligand trimethylamine-N-oxide showed less pronounced piston motions (Moore and Hendrickson, 2012). These structural changes are dominated by changes in helix tilt and bundle radius; changes in the z parameter for each helix were below 0.3 Å, ruling out a shift as large as was observed in Tar.

Our analysis of the HAMP(Af1503)/DHp(EnvZ) chimera is also consistent with the analysis of Ferris *et.al.*, who proposed a helical rotation as well as other changes in packing. Structures of the chimeric "WT" were compared to those of single-site mutants in the HAMP domain that strongly modulate the degree of activation when introduced into the corresponding full-length hybrid receptors. In accord with the gearbox model, we compute that the helices indeed move in a con-rotary manner as anticipated by the gearbox motion (**Table 2.2**). These changes in helical phase underlie large structural changes, which our analysis associates with helical tilting about (φ_1) and rigid body shifts that change the radius of the bundle (**Fig. 2.5**). Together, these changes lead to differences in interhelical distance as large as 5 – 6 Å near the end of the bundle leading to the neighboring DHp domain.

The largely symmetric displacements of the diagonally opposed helix 2 and 2' of the HAMP leads to asymmetric buckling of the C-terminal helices at the junction between the HAMP and DHp domains, and remodeling of the N-terminal end of the DHp domain. The changes in bundle geometry reflect large changes in helical tilt (φ_1, φ_2) and bundle radius (r) (**Fig. 2.5A**, middle), which are believed to control the ability of the ATP-binding domain to dock onto and phosphorylate the His residue on the surface of the DHp domain (Ferris et al., 2012), as seen in recent structures of DHp-CA domains (Diensthuber et al., 2013; Mechaly et al., 2014; Wang et al., 2013).

Moreover, similar changes were observed for the DesK DHp domains. Lateral translations consistently dominate, with the exception of a single pair of helices in DesK, where helical rotations dominate and lateral translations come in a close second (**Table 2.2**). However, this helical rotation is only observed in a mutant in which the His residue involved in phosphotransfer is mutated to Val, and is not seen when the mutation is to Glu.

In summary, in the majority of the domains we studied, one of the two largest changes is either a radial displacement (r) or towards tilt (φ_1), both resulting in displacements of the helices relative to the bundle axis. However, this does not imply that the relatively small gearbox and/or piston shift motions might propel the larger changes in other degrees of freedom.

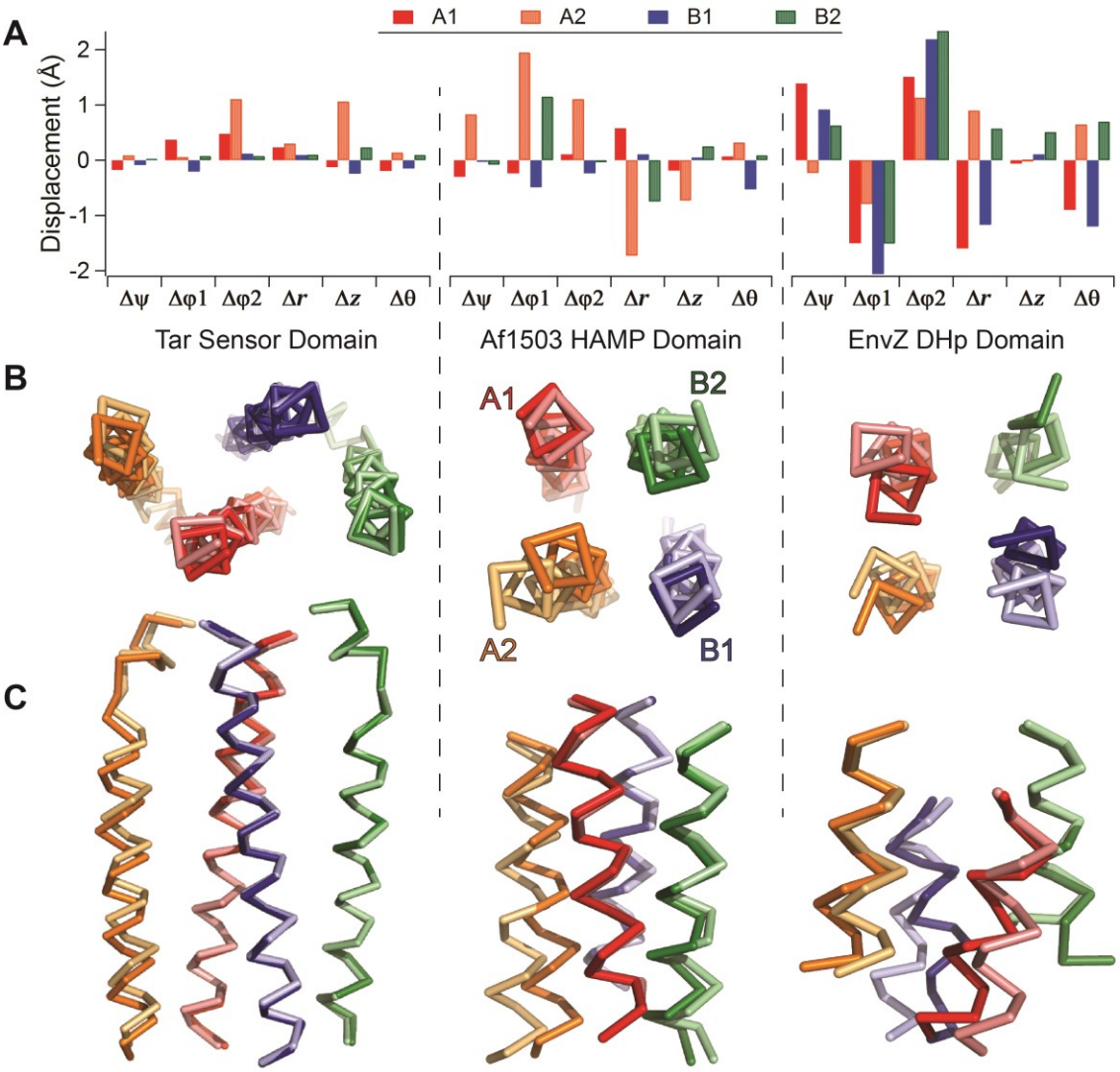


Figure 2.5 – Comparison of displacements between crystal structures in different states. (A) Graph of measured displacements for Tar Sensor (2LIG-1LIH), Af1530 HAMP (3ZRX-3ZRV), and EnvZ DHp (3ZRX-3ZRV) domains. Each of the four helices in the bundle is color coded differently. (B) Head-on view of 4-helix bundles. The helices in state 1 (PDBs 2LIG and 3ZRX) are colored vividly, while state 2 (1LIH and 3ZRV) are colored a corresponding lighter shade. (C) Side-view of 4-helix bundles, with helices colored as in (B).

2.5 Discussion

We present a model of HK signal transduction through the membrane, the first to utilize structural data taken from the full-length, dimeric protein in a native membrane. We generated this model using disulfide scanning mutagenesis and existing homologous crystal structures. Disulfide scanning mutagenesis allowed us to study the full-length protein in its native membrane environment, without relying on isolated domains in micelles or other membrane mimetics. We found good agreement between the crosslinking data and existing structures of HAMP, TM, and periplasmic domains, indicating that the isolated domain structures are reasonable models for the corresponding domains within the full-length protein in a native membrane environment, as well as that the cross-linking data is accurate. The crosslinking data are almost 180° out of phase with distances derived from homologous crystal structures (**Table 2.1**), which is expected, as crosslinking efficiency decreases over greater distances.

The crosslinking data covers the juxtamembrane regions connecting the TM domain to the sensor and HAMP domains (**Fig. 2.2**). These data provide the first evidence for an uninterrupted helix spanning TM1 to the N-terminal helix of the sensor domain. Additionally, the crosslinking data spanning the TM2-HAMP boundary indicates a possible interruption, which may be either a kinked helix or a disordered linker connecting the two domains. This interrupted structure may be necessary to form the previously described water hemichannel on the cytoplasmic face of the TM four-helix bundle (Goldberg et al., 2010).

Bayesian modeling revealed that the crosslinking experiments likely probed two structural states. We anticipated at least two states for the following two reasons. First, PhoQ must respond to its environment by relying on a thermodynamic equilibrium between its two signaling states, a prediction that is consistent with the similar proportions of the two modeled states present in the sample (40.5% State-1 and 59.5% State-2). Experimental data also supports signaling states near equilibrium, where we see a degree of activation of only 2-5 fold in low Mg^{2+} concentrations. These results are in agreement with the EPR studies of Trg from *E. coli*, which identified a dynamic and loosely packed TM domain (Barnakov et al., 2002). Second, two structural states

could explain conflicting crosslinking data within the TM domain, where TM1-TM1' crosslinks are sterically inconsistent with TM2-TM2' crosslinks.

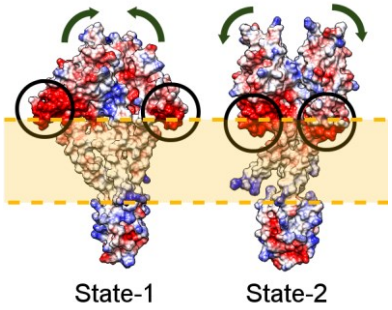


Figure 2.6 – Cation-binding, acidic patch movements predicted by the Bayesian multi-state modeling. (A) Electrostatic surface representation of the two states of the acidic patch as it moves out of (State-1) and in to (State-2) the membrane bilayer. Surface representation made with UCSF Chimera (Pettersen et al., 2004).

These two alternative conformations suggest large displacements of the sensor domains that insert or remove the periplasmic acidic patch within the membrane (**Fig. 2.6**). This patch is known to bind divalent cations (Waldburger and Sauer, 1996), which are believed to bridge to acidic lipids in the membrane (Cho et al., 2006), allowing the patch to insert into the membrane in the high Mg^{2+} signaling state. This insertion is coupled to scissoring transitions in the sensor, and remodeling of the helical bundles in the TM, HAMP, and DHp domains (**Fig. 2.7**), ultimately changing catalytic domain activity. While the highly simplified diagram in **Fig. 2.7A** is symmetrical, asymmetric version are equally likely, particularly in cases of negative cooperativity. Also, a transition within one domain need not require an all-or-nothing transition in the neighboring subunit, as is the case with rigid coupling. Rather, a structural transition changes the energetics or probability that the neighboring domain will transition from one state to another. The modeling predicts conformational rearrangements in the TM domain motion (**Fig. 2.3B** and 2.**3E**), in which two opposing helices move inward and displace the other two opposing helices that move outward (**Fig. 2.7A)**. We observed similar diagonal displacements across several two-component systems (**Fig. 2.7B-D**). Diagonally opposing displacements were observed in sensor, HAMP, and DHp domains; furthermore, these motions are consistent with the torque motion proposed recently for the blue-light sensing HK, YF1 (Diensthuber et al., 2013). Indeed, the HAMP domain, in which the gearbox model was discovered, also exhibits large lateral changes (Dunin-Horkawicz and Lupas, 2010) that induce correlated motions in the phosphor-accepting DHp domain. The

diagonal displacements can involve translation of helices within the bundle, bending, or tilting motions. When they involve a change of the crossing angle, the cores of the domains can remain relatively fixed between different states, engaging in limited motions that amplify near the ends of the helices to propagate into the neighboring domains. Indeed, the transmitted conformational changes are largest near the N-terminal end of a DHp bundle adjacent to the phosphorylated His residue (as in **Fig. 2.5**, EnvZ DHp).

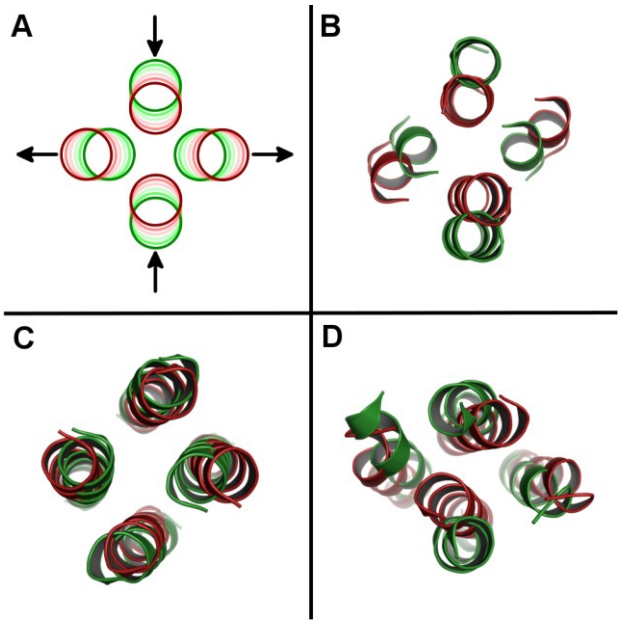


Figure 2.7 – Diagonal scissoring motions across several two-component domains. (A) A helix bundle exhibits orthogonal scissoring if two opposing helices move inward and the other two opposing helices simultaneously move outwards. (B) Citrate sensor domain, residues 12-25 and 45-51 (Green: 1P0Z, Red: 2J80). (C) HAMP domain from AF1503-EnvZ chimera, residues 283-297 (Green: 2L7H, Red: 2Y21).
(D) DesK DHp domain, residues 182-198 and 224-238 (Green: 3GIG, Red: 3EHH). For graphical representation of helix displacements see **Fig. S2.5**.

We place these qualitative observations on a more quantitative footing by measuring the variation between pairs of structures along six orthogonal degrees of freedom representing: 1) gearbox rotation about the helix axis; 2) piston shifts that vertically displace helices; 3 & 4) tilting towards and perpendicular to the bundle axis; 5) radial displacement of the helix from the bundle axis; and 6) rotation of the individual helices relative to the others about the bundle axis. In every examined case, we find that these domains, including the 2-state model, are not purely described by one pure motion, yet the tilting and radial displacements are the dominant change in almost every two-component domain that we analyzed (**Table 2.2**).

While the piston shift mechanism is well documented in the CheA-linked chemoreceptor, Tar, and our independent component analysis is in complete agreement with the original analysis of Falke and coworkers, this motion was not observed to contribute significantly to two-component HK

receptor proteins we examined. While there were many similarities between our cross linking profiles and those of Tar, there were also many significant differences. Both TM1 and TM2 formed inter-subunit crosslinks near the periplasmic end of the bundle, but inter-subunit crosslinks were not observed at corresponding positions in TM2 of Tar (Chervitz and Falke, 1996; Pakula and Simon, 1992). We observed a loss of crosslinking near the cytoplasmic half of the membrane for PhoQ, while the opposite behavior was observed for Tar. Moreover, the efficiency of crosslinking in the periplasmic helix of PhoQ showed a strong dependence on the concentration of its signaling ligand (Mg^{2+}, **Fig. 2.4**), but not for the Tar helix (α1-α1') and varying amounts of aspartate (Chervitz et al., 1995). These proteins have vastly different *in vivo* interaction partners, where PhoQ is regulated by some small proteins (Eguchi et al., 2012; Lippa and Goulian, 2009) and Tar forms large hexagonal lattices with CheA and CheW (Liu et al., 2012). Thus, the mechanism by which information transmits from the sensor domain to the HAMP must be significantly different for the two proteins, and it is unlikely that a piston shift is a significant component of the signaling transition for the HK receptor structures studied here.

Finally, it is important to end on a note of caution against over-interpretation of our analysis. Ironically, the helical scissoring motion seen here in the periplasmic domain of PhoQ is similar to the mechanism initially proposed by Koshland, Kim and coworkers for the Tar receptor (Milburn et al., 1991). This mechanism, however, fell out of favor following several studies: 1) crystallographic studies from the same investigators of a disulfide crosslinked mutant of Tar that was fully functional but did not show the large scissor-like motion seen in earlier constructs (Yeh et al., 1993); 2) an improved method of coordinate analysis introduced by Falke (Chervitz and Falke, 1996); 3) the observation that disulfide bonds across the subunit interface do not perturb receptor function (Chervitz et al., 1995); and 4) the lack of effect of attractant binding on disulfide formation rates across the subunit interface (Hughson and Hazelbauer, 1996). Similar caveats also apply to our analysis. The isolated domains that we analyze here might have similar flexibility unrelated to function. Moreover, while our crosslinking is carried out on full-length protein, our structural analysis is intrinsically coarse grained due to the errors associated with Cys disulfide crosslinking, particularly in flexible domains (Careaga et al., 1995). Nevertheless, diagonal

displacements at the dimer interface are a common feature of many recent symmetric and asymmetric models, including inter-helical torqueing (Diensthuber et al., 2013), helix-bending (Wang et al., 2013), DHp domain cracking (Dago et al., 2012), or a combination of these motions (Casino et al., 2009). More generally, helical bundle remodeling provides a mechanism for interdomain communication of a receptor protein and provides a signal transduction pathway from the outside of the cell to the phospho-accepting response regulator.

2.6 Materials and Methods

Plasmids. *pho*Q-His_6 Cys-mutant plasmids were created and then transformed into a Δ*phoQ* Δ*lacZ* strain (TIM206) as described previously (Goldberg et al., 2010).

Cell propagation. For crosslinking reactions in the TM and HAMP domains, cells were grown on LB agar or in LB medium at 37°C. For periplasmic mutants, cells were grown in MOPS minimal medium (Neidhardt et al., 1974) supplemented with 0.4% glucose, MEM vitamins, 0.2% casamino acids at 37°C. In both cases, the plasmid was maintained with 100μg/mL ampicillin.

Envelope preparations. Freshly plated colonies were picked by sterile loop and used to inoculate 5mL LB + 100μg/mL ampicillin. Cultures were grown at 37°C for 24 hours with vigorous shaking (220 rpm) and pelleted by centrifugation at 3700 x g for 10 minutes at 4°C. Cells were washed by resuspension in 30mM Tris, pH 8 and pelleted as above. Next, cells were treated with 20% sucrose in 30mM Tris, pH 8 for osmotic shock and 10mg/mL lysozyme to remove the cell wall. After 30 minutes incubation at 4°C, the cell envelopes were resuspended in 3mL of 3mM EDTA, pH 8 and sonicated briefly. TM and HAMP samples were spun at 16000 x g for 30 minutes at 4°C to pellet membranes. The membrane fraction was resuspended in 200μL of 2mM Tris, pH 7.5 and stored for use at -80°C. Periplasmic mutants were collected with a 10 minute ultracentrifuge spin (489000 x g) and then resuspended in 150μL of 8M Urea, 20mM NEM. Samples were stored at -80°C until run.

Crosslinking reactions. The oxidative catalyst, Cu(II)(1,10-phenanthroline)$_3$, a small, membrane-permeable reagent was used to efficiently catalyzes disulfide bond formation in the TM and HAMP domains (Lynch and Koshland, 1991). We combined a 10μL sample of cell envelopes with 10μL of buffer containing 2mM or 0.2mM CuPhen for a final concentration of either 1mM or 0.1mM. Reactions proceed for 30 minutes at 25°C. Reactions were stopped with of 20mM N-Ethyl Maleimide (NEM) and 20mM EDTA, and reactions were spun at 16,000 x g at 4°C to concentrate membranes. For the periplasmic domain mutants, we used the natural oxidizing environment of the periplasm to promote disulfide bond formation.

Western blotting and analysis. Oxidized membranes were reconstituted in 20μL of loading buffer (Invitrogen LDS buffer, 8M urea, 0.5M NEM) and heated for 10 minutes at 70°C. 5μL of

sample were loaded onto either a 7% or 3-8% gradient Tris Acetate gel (NuPage®, Invitrogen). Proteins were separated by electrophoresis and dry transferred to a nitrocellulose membrane (iBlot®, Invitrogen). For crosslinking reactions in the TM region, membranes were washed with TBST buffer (10mM Tris, pH 7.5, 2.5mM EDTA, 50mM NaCl, 0.1% Tween 20) and blocked with 3% BSA in TBST. PhoQ was probed using a penta-His antibody (Qiagen). The antibody was probed with HRP-conjugated sheep anti-mouse IGg (Pierce). Proteins were visualized by exposure to ECL reagent (Amersham, GE health sciences) for 1 minute and exposure to film for 30-60 seconds. For crosslinking reactions in the periplasmic region, membranes where blocked with TBST and 1% BSA (SNAP i.d.®, Millipore), then probed with penta-His HRP conjugate (Qiagen). Pixel density histograms were generated using the ImageJ software, freely available from the NIH (Abràmoff et al., 2004), and crosslinking efficiency was determined using the ratio of crosslinked dimer to total visible protein (dimer/(dimer+monomer)).

Multi-State Bayesian Modeling. The modeling and analysis were carried out with the open source *Integrative Modeling Platform* package (IMP; http://www.integrativemodeling.org) (Alber et al., 2007; Russel et al., 2012). IMP can construct structural models of macromolecular protein complexes by satisfaction of spatial restraints from a variety of experimental data. Model analysis is described in Supplemental Information.

Representation of the system and initial model. We generated a Cα model of the PhoQ dimer by assembling the models of HAMP, TM, and periplasmic domain. A comparative model of the HAMP domain dimer was created by using the dimeric HAMP-DHp fusion A291V mutant (PDB ID: 3ZRW) as a template. A comparative model of the TM monomer was built by using the two helices in the crystal structure of HtrII (PDB ID: 1H2S), corresponding to residues 23-82 of chain B, as a template. The model of the TM dimer was then obtained by applying the crystallographic C_2 symmetry about the dimer axis, observed in 1H2S. The dimer models of the three domains were positioned relative to each other into an initial dimer model of the whole PhoQ using UCSF Chimera (Pettersen et al., 2004), subject to the polypeptide chain connectivity between the three domains in each monomer (**Fig. 2.1A**). For the subsequent sampling, each monomer was decomposed into 6 rigid bodies and 5 short intervening flexible segments. Rigid bodies included

the following segments: 13-41 (TM1), 45-184 (periplasmic rigid body), 194-205 (N-terminus of TM2), 208-217 (C-terminus of TM2), 220-233 (N-terminal HAMP domain rigid body), and 245-265 (C-terminal HAMP domain rigid body). TM2 was divided into two rigid bodies due to a potential kink at P208. The two chains of the PhoQ dimer were sampled without enforcing any symmetry.

Bayesian model of cysteine crosslink data. The Bayesian approach (Habeck et al., 2006) estimates the probability of a model, given information available about the system, including both prior knowledge and newly acquired experimental data. When modeling multiple structural states of a macromolecular system, the model M includes a set X of *N* modeled structures $\{X_i\}$, their population fractions in the sample $\{w_i\}$, and the additional parameters $\{\alpha_n\}$ defined below. Using Bayes theorem, the posterior probability $p(M|D,I)$ of model M, given data D and prior knowledge I, is

$$p(M|D,I) \propto p(D|M,I) \cdot p(M|I)$$

where the likelihood function $p(D|M,I)$ is the probability of observing data D, given M and I; and the prior $p(M|I)$ is the probability of model M, given I. To define the likelihood function, one needs a forward model $f(X)$ that predicts the data point that would have been observed for structure(s) X, and a noise model that specifies the distribution of the deviation between the observed and predicted data points. The Bayesian and likelihood scores are the negative logarithm of $p(D|M,I) \cdot p(M|I)$ and $p(D|M,I)$, respectively. Detailed methods on the forward model, likelihood function, and prior information are described in Supplemental Information.

Sampling. A Gibbs sampling scheme based on Metropolis Monte Carlo (Rieping et al., 2005) enhanced by replica exchange was used to generate a sample of coordinates $\{X_i\}$ as well as parameters α_n and w_i from the posterior distribution of a given number of structures (*N*). The moves for $\{X_i\}$ included random translation and rotation of rigid parts (0.15 Å and 0.03 radian maximum, respectively), random translation of individual beads in the flexible segments (0.15 Å maximum), as well as normal perturbation of the parameters α_n and w_i. To facilitate the sampling of the posterior probability, we eliminated its dependence on the uncertainties σ_n by numerical marginalization (Sivia and Skilling, 2006).

Quantitative Structural Analysis. We gathered structures from TCSs with multiple structures of the same domain, listed in **Table 2.2**. For each domain, we define the bundle axis by first selecting two pairs of equivalent residues, one from each chain, calculating the Cα-Cα vector between those two residues for both chains, and then summing these two vectors to create the axis vector. We define the bundle axis vector for one structure (the first PDB ID in each row) to be the z axis, arbitrarily specify an x axis orthogonal to the z axis, and define the y axis perpendicular to x and z, using a right-handed coordinate system. We then align the remaining domains to the first structure using CEAlign (Shindyalov and Bourne, 1998) along the domain boundaries listed in **Table S2.2**.

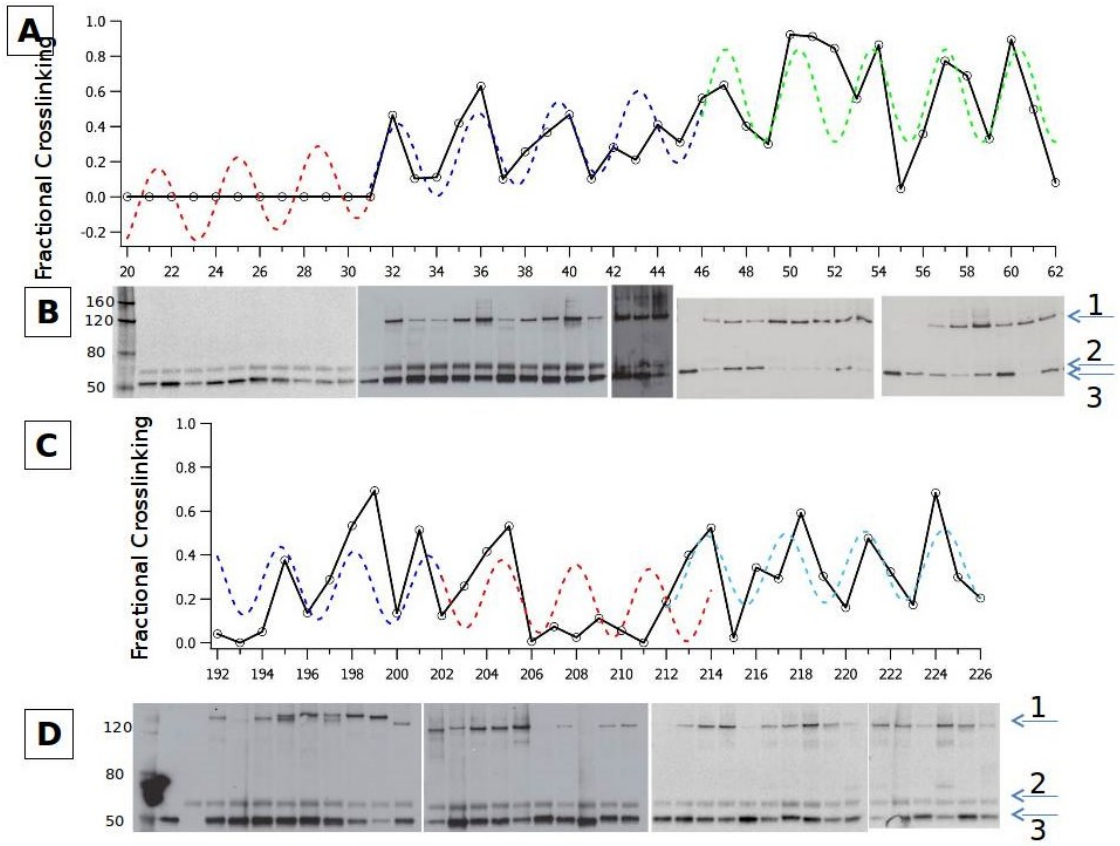
We fit each helix to a straight ideal helix (2.3 Å α-carbon radius, 3.6 residues/turn, 1.5 Å rise/residue) and extract six geometric parameters that define the helix's position and orientation by fitting a sequence of six motions (**Fig. S2.4**) using the Levenberg-Marquardt algorithm from the GNU scientific library (Galassi and Gough, 2005) (10^{-4} absolute tolerance, 10^{-4} relative tolerance, maximum 20 iterations). For each helix and for all six motions, we measured the maximum variation (range) of the fitted parameter. We normalize all rotations to distances by converting degrees to subtended arcs using a radius equivalent to the distance of an ideal β-carbon at a helix endpoint from the focal point of the rotation. This corresponds to an arc radius of 4 Å for rotations about the helix axis and an arc radius of (1.5 Å x (# of helix residues) / 2) for tilting motions. The full set of calculated displacements is given in **Table S2.3**.

2.7 Acknowledgements

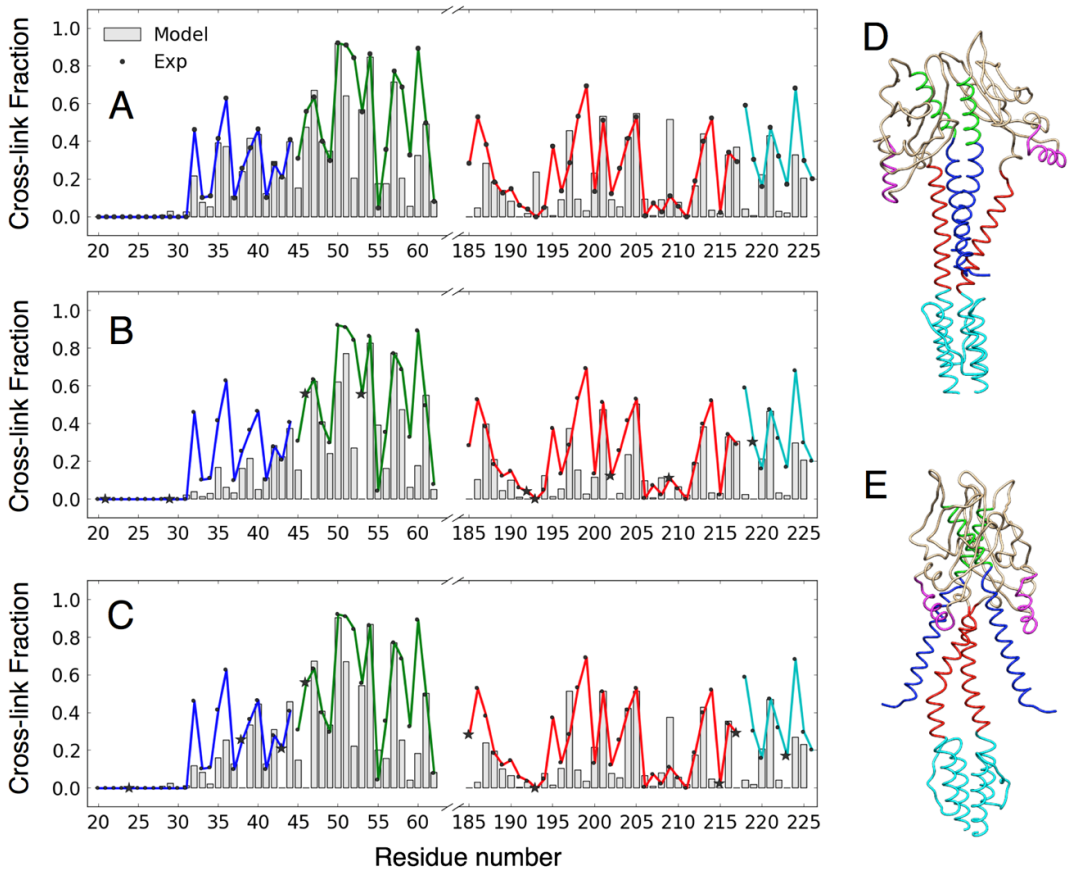
The DeGrado lab acknowledges support from grants from the NIH, GM54616 and AI074866, as well as support from the MRSEC program of NSF (DMR-1120901). The Sali lab acknowledges support from grants from the NIH (NIGMS U54 RR022220, R01 GM083960 and U54 GM074929). R. Pellarin was supported by grants from the Swiss National Science Foundation (PA00P3-139727 and PBZHP3-133388). The authors also thank Dr. Joseph Falke for critical reading of this manuscript, and many helpful suggestions.

2.8 Supplemental material

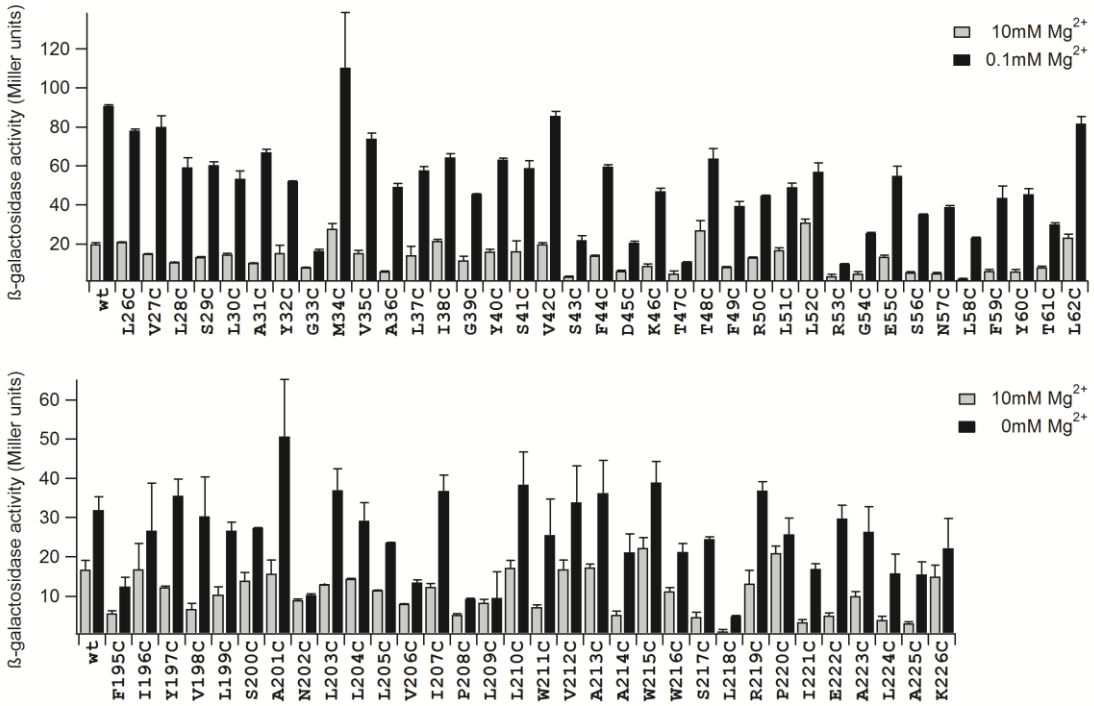
Figures



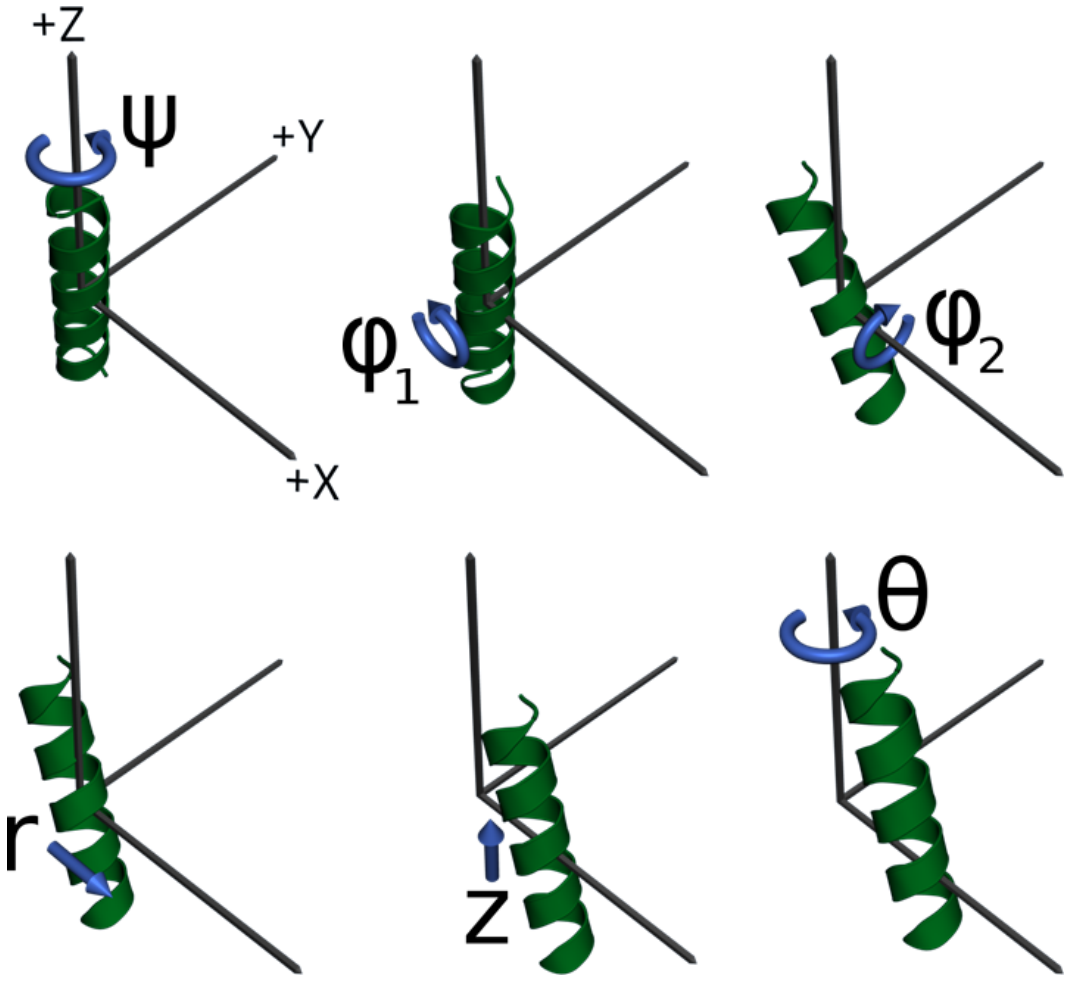
Supplementary Figure 2.1 (Related to Figure 2.2) – Analysis of the fractional crosslinking of PhoQ residues. (A) Fractional crosslinking of PhoQ residues 20-62 (black lines with circles) are fitted using a sine wave over the regions that correspond to the domains of PhoQ (dashed lines) the colors are maintained from **Figure 2.2**: dark blue is the well-packed domain, red are residues that line the cytoplasmic cavity and green are the HAMP residues. These data demonstrate a right-handed helix (ω=3.62) for TM1 that is in phase with the previously reported data. (B) Representative western blots of PhoQ residues reported in A. Each lane represents the data point directly above it in A. Arrows on the right of the figure indicate 1) the crosslinked PhoQ dimer band 2) an E coli lysate band 3) PhoQ monomer band. (C) Fractional crosslinking of PhoQ residues 192-226 are fitted using a sine wave over the regions that correspond to the domains of PhoQ (dashed lines) the colors are maintained from **Figure 2.2**: dark blue is the well-packed domain, red are residues that line the cytoplasmic cavity and light blue is the HAMP. These data demonstrate TM2 is a left-handed helix (ω=3.29) with a striking lack of continuity with the first HAMP helix due to a disturbance in the phase of the helix which arises from residue P208. (D) Representative western blots of PhoQ residues reported in (C). The numbering of the arrows on the right is identical to (B).



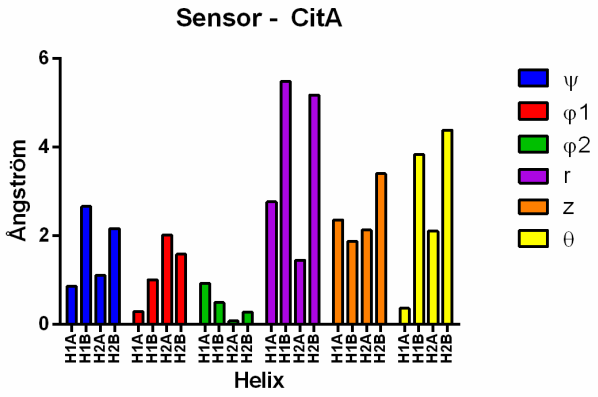
Supplementary Figure 2.2 (Related to Figure 2.3) – Assessing the robustness of the two-state modeling Bayesian approach by data jackknifing. (A-C) Data predicted by the model (grey bars) with highest likelihood in the most populated cluster using the whole dataset of 85 cross-links (A) and randomly discarding 10% of the data points (9 data points in panel B and 8 data points in panel C, represented by black stars). Experimental data is represented by continuous line color-coded according to the PhoQ domains defined in the caption of Fig. 3 in the main text. (D, E) Two-state structural model with highest likelihood in the most populated cluster obtained by removing 10% of data points. The two states correspond to State-1 (D) and State-2 (E) in Fig. 3 in the main text. The inferred population fractions of State-1 and -2 are 25% and 75%, respectively.

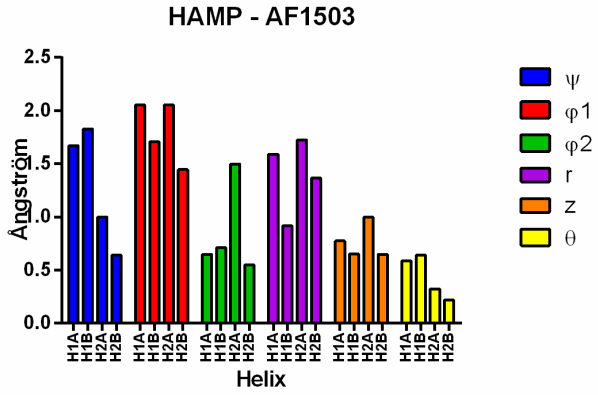


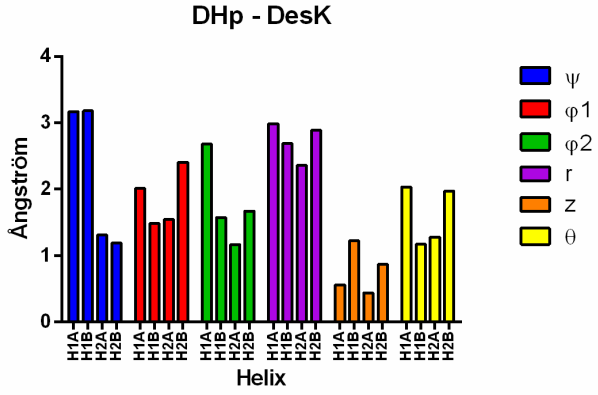
Supplementary Figure 2.3 (Related to Figure 2.4) – Phenotypic changes in response to Cys mutations in PhoQ. We assessed the activity of Cys mutants by the Miller assay (Miller, 1972). TIM206 (*mgtA::lacZ* Δ*phoQ*) cells were transformed with a plasmid encoding PhoQ and a Cys mutation at a single position. Mutation to Cys was tolerated at most positions. Positions known to have a critical function also have no activity when mutated (*e.g.,* 202). Many positions where the Bayesian model does not explain the data (residues 195,208, 209, and 218) also do not respond to Mg^{2+}.



Supplementary Figure 2.4 (Related to Table 2.2) – The six degrees of motion in the order they are applied to fit any given helix: ψ: first rotation about Z axis; φ1: rotation about Y axis; φ2: rotation about X axis; r: translation along X axis; z: translation along Z axis; θ: second rotation about Z axis.







Supplementary Figure 2.5 (Related to Figure 2.7): Measured differences between equivalent helices in two component systems. Each chart corresponds to a single set of domains referenced by domain type (e.g. Sensor or HAMP) and protein (e.g. AF1503 or DesK). Each chart groups the differences by direction, with four measured differences per direction, one for each helix in a four-helix bundle: H1A) Helix 1 - Chain A, H1B) Helix 1 – Chain B, H2A) Helix 2 – Chain A, H2B) Helix 2 – Chain B. All measurements are calculated maximum displacements along each degree of freedom for an ideal β-carbon at a helix endpoint.

Tables

Supplemental Table 2.1 (Related to Figure 2.3): Properties of the clusters with population greater than 3% found with 1-state, 2-state and 3-state modeling: cluster population, average and best χ^2 and likelihood score (-log p(D|M,I)).

Number of states	Cluster id	Cluster population	Center		Best	
			χ^2	*L*	χ^2	*L*
1	1	0.16	0.85	-33.70	0.75	-46.64
	2	0.06	0.94	-30.47	0.79	-39.69
	3	0.05	0.91	-35.26	0.80	-43.99
	4	0.04	0.78	-43.38	0.66	-57.33
	5	0.03	0.79	-39.72	0.71	-50.21
2	1	0.12	0.65	-59.60	0.54	-67.78
	2	0.10	0.84	-29.91	0.70	-49.31
	3	0.06	0.73	-38.59	0.64	-57.11
	4	0.05	0.83	-31.29	0.69	-44.36
	5	0.04	0.71	-48.21	0.59	-61.72
	6	0.04	0.79	-26.59	0.69	-44.70
	7	0.04	0.77	-33.16	0.67	-44.82
3	1	0.16	0.79	-23.85	0.69	-43.50
	2	0.06	0.64	-48.47	0.55	-62.76
	3	0.06	0.84	-31.42	0.65	-53.75
	4	0.05	0.82	-16.56	0.73	-45.09
	5	0.04	0.76	-36.36	0.65	-51.36

Supplemental Table 2.2 (Related to Table 2.2). Parameters used for domain fitting.

Domain	Protein	Organism	Domain Boundary	Helix Residues	Aligning Residues
Sensor domain	Tar	*S. typhimurium*	42-174	42-57, 155-174	50, 61
Sensor domain	TorS	*V. parahaemolyticus*	283-318	52-67, 300-317	297, 308
HAMP	Af1503	*A. fulgidus*	279-328	280-297, 310-328	284, 295
DHp	EnvZ	*S. flexneri*	335-385	333-345, 373-385	376, 383
DHp	DesK	*B. subtilis*	191-232	182-198, 224-238	187, 198

Supplemental Table 2.3 (Related to Table 2.2). Geometric parameters[a] describing positions of individual helices relative to the bundle axis.

		PDB ID	Chain	Helix	ψ (°)	ϕ_1 (°)	ϕ_2 (°)	r (Å)	z (Å)	θ (°)
Sensor Domains	Tar	2LIG	A	1 (42-57)	123	2.58	-6.13	6.21	-8.80	53.5
			B	1 (42-57)	127	-0.06	-7.39	6.39	-8.87	-129
			A	2 (155-174)	144	2.98	-7.16	11.9	2.69	149
			B	2 (155-174)	141	3.55	-10.4	11.6	1.92	-29.2
		1LIH	A	1 (42-57)	125	0.80	-8.42	5.98	-8.68	55.3
			B	1 (42-57)	128	0.92	-7.96	6.30	-8.62	-127
			A	2 (155-174)	142	2.75	-11.4	11.6	1.63	149
			B	2 (155-174)	141	3.27	-10.7	11.5	1.68	-29.6
	TorS	3O1H	A	1 (52-67)	-98.3	-0.46	-6.65	6.53	-44.2	-105
			B	1 (52-67)	-98.3	-0.46	-6.65	6.53	-44.1	74.8
			A	2 (300-317)	2.86	-3.21	2.64	8.80	48.5	128
			B	2 (300-317)	2.86	-3.21	2.64	8.80	48.5	-51.6
		3O1I	A	1 (52-67)	-96.8	-0.29	-8.48	7.23	-44.4	-99.3
			B	1 (52-67)	-104	-0.63	-5.96	5.70	-44.0	78.4
			A	2 (300-317)	-2.92	-6.25	0.29	8.99	48.5	127
			B	2 (300-317)	6.02	-4.13	-0.52	9.01	48.3	-55.6
HAMP Domain	AF1503	3ZRV	A	1 (280-297)	-22.2	-2.98	-8.37	6.80	-0.30	-122
			B	1 (280-297)	-29.2	-2.35	-8.71	7.32	-0.73	63.4
			A	2 (310-328)	-166	-1.55	-9.74	7.77	3.05	-24.3
			B	2 (310-328)	-158	3.15	-6.30	6.81	2.16	-202
		3ZRW	A	1 (310-328)	-19.7	-2.46	-7.85	7.12	-0.12	-121
			B	1 (310-328)	-19.9	-2.69	-7.96	7.14	-0.29	58.6
			A	2 (280-297)	-164	1.03	-7.45	7.18	2.47	-22.0
			B	2 (280-297)	-164	1.15	-7.56	7.14	2.30	-202
		3ZRX	A	1 (280-297)	-26.4	-4.01	-7.91	7.38	-0.49	-122
			B	1 (280-297)	-29.6	-4.41	-9.74	7.43	-0.68	59.3
			A	2 (310-328)	-154	6.30	-5.27	6.04	2.31	-21.6
			B	2 (310-328)	-159	7.79	-6.47	6.06	2.41	-201

DHp Domains	EnvZ	3ZRV	A	1 (333-345)	41.3	2.18	10.3	7.50	-3.74	-55.1
			B	1 (333-345)	28.5	3.09	-8.37	7.57	-4.62	136
			A	2 (373-385)	101	-5.73	-9.28	7.43	-1.32	63.5
			B	2 (373-385)	86.3	0.00	-10.2	7.69	-2.06	-120
		3ZRW	A	1 (333-345)	40.9	-0.63	0.97	8.28	-3.59	-46.7
			B	1 (333-345)	40.1	-1.32	-0.06	8.30	-4.32	133
			A	2 (373-385)	90.8	-1.26	-7.10	7.04	-1.26	64.3
			B	2 (373-385)	91.0	-1.09	-5.79	7.06	-1.34	-116
		3ZRX	A	1 (333-345)	61.6	-6.65	19.1	5.90	-3.81	-62.9
			B	1 (333-345)	41.7	-9.11	4.64	6.40	-4.52	126
			A	2 (373-385)	98.0	-10.4	-2.64	8.34	-1.33	68.3
			B	2 (373-385)	95.4	-8.88	3.67	8.26	-1.55	-115
	DesK	3EHF	A	1 (182-198)	87.7	4.58	-3.50	7.49	-3.13	96.6
			B	1 (182-198)	87.9	-1.66	-5.73	6.35	-3.62	-82.8
			A	2 (224-238)	-115	-0.52	-7.73	7.15	0.50	4.24
			B	2 (224-238)	-111	-4.98	-9.51	7.23	0.60	178
		3EHH	A	1 (182-198)	47.6	-0.23	-11.8	5.13	-3.59	105
			B	1 (182-198)	131	-3.32	-12.1	4.40	-2.77	-74.1
			A	2 (224-238)	-132	-9.40	-11.1	9.51	0.11	-0.52
			B	2 (224-238)	-128	-10.7	-15.6	9.75	-0.13	176
		3GIF	A	1 (182-198)	93.0	8.77	3.72	8.12	-3.03	87.2
			B	1 (182-198)	85.9	-8.59	-9.28	5.54	-4.00	-83.1
			A	2 (224-238)	-123	-1.66	-6.59	7.39	0.39	4.70
			B	2 (224-238)	-120	-6.93	-9.17	7.70	0.40	179
		3GIG	A	1 (182-198)	90.4	-2.86	1.26	6.98	-3.11	93.3
			B	1 (182-198)	89.4	-0.86	-3.04	7.08	-3.60	-82.3
			A	2 (224-238)	-113	-7.45	-13.3	7.84	0.06	-4.35
			B	2 (224-238)	-116	-17.6	-18.8	8.46	0.74	166

[a] parameters were defined as described in **Supplementary Figure 2.4**

Methods

Forward model. The forward model predicts the cross-linked fraction of cysteine pair n after a reaction time t, for a mixture of *N* states $\{X_i\}$:

$$f_n(\{X_i, w_i\}) = \sum_{i=1}^{N} w_i \left(1 - e^{-\alpha_n \rho(r_n)}\right)$$

where $\alpha_n = k_n\, t$ is the product of the unknown intrinsic reaction rate of cysteine pair n and the total reaction time. $\rho(r_n)$ is an efficiency term that depends on the distance r_n between the cysteine Cα atoms and it is computed by considering (i) the uncertainty in the position of the residue centroids along the main chain due to the limited precision in determining the position of the residues, (ii) the cost of having a disulfide bond geometry far from the ideal one, and (iii) the reduction of the reaction volume due to the presence of proximal components and moieties.

Likelihood function. The likelihood function $p(D|M, I)$ for dataset $D = \{d_n\}$ of N_{XL} independently measured cross-linked fractions is a product of likelihood functions for each data point. Because the cross-linked fractions vary between 0 and 1, we modeled the noise with a normal distribution truncated to this interval. The likelihood for data point d_n can thus be written as:

$$p(d_n|\{X_i, w_i\}, \alpha_n, \sigma_n) = Z^{-1}\exp\left(-\frac{[d_n - f_n(\{X_i, w_i\})]^2}{2\sigma_n^2}\right)$$

where the uncertainty σ_n shapes the likelihood function and Z is the normalization factor. To account for varying levels of noise in the data, each data point has an individual σ_n.

Furthermore, to encode template structure information for the HAMP dimer domain (residues 227-263), a likelihood function with log-normal noise was defined based on the distances r_{jk} between all Cα atoms that are below 8 Å in the template (PDB code 2Y20):

$$p\left(r_{jk}\middle|\{X_i\}, \sigma_H\right) = \prod_i Z^{-1}\exp\left(-\frac{\log^2 r_{jk}/r_{jk,i}}{2\sigma_H^2}\right),$$

where $r_{jk,i}$ is the distance between atom j and k in the modeled structure X_i and σ_H is the uncertainty.

Prior Information. The prior on a structure is defined as $p(\{X_i\}) \propto \exp(-\sum_i V(X_i))$ where V is a sum of spatial restraints: $V = V_{excl.\ vol.} + V_{C_\alpha\ bonds} + V_{C_\alpha\ angles} + V_{C_\alpha\ dihedrals} + V_{Ez} + V_{layer}$.

The excluded volume restraint $V_{excl.vol.}$ was implemented as a pairwise hard-sphere repulsive potential, where the volume of each Cα particle equals the volume of the corresponding amino acid residue (Pontius et al., 1996). The bond, angle, and dihedral terms $V_{Cα\ bonds}$, $V_{Cα\ angles}$, and $V_{Cα\ dihedrals}$, respectively, are statistical potentials that enforce the correct stereochemistry, as well as the correct secondary structure propensity, of the flexible backbone [see below]. The V_{Ez} potential (Senes et al., 2007) was used to model the membrane environment. Furthermore, residues F17 of the two PhoQ chains were confined inside a layer representing the inner leaflet of the membrane, by using a flat bottom harmonic restraint acting on the z coordinate between -17 Å and -13 Å, V_{layer}.

Crosslinking data was collected in three separate experiments for the periplasmic, membrane, and cytoplasmic domains. We used three α_n parameters to model experimental variation between these three data subsets. The priors for α_n are bounded uniform distributions: the lower bound was determined by the highest observed fraction in the subset and the upper bound by the highest detectable fraction. The priors for σ_n are unimodal distributions (Sivia and Skilling, 2006): $p(\sigma_n|\sigma_0) = \frac{2\sigma_0}{\sqrt{\pi}\sigma_n^2}\exp\left(-\frac{\sigma_0^2}{\sigma_n^2}\right)$, where σ_0 is an unknown experimental uncertainty; the heavy tail of the distribution allows for outliers. The priors for w_i were uniform distributions over the range from 0 to 1, with the constraint $\sum_i w_i = 1$. Furthermore, a lower bound at 0.2 was enforced on each w_i to avoid visiting conformations already sampled at smaller *N* values. A Jeffrey's prior $p(\sigma_H) = 1/\sigma_H$ was used for the uncertainty parameter of the likelihood used to incorporate template structure information.

Analysis. The set of sampled models $\{M_j\}$ were clustered (Daura et al., 1999) based on the value of the forward model $f_n(M)$, using the following data-based metric:

$$\left|\left|M_1 - M_2\right|\right|^2 = \frac{1}{N_{XL}}\sum_{n=1}^{N_{XL}}\frac{[f_n(M_1) - f_n(M_2)]^2}{\sigma_{n,1}^2 + \sigma_{n,2}^2}$$

where $\sigma_{n,j}$ is the inferred measurement error associated with data point n in model j, and N_{XL} is the total number of crosslinks. A cutoff of 0.05 was used. In multi-state modeling, data-based clustering is preferred to structure-based clustering (*e.g.,* using Cα-RMSD as the distance metric)

because it reflects the degeneracy of models that would generate the same data and because it provides a natural way of mixing X_i, w_i, and σ_n^E that is not possible in structure-based clustering. Because the sample is drawn from the posterior distribution, the cluster population is proportional to the average posterior probability of its members. We focused our analysis on the clusters with a population greater than 3%. The structural model precision of a given cluster was defined as the median of the RMSD distribution calculated on all pairs of cluster members.

Stereochemistry scoring terms. The bond, angle, and dihedral terms $V_{C\alpha\ bonds}$, $V_{C\alpha\ angles}$, and $V_{C\alpha\ dihedrals}$, respectively, are statistical potentials that enforce the correct stereochemistry, as well as the correct secondary structure propensity, of the flexible backbone. The input information is the predicted secondary structure (using DSSP secondary structure symbols (Kabsch and Sander, 1983)) for a given protein segment. These terms were calculated by estimating the probability that residues in a given secondary structure sequence adopt a given configuration, defined by distances, angles and torsion angles. The probabilities are derived from the MRS database of crystallographic structures with assigned secondary structure (Hekkelman and Vriend, 2005). For each sequence-contiguous residue pair $(n, n+1)$, triplet $(n, n+1, n+2)$ and quintuplet $(n, n+1, n+2, n+3, n+4)$, the potentials are calculated as:

$$V_{C_\alpha-Bonds}(r_{n,};\ S_n, S_{n+1}) = -log\left(\frac{\sum_k \sum_i \delta(r_n - r_i^{(k)})\delta_{S_n,S_i^{(k)}}\delta_{S_{n+1},S_{i+1}^{(k)}}}{\sum_k \sum_i \delta_{S_n,S_i^{(k)}}\delta_{S_{n+1},S_{i+1}^{(k)}}}\right)$$

$$V_{C_\alpha-Angle}(\alpha_n;\ S_n, S_{n+1}, S_{n+2}) = -log\left(\frac{\sum_k \sum_i \delta(\alpha_n - \alpha_i^{(k)})\delta_{S_n,S_i^{(k)}}\delta_{S_{n+1},S_{i+1}^{(k)}}\delta_{S_{n+2},S_{i+2}^{(k)}}}{\sum_k \sum_i \delta_{S_n,S_i^{(k)}}\delta_{S_{n+1},S_{i+1}^{(k)}}\delta_{S_{n+2},S_{i+2}^{(k)}}}\right)$$

$$V_{C_\alpha-Dihedral}(\tau_n, \tau_{n+1};\ S_n, S_{n+1}, S_{n+2}, S_{n+3}, S_{n+4}) =$$

$$= -log\left(\frac{\sum_k \sum_i \delta(\tau_n - \tau_i^{(k)})\delta(\tau_{n+1} - \tau_{i+1}^{(k)})\delta_{S_n,S_i^{(k)}}\delta_{S_{n+1},S_{i+1}^{(k)}}\delta_{S_{n+2},S_{i+2}^{(k)}}\delta_{S_{n+3},S_{i+3}^{(k)}}\delta_{S_{n+4},S_{i+4}^{(k)}}}{\sum_k \sum_i \delta_{S_n,S_i^{(k)}}\delta_{S_{n+1},S_{i+1}^{(k)}}\delta_{S_{n+2},S_{i+2}^{(k)}}\delta_{S_{n+3},S_{i+3}^{(k)}}\delta_{S_{n+4},S_{i+4}^{(k)}}}\right)$$

where r_n , α_n and τ_n are respectively the distance, the angle and the torsion angle between sequence-contiguous residue pairs, triplets and quadruplets starting from residue n; n , i and k are respectively indexes for the residue in the model, the residue in the database structure and the structure number in the database; δ is the Kronecker delta function; $S_n \in \{H, E, C\}$ is the

secondary structure symbol for residue n, where H, E, C correspond to helical, beta and random coil. The denominator in the left side of each equation is the normalization term over the given secondary structure sequence.

The dihedral term corresponds to the joint probability of having the torsion angles τ_n and τ_{n+1} at given values, given that the secondary structure sequence is $S_n, S_{n+1}, S_{n+2}, S_{n+3}, S_{n+4}$. This term enforces the secondary structure geometry on the C_α model more effectively that a term that depend on a single torsion angle τ_n.

Sequence-structure threading and model manipulation. To generate electrostatic maps, we threaded PhoQ's sidechains on to our 2-state models using Scwrl (Canutescu et al., 2003) and minimized the side-chain using Rosetta fast-relax (Khatib et al., 2011). Structure visualization and manipulation was performed using PyMol molecular viewer (Schrodinger).

2.9 Appendix to Chapter Two

In this appendix, I detail some of the experiments that were done in order to explain the conflicts between experimental data and the Bayesian models. The result has been presented in the previous chapter; here I outline the specifics of how we got to that result. Also, we consider crosslinking data in an MgrB knockout strain.

In the periplasm of bacteria, disulfide crosslinking appeared to change with time; therefore, we hypothesized that growth rate and conditions could affect the degree of crosslinking. I first sought to recapitulate the original data to rule out the possibility of a mutant being incorrectly assigned. The repeat of the data proved to be straightforward with the same pattern observed over these three residues (Fig. A2.1).

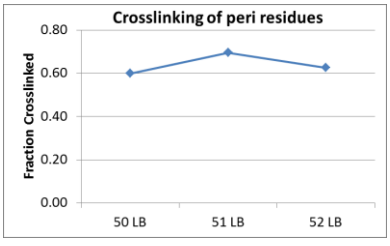


Figure A2.1 – Comparison of disulfide crosslinking of mutant PhoQ carried out in a PhoQ knockout strain (TIM206), grown in LB, overnight, at 37°C.

We hypothesized that the crosslinking was time dependent. Here, I looked at crosslinking for the three residues in question over different phases of growth. As expected the degree of crosslinking depended greatly on bacterial growth phase and for particular residues (L50) the growth phase mattered more than others. With this validation that the spurious crosslinks are likely due to long incubation times, I began looking for differences in crosslinking under different activation conditions. Using high and low concentrations of magnesium (0.1mM and 20mM Mg^{2+}) in LB broth the results were inconsistent (*data not shown*). I moved to minimal media broth (MOPS) were I could more accurately control the divalent ion concentration. With these conditions I was able to produce periodic crosslinking that better approximated a helix (Fig. A2.2). If we compare the extent of crosslinking for mutants 51-53, the relative values are constant across the various growth times or the types of media. For F49C and R50C, there seems to be an additional dependence. At longer times in LB media, R50C is capable of making a crosslink that is not physiologically relevant. Similarly, the F49C mutant will begin to crosslink at long incubation times in minimal media.

This result suggests that perhaps positions R50 and F49 are targeted by native periplasmic isomerases and over time, preferentially form these spurious crosslinks. Perhaps MgrB has a role to play in this because it is known to contain two periplasmic cysteine residues that can interact with PhoQ's sensor domain.

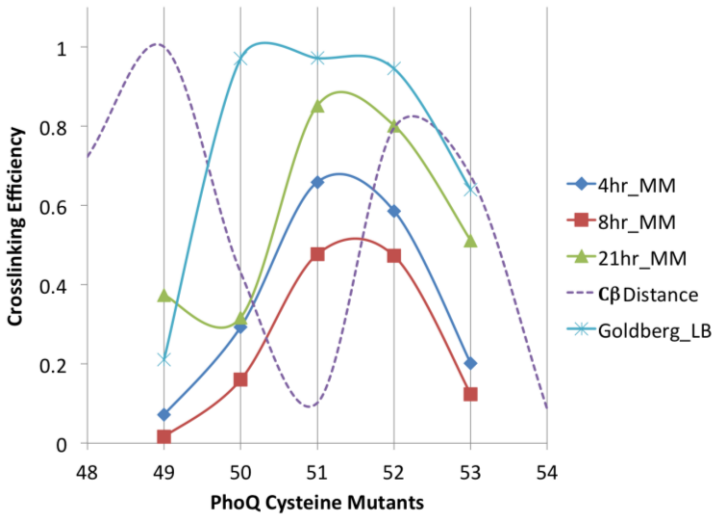


Figure A2.2 – Disulfide crosslinking of periplasmic cysteine mutants in 0.1mM Mg^{2+}. Plasmids containing Cys mutant PhoQ at positions 49-54 in the TIM206 strain were inoculated 1:1000 into 100mL of MOPS minimal media with 0.1mM Mg^{2+}. At noted time intervals, 15mL aliquots were removed and crosslinking was halted with NEM. Crosslinking efficiency was determined using western blot. The dotted line is the normalized distance between Cβ carbons at each position in 3BQ8. Data in LB media was collected previously (Goldberg et al., 2008).

I preformed crosslinking studies of the remainder of the periplasmic helix using minimal media and halting the experiment at mid-log phase growth in order to generate a set of data at high and low concentrations of the signaling molecule, magnesium. All of the previous LB data was done only under low magnesium conditions. While collecting this new set of data, I also collected mutants the absence of the periplasmic repressor protein, MgrB (Figure A2.3) (Lippa and Goulian, 2009, 2012).

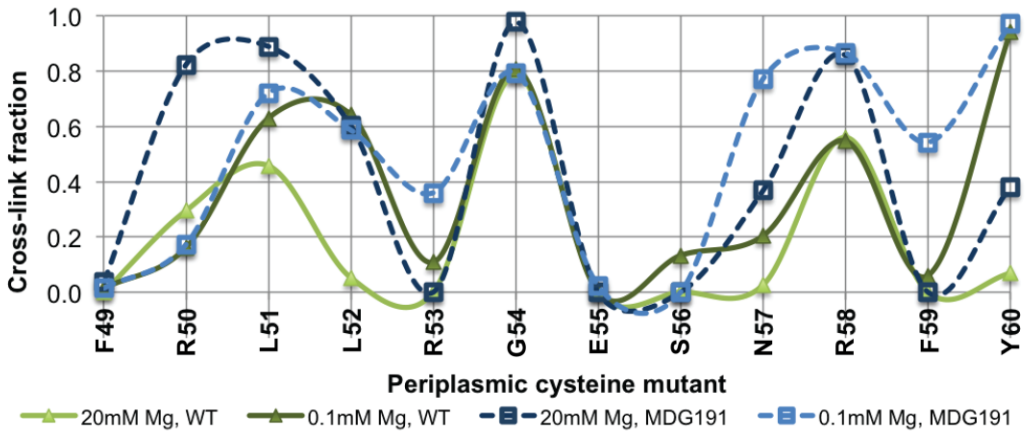


Figure A2.3 – Revised crosslinking in the PhoQ periplasmic domain. All data points were determined by western blot and with cultures grown in minimal media in the midlog phase. The *mgrB* deletion strain (MDG191) was provided by the Goulian lab.

In these experiments we are changing PhoQ activation using both magnesium as well as by removing the negative regulator MgrB. Activity changes due to magnesium are well characterized (Cho et al., 2006). Activity changes due to MgrB have been characterized in the Goulian lab as well as via LacZ reporter assays (data not shown). Using these tools the upper limit of PhoQ activation would be testing the protein in the ΔMgrB strain under low magnesium conditions. By contrast, the most repressed state assayed is the protein in a wild-type strain in the presence of high magnesium. Here, we consider whether the expression of PhoQ on a plasmid is equivalent to the native expression. Since we know that the leaky expression of the plasmid does exceed native levels, it is likely that over expression of MgrB would also be needed to fully repress PhoQ activity. The Goulian lab has over-expressed MgrB and investigating the changes in the crosslinking in that case (*unpublished data*).

Twelve periplasmic cysteine mutants were tested over all four experimental conditions (1. -MgrB, low [Mg^{2+}]; 2. -MgrB, high [Mg^{2+}]; 3. +MgrB, low [Mg^{2+}]; 4. +MgrB, high [Mg^{2+}]) and the results are plotted in Figure A. 3. Experiments in the presence of MgrB were done in triplicate, while the data in the absence of MgrB is only a single measurement.

There are interesting similarities across all four experimental conditions. Positions at the center of the periplasmic helix (R53, G54, E55 and S56) were not easily perturbed. Glycine is known to mediate helix-helix interactions (Senes et al., 2004) so it is not surprising that the crosslinking is extremely high under all conditions assayed. It also follows that for the flanking residues crosslinking would be limited because the Cys is prevented from making dimer interactions when the glycine is mediating interactions. When you couple this crosslinking data with subsequent activity data, we can postulate the crosslinking is so consistent at position 54 because the activity is profile is almost locked in the phosphatase state. This means that we are not actually sampling both conformational states of the protein.

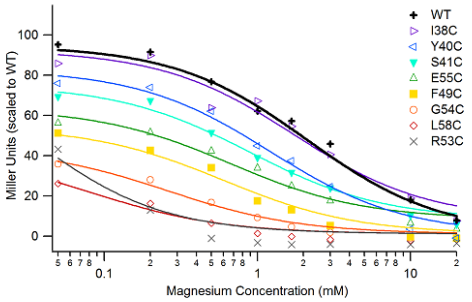


Figure A2.4 - Fits of LacZ activity data for a handful of periplasmic cysteine mutants, four of which are part of this crosslinking study. LacZ data and fits are explained fully in Chapter 3.

In contrast, certain residues seem to be quite sensitive to the experimental conditions. The best example of this can be found in N57; the more activating condition that PhoQ experiences, the higher the crosslink-fraction of this residue. Other examples are found in residues L52 and F59, though these positions are not as sensitive as N57 and only show a difference at the most inactive or most active condition. The last group of residues (L51, 58 and Y60) has a shift crosslink fraction that is dependent on MgrB. It is easiest to see in 58 where the addition of MgrB shifts the crosslink fraction down to 50% from 90%. The shift in Y60 is only apparent for the conditions at high Mg because the low Mg condition is already at the upper limit (100%).

These results showed a magnesium dependent change of the interfacing helix of PhoQ is unlike the results from other crosslinking studies of the well-studied chemoreceptor, Tar (Butler and Falke, 1998; Chervitz and Falke, 1996; Chervitz et al., 1995). The result is recapitulated in the +/- MgrB data where we also see peptide dependent changes in crosslinking. Together these suggest that the sensor domain of PhoQ is a lot more flexible than the chemotaxis sensor. This difference can explained by the differences in the secondary structure and packing of the periplasmic domains such as Tar or Tsr (Studdert and Parkinson, 2007; Swain et al., 2009). In that system, the dimeric interface is a four-helix bundle. Those results showed clearly that the interfacing helix (P1) was static between on and off states of the protein, where instead aspartate dependent movement was observed in a neighboring helix (P4). Additionally, chemotaxis proteins form large lattice networks that are critical to signaling (Briegel et al., 2012; Hall et al., 2012)}. It is likely that the rigidity of this structure prevents any signal dependent structural changes in the periplasmic domains.

2.10 References

Abràmoff, M.D., Magalhães, P.J., and Ram, S.J. (2004). Image processing with ImageJ. Biophotonics international *11*, 36-42.

Albanesi, D., Martín, 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*, 16185-16190.

Alber, F., Dokudovskaya, S., Veenhoff, L., Zhang, W., Kipper, J., Devos, D., Suprapto, A., Karni-Schmidt, O., Williams, R., Chait, B.*, et al.* (2007). The molecular architecture of the nuclear pore complex. Nature *450*, 695-701.

Barnakov, A., Altenbach, C., Barnakova, L., Hubbell, W.L., and Hazelbauer, G.L. (2002). Site-directed spin labeling of a bacterial chemoreceptor reveals a dynamic, loosely packed transmembrane domain. Protein Sci *11*, 1472-1481.

Briegel, A., Li, X., Bilwes, A.M., Hughes, K.T., Jensen, G.J., and Crane, B.R. (2012). Bacterial chemoreceptor arrays are hexagonally packed trimers of receptor dimers networked by rings of kinase and coupling proteins. Proc Natl Acad Sci U S A *109*, 3766-3771.

Butler, S.L., and Falke, J.J. (1998). Cysteine and Disulfide Scanning Reveals Two Amphiphilic Helices in the Linker Region of the Aspartate Chemoreceptor. Biochemistry *37*, 10746-10756.

Canutescu, A.A., Shelenkov, A.A., and Dunbrack, R.L. (2003). A graph-theory algorithm for rapid protein side-chain prediction. Protein Sci *12*, 2001-2014.

Careaga, C.L., and Falke, J.J. (1992). Thermal motions of surface alpha-helices in the D-galactose chemosensory receptor. Detection by disulfide trapping. J Mol Biol *226*, 1219-1235.

Careaga, C.L., Sutherland, J., Sabeti, J., and Falke, J.J. (1995). Large amplitude twisting motions of an interdomain hinge: a disulfide trapping study of the galactose-glucose binding protein. Biochemistry *34*, 3048-3055.

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

Chervitz, S.A., and Falke, J.J. (1995). Lock on/off disulfides identify the transmembrane signaling helix of the aspartate receptor. J Biol Chem *270*, 24043-24053.

Chervitz, S.A., and Falke, J.J. (1996). Molecular mechanism of transmembrane signaling by the aspartate receptor: a model. Proc Natl Acad Sci U S A *93*, 2545-2550.

Chervitz, S.A., Lin, C.M., and Falke, J.J. (1995). Transmembrane signaling by the aspartate receptor: engineered disulfides reveal static regions of the subunit interface. Biochemistry *34*, 9722-9733.

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

Cheung, J., and Hendrickson, W.A. (2009). Structural analysis of ligand stimulation of the histidine kinase NarX. Structure (London, England: 1993) *17*, 190-201.

Cho, U.S., Bader, M.W., Amaya, M.F., Daley, M.E., Klevit, R.E., Miller, S.I., and Xu, W. (2006). Metal Bridges between the PhoQ Sensor Domain and the Membrane Regulate Transmembrane Signaling. J Mol Biol *356*, 1193-1206.

Dago, A.E., Schug, A., Procaccini, A., Hoch, J.A., Weigt, M., and Szurmant, H. (2012). Structural basis of histidine kinase autophosphorylation deduced by integrating genomics, molecular dynamics, and mutagenesis. Proc Natl Acad Sci U S A *109*, 1733-1742.

Daura, X., Gademann, K., Jaun, B., Seebach, D., van Gunsteren, W.F., and Mark, A.E. (1999). Peptide folding: when simulation meets experiment. Angewandte Chemie International Edition *38*, 236-240.

Diensthuber, R.P., Bommer, M., Gleichmann, T., and Möglich, A. (2013). Full-Length Structure of a Sensor Histidine Kinase Pinpoints Coaxial Coiled Coils as Signal Transducers and Modulators. Structure *21*, 1127-1136.

Dunin-Horkawicz, S., and Lupas, A.N. (2010). Comprehensive analysis of HAMP domains: implications for transmembrane signal transduction. J Mol Biol *397*, 1156-1174.

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

Eguchi, Y., Ishii, E., Yamane, M., and Utsumi, R. (2012). The connector SafA interacts with the multi-sensing domain of PhoQ in Escherichia coli. Mol Microbiol *85*, 299-313.

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

Ferris, Hedda U., Dunin-Horkawicz, S., Hornig, N., Hulko, M., Martin, J., Schultz, Joachim E., Zeth, K., Lupas, Andrei N., and Coles, M. (2012). Mechanism of Regulation of Receptor Histidine Kinases. Structure *20*, 56-66.

Galassi, M., and Gough, B. (2005). GNU scientific library: reference manual (Network Theory).

Galperin, M.Y., Nikolskaya, A.N., and Koonin, E.V. (2001). Novel domains of the prokaryotic two-component signal transduction systems. FEMS Microbiol Lett *203*, 11-21.

Gao, R., and Lynn, D.G. (2005). Environmental pH sensing: resolving the VirA/VirG two-component system inputs for Agrobacterium pathogenesis. J Bacteriol *187*, 2182-2189.

García Véscovi, E., Soncini, F.C., and Groisman, E.A. (1996). Mg2+ as an extracellular signal: environmental regulation of Salmonella virulence. Cell *84*, 165-174.

Goldberg, S.D., Clinthorne, G.D., Goulian, M., and DeGrado, W.F. (2010). Transmembrane polar interactions are required for signaling in the Escherichia coli sensor kinase PhoQ. Proc Natl Acad Sci U S A *107*, 8141-8146.

Goldberg, S.D., Soto, C.S., Waldburger, C.D., and DeGrado, W.F. (2008). Determination of the Physiological Dimer Interface of the PhoQ Sensor Domain. J Mol Biol *379*, 656-665.

Gordeliy, V.I., Labahn, J., Moukhametzianov, R., Efremov, R., Granzin, J., Schlesinger, R., Büldt, G., Savopol, T., Scheidig, A.J., Klare, J.P.*, et al.* (2002). Molecular basis of transmembrane signalling by sensory rhodopsin II–transducer complex. Nature *419*, 484-487.

Habeck, M., Rieping, W., and Nilges, M. (2006). Weighting of experimental evidence in macromolecular structure determination. Proc Natl Acad Sci U S A *103*, 1756-1761.

Hall, B.A., Armitage, J.P., and Sansom, M.S. (2012). Mechanism of bacterial signal transduction revealed by molecular dynamics of Tsr dimers and trimers of dimers in lipid vesicles. PLoS Comput Biol *8*, e1002685.

Hazelbauer, G.L. (2012). Bacterial chemotaxis: the early years of molecular studies. Annu Rev Microbiol *66*, 285-303.

Hekkelman, M.L., and Vriend, G. (2005). MRS: a fast and compact retrieval system for biological data. Nucleic Acids Res *33*, W766-769.

Hughson, A.G., and Hazelbauer, G.L. (1996). Detecting the conformational change of transmembrane signaling in a bacterial chemoreceptor by measuring effects on disulfide cross-linking in vivo. Proc Natl Acad Sci U S A *93*, 11546-11551.

Hulko, M., Berndt, F., Gruber, M., Linder, J.U., Truffault, V., Schultz, A., Martin, J., Schultz, J.E., Lupas, A.N., and Coles, M. (2006). The HAMP Domain Structure Implies Helix Rotation in Transmembrane Signaling. Cell *126*, 929-940.

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. Journal of molecular graphics *14*, 33-38.

Kabsch, W., and Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers *22*, 2577-2637.

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

Khatib, F., Cooper, S., Tyka, M.D., Xu, K., Makedon, I., Popović, Z., Baker, D., and Players, F. (2011). Algorithm discovery by protein folding game players. Proc Natl Acad Sci U S A *108*, 18949-18953.

Kindrachuk, J., Paur, N., Reiman, C., Scruten, E., and Napper, S. (2007). The PhoQ-Activating Potential of Antimicrobial Peptides Contributes to Antimicrobial Efficacy and Is Predictive of the Induction of Bacterial Resistance? Antimicrob Agents Chemother *51*, 4374-4381.

Lee, A.I., Delgado, A., and Gunsalus, R.P. (1999). Signal-Dependent Phosphorylation of the Membrane-Bound NarX Two-Component Sensor-Transmitter Protein ofEscherichia coli: Nitrate Elicits a Superior Anion Ligand Response Compared to Nitrite. J Bacteriol *181*, 5309-5316.

Lemmin, T., Soto, C.S., Clinthorne, G., DeGrado, W.F., and Dal Peraro, M. (2013). Assembly of the Transmembrane Domain of E. coli PhoQ Histidine Kinase: Implications for Signal Transduction from Molecular Simulations. PLoS Comput Biol *9*, e1002878.

Lippa, A.M., and Goulian, M. (2009). Feedback Inhibition in the PhoQ/PhoP Signaling System by a Membrane Peptide. PLoS Genet *5*, e1000788.

Lippa, A.M., and Goulian, M. (2012). Perturbation of the oxidizing environment of the periplasm stimulates the PhoQ/PhoP system in Escherichia coli. J Bacteriol *194*, 1457-1463.

Liu, J., Hu, B., Morado, D.R., Jani, S., Manson, M.D., and Margolin, W. (2012). Molecular architecture of chemoreceptor arrays revealed by cryoelectron tomography of Escherichia coli minicells. Proc Natl Acad Sci U S A *109*, E1481-1488.

Lombardi, A., Summa, C.M., Geremia, S., Randaccio, L., Pavone, V., and DeGrado, W.F. (2000). Retrostructural analysis of metalloproteins: application to the design of a minimal model for diiron proteins. Proc Natl Acad Sci U S A *97*, 6298-6305.

Lowe, E.C., Basle, A., Czjzek, M., Firbank, S.J., and Bolam, D.N. (2012). A scissor blade-like closing mechanism implicated in transmembrane signaling in a Bacteroides hybrid two-component system. Proc Natl Acad Sci U S A *109*, 7298-7303.

Lynch, B.A., and Koshland, D. (1991). Disulfide cross-linking studies of the transmembrane regions of the aspartate sensory receptor of Escherichia coli. Proc Natl Acad Sci U S A *88*, 10402-10406.

Mascher, T., Helmann, J.D., and Unden, G. (2006). Stimulus Perception in Bacterial Signal-Transducing Histidine Kinases. Microbiol Mol Biol Rev *70*, 910-938.

Maslennikov, I., Klammt, C., Hwang, E., Kefala, G., Okamura, M., Esquivies, L., Mörs, K., Glaubitz, C., Kwiatkowski, W., Jeon, Y.H.*, et al.* (2010). Membrane domain structures of three classes of histidine kinase receptors by cell-free expression and rapid NMR analysis. Proc Natl Acad Sci U S A *107*, 10902-10907.

Mechaly, A.E., Sassoon, N., Betton, J.M., and Alzari, P.M. (2014). Segmental helical motions and dynamical asymmetry modulate histidine kinase autophosphorylation. PLoS Biol *12*, e1001776.

Metcalf, D.G., Kulp, D.W., Bennett, J.S., and DeGrado, W.F. (2009). Multiple approaches converge on the structure of the integrin αIIb/β3 transmembrane heterodimer. J Mol Biol *392*, 1087-1101.

Milburn, M.V., Prive, G.G., Milligan, D.L., Scott, W.G., Yeh, J., Jancarik, J., Koshland, D.E., Jr., and Kim, S.H. (1991). Three-dimensional structures of the ligand-binding domain of the bacterial aspartate receptor with and without a ligand. Science *254*, 1342-1347.

Miller, J.H. (1972). Experiments in molecular genetics (Cold Spring Harbor Laboratory).

Miller, S.I., Kukral, A.M., and Mekalanos, J.J. (1989). A two-component regulatory system (phoP phoQ) controls Salmonella typhimurium virulence. Proc Natl Acad Sci U S A *86*, 5054-5058.

Molnar, K.S., Bonomi, M., Pellarin, R., Clinthorne, G.D., Gonzalez, G., Goldberg, S.D., Goulian, M., Sali, A., and DeGrado, W.F. (2014). Cys-scanning disulfide crosslinking and bayesian modeling probe the transmembrane signaling mechanism of the histidine kinase, PhoQ. Structure *22*, 1239-1251.

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

Moore, J.O., and Hendrickson, W.A. (2012). An Asymmetry-to-Symmetry Switch in Signal Transmission by the Histidine Kinase Receptor for TMAO. Structure *20*, 729-741.

Neidhardt, F.C., Bloch, P.L., and Smith, D.F. (1974). Culture medium for enterobacteria. J Bacteriol *119*, 736-747.

Pakula, A.A., and Simon, M.I. (1992). Determination of transmembrane protein structure by disulfide cross-linking: the Escherichia coli Tar receptor. Proc Natl Acad Sci U S A *89*, 4144-4148.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem *25*, 1605-1612.

Pontius, J., Richelle, J., and Wodak, S.J. (1996). Deviations from standard atomic volumes as a quality measure for protein crystal structures. J Mol Biol *264*, 121-136.

Rieping, W., Habeck, M., and Nilges, M. (2005). Inferential structure determination. Science *309*, 303-306.

Robinson, V.L., Buckler, D.R., and Stock, A.M. (2000). A tale of two components: a novel kinase and a regulatory switch. Nat Struct Biol *7*, 626-633.

Royant, A., Nollert, P., Edman, K., Neutze, R., Landau, E.M., Pebay-Peyroula, E., and Navarro, J. (2001). X-ray structure of sensory rhodopsin II at 2.1-Å resolution. Proc Natl Acad Sci U S A *98*, 10131-10136.

Russel, D., Lasker, K., Webb, B., Velazquez-Muriel, J., Tjioe, E., Schneidman-Duhovny, D., Peterson, B., and Sali, A. (2012). Putting the pieces together: integrative modeling platform software for structure determination of macromolecular assemblies. PLoS Biol *10*, e1001244.

Schultz, J.E., and Natarajan, J. (2013). Regulated unfolding: a basic principle of intraprotein signaling in modular proteins. Trends Biochem Sci *38*, 538-545.

Senes, A., Chadi, D.C., Law, P.B., Walters, R.F.S., Nanda, V., and DeGrado, W.F. (2007). E-z, a depth-dependent potential for assessing the energies of insertion of amino acid side-chains into membranes: Derivation and applications to determining the orientation of transmembrane and interfacial helices. J Mol Biol *366*, 436-448.

Senes, A., Engel, D.E., and DeGrado, W.F. (2004). Folding of helical membrane proteins: the role of polar, GxxxG-like and proline motifs. Curr Opin Struct Biol *14*, 465-479.

Shindyalov, I.N., and Bourne, P.E. (1998). Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. Protein Eng *11*, 739-747.

Sivia, D., and Skilling, J. (2006). Data analysis: a Bayesian tutorial.

Soncini, F.C., García Véscovi, E., Solomon, F., and Groisman, E.A. (1996). Molecular basis of the magnesium deprivation response in Salmonella typhimurium: identification of PhoP-regulated genes. J Bacteriol *178*, 5092-5099.

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

Studdert, C.A., and Parkinson, J.S. (2007). In vivo crosslinking methods for analyzing the assembly and architecture of chemoreceptor arrays. Methods Enzymol *423*, 414-431.

Summa, C.M., Lombardi, A., Lewis, M., and DeGrado, W.F. (1999). Tertiary templates for the design of diiron proteins. Curr Opin Struct Biol *9*, 500-508.

Swain, K.E., Gonzalez, M.A., and Falke, J.J. (2009). Engineered socket study of signaling through a four-helix bundle: evidence for a yin-yang mechanism in the kinase control module of the aspartate receptor. Biochemistry *48*, 9266-9277.

Waldburger, C.D., and Sauer, R.T. (1996). Signal detection by the PhoQ sensor-transmitter. Characterization of the sensor domain and a response-impaired mutant that identifies ligand-binding determinants. J Biol Chem *271*, 26630-26636.

Wang, C., Sang, J., Wang, J., Su, M., Downey, J.S., Wu, Q., Wang, S., Cai, Y., Xu, X., and Wu, J. (2013). Mechanistic Insights Revealed by the Crystal Structure of a Histidine Kinase with Signal Transducer and Sensor Domains. PLoS Biol *11*, e1001493.

Xie, W., Dickson, C., Kwiatkowski, W., and Choe, S. (2010). Structure of the Cytoplasmic Segment of Histidine Kinase Receptor QseC, a Key Player in Bacterial Virulence. Protein Pept Lett *17*, 1383-1391.

Yeh, J.I., Biemann, H.P., Pandit, J., Koshland, D.E., and Kim, S.H. (1993). The three-dimensional structure of the ligand-binding domain of a wild-type bacterial chemotaxis receptor. Structural comparison to the cross-linked mutant forms and conformational changes upon ligand binding. J Biol Chem *268*, 9787-9792.

Yohannan, S., Faham, S., Yang, D., Whitelegge, J.P., and Bowie, J.U. (2004). The evolution of transmembrane helix kinks and the structural diversity of G protein-coupled receptors. Proc Natl Acad Sci U S A *101*, 959-963.

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*, 801-814.

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

CHAPTER THREE

3.1 Abstract

Bacteria transduce signals across the membrane using two-component systems, consisting of a membrane-spanning sensor histidine kinase and a cytoplasmic response regulator. The histidine kinase, PhoQ, serves as a master regulator of virulence response in S. typhimurium and E. coli. It also is inhibited by divalent cations, particularly Mg2+. While the periplasmic sensor domain of this protein has a unique function, the cytoplasmic portion of this modular protein is made of structurally conserved domains found in many other bacterial sensor kinases. Signal transduction through these conserved domains is thought to be universal; however, the structural and energetic rearrangements that occur during signaling have generated numerous models.

To unify the various signaling mechanisms that exist for the different protein domains, we built a thermodynamic model based on Linked Equilibrating Domains (LED). We used this model to quantitatively interpret functional data of single-point Ala, Phe and Cys mutants throughout the signal transducing regions of PhoQ. Data from 35 mutants, including both activating and deactivating phenotypes, were globally fit using LED and gross features such as V_{max} and K_d were related to more nuanced population distributions and thermodynamic coupling. LED analysis highlights the principles by which individual signaling domains can be connected to create a functional signal transducer. These principles allow us to quantitatively explain signaling in histidine kinases and are likely to be broadly applicable to many other signal transduction proteins.

3.2 Introduction

Sensor histidine kinases sample require at least two states to control signaling. Defining a mechanism for these bifunctional proteins requires defining both the kinase state that is competent for autophosphorylation and a phosphatase state that is able to bind and dephosphorylate the response regulator. In the previous chapter, we show that HK proteins, like

PhoQ, sample distinct structural states during the catalytic cycle (Molnar et al., 2014). In this chapter, we build a framework to describe the populations of these critical kinase and phosphatase states.

We undertook a mutational analysis of PhoQ to perturb the signal transduction helices of PhoQ. From our own work and others, we know that a single point mutation can affect the signaling of the entire protein (Goldberg et al., 2010; Sanowar et al., 2003). Having established that these mutants are well folded and properly inserted into the membrane, we sought to understand the mechanistic basis of how these mutations alter signaling. Is there are larger framework within which varied mutant phenotypes can be clustered and explained? What do these results tell us about how PhoQ transmits a signal across the membrane? To answer these questions we expanded upon our previous cysteine mutant scanning of the periplasmic helical interface, and scanned the transmembrane region using tryptophan, the HAMP domain using both alanine and phenylalanine, and finally, the S-helix link using a leucine scan. Variety in the mutant reside adds depth to the analysis beyond a straightforward Ala-scan, and is critical to the success of this analysis, which spans topologically different parts of this large multi-domain protein.

In order to quantitatively describe the results from the mutational scan we develop a thermodynamic framework to rigorously describe signaling though multi-domain proteins that goes beyond the current models. Previous models have gravitated towards a single physical motion (e.g. piston shift or gear rotation) to transmit a signal (Chervitz and Falke, 1996; Falke and Hazelbauer, 2001; Ferris et al., 2012; Hazelbauer, 2012). Instead, we postulate that each domain acts as an independent folding unit with two distinct states. Switching of the sensor domain from a kinase-permissive to a phosphatase-permissive state influences the equilibrium constant for the switching of the subsequent connected domains (TM, HAMP, and S-helix, Fig. 3.1). Finally, the S-helix transmits the signal to the DHp domain, which undergoes a conformational change that is coupled to the energetics of forming the Michaelis complex with the catalytic ATP-binding domain (Mechaly et al., 2014). The architecture of histidine kinases, a sequential connection of multiple, independent domains, lends itself to description by a series of thermodynamically linked equilibria

(Grigoryan and DeGrado, 2008). Here, energetic effect of one domain on the neighboring domain is considered as the coupling energy between domains. Signal transduction can propagate through this coupled system in two ways. Under a rigid coupling limit the coupling interacting effectively reduces a multi-state system into a two-state (on and off). In a weak coupling limit the system can populate various intermediates which can affect overall activity.

We used this framework to globally analyze 42 sets of magnesium dependent LacZ reporter assay curves from 35 different PhoQ mutants (the remaining 7 were wild-type) to obtain estimates for the internal equilibrium constants and coupling terms. With this method, we explore the energetic coupling between transitions within the domains, and the critical role played by interconnecting helices or linking sequences. This model connects the energetic transition of each individual domain as interconnecting parts to the whole signaling molecule, and allows for the articulation of these relationships; therefore, we can now, for the first time, quantitatively explain signaling in this class of proteins.

This work required careful optimization of the conditions for measurement of transcriptional activity of PhoP/Q, construction of hundreds of mutants, measurement of their activities, and the development of a mathematical model to describe inter-domain signaling. I performed each aspect of this work independently, with the exception of the construction of a number of mutants, which were cloned by Graham Clinthorne.

3.3 Results

Selection of mutants. Mutations were made at positions along the signaling pathway. We define this pathway as the helices that line the dimer interface, and it includes: the P-helix along the external PAS domain, the helices comprising the TM and HAMP helix bundles, and the S-helix. We specifically excluded the sensor domain (except for dimeric interfacing P-helix that connects to TM1), DHp, and catalytic domains. With these exclusions we are left with about 180 AAs to consider for mutation. Along the signaling pathway, 125 unique mutations were made on the pTrc plasmid that has been described in the lab's previous work, which relies on leaky expression of the plasmid to produce protein (Goldberg et al., 2008). Mutant plasmids were transformed into a

PhoQ knock-out strain (TIM206) from the Goulian lab to remove any contribution from endogenous protein.

We typically evaluated three different mutations: Ala, Cys, and Phe at each position in the regions of interest. Ala was chosen as a small residue that is standardly used to test the role of specific residue's contributions to structure and function, Cys was used in cross-linking studies so it was important to test the effect of mutation on function, and Phe was used to test the effect of adding a large residue. For regions such as the TM or the S-helix we mutated residues to produce a desired perturbation. Tryptophan is known to preferentially reside at the boundary between lipid head groups and the aliphatic tails, and leucine mutation in the S-helix was chosen to stabilize a coiled coil conformation, which might occur in one state of the protein.

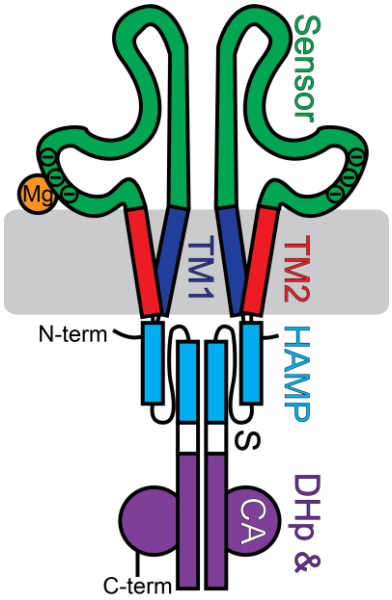


Figure 3.1 – PhoQ dimer topology. The portion of the protein that was mutated is along the dimeric interface, specifically, the dark blue TM1, green sensor p-helix, red TM2, teal HAMP domain, and white S-helix residues.

All of the phenotypes generated from the point mutants can be classified along a phosphatase to kinase spectrum. Where the WT protein generates a large fold-change between low and high magnesium conditions (approximately 10-fold change or larger), and the mutants can alter this into two directions. When the mutation drives PhoQ to a kinase state, the activity at high magnesium increases to a limit where the protein is active yet unresponsive to magnesium (Fig. 3.2, E261A and N255F). The reverse is true in the phosphatase state (Fig. 3.2, L247A, E233F).

The spectrum of activities of the various PhoQ proteins varies with respect to the nature of the scanning mutation (Fig. 3.3). For the cysteine scan, these activity assays give us an idea of what states of the protein the mutant represents, and therefore what state of the protein we are probing with cross-linking. For example, V121C has very little activity with or without magnesium. This suggests that any crosslinking data we get form this particular mutant will only represent the

phosphatase or off-state of the protein. This is in contrast to a mutant K186C that has completely WT-like activity.

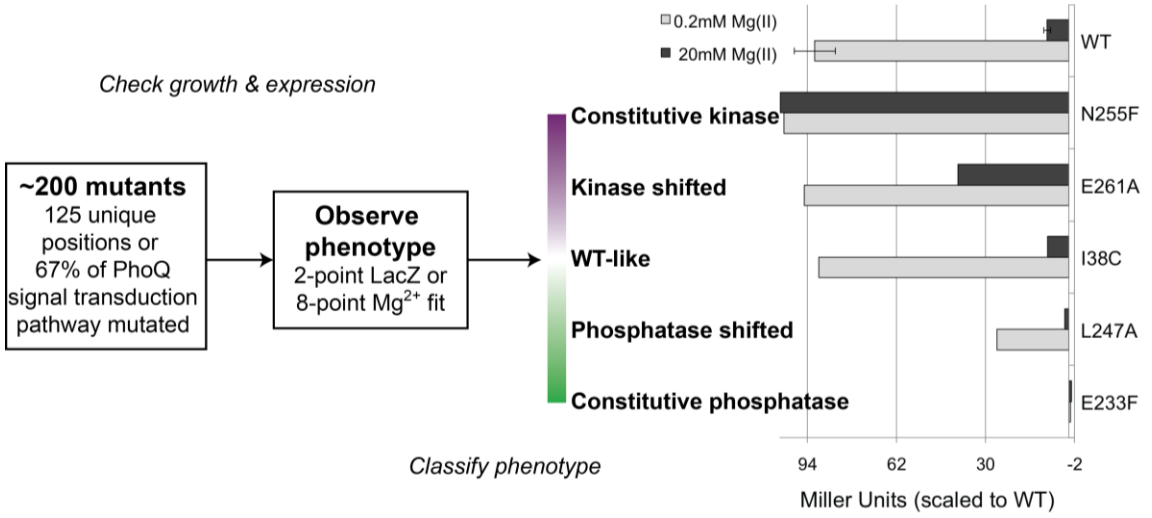


Figure 3.2 – Flow chart of PhoQ mutations. Over 200 mutations were made along the signaling pathway. For the majority of those a 2-point LacZ assay was done to observe the general phenotype. Only 35 mutants were used in the 8-point Mg^{+2} curves. Illustrations of the different phenotypes from a 2-point LacX assay are shown at left.

The next scan was a tryptophan scan at the edges of the TM helices. The propensity of this mutation to selectively partition in the bilayer could nudge the TM helices into or out of the membrane. We first noticed there were many positions in and surrounding PhoQ TM2 that were already Trp (5 residues), and we also observed many positions were able to tolerate Trp mutations in both TM1 and TM2 (9 residues). While Trp-scans such as these are frequently analyzed in terms of the influence of the Trp residue on the location of the TM helices within the membrane (Adase et al., 2012; Bissonnette et al., 2009; Draheim et al., 2005; Lin et al., 2014), the analysis is complicated by the fact that Trp is a large bulky amino acid.

The Ala/Phe scan of the HAMP was performed to assess the effects of large and small mutations. In the case of S217, a mutation to Ala slightly reduces the activity, whereas a Phe mutation completely ablates signaling. At this residue we have modeled the dimer interface to be tightly packed, so it seems plausible that a steric clash could interrupt signaling. However, the S217W mutation increases activity. Finally, a small Leu scan of the S-helix was carried out to test the dimer interface. We selected residues located at helical wheel positions a and d by homology

to CpxA (4BIV/4BIU) (Mechaly et al., 2014). Many of these mutants do not exhibit WT-like activity profiles though this is expected because we selected residues at the dimer interface.

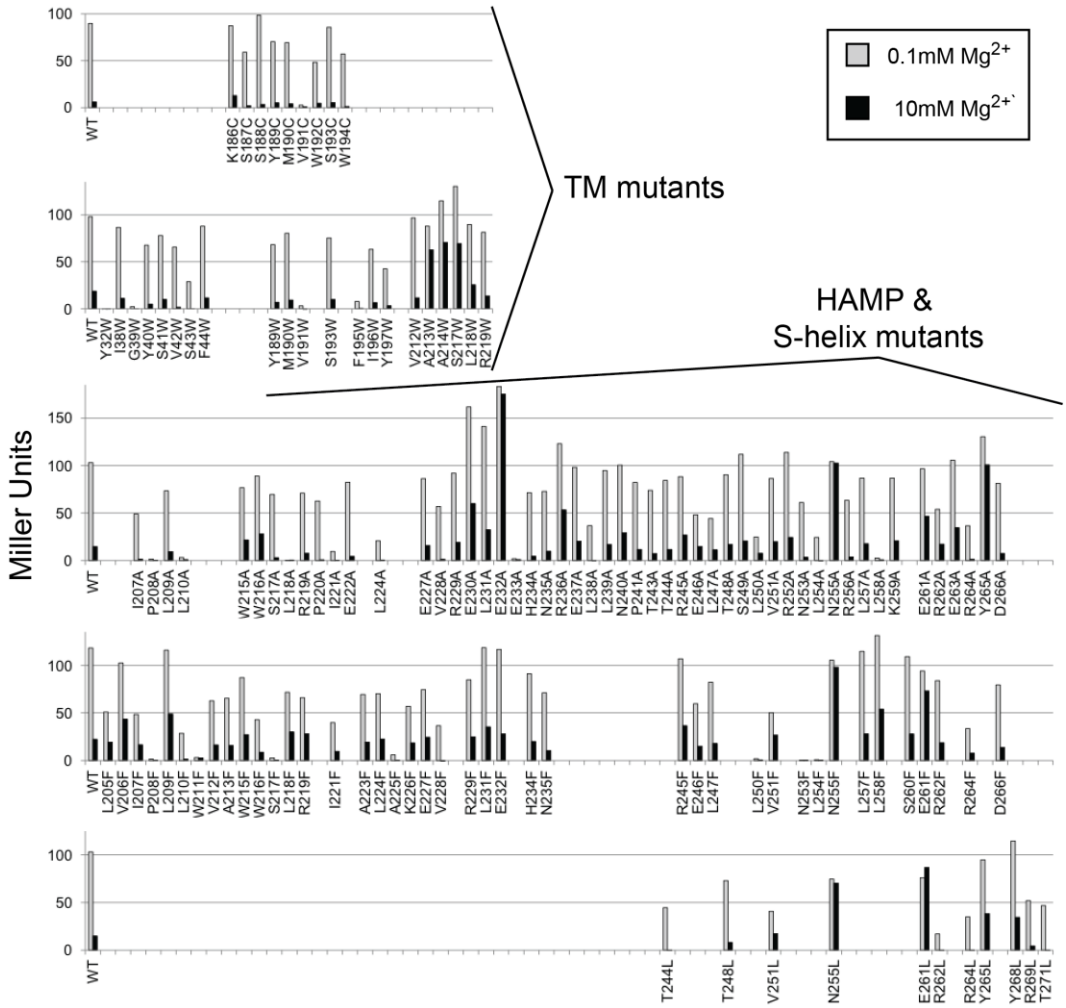


Figure 3.3 – 2-point LacZ assays of various mutant scans. Top most bar graph is a Cys scan, followed by Trp, Ala, Phe and Leu at the bottom. The WT results are at the beginning of each graph.

The mutational landscape from the 2-point LacZ shows the largest population of mutants displays a muted activity where the signal at low magnesium is reduced between 25-75% from the WT strain and the high magnesium signal might also be reduced, but it is not always perturbed. The second group of mutants can maintain near WT-like signaling. Finally the remaining mutants have the most severe phenotypes. These can be classified as a greater than 75% loss of signal at low magnesium (L210A) or a large gain in signal at high magnesium (E232A). Positions that were tested with multiple mutants also gave us trends. As expected, a

TM2 proline, residue 208, cannot tolerate Ala or Phe and all signaling is lost. Surprisingly, residue 255 of the HAMP helix 2 cannot tolerate changes (A, F, L) and becomes signal blind but completely active.

Beyond this general classification, it is difficult to say more about these particular mutants from just the 2-point data. We therefore initiated a study of select mutants, representing each phenotype at multiple Mg^{2+} concentrations to see how the mutants varied in respect to their avidity for this ion and how this related to their enzymatic activities.

Response of mutants to variable concentrations of Mg2+. With a general idea of mutant phenotype, we selected mutants in that would probe each part of the linking domains (i.e. HAMP H1, loop, and H2 mutants), and we avoided mutants that were known to have a constitutive phenotype (ON or OFF) as those curves would be difficult to fit. Activity based on reporter assays was measured over the full range of Mg^{2+} concentrations for selected mutants. Curves for 35 mutants plus seven repetitions of WT were evaluated using a modified Hill equation (Eqn. 7). The derived parameters varied quite significantly, in terms of the degree of activity extrapolated to very low and high [Mg^{2+}] as well as the midpoint for Mg^{2+} (K_{app}). The variation in K_{app} was quite significant, even for mutants with replacements in the intracellular domains, which are far removed from the extracellular Mg^{2+}-binding site. Again, we see examples of mutants with large activity changes from the linking regions of the protein, TM and HAMP (Fig. 3.4). Additionally, a

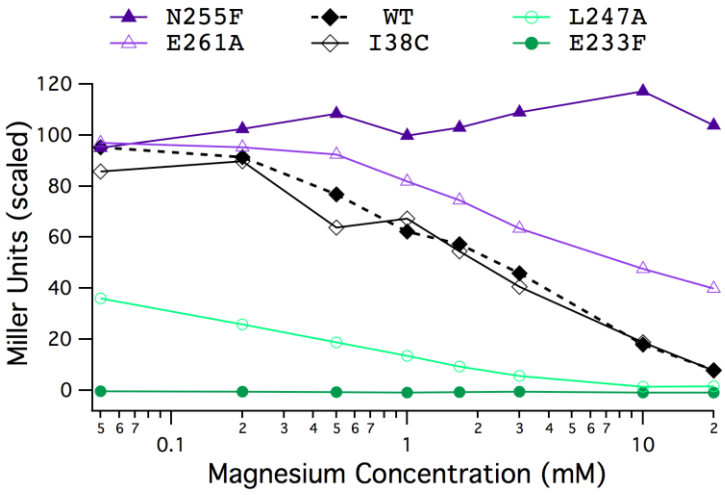


Figure 3.4 – Diversity of Mg binding curves generated by point mutants. WT reference curve is shown in dashed line (black). Mutants and color-coding match the previous figure 3.2.

WT-like mutant from the periplasmic domain is included to demonstrate that extracellular mutations are not always involved with Mg^{2+} binding.

This diversity causes the apparent binding of magnesium to shift quite drastically, despite selecting intracellular mutants that are distant from both the magnesium binding loops and the catalytic domain. The average K_{app} for the WTs is 2mM and the range spans four orders of magnitude from 4μM to 40mM. We thought it unlikely that mutations between 50 and 70Å away from the outside of the periplasmic membrane would be able to alter magnesium affinity that drastically. More likely the altered phenotype of these mutations was a result of perturbing the *thermodynamic coupling* between the Mg^{2+}-binding sensor domain and the other signal-transducing domains. In this model, the sensor domain changes conformation upon binding Mg^{2+}, and this information is relayed to the DHp/kinase domain via the intermediate linking domains. Thus, changes that stabilize the phosphatase-competent forms of the signal-transducing domains will increase the avidity with which the sensor domains bind Mg^{2+}.

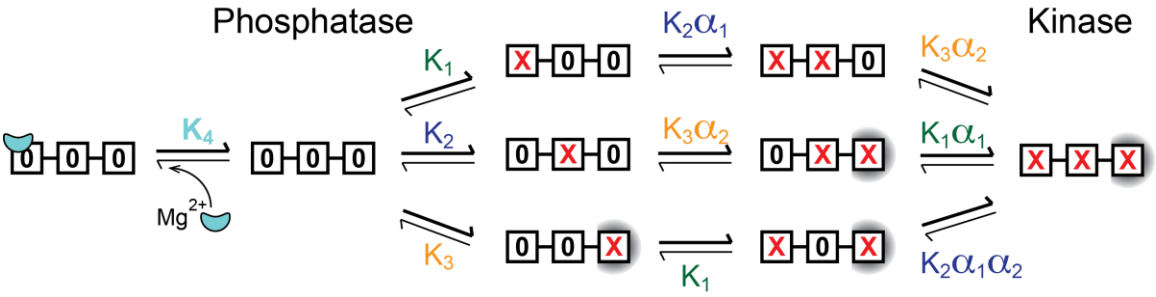


Figure 3.5 – Linked Equilibrating Domains (LED) model. We represent PhoQ as a three coupled domains. Each of the domains can switch between phosphatase (0) or kinase (X) state. Here we define the various pathways and the equilibrium constants involved in the pathway. We assume that only when the first domain is in a phosphatase state can it interact with magnesium. Here we show only state [000] interacting with magnesium for clarity.

Thermodynamic framework. To parse the contributions of the various domains and their ability to communicate with each other, we defined an equilibrium framework to describe the system. We know the catalytic region of the protein minimally can access two distinct states (phosphatase and kinase), and the sensor domain can likewise have two states (Mg^{2+} bound and free). We similarly postulate two-state behavior to each of the individual domains of PhoQ. We can then link

these domains together and define a term that accounts for the thermodynamic coupling between the domains. While this large protein could be divided up into many different domains, a three-domain model was selected for three basic reasons: 1) to differentiate sensing/binding, linking, and catalytic parts of the protein; 2) to build in enough complexity to the model so the equilibrium constants gain physiological meaning; 3) to remain simple enough to avoid over-fitting the data.

From the diagram (Fig. 3.7) we assume the protein has two conformations (X and 0) for each domain, designated from left to right as the sensor, linking, and DHp/catalytic domain. By definition, the phosphatase-promoting X state of the sensor, is the Mg^{2+}-bound state, and the kinase-promoting 0 state is the apo-state. The phosphatase and kinase-permissive conformations of the middle linker and DHp and catalytic domains receive the same designation. The dimensionless constants, $K_1 – K_3$ are defined as microequilibrium constants describing the ratio of the equilibrium population of the 0 vs. X state for a given domain when the neighbors on either side are both in the 0 state. If the neighboring domain(s) are in the X state, the equilibrium constants are multiplied by a second term α_i which considers whether the neighbor to the left is in the X state. For example, $K_2\alpha_1$ represents both a ratio of populations of the second transducing domain when the first sensor domain is already in the X state; in this expression α_1 represents the coupling of the domains. Finally, constant K_4 is defined as the dissociation constant (1 M standard state) for dissociation of Mg^{2+} from the sensor domains in the phosphatase-promoting 0 state.

There are several assumptions (Eqn. 1) associated with this equilibrium model. From the analysis of the Hill coefficient it is clear that the binding of a single Mg(II) to the dimer is sufficient to strongly stabilize the phosphatase permissive state. Additional ions might, however, bind with lower affinities, although the lack of strong cooperativity would suggest that they are not critically involved in signaling. We also assume that Mg(II) binds to the sensor only when it is in the phosphatase-promoting 0 state. Therefore, only states where there is a (0) in the first box will bind Mg^{2+} (Eqn. 1). Moreover, because the mutations are made distant from the Mg(II)-binding site we assume that the intrinsic constant associated with Mg(II), K_4, should be the same for all the mutants (i.e. changes in Mg(II) affinity relate to indirect effects associated with the energetic

transition of the sensor from the X to 0 state). We will see that this is true for all the mutations except for three adjacent to the Mg(II)-binding acidic loops of the sensor domain.

Equation (1)

$$K_4 = \frac{[Mg^{2+}][000]}{[000 \cdot Mg^{2+}]} = \frac{[Mg^{2+}][0X0]}{[0X0 \cdot Mg^{2+}]} = \frac{[Mg^{2+}][00X]}{[00X \cdot Mg^{2+}]} = \frac{[Mg^{2+}][0XX]}{[0XX \cdot Mg^{2+}]}$$

From this equation we can relate the magnesium concentration to the stability of specific states. Next, we would like to relate the states in this equilibrium (Fig. 3.5) to the signal that we see from the reporter assay. The fraction of the protein in a kinase state is obtained from the summation of all catalytically active states (all states with (X) in the third box) over the summation all possible states (Eqn. 2).

Equation (2)

$$f(kinase) = \frac{\sum kinase\ states}{\sum all\ states} = \frac{[00X] + [0XX] + [X0X] + [XXX]}{[000] + [X00] + [0X0] + [00X] + [XX0] + [0XX] + [X0X] + [XXX]}$$

We can relate the all of the states to one reference state, [000], via the fitting parameters and a bit of algebra. Once each state is written in terms of the reference state, shown in Eqn. 3 (and Methods), the reference state will cancel from both numerator and denominator and Eqn. 2 will be in terms of the fitting parameters.

Equation (3)

$$K_2 = \frac{[0X0]}{[000]} \qquad [0X0] = K_2[000]$$

We introduce our final fitting parameter in the final fitting equation (Eqn. 4). The amount of signal that is produced in the reporter assay will be the fraction of the population in the kinase state multiplied by V_{max}, the hypothetical maximal value for the system when the DHp/catalytic domain if fully in the active X state under the conditions of our assay.

Equation (4)

$$MU = V_{max} * f(kinase)$$

Insertion of equation 2 into equation 4, after introduction of the various equilibrium constants, generates Equation 5, which describes the activity of a given mutant in terms of six variables consisting of V_{max}, $K_1 - K_3$, α_1, and α_1.

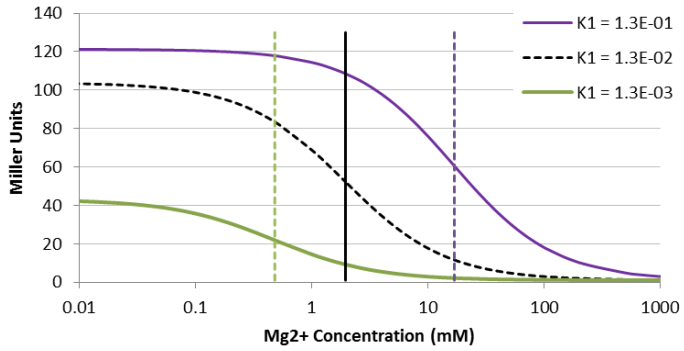
Equation (5)

$$MU = \left(\frac{V_{max}\ K_3\alpha_2 \left(K_1K_2\alpha_1 + K_2 \left(1 + \frac{[Mg^{2+}]}{K_4} \right) \right)}{1 + \frac{[Mg^{2+}]}{K_4} + K_2 \left(1 + \frac{[Mg^{2+}]}{K_4} \right) + K_2K_3\alpha_2 \left(1 + \frac{[Mg^{2+}]}{K_4} \right) + K_1 + K_1K_2\alpha_1 + K_1K_2\alpha_1K_3\alpha_2} \right)$$

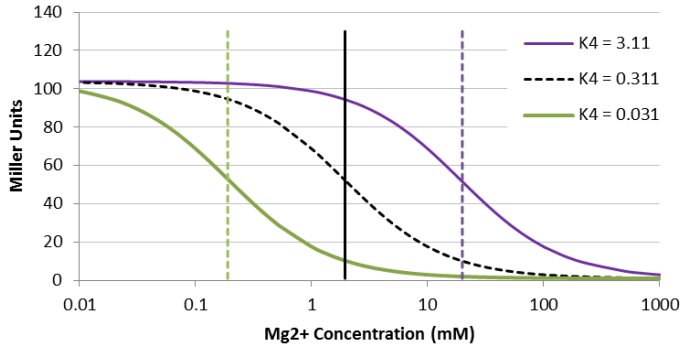
Theoretical curves. To explain this model, it is important to assess how individual parameters affect the midpoint and maximal/minimal values of theoretical curves. Using Eqn. 5, each of the parameters was varied individually while holding the rest constant. To compare how each coefficient can change the fit, the same theoretical trace (dotted, black curve) was plotted on each graph in Fig. 3.6. Midpoints of the graphs are designated with a vertical line.

The simplest parameter to interpret is K_4 (the intrinsic Mg^{2+} binding constant), which moves the midpoint of the dose response curves to the right or left without changing the upper or lower limits of the curves (Fig. 3.6, B). By contrast, varying K_1 (also, $K_2\alpha_1$ or $K_3\alpha_2$) while holding the remaining coefficient constant, shifts the midpoint to the right or left as the constants are increased or decreased, but the limits of the curve also change. Most drastically the limit at low Mg^{2+} changes with these particular coefficients (Fig. 3.6, A). Still other parameters change the curves in additional unique ways. K_2 changes the lower limit and slightly changes the midpoint, while V_{max} changes the upper limit of the plot, but the midpoint remains constant (Fig. 3.6, C & D). For each individual mutant, we designate which of these coefficients to vary. The initial estimate is based on the location in the protein; however, in instances where that designation leads to a poor fit, we can use the behaviors seen in Fig. 3.6 to select an additional variable coefficient.

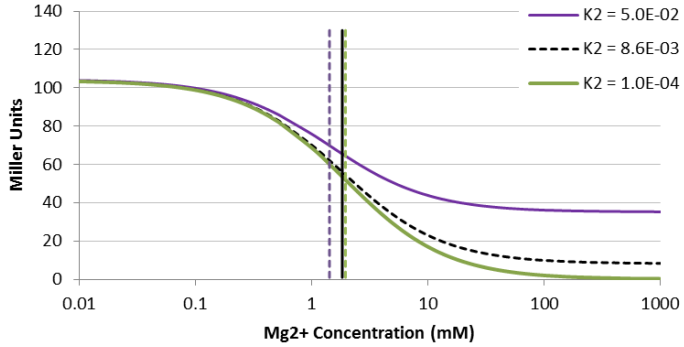
A



B



C



D

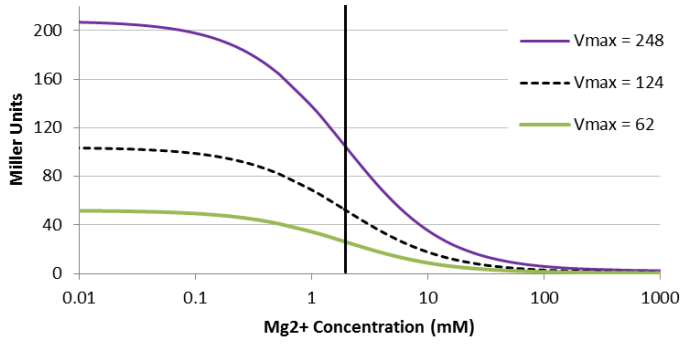


Figure 3.6 – Theoretical curves varying four equilibrium constants over several orders of magnitude. In all cases, values similar to WT are in black. Purple traces show effects of increased coefficients and green show effects of decreasing.

(A) Perturbations in K_1, which are similar to what we expect for other coefficients (specifically, $K_2\alpha_1$ and $K_3\alpha_2$)

(B) Perturbations in K_4. In contrast to the top graph, we see the limits in the graph are the same and only the mid point changes.

(C) The perturbations in K_2 coefficients are varied slightly more than 10-fold for visualization purposes. This constant sets the lower limit of the graph.

(D) The perturbations in V_{max} do not alter the midpoint and mostly affect the values at low magnesium.

Table 3.1 – Coefficients for the LED fit

Mutant	$K_1K_2\alpha_1$	error	K_2	error	$K_3\alpha_2$	error	Location	$\Delta\Delta G$
L28C	0.24	0.074	1.0E-04	held	8.17	2.25	Peri	-0.81
Y32C	0.21	0.066	1.0E-04	held	8.17	2.25	Peri	-0.88
M34C	0.24	0.076	1.0E-04	held	8.17	2.25	Peri	-0.79
A36C	0.18	0.055	1.0E-04	held	8.17	2.25	Peri	-0.98
I38C	0.81	0.273	1.0E-04	held	8.17	2.25	Peri	-0.04
Y40C	0.42	0.134	1.0E-04	held	8.17	2.25	Peri	-0.46
S41C	0.32	0.102	1.0E-04	held	8.17	2.25	Peri	-0.62
F49C	0.13	0.038	1.0E-04	held	8.17	2.25	Peri	-1.20
R53C	31.43	2650	1.0E-04	held	8.17	2.25	TM1	-4.42
G54C	0.07	0.027	1.0E-04	held	8.17	2.25	TM1	-3.52
E55C	0.20	0.063	1.0E-04	held	8.17	2.25	TM1	-0.90
L58C	0.07	0.041	1.0E-04	held	8.17	2.25	TM1	-4.23
I196W	0.30	0.094	1.0E-04	held	8.17	2.25	TM2	-0.67
I207A	0.18	0.054	1.0E-04	held	8.17	2.25	TM2	-0.99
L209A	0.61	0.202	1.0E-04	held	8.17	2.25	TM2	-0.22
A214W	1.52	0.671	1.4E-01	4.0E-02	8.17	2.25	TM2	0.35
S217A	0.25	0.079	1.0E-04	held	8.17	2.25	TM2-link	-0.77
S217W	4.01	1.731	2.4E-02	1.6E-02	8.17	2.25	TM2-link	0.96
L218F	0.32	0.102	1.0E-04	held	8.17	2.25	TM2-link	-0.62
I221F	0.86	0.285	1.0E-04	held	1.65	0.41	HAMP H1	-1.00
E222A	0.86	0.285	1.0E-04	held	5.38	1.44	HAMP H1	-0.26
L224F	0.86	0.285	1.0E-04	held	3.77	0.98	HAMP H1	-0.48
K226F	0.86	0.285	1.0E-04	held	3.62	0.93	HAMP H1	-0.51
R229F	0.86	0.285	1.4E-02	7.3E-03	5.20	1.37	HAMP H1	-0.28
E232F	0.86	0.285	2.2E-02	8.6E-03	7.33	2.03	HAMP H1	-0.07
R236A	21.51	10.427	1.9E-02	1.3E-01	1.43	0.17	HAMP L	0.74
N240A	0.86	0.285	1.9E-02	8.0E-03	6.90	1.90	HAMP L	-0.11
R245F	0.86	0.285	6.8E-03	4.7E-03	28.0	10.2	HAMP H2	0.77
L247A	0.86	0.285	1.0E-04	held	0.91	0.23	HAMP H2	-1.37
V251A	0.86	0.285	1.0E-04	held	3.28	0.84	HAMP H2	-0.57
R252A	0.86	0.285	1.0E-04	held	4.26	1.11	HAMP H2	-0.41
R256A	0.86	0.285	1.0E-04	held	2.11	0.52	HAMP H2	-0.84
E261A	0.86	0.285	3.3E-02	1.2E-02	12.6	3.93	S-helix	0.27
Y265A	0.86	0.285	3.2E-02	1.7E-02	38.7	17.9	S-helix	0.97
R269L	0.86	0.285	1.0E-04	held	2.32	0.58	S-helix	-0.79
WT1-7	0.86	0.285	1.0E-04	held	8.17	2.25	-	-

Green shading indicatates for which mutants the parameter was involved in the global fit.
V_{max}(global fit)= 120 $\pm$ 3.4; K_4(global fit)= 0.24 $\pm$ 0.03
K_4(R53C)= 1E04 $\pm$ 1E02; K_4(G54C) = 0.17 $\pm$ 0.08; K_4(L58C)= 0.06 $\pm$ 0.06

Global fits of activity data to thermodynamic model. We next fit the parameters in Eqn. 5 to the experimental data from 42 curves (7 wild-type and 35 mutants). To allow fitting of the entire data set we used global analysis, in which the same value of a given parameter is required to fit all the curves for which it is globally fit. For any mutation no more than two parameters were allowed to vary locally, while the remainders were treated as global variables. The position of the mutation in the protein determines which parameters are fit globally versus locally (Table 2, Methods). For

example, a mutant in the periplasmic domain that shifts the apparent binding of magnesium likely does that by changing the equilibrium constant K_1 as opposed to $K_3\alpha_2$. By contrast a S-helix mutant is likely to perturb $K_3\alpha_2$ and we expect that mutant to have a WT K_1 value. Table 1 lists the 6 parameters for the LED fit, the location of the mutation, the change in volume, and an apparent $\Delta\Delta G$ for each residue's contribution to the overall process of proceeding from the 000 to the XXX state. By calculating the values of $\Delta\Delta G$ in terms of the overall process, we circumvent errors associated with covariance in individual terms.

An outstanding fit was obtained using this model as shown in Figure 3.7. Despite the large number of curves analyzed, some parameters were individually not well defined because they covaried with a second parameter (see Methods). For example, the product $K_1{*}K_2\alpha_1$ was well defined in the fitting procedure even though the individual parameters were not uniquely defined. In total, five parameters were fit for both mutant and WT data: V_{max}, $K_1K_2\alpha_1$, $K_3\alpha_2$, and K_2. We discuss their qualitative meaning below. For an in depth discussion of parameter selection and for the final fitting expression (Eqn. 10), see methods.

V_{max} could be globally defined, which implies that the DHp-catalytic domain is functionally the same in all the mutants described here, and the degree of activity was defined only by the fraction that it assumed the kinase-on configuration. K_4 defines the equilibrium constant for binding of Mg^{2+} to the sensor in the phosphatase-promoting O state. As expected, it could be globally fit for all except a few mutants, all of which were in the P-helix immediately adjacent to the sensor domain.

K_2 defines the stability of the signal-transducing TM-HAMP domain. As seen in Fig. 3.6 (bottom graph) K_2 leads to an increase in the minimum at high Mg2+ concentration as its value approaches 1. This can be understood as functional decoupling, as TM-HAMP can achieve the kinase-on state even when the sensor is in the ligand-free O state. Accordingly K_2 was found to be less than 0.001 for all WT mutants. For the partially to fully decoupled "kinase-on" mutants its value was well defined (Table 1).

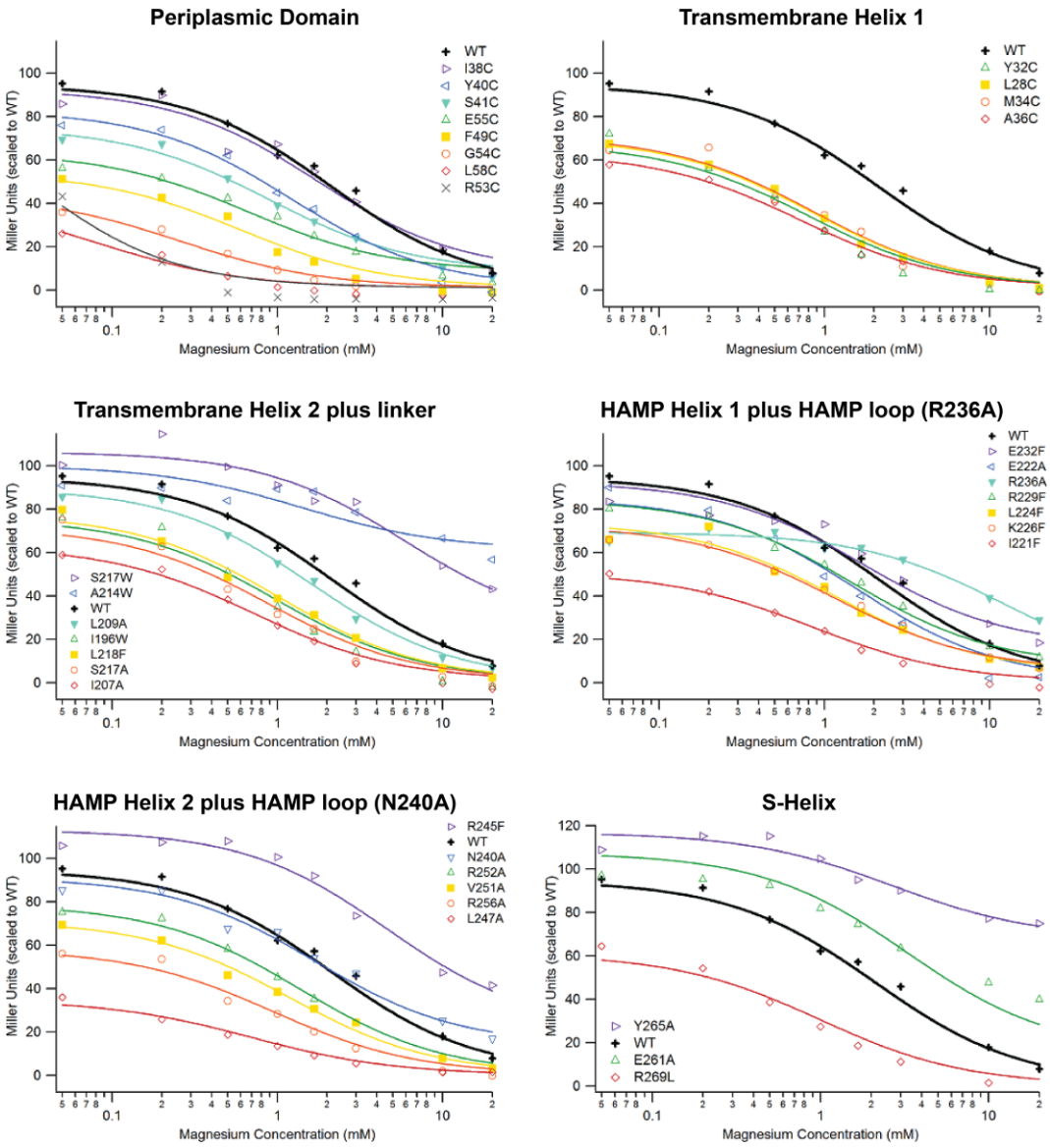


Figure 3.7 – LED fits to raw data grouped by region. On each graph the WT value is shown in bold black. A color-coded key on each graph shows which mutant data that is fit.

$K_1K_2\alpha_1$ defines the stability of the XX0 state relative to 000, and $K_3\alpha_2$ defines the stability of XXX relative to XX0 (Eqn. 8 & 9, Methods). The WT value of $K_1K_2\alpha_1$ is 0.86, which suggests that protein slightly favors the phosphatase state even in the absence of Mg^{2+}. $K_3\alpha_2$ is 8, which indicates relatively strong coupling between the XX0 and XXX states for the WT. Mutants that stabilize kinase-on state tend to have increased values of these two parameters and higher midpoints in the empirical dose-response curves. Conversely, phosphatase-promoting mutations

lead to decreases in these parameters and lead to a shift to lower midpoints in the dose-response curves.

3.4 Discussion

PhoQ has been extensively studied by site-directed mutagenesis, in which interesting mutants were selected based on activity screens or interest in probing the importance of residues in various portions of the protein. Generally, the mutations are analyzed by 2-point LacZ assays yielding limited conclusions. The effects of substitutions are generally discussed as *decoupled from metal ion binding*, or constitutively partially active, or lower/higher affinity and so on. These approaches have limited ability to answer a number of important questions: Does a *decoupled* mutant fail to bind metal ions? show less active effective coupling to the kinase domain? or is it so far "ON" or "OFF" that, while the energetic coupling is intact, it failed to cause a significant change in population? In addition to these questions, it remains unclear whether mutants induce distinct and varied conformations of the DHp/catalytic domains with differing enzymatic activities, or whether mutants simply changed to populations of the two states. Many models are inherently limited by data from isolated domain and cannot take into account the larger protein and membrane neighborhood.

Here we examine far larger numbers of mutants and analyze the most interesting ones over a full range of Mg^{2+} concentrations. It was immediately clear that mutants up to a hundred angstrom away from the magnesium binding site had a large effect on the cation binding curve as well as the degree of activation at low and high metal ion. Because of the location of the tested mutations, the ability to change the apparent binding of Mg(II) has little to do with changing direct binding or direct catalytic activity. We hypothesize that the perturbation we see in the curves relate to changes in the energy gap between kinase and phosphatase states and therefore the equilibrium population of both states.

With the fitting parameters in hand we can now generate population curves for states that were previously invisible. Using the kinase state, [XXX], as a proxy for the experimentally determined data, we can graph the populations of an initial phosphatase state, [000], along with

an intermediate phosphatase state, [XX0], over the range of Mg^{2+} concentrations, for both phenotypes (Fig. 3.8). We selected the intermediate, [XX0], because it is on the canonical pathway of activation. Additionally, we found that for specific states in the thermodynamic equilibrium the population of that state is estimated to be negligible (see Methods, Fig. 3.14).

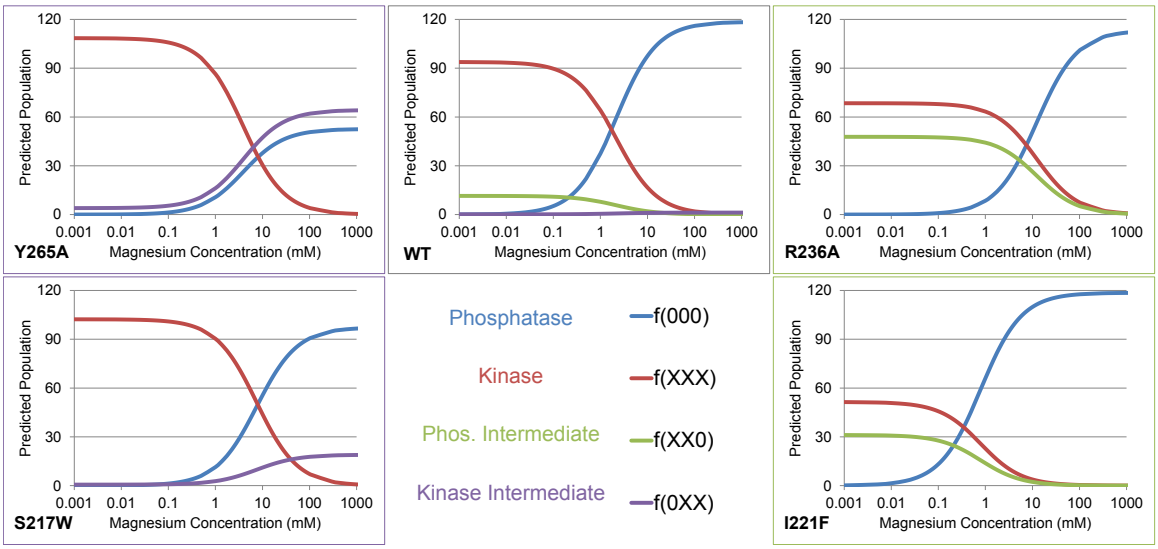


Figure 3.8 Population curves based on LED fit parameters. The middle graph shows the WT population curves for four states listed in the legend. (Left) Mutants with significant populations of the kinase intermediate (S217W, Y265A), (Right) mutants with significant populations of phosphatase intermediate (R236A, I221F).

First, we see the WT strongly favors the phosphatase state until about 2mM Mg^{2+} and switch almost entirely to the kinase state at low Mg^{2+}, but with a small contribution from the phosphatase intermediate. A similar plot can be made for mutants that show both a kinase and phosphatase phenotypes (right most graphs R236A and I221F), yet here the phosphatase intermediate is significant. For the R236A mutant (K_{app} is shifted right), the buildup of this intermediate prevents the mutant from reaching WT levels of activation. We can relate the I221F mutant to the previous mutant, despite the fact that it has a completely different phenotype (K_{app} is shifted left), because it also suffers from the same buildup of phosphatase active intermediate at low magnesium. Here the low values for $K_3\alpha_2$ (~1.5 vs. WT 8.2) constrict signaling out of the HAMP domain and cause a large buildup of intermediate. For other mutants (left most graphs Y265A and S217W), a kinase active intermediate is contributing to the equilibrium. In the first case, Y265A, the S-helix mutant has a $K_3\alpha_2$ 5-fold larger than WT and this favors the kinase state

under activating conditions, but at high magnesium that large parameter favors the kinase active intermediate. In the second case, S217W, the TM/HAMP linker mutant has a $K_1K_2\alpha_1$ 5-fold larger than WT. This example shows how the different mutants can achieve similar phenotypes through different means. It is from plots like these that we can extract a more intuitive result than a table of coefficients.

For another view of the data, we consult fits of the model to raw experimental data and grouped the fits by the various locations of the mutations in the protein (Fig. 3.7). Two conclusions are immediately apparent. Mutations to the TM1 and the P-helix lead only decreased kinase activity and shifts towards higher avidity for Mg^{2+}. By contrast, variants in the remaining domains can exhibit either increase or decreases in kinase activities and coordinated shifts to higher or lower kinase activities with concomitant shifts in to lower and higher apparent avidities for Mg^{2+}. The differences become even more apparent if one examines apparent $\Delta\Delta G$ for these mutants as a function of the position of the substitution along the kinase Z-axis from the extracellular sensor to the catalytic domains (Fig. 3.9). The apparent $\Delta\Delta G$ are arranged along the Z-axis coordinate of the homology model developed by a computational scientist in the lab, Dr. Tomas Lemmin. Using the known crystal structures for PhoQ (3BQ8) (Cheung et al., 2008) along with structures for homologous proteins, like CpxA which shares a HAMP and DHp (Mechaly et al., 2014), he generated a full-length model of the protein (also in Fig. 3.11). The Z-axis is placed at the center of the four helix bundle that forms the dimeric interface. Along with calculated $\Delta\Delta G$ from the LED fit, I added constitutively active mutants that I had tested in either the 2-point or 8-point reporter assays (Fig. 3.9, light/dark blue bars). We see clusters of very active mutants in the lower portion of the HAMP domain.

I also incorporated some of the well-studied mutants that control activity in the periplasmic domain. At position T48 saturation mutagenesis has been done for both *S. typhimurium* (Sanowar et al., 2003) and *E. coli* (Regelmann et al., 2002). Mutant T48Y displayed larger than WT activation for the low magnesium condition, but did not elevate the activity in the high magnesium condition. Because this mutation does not shift the baseline phosphates state,

like the other activating mutants in the cytoplasm, it would not shift the K_{app} for magnesium binding. Instead this mutant likely increases V_{max}, which is consistent with increasing the population of the kinase state without altering K_{app}. A mutation at one periplasmic position (D179L/A) is known to lock the protein in a kinase active state and has been shown to make a critical inter-dimeric salt bridge between the PAS domain and P-helix (Cheung et al., 2008; Minagawa et al., 2005). Finally, there is 2-point data (Fig. 3.3) that suggest that L51C and L52C increase the activity at high magnesium and could be a hotspot for kinase active mutants in the periplasm.

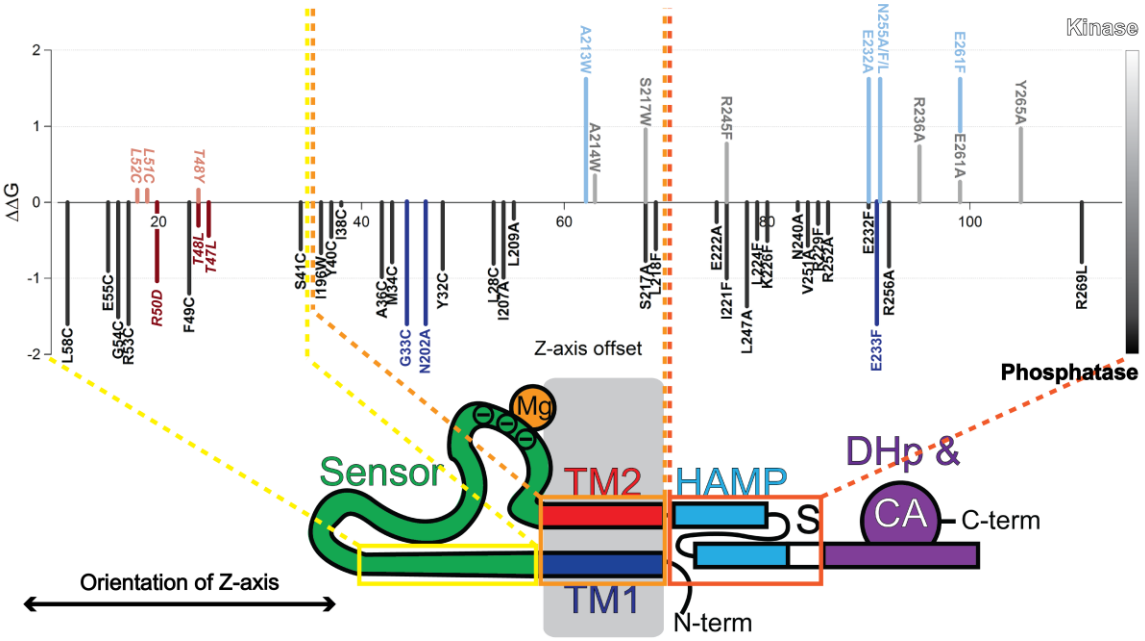


Figure 3.9 – Apparent ΔΔG plotted along the Z-axis of PhoQ. The PhoQ sensor domain was placed at z = 0Å and the last mutant is ~120Å. Grey/black bars represent the calculated ΔΔG from the LED fits, and the K_4 mutants (L58C, G54C, R53C) were truncated and shown simply as <1.5 kcal. Light/dark blue bars represent Mg(II) independent mutants and the ΔΔG was estimated as a max or min value. Light/dark red bars represent ΔΔG predictions for interesting mutants from the literature.

These results imply that the outward half of the protein is strongly restrained by a series of interactions that are easily disrupted by mutations. The crystal structure of the kinase-on state of PhoQ's sensor domain has been determined. It is highly asymmetric and stabilized by a series of intricate salt bridges and other polar interactions (Fig. 3.10) that are presumably broken as the

protein transitions to the Mg^{2+} bound state. Thus, the upper half of the transduction pathway has a relatively tightly restrained sequence-function landscape in which mutations from the native sequence skew towards lower kinase activity and higher net avidity for Mg^{2+}.

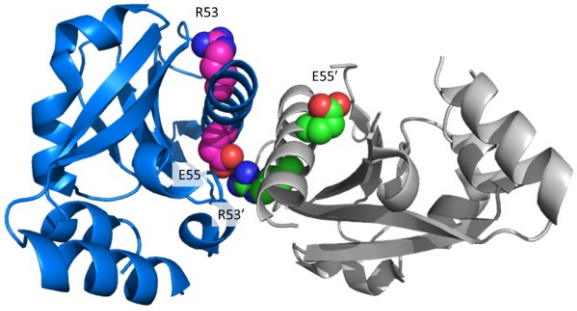


Figure 3.10 – Asymmetric crosslinking in the periplasmic domain. Cartoon view of 3BQ8 with spheres for the critical residues that form an asymmetric inter-dimer salt bridge. Perspective is looking towards the top of the sensor domain from the membrane. Image made in PyMol.

By contrast the lower half of the signal-transducing domains spanning from the middle of the bilayer downward would appear to be evolutionarily optimized to achieve an energetic *balance* that can be tipped in either direction by mutations (Fig. 3.11). This makes sense with respect to information flow through the system, beginning with highly specific and evolutionarily tuned closest to the binding site to more tunable, particularly in the HAMP and S-helix domains. These highly interchangeable domains along with other linkers that include PAS and GAF domains are frequently inserted during the evolution of HKs. This work shows how changes within these signal transducing domains can shift and shape the response curves of the output kinase domains.

Several of the most strongly activating mutations in the cytoplasmic domains deserve particular notice. In our model for PhoQ, which is based on the closely related crystal structure of the cytoplasmic domain of CpxA, Asn255 forms a bidentate hydrogen bond to the backbone of the loop in the HAMP domain. We propose it plays a critical role in stabilizing a given conformation of this domain. Glu232 in the HAMP domain is a particularly interesting mutant in that a homologous residue in CpxA makes a hydrogen bond to its catalytic domain. Thus, this mutation might remove a tertiary interaction that is important for maintaining the kinase in its phosphatase form. The S-helix connecting the HAMP helix 2 to the DHp is another hot spot for activating mutations. Mutations to this region were designed to stabilize the packing between the

helices when the S-helix is a continuous extension of the HAMP helix 2, and indeed led to a strong activation. Finally, mutations of residues to Trp near the TM helix2/HAMP boundary also were strongly activating. With the LED model in hand, these mutants can be combined to give a greater understanding of the signaling landscape of these multi-domain proteins, and in the future help define the signal transduction pathway in bacterial histidine kinases.

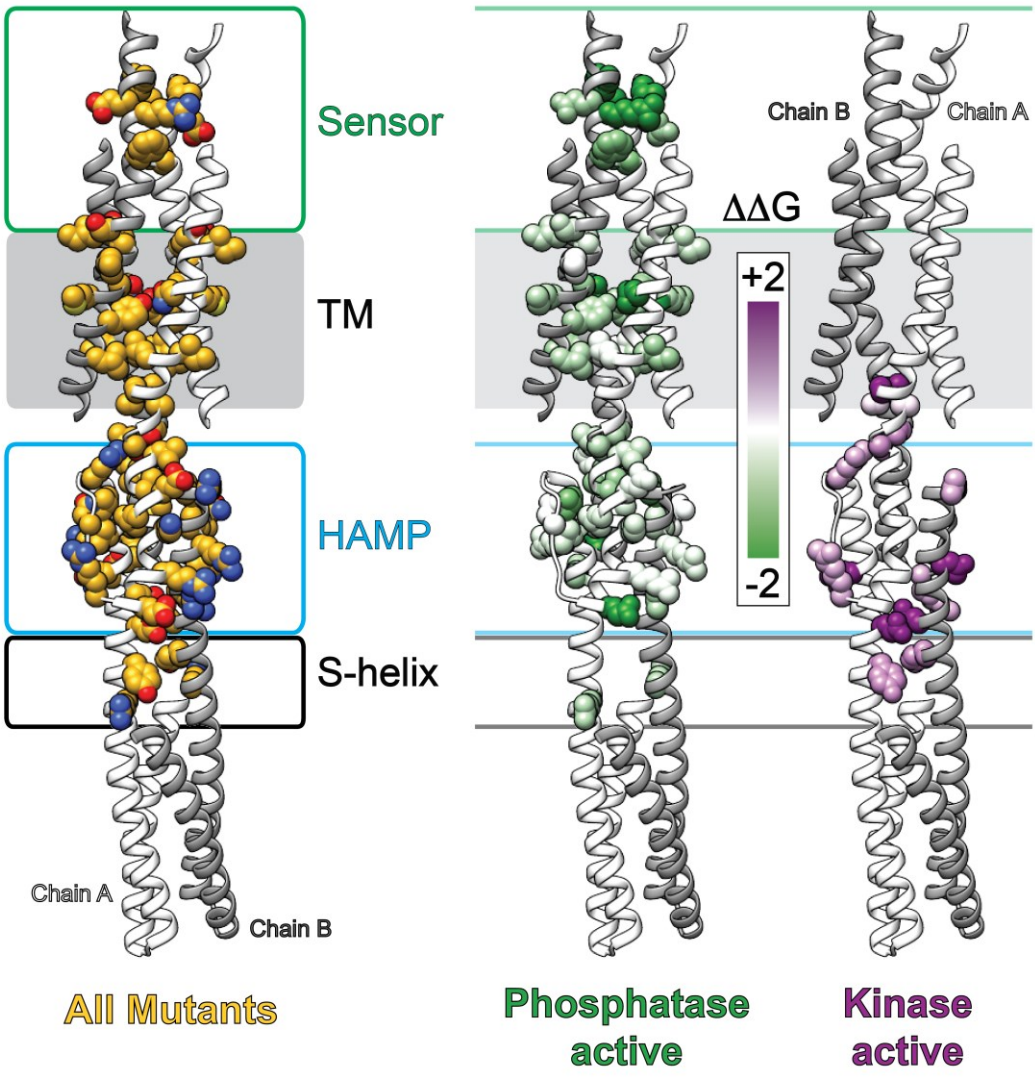


Figure 3.11 – Homology model for dimeric core of PhoQ. Displaying the Lemmin full-length homology model without the periplasmic lobes or the catalytic domain. Chain A is shown in white and Chain B in grey. (Left) Sphere filled mutants are a summary of all of the 35 mutants involved in the fit plus a few known activating/inactivating mutants. (Middle) Mutants found to be phosphatase active painted via the computed ΔΔG on a gradient shown in the key, along with (right) kinase active mutants. Images were made in UCSF Chimera (Pettersen et al., 2004).

3.5 Methods

Mutation LacZ reporter assay. Mutations were made using the Quickchange protocol, where primers were designed on Agilent website and ordered from IDT.

This assay depends on the reporter strain, TIM206, where the LacZ gene is under control of the *mgtA* promoter and the expression is hence governed by phosphorylated Pho-P. Each of the desired mutant plasmids were grown in reporter strain to a desired optical density. The cells were then processed as described in chapter 2 to allow readout of the extent of transcription of the LacZ gene measured in Miller Units (equation 6 in which *ratio* refers to the volume of culture to volume of enzymatic solution used and $t_{incubation}$ refers to the time of the enzymatic incubation following the addition of the substrate, in minutes).

Equation (6)

$$MU = \frac{1000 * (OD420 - 1.5 * OD500)}{OD600 * ratio * t_{incubation}}$$

2-point assays. All of the mutants underwent a 2-point activity assay, where the activity of the mutant protein was tested at both high and low Mg(II) concentrations. The 2-point activity test is a reliable and quick assay to determine the general phenotype of the mutant. Many of these tests were done in MOPS minimal media, however some mutants were only tested in LB.

For these assays, 3mL of media was inoculated 1:1000 from an overnight culture grown at low magnesium concentration. The growth is stopped at an optical density (OD600) between 0.2 and 0.3 and processed as described below.

Variability of wild type controls in 2-point assays. From these assays we noted that repeat measurements for WT and many mutants were not as reproducible as needed for a quantitative analysis of signaling. Specifically, we sought to obtain less than 10% error in the observed Miller Units. This observation began a series of experiments to reduce fluctuations in the WT. One observation was that Miller unit values varied depending on the growth conditions. These conditions include: tube size, volume of media in the tube, angle of aeration of the tube, and shaker RPM. To reduce this variation, optimal growth conditions were selected and adhered to.

The selected growth conditions are: 3mL of MOPS minimal media, in a 14 mL tube, aerated at 45° tilt, and shaken at 225 RPM.

Unfortunately, these improvements did not account for all of the variation. Next we suspected the magnesium concentration of the overnight culture inoculum. To test if the Miller unit signal depended on the concentration of Mg^{2+} in overnight, we started low magnesium samples from an overnight grown at high Mg^{2+} and vice versa. Next we tracked the Miller unit values over time (Fig. 3.12). At early time points, we saw elevated signal in the high magnesium samples, when started from low magnesium overnights that persisted over 4 hours of growth. This persistent signal has a very large effect in the high magnesium conditions because the Miller unit values are generally small. Previous to these experiments, we were stopping the reactions around the 3 or 4-hour mark when the signal has not yet reached a standard state. Additionally, we saw a time dependent rise of the signal in the low magnesium samples, where the readings would level out after 4 hours of growth. This time dependent change in signal affected the low magnesium condition the most, which is evident in the large error standard deviations.

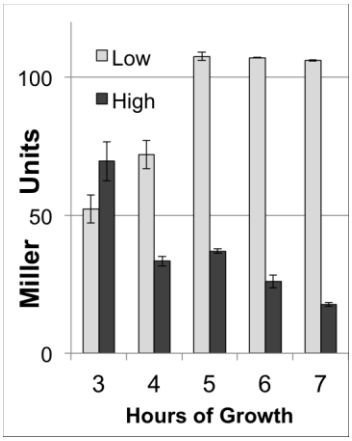


Figure 3.12 – Duplicate WT O/N cultures were started in both low and high Mg^{2+}. Duplicate cultures were inoculated 1:1000 (low into high; high into low) and grown over 7 hours. Error bars indicate standard deviation.

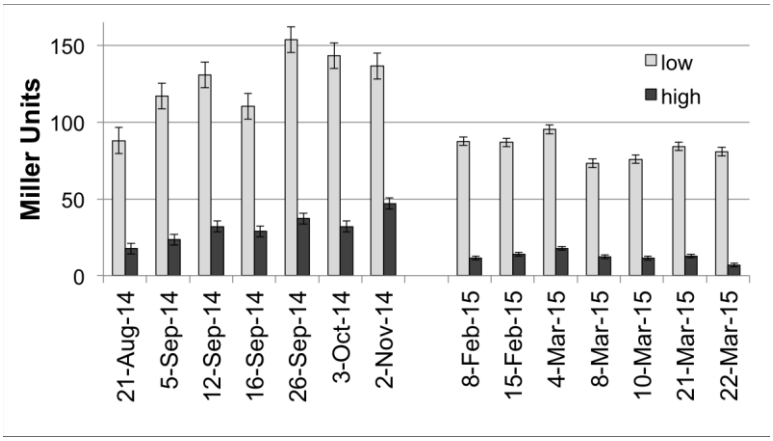


Figure 3.13 – Variability in the WT values. (Left) Seven WT measurements done over the course of several months where growth conditions were optimized to reduce variability. (Right) Seven WT measurements done over two months where the O/N cultures was started in 1mM Mg^{2+} and the measurement was taken after 5 hours of growth. Error bars indicate standard error for the seven measurements.

From these experiments, we made two major changes to the protocol. First, to fix the elevated signal at high Mg, I grew overnights at 1mM Mg^{2+}, and second, to fix the time dependent signal at low Mg, I extended the growth time by two hours. These changes reduced the standard error at the low magnesium condition from 8.4 to 2.8 Miller Units and the error at high magnesium was reduced from 3.6 to 1.2 Miller Units (Fig. 3.13). It is clear from the bar graph that the overall signal also dropped about 30%. This decrease in signal is due to a combination of factors, such as higher optical density of the bacteria required for the signal to equilibrate, effects from stationary phase, etc.

8-point assays. With a more rigorous protocol in hand, forty-one mutants were chosen for the 8-point Mg(II) curve where the activity was determined at concentrations of magnesium varying over a 400-fold range (0.05, 0.2, 0.5, 1, 1.67, 3, 10, and 20mM Mg^{2+}). Cell growth was done in MOPS minimal media with magnesium supplemented in the form of $MgCl_2$. Tubes were diluted 1:1000 from overnight cultures growing in 1mM Mg^{2+} into pre-warmed MOPS minimal media and activity readings were done after 5 hours of growth at 37°C. A WT sample was included as a positive control on each experimental day. These data were used to subsequently scale all of the data to the one WT value to account for variation from day to day.

Western blotting. Western blots were carried out to verify the expression of low activity mutants. Cultures of 3 mL, in MOPS minimal media, grown at 37°C, were grown to mid log phase and a pellet was collected and stored at -20°C. These pellets were membrane prepped as described previously. Samples were run on a 7% tris-acetate gel and immediately transferred to nitrocellulose membrane via an iBlot dry transfer system. The membrane was blotted with anti 6xHis-HRP and visualized with ECL reagent on the Kodak imager. Bands were quantified in ImageJ.

Data Fitting. Initially, each individual curve was analyzed using a version of the Hill equation (Eqn. II) to assess the degree of cooperatively and midpoint for magnesium activation. The cooperatively values (n) were determined to be close to unity for a large number of mutants, and was subsequently held at that value in all other fits. From these fits we also obtain a $K_{apparent}$ for each of the curves. There was a weak correlation between the $K_{apparent}$, which is a measure of the

avidity for magnesium outside the cell, and MU_{Inf}, which is a measure of kinase activation within the cell (Fig. 3.14).

$$f([Mg^{2+}]) = MU_{inf} + \frac{(MU_0 - MU_{inf})}{1 + \left(\frac{K_{app}}{[Mg^{2+}]}\right)^n}$$

Equation (7) – Hill equation for PhoQ activation with loss of magnesium

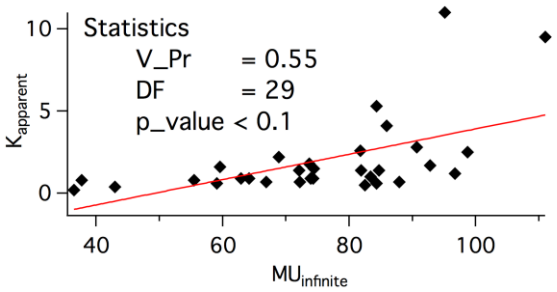


Figure 3.14 – Linear correlation between K_{app} and MU_{inf} from one WT fit and 30 mutants fits. P-value Statistics from Pearson's r for linear fit (Igor) and Degrees of Freedom = 30+1-2 = 29.

Although not reaching a high level of statistical significance, the correlation encouraged us to explore models that evoke thermodynamic coupling between the binding of Mg(II) and the fraction activation of the kinase domain.

The data were therefore fit to the simplest model that was physically reasonable as described in (Fig 3.5, Eqn. 5). This model defines allows fitting of the curves using six adjustable parameters. These parameters are explicitly described in the following equations, with the exception of K_2 and K_4 which were defined previously. These definitions were subsequently plugged in to the expression for fraction kinase in order to generate an expression for the fit.

Equation (8)

$$K_1 K_2 \alpha_1 = \frac{[XX0]}{[000]}$$

Equation (9)

$$K_3 \alpha_2 = \frac{[XXX]}{[XX0]} = \frac{[0XX]}{[0X0]}$$

In the fits, no more than two parameters were allowed to vary. Table 3.2 outlines which parameters were varied locally versus globally, which depends on the position of the mutant in the protein.

Table 3.2 – Variation of coefficients for LED fit

Mutation location	K_4	$K_1K_2\alpha_1$	K_2	$K_3\alpha_2$	V_{max}
Periplasm	*Local*	*Local*	Global	Global	Global
TM1/2 and linkers	Global	*Local*	*Local*	Global	Global
HAMP H1/Loop/H2	Global	Global	*Local*	*Local*	Global
S-helix	Global	Global	Global	*Local*	Global

In spite of the large number of datasets, we found that some parameters could not be individually defined due to covariation of two parameters. For example, we found that the product of K_1 and $K_2\alpha_1$ was well defined by the individual curves, but the individual constants could co-vary; so long as the value of K_1 << $K_2\alpha_1$ a reasonable fit could be obtained. This finding indicates that the formation of 000 from XX0 (see Fig. 3.5) is positively cooperative, i.e. that the concentration of X00 is much smaller than either 000 or XX0 for the mutants that we examined. Thus, we neglect the very small contribution of this state (Fig 3.14) and obtain a simplified LED model. The global fit parameters to this simpler model are listed in Table 1, and were obtained using a global fitting procedure (Eqn. 8) on the data using the linked equilibrium model as implemented in Igor.

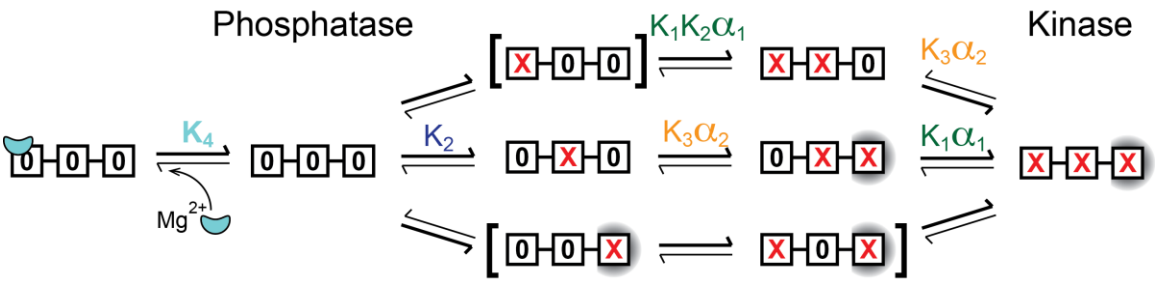


Figure 3.15 – Simplified LED model. The populations in brackets were found to not contribute significantly to the fit. Once you remove those populations, the equation simplifies from 7 parameters to 5 (including V_{max}, which is not shown in this schematic).

Equation (10)

$$MU = \left(\frac{V_{max}\ K_3\alpha_2 \left(K_1K_2\alpha_1 + K_2 \left(1 + \frac{[Mg^{2+}]}{K_4} \right) \right)}{1 + \frac{[Mg^{2+}]}{K_4} + K_2 \left(1 + \frac{[Mg^{2+}]}{K_4} \right) + K_2K_3\alpha_2 \left(1 + \frac{[Mg^{2+}]}{K_4} \right) + K_1K_2\alpha_1 + K_1K_2\alpha_1K_3\alpha_2} \right)$$

3.6 References

Adase, C.A., Draheim, R.R., and Manson, M.D. (2012). The residue composition of the aromatic anchor of the second transmembrane helix determines the signaling properties of the aspartate/maltose chemoreceptor Tar of Escherichia coli. Biochemistry *51*, 1925-1932.

Bissonnette, M.L., Donald, J.E., DeGrado, W.F., Jardetzky, T.S., and Lamb, R.A. (2009). Functional analysis of the transmembrane domain in paramyxovirus F protein-mediated membrane fusion. J Mol Biol *386*, 14-36.

Chervitz, S.A., and Falke, J.J. (1996). Molecular mechanism of transmembrane signaling by the aspartate receptor: a model. Proc Natl Acad Sci U S A *93*, 2545-2550.

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

Draheim, R.R., Bormans, A.F., Lai, R.Z., and Manson, M.D. (2005). Tryptophan residues flanking the second transmembrane helix (TM2) set the signaling state of the Tar chemoreceptor. Biochemistry *44*, 1268-1277.

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

Ferris, Hedda U., Dunin-Horkawicz, S., Hornig, N., Hulko, M., Martin, J., Schultz, Joachim E., Zeth, K., Lupas, Andrei N., and Coles, M. (2012). Mechanism of Regulation of Receptor Histidine Kinases. Structure *20*, 56-66.

Goldberg, S.D., Clinthorne, G.D., Goulian, M., and DeGrado, W.F. (2010). Transmembrane polar interactions are required for signaling in the Escherichia coli sensor kinase PhoQ. Proc Natl Acad Sci U S A *107*, 8141-8146.

Goldberg, S.D., Soto, C.S., Waldburger, C.D., and DeGrado, W.F. (2008). Determination of the Physiological Dimer Interface of the PhoQ Sensor Domain. J Mol Biol *379*, 656-665.

Grigoryan, G., and DeGrado, W.F. (2008). Modest membrane hydrogen bonds deliver rich results. Nat Chem Biol *4*, 393-394.

Hazelbauer, G.L. (2012). Bacterial chemotaxis: the early years of molecular studies. Annu Rev Microbiol *66*, 285-303.

Lin, Y.H., Pierce, B.D., Fang, F., Wise, A., Binns, A.N., and Lynn, D.G. (2014). Role of the VirA histidine autokinase of Agrobacterium tumefaciens in the initial steps of pathogenesis. Frontiers in plant science *5*, 195.

Mechaly, A.E., Sassoon, N., Betton, J.M., and Alzari, P.M. (2014). Segmental helical motions and dynamical asymmetry modulate histidine kinase autophosphorylation. PLoS Biol *12*, e1001776.

Minagawa, S., Okura, R., Tsuchitani, H., Hirao, K., Yamamoto, K., and Utsumi, R. (2005). Isolation and molecular characterization of the locked-on mutant of Mg2+ sensor PhoQ in Escherichia coli. Biosci Biotechnol Biochem *69*, 1281-1287.

Molnar, K.S., Bonomi, M., Pellarin, R., Clinthorne, G.D., Gonzalez, G., Goldberg, S.D., Goulian, M., Sali, A., and DeGrado, W.F. (2014). Cys-scanning disulfide crosslinking and bayesian modeling probe the transmembrane signaling mechanism of the histidine kinase, PhoQ. Structure *22*, 1239-1251.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem *25*, 1605-1612.

Regelmann, A.G., Lesley, J.A., Mott, C., Stokes, L., and Waldburger, C.D. (2002). Mutational analysis of the Escherichia coli PhoQ sensor kinase: differences with the Salmonella enterica serovar typhimurium PhoQ protein and in the mechanism of Mg2+ and Ca2+ sensing. J Bacteriol *184*, 5468-5478.

Sanowar, S., Martel, A., and Moual, H.L. (2003). Mutational analysis of the residue at position 48 in the Salmonella enterica serovar Typhimurium PhoQ sensor kinase. J Bacteriol *185*, 1935-1941.