Structure

Article

Cys-Scanning Disulfide Crosslinking and Bayesian Modeling Probe the Transmembrane Signaling Mechanism of the Histidine Kinase, PhoQ

Kathleen S. Molnar,[1,2,6] Massimiliano Bonomi,[5] Riccardo Pellarin,[5] Graham D. Clinthorne,[3] Gabriel Gonzalez,[2,6] Shalom D. Goldberg,[1,7] Mark Goulian,[4] Andrej Sali,[5] and William F. DeGrado[6,*]
[1]Department of Biochemistry and Biophysics
[2]Biochemistry and Molecular Biophysics Graduate Group19104, USA
[3]Pharmacological Sciences Graduate Group
[4]Department of Biology
University of Pennsylvania, Philadelphia, PA 19104, USA
[5]Department of Bioengineering and Therapeutic Sciences, California Institute for Quantitative Biosciences
[6]Department of Pharmaceutical Chemistry, Cardiovascular Research Institute
University of California, San Francisco, San Francisco, CA 94158, USA
[7]Present address: Centyrex, Janssen Research & Development, LLC, Spring House, PA 19002, USA
*Correspondence: william.degrado@ucsf.edu
http://dx.doi.org/10.1016/j.str.2014.04.019

SUMMARY

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. Because a similar scissors model has been ruled out in CheA-linked chemoreceptors, the evidence suggests that sensor His-kinase and CheA-linked receptors possess different signaling mechanisms.

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 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 two-component systems (TCSs) are diverse (Dutta et al., 1999), the topology of a canonical HK receptor (Figure 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 *h*istidine 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 dimerization and histidine phosphotransfer 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 (Figure 1B) with an extended dimer interface along the entire length of the molecule (Goldberg et al., 2008). Because 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). Nuclear magnetic resonance (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).
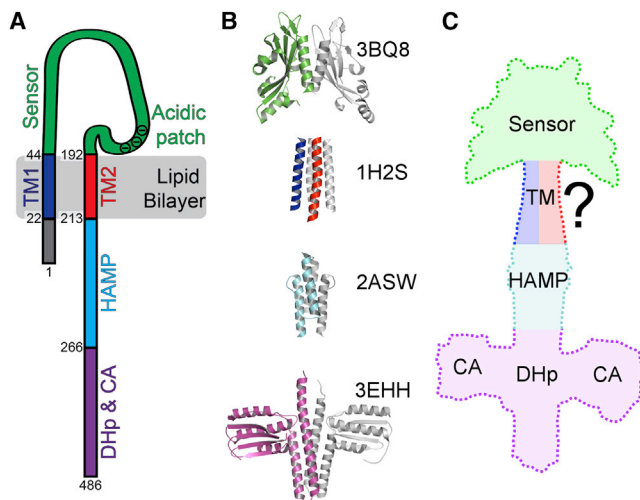
Figure 1. Structural Representations of PhoQ

(A) Schematic of the topology of a PhoQ monomer. The numbers indicate residue numbers for *E. coli* PhoQ (UniProt ID: 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 gray.

(C) A model of the first three domains of PhoQ: sensor, transmembrane (TM), and HAMP domains. The dimerization and histidine phosphotransfer domain (DHp) and catalytic domain (CA) are added for clarity but were not modeled.

Many recent multidomain 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. Chervitz and Falke 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 interhelical 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 used 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 hemichannel 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 cross-linking data (Lemmin et al., 2013). To elucidate larger scale changes across the membrane, we incorporate additional cross-linking data in the HAMP and juxtamembrane regions of PhoQ with previous data and analyze it using multistate Bayesian modeling (Rieping et al., 2005). This approach provides an investigation into the structures of the two signaling states of PhoQ, which interconvert through a motion in which opposing helices move toward or away from the bundle axis. Our subsequent quantitative structural analysis of additional receptor HK domains also divulges 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.

RESULTS

We probed the TM domain and the neighboring HAMP and periplasmic domains of PhoQ using disulfide-scanning mutagenesis. Building on our analysis of the periplasmic helix at the dimer
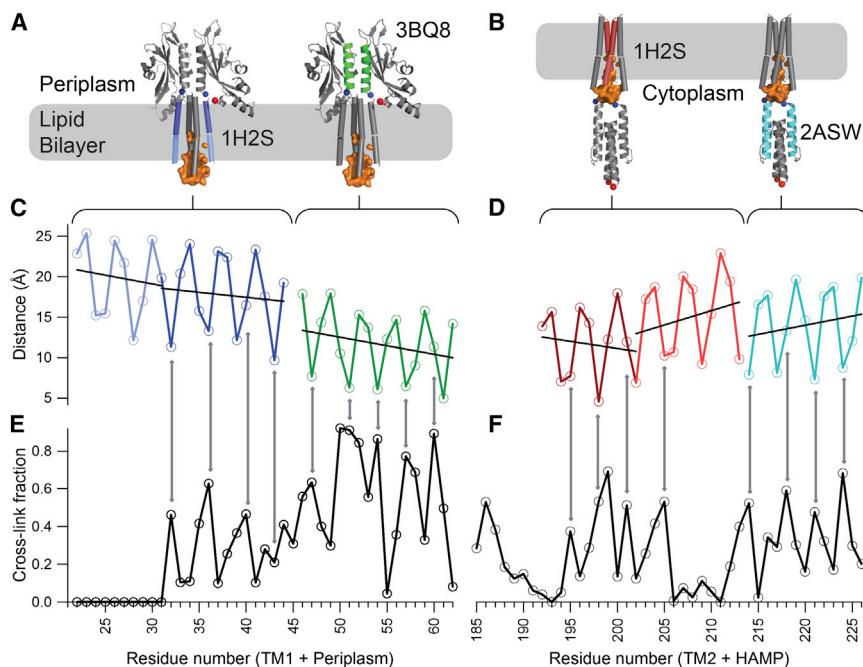
Figure 2. Comparison of the Crosslinking Efficiency with Structural Models

(A) PhoQ TM1-Periplasm-TM2 model in a lipid bilayer. 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) Intermonomer distances 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 and Cβ′ of corresponding residues (or Cα–Cα′ for glycine). Black lines indicate linear fits to each helical segment.

(D) Intermonomer distances 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 22–61.

(F) Crosslinking data from the full-length PhoQ protein in a native membrane for cysteine mutants 185–226.

Western blot data from (E) and (F) are shown in Figure S1.

interface (Goldberg et al., 2008), new single cysteine residue mutations were introduced along the TM helices and at selected positions within the HAMP domain (Figures 2A and 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* (Figure 1B). These comparisons test how faithfully these individual domains represent the full-length structure of PhoQ (Figure 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 (residues 42–49), the crosslinking efficiency maintains a sinusoidal variation with a consistent phase (Figures S1A and

S1B available online), suggesting an uninterrupted helix. The similarity in the phases can be seen qualitatively (Figure 2E), but we also fit these data to a sinusoidal function (Figure S1A). 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 (residues 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 (Figures S1C and S1D), suggesting this linker can access a distorted helical or nonhelical geometry.

To check how faithful a model HtrII is for the PhoQ TM bundle, we compared the interresidue distances to corresponding experimental crosslinking data (Figure 2). The topology of the HtrII bundle is like other HKs, where the helices are antiparallel and the N terminus of the first TM is toward the cytoplasm (Figure 1A). The crosslinking fractions agree qualitatively with bundle being well packed near the periplasm, but diverge slightly near the cytoplasm. At the periplasmic side (Figure 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 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 (Figure 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,

Table 1. Least-Squares Fitting of a Sinusoidal Function to the Crosslinking Efficiency of PhoQ and the Interresidue Distances of HtrII and Af1503 Crystal Structures

Helix	Period[a]	Phase Offset (°)[b]
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	

[a]Number of residues per repeat.
[b]Differences in phase for the fitted sinusoidal waves between the experimental crosslinking data and the intermonomer distance data (Cβ–Cβ′ distance or Cα–Cα′ for Gly) taken from corresponding crystal structure.

shown as solvent accessible surface in orange in Figure 2. However, the complete lack of crosslinking on the cytoplasmic side of PhoQ TM1 helices suggests a larger separation in the PhoQ hemichannel 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.

Multistate 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; Figures 2E and 2F) without introducing steric clashes. Consequently, we hypothesize the presence of multiple, distinct structural states in the sample. We used a multistate 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 Experimental Procedures).

We divided the PhoQ dimer into six rigid bodies for each monomer, for a total of 12 rigid bodies in the dimer (Experimental Procedures). 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 crosslinked 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 (Figure 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 crosslinking data and domain models. Cluster representatives and predicted cross-linked fractions for all clusters with a population greater than 3% are reported in Table S1. To predict the minimal number of states that best explain the crosslinking data, the Bayesian approach was applied independently for one, two, and three states.

One-State Modeling

The cluster analysis of the sampled models (Table S1) revealed that the experimental data could not be fully explained by a single structure. The one-state model was in good agreement with the predicted crosslinked fractions in the periplasmic side of TM1 (residues 13–45) and cytoplasmic side of TM2 (residues 205–215). However, the model does not match a large region of data with a high crosslinked fraction—the periplasmic side of TM2 (residues 195–205). 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 crosslinked fractions equal to zero. Therefore, in the one-state model, proximity between the periplasmic region of TM2 and TM2′ is not observed due to steric exclusion by the TM1 and TM1′ helices.

Two-State Modeling

The most populated cluster of two states found by the two-state modeling approach explained crosslinking data better than one-state modeling, as shown by the lower likelihood score (Table S1) and the improved agreement between the model and the data for the periplasmic TM2 region and surrounding residues (185–205; Figure 3). By postulating the existence of a mixture of states, the two-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 with data jackknifing (Figure S2).

For the periplasmic region, State-2 resembles the crystallographic structure of the PhoQ sensor domain, 3BQ8 (Cα; root-mean-square deviation = 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 two-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 Figures 3B
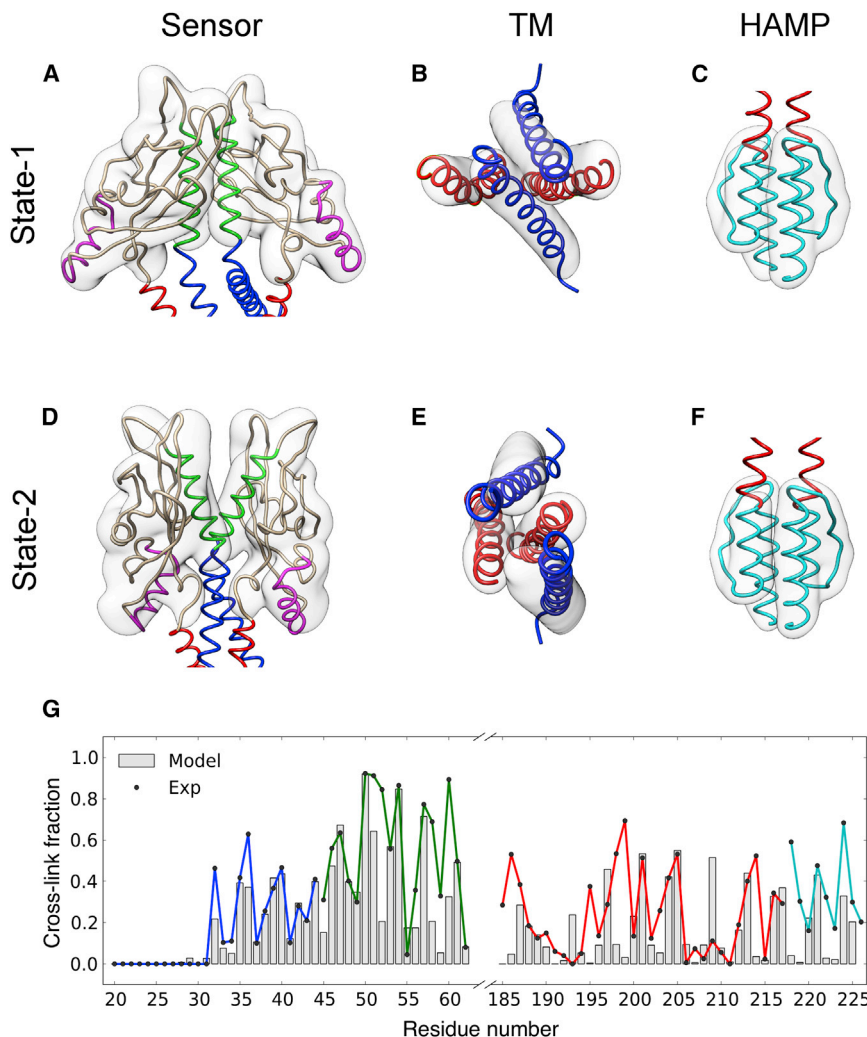
Sensor   TM   HAMP

State-1

State-2

G

Figure 3. Analysis of the Most Populated Cluster Found in Two-State Modeling
(A–F) 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: periplasmic sensor helices (residues 45–61) are green; the Mg^{2+}-binding, acidic patch (residues 137–150) is magenta; the TM1 and TM2 helices of the TM domain are blue and red, respectively; and the HAMP domain is cyan.
(G) Overlay of model data, predicted by the highest likelihood model of the cluster (gray bars), and experimental crosslinked fractions, color-coded by the domain definition above.
Model robustness assessment is summarized in Figure S2.

cated by the average and best likelihood scores for the clusters (Table S1). The previously identified models were not found here because we imposed a lower bound of 0.2 on the individual population fractions w_i (Supplemental Experimental Procedures).

Selecting the Best Model
The two-state model fits the data significantly better than the one-state model, whereas the three-state model does not improve the fit. Therefore, the two-state model is the most parsimonious explanation of the data.

and 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), whereas in State-2, the TM2 and TM2′ helices move toward 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 toward 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 (Figures 3C and 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).

Three-State Modeling
Models in the most populated clusters in three-state modeling do not improve the fit relative to the two-state model, as indi-

Functional Measurements of Cys Mutants Explain Most of the Deviations between Crosslinking Data and the Two-State Model
Although the two-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 (Figure 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 nonnative 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 (Figure 4C and Figure S3), 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
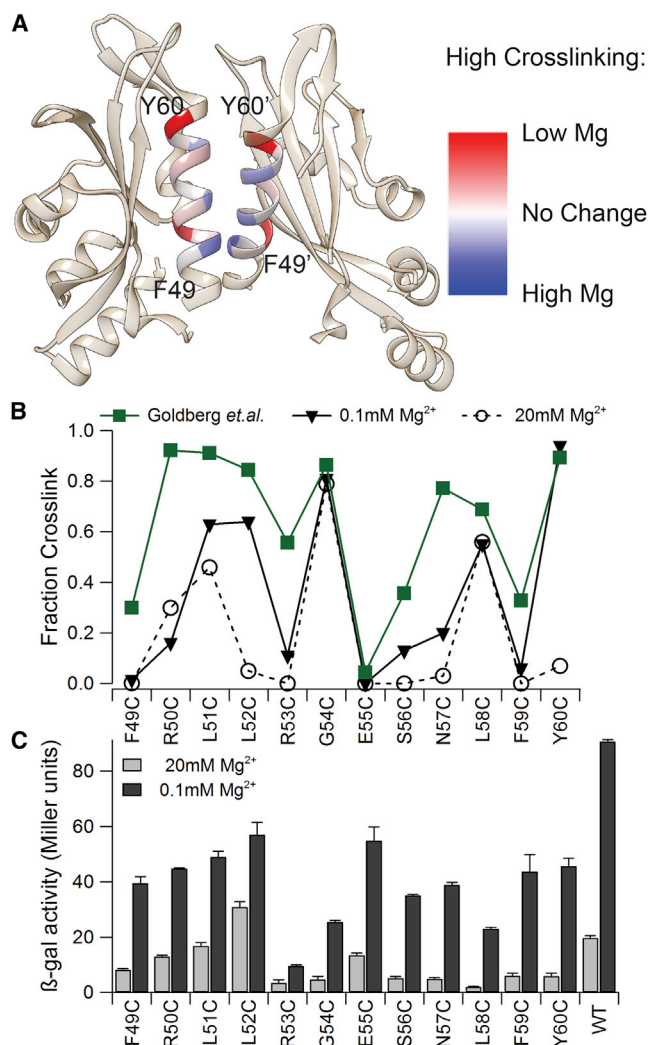
Figure 4. Change in Crosslink Fraction for the Periplasmic Helix of PhoQ at Low and High [Mg²⁺]

(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 Mg²⁺ conditions, white indicates no Mg-dependent change, and blue represents higher crosslinking under high Mg²⁺ (key at right).

(B) Crosslinking data from the full-length PhoQ protein in a native membrane for cysteine mutants 49–60 under both high and low [Mg²⁺] conditions at mid-log phase growth. The green curve is data reproduced elsewhere (Goldberg et al., 2008).

(C) Activity response of β-galactosidase reporter in whole cells for the mutants involved (data for extended region in Figure S3).

Error bars represent the SE of three replicate experiments.

concentrations of Mg²⁺. In contrast, the wild-type protein activity changes 2- to 5-fold between these Mg²⁺ 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 Mg²⁺ 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 195–199). TM2 was modeled as a rigid body extending from 194 to 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 two-state model; it is also the first of three consecutive residues that shows high crosslinking (Figure 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 Luria broth (LB) media. However, when we incubated for shorter periods (to avoid spurious crosslinks) in minimal media (for precise control of Mg²⁺ concentration), we found that residue 50 does not crosslink nearly as much as residue 51, and the extent of crosslinking at 52 depended on the Mg²⁺ concentration (Figure 4B). Activity measurements also showed that, although the mutants had somewhat lower transcriptional activity than wild-type, they were still responsive to changes in the Mg²⁺ 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 Mg²⁺, particularly at residues 52 and 60 (Figure 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, and 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 downweighted in the construction of the model. In contrast to traditional modeling, the two-state model motivated additional functional experiments to explain the large differences between the observed and predicted data.

Structural Variation between Signaling States

The two-state model proposes conformational changes between State-1 and State-2 in which the helices show larger

Table of coordinate displacements between 4-helix bundle crystal structures

		Comparison	PDB IDs	Chain	Helix	Displacements in coordinate system (Å)					
						$\Delta\psi$ [a]	$\Delta\varphi_1$ [b]	$\Delta\varphi_2$ [b]	Δr	Δz	$\Delta\theta$ [c]
Sensor Domains	**Tar**	Asp bound to apo	2LIG-1LIH	A	1 (42-57)	-0.18	0.37	0.48	0.24	-0.12	-0.20
				B	1 (42-57)	-0.09	-0.20	0.12	0.09	-0.25	-0.15
				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	TorT/TMAO bound to apo TorT	3O1H-3O1I	A	1 (52-67)	-0.10	-0.04	0.38	-0.71	0.21	-0.71
				B	1 (52-67)	0.38	0.04	-0.14	0.83	-0.14	-0.38
				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 Domain	**AF1503**	WT to A291F	3ZRX-3ZRV	A	1 (280-297)	-0.30	-0.24	0.11	0.58	-0.19	0.07
				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
		WT to A291V	3ZRX-3ZRW	A	1 (280-297)	-0.47	-0.36	-0.01	0.27	-0.37	-0.06
				B	1 (280-297)	-0.67	-0.40	-0.42	0.29	-0.39	0.09
				A	2 (310-328)	0.68	1.31	0.54	-1.14	-0.16	0.05
				B	2 (310-328)	0.33	1.65	0.27	-1.08	0.11	0.10
DHp Domains	**EnvZ**	WT to A291F	3ZRX-3ZRV	A	1 (333-345)	1.39	-1.50	1.51	-1.60	-0.06	-0.90
				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
		WT to A291V	3ZRX-3ZRW	A	1 (333-345)	1.42	-1.02	3.04	-2.38	-0.22	-1.97
				B	1 (333-345)	0.11	-1.32	0.80	-1.90	-0.20	-0.92
				A	2 (373-385)	0.51	-1.54	0.76	1.29	-0.07	0.53
				B	2 (373-385)	0.31	-1.32	1.60	1.21	-0.21	0.21
	DesK	WT to WT +ADP +AMPPCP	3EHF-3GIG	A	1 (182-198)	-0.19	1.65	-1.06	0.51	-0.02	0.41
				B	1 (182-198)	-0.10	-0.18	-0.60	-0.74	-0.03	-0.06
				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
		WT to H188V +ADP +ADP	3EHF-3EHH	A	1 (182-198)	2.58	1.07	1.84	2.36	0.46	-0.89
				B	1 (182-198)	-2.76	0.37	1.41	1.95	-0.86	-0.81
				A	2 (224-238)	1.12	1.74	0.65	-2.36	0.39	0.69
				B	2 (224-238)	1.17	1.11	1.20	-2.53	0.73	0.31
		WT to H188E +ADP +ADP	3EHF-3GIF	A	1 (182-198)	-0.37	-0.93	-1.60	-0.63	-0.10	1.27
				B	1 (182-198)	0.14	1.54	0.79	0.81	0.37	0.03
				A	2 (224-238)	0.52	0.22	-0.22	-0.25	0.11	-0.06
				B	2 (224-238)	0.63	0.38	-0.07	-0.47	0.20	-0.17

Figure 5. Measure of Coordinate Displacements between Four-Helix Bundle Crystal Structures

Each helix was parameterized as described in Figure S4. For each domain, we analyzed pairs of helical bundles believed to correspond to different signaling states, and the differences in the computed parameters associated with helical phase, tilts, rotation, and displacements, and translations were compared. Angular valves for $\Delta\psi$, $\Delta\varphi_1$, $\Delta\varphi_2$, and $\Delta\theta$ are listed in Table S4. The coordinate displacement associated with these angular displacements was determined from these following equations:
[a] $4 * \sin(\Delta\psi)$
[b] $1.5 * (\text{helix length})/2 * \sin(\Delta\varphi)$
[c] (Average radius) $* \sin(\Delta\theta)$
The cell color indicates the largest (red or blue) and second largest (salmon or light blue) value in each row.
See also Figure S4 and Tables S2 and S3.

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 four-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 (Figure 5). 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 Figure S4. 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 around the helix axis (helix phase = ψ) corresponds to the gearbox model. The remaining four degrees of freedom are helix tilt toward the bundle axis (toward tilt = φ_1), helix tilt perpendicular to the toward 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; Figure S4). Of these degrees of freedom, radial displacements (r), toward 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 S3, and the displacements along these modes are listed in Figure 5.

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 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.
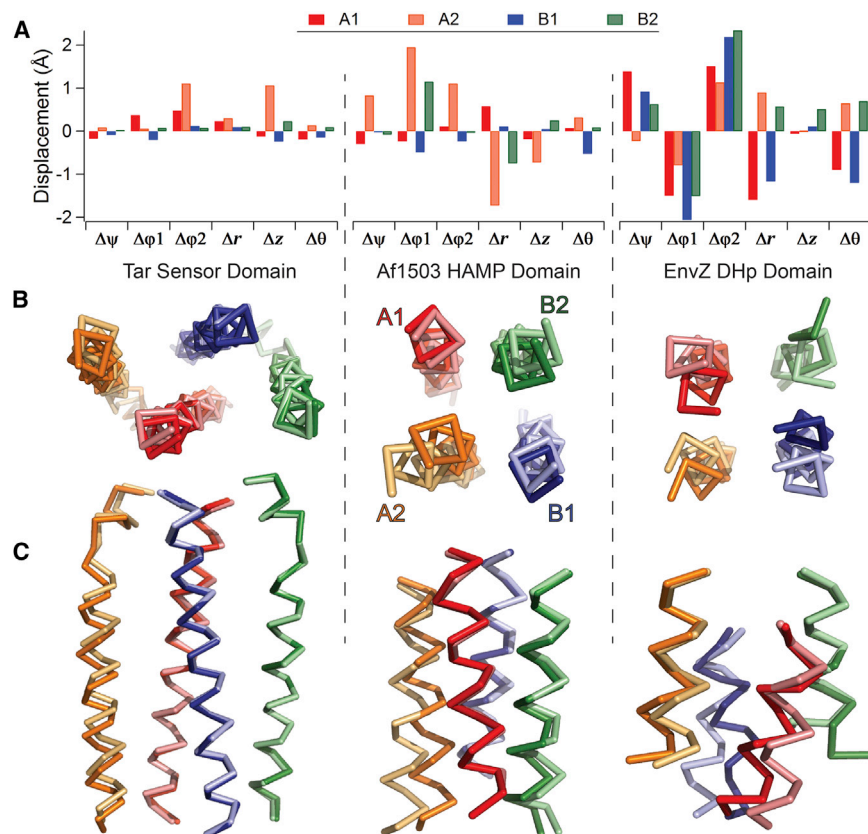
Figure 6. Comparison of Displacements between Crystal Structures in Different States

(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 four-helix bundles. The helices in state 1 (2LIG and 3ZRX) are colored vividly, whereas state 2 (1LIH and 3ZRV) are colored a corresponding lighter shade.

(C) Side-view of four-helix bundles, with helices colored as in (B).

exception of a single pair of helices in DesK, where helical rotations dominate and lateral translations come in a close second (Figure 5). However, this helical rotation is 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 toward 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.

Our analysis of the HAMP(Af1503)/DHp(EnvZ) chimera is also consistent with the analysis of Ferris and colleagues, who proposed a helical rotation as well as other changes in packing. Structures of the chimeric wild-type 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 (Figure 5). These changes in helical phase underlie large structural changes, which our analysis associates with helical tilting around (φ_1) and rigid body shifts that change the radius of the bundle (Figure 6). 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 lead 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) (Figure 6A, 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

DISCUSSION

We present a model of HK signal transduction through the membrane and use 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 crosslinking data are accurate. The crosslinking data are almost 180° out of phase with distances derived from homologous crystal structures (Table 1), which is expected, because crosslinking efficiency decreases over greater distances.

The crosslinking data cover the juxtamembrane regions connecting the TM domain to the sensor and HAMP domains (Figure 2). These data provide 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.
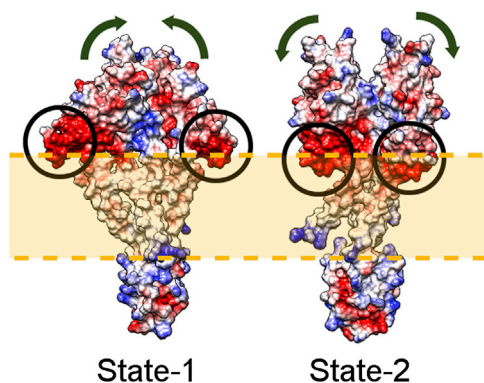
Figure 7. Cation-Binding, Acidic Patch Movements Predicted by the Bayesian Multistate Modeling

Electrostatic surface representation of the two states of the acidic patch as it moves out of (State-1) and into (State-2) the membrane bilayer. Surface representation made with UCSF Chimera (Pettersen et al., 2004).
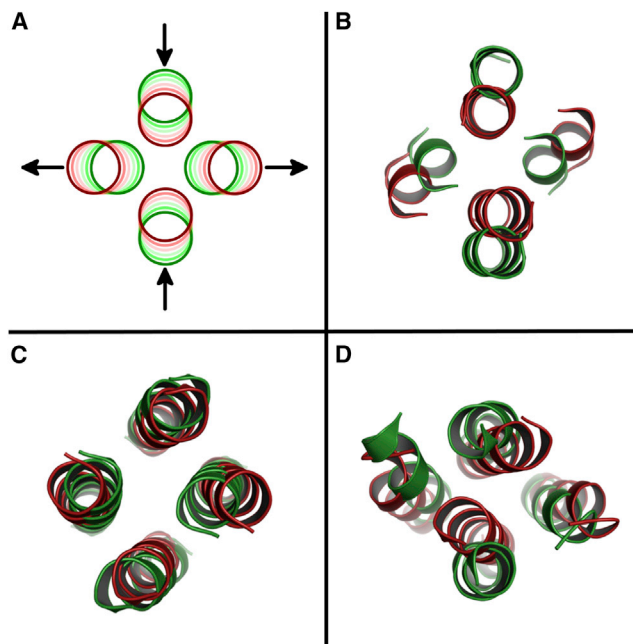

Figure 8. Diagonal Scissoring Motions across Several Two-Component Domains

(A) A helix bundle exhibits orthogonal scissoring if two opposing helices move inward and the other two opposing helices simultaneously move outward.
(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 Figure S5.

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 support signaling states near equilibrium, where we see a degree of activation of only 2- to 5-fold in low Mg^{2+} concentrations. These results are in agreement with the electron paramagnetic resonance studies of Trg from *Escherichia 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.

These two alternative conformations suggest large displacements of the sensor domains that insert or remove the periplasmic acidic patch within the membrane (Figure 7). 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 (Figure 8), ultimately changing catalytic domain activity. Whereas the highly simplified diagram in Figure 8A is symmetrical, asymmetric version are equally likely, particularly in cases of negative cooperativity. In addition, 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 (Figures 3B and 3E), in which two opposing helices move inward and displace the other two opposing helices that move outward (Figure 8A). We observed similar diagonal displace-

ments across several two-component systems (Figures 8B–8D). 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 Figure 6, EnvZ DHp).

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 around the helix axis, (2) piston shifts that vertically displace helices, (3 and 4) tilting toward 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 around the bundle axis. In every examined case, we find that these domains, including the two-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 (Figure 5).

Whereas 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. Although there were many similarities between our cross linking profiles and those of Tar, there were also many significant differences. Both TM1 and TM2 formed intersubunit 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, whereas 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+}, Figure 4), but not for the Tar helix ($\alpha 1$-$\alpha 1'$) 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 and coworkers for the Tar receptor (Milburn et al., 1991). This mechanism, how-ever, 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 interhelical 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.

EXPERIMENTAL PROCEDURES

Plasmids
phoQ-His$_6$ Cys mutant plasmids were created and then transformed into a $\Delta phoQ$ $\Delta lacZ$ 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, and 0.2% casamino acids at 37°C. In both cases, the plasmid was maintained with 100 μg/ml ampicillin.

Envelope Preparations
Freshly plated colonies were picked by sterile loop and used to inoculate 5 ml LB + 100 μg/ml ampicillin. Cultures were grown at 37°C for 24 hr with vigorous shaking (220 rpm) and pelleted by centrifugation at 3,700 × *g* for 10 min at 4°C. Cells were washed by resuspension in 30 mM Tris, pH 8 and pelleted as above. Next, cells were treated with 20% sucrose in 30 mM Tris, pH 8 for osmotic shock and 10 mg/ml lysozyme to remove the cell wall. After 30 min incubation at 4°C, the cell envelopes were resuspended in 3 ml of 3 mM EDTA, pH 8, and sonicated briefly. TM and HAMP samples were spun at 16,000 × *g* for 30 min at 4°C to pellet membranes. The membrane fraction was resuspended in 200 μl of 2 mM Tris, pH 7.5, and stored for use at −80°C. Periplasmic mutants were collected with a 10 min ultracentrifuge spin (489,000 × *g*) and then resuspended in 150 μl of 8 M urea, and 20 mM N-ethylmaleimide (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 10 μl sample of cell envelopes with 10 μl of buffer containing 2 mM or 0.2 mM CuPhen for a final concentration of either 1 mM or 0.1 mM. Reactions proceed for 30 min at 25°C. Reactions were stopped with 20 mM NEM and 20 mM EDTA, and reactions were spun at 16,000 × *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 20 μl of loading buffer (Invitrogen LDS buffer, 8 M urea, 0.5 M NEM) and heated for 10 min at 70°C. Five microliters of sample was 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 Tris-buffered saline with Tween (TBST) buffer (10 mM Tris, pH 7.5, 2.5 mM EDTA, 50 mM 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 antimouse IGg (Pierce). Proteins were depicted by exposure to ECL reagent (Amersham, GE Health Sciences) for 1 min and exposure to film for 30–60 s. 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)).

Multistate 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 macro-molecular protein complexes by satisfaction of spatial restraints from a variety of experimental data. Model analysis is described in the Supplemental Experimental Procedures.

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 (Protein Data Bank [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 C2 symmetry around 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 (Figure 1A). For the subsequent sampling, each monomer was decomposed into six rigid bodies and five 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 $\{w_i\}$, the calibration parameters $\{\alpha_n\}$, and the uncertainty in the data $\{\sigma_n\}$. 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 the Supplemental Experimental Procedures.

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 Figure 5. 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-axes, 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.

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 (Figure S4) using the Levenberg-Marquardt algorithm from the GNU scientific library (Gal-

assi 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 around the helix axis and an arc radius of 1.5 Å × (# of helix residues)/2 for tilting motions. The full set of calculated displacements is given in Table S3.

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. USA 106, 16185–16190.

Alber, F., Dokudovskaya, S., Veenhoff, L.M., Zhang, W., Kipper, J., Devos, D., Suprapto, A., Karni-Schmidt, O., Williams, R., Chait, B.T., 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.

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. USA 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., and Hendrickson, W.A. (2009). Structural analysis of ligand stimulation of the histidine kinase NarX. Structure 17, 190–201.

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

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

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. USA 109, E1733–E1742.

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, H.U., Dunin-Horkawicz, S., Hornig, N., Hulko, M., Martin, J., Schultz, J.E., Zeth, K., Lupas, A.N., and Coles, M. (2012). Mechanism of regulation of receptor histidine kinases. Structure 20, 56–66.

Galassi, M., and Gough, B. (2005). GNU Scientific Library: Reference Manual. (Network Theory), http://www.gnu.org/software/gsl/manual/gsl-ref.html.

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., 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.

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. USA 107, 8141–8146.

Gordeliy, V.I., Labahn, J., Moukhametzianov, R., Efremov, R., Granzin, J., Schlesinger, R., Büldt, G., Savopol, T., Scheidig, A.J., Klare, J.P., and Engelhard, M. (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. USA 103, 1756–1761.

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

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. USA 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. J. Mol. Graph. 14, 33–38, 27–28.

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.

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 of Escherichia 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.

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. USA 109, E1481–E1488.

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. USA 97, 6298–6305.

Lowe, E.C., Baslé, 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. USA 109, 7298–7303.

Lynch, B.A., and Koshland, D.E., Jr. (1991). Disulfide cross-linking studies of the transmembrane regions of the aspartate sensory receptor of Escherichia coli. Proc. Natl. Acad. Sci. USA 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., and Choe, S. (2010). Membrane domain structures of three classes of histidine kinase receptors by cell-free expression and rapid NMR analysis. Proc. Natl. Acad. Sci. USA 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 alphaIIb/β3 transmembrane heterodimer. J. Mol. Biol. 392, 1087–1101.

Milburn, M.V., Privé, 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, 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. USA 86, 5054–5058.

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. USA 89, 4144–4148.

CellPress

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.

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-A resolution. Proc. Natl. Acad. Sci. USA *98*, 10131–10136.

Russel, D., Lasker, K., Webb, B., Velázquez-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.

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. (New York: Oxford University Press).

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.

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.

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., Wu, J., et al. (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. USA *101*, 959–963.

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.

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.
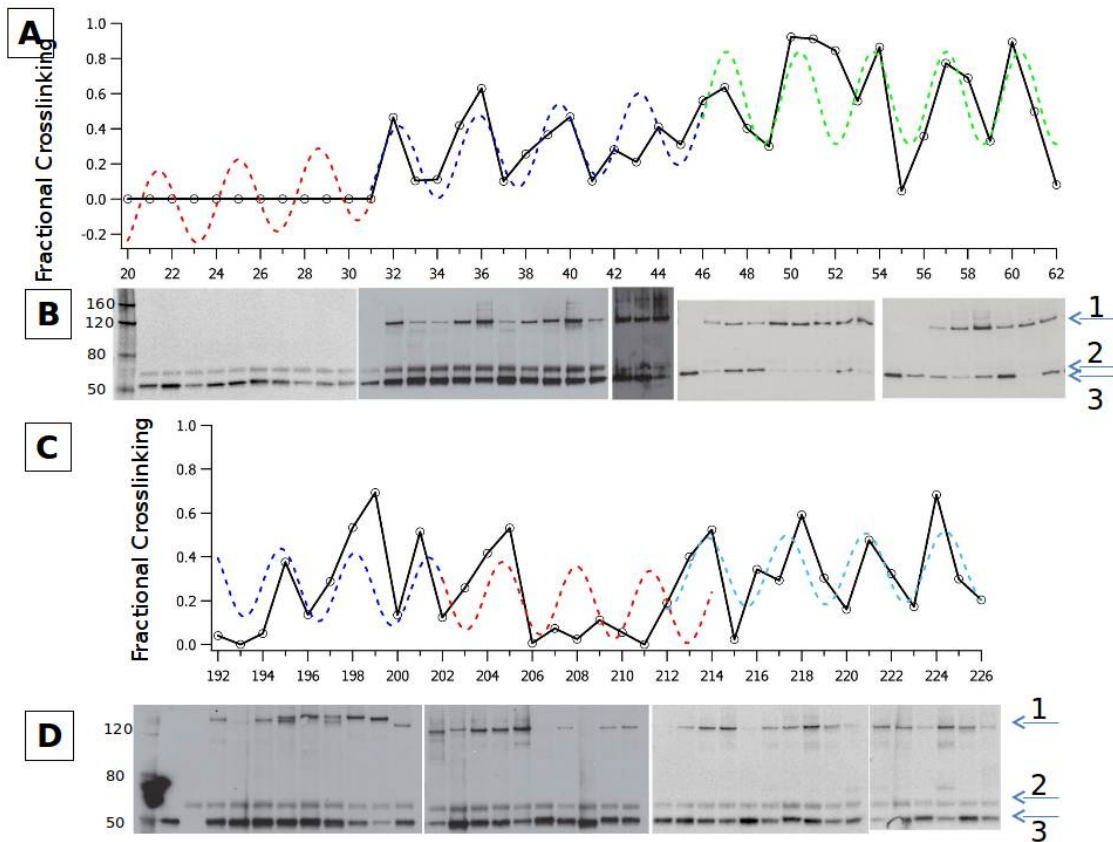
Cys-Scanning Disulfide Crosslinking and Bayesian

Modeling Probe the Transmembrane Signaling

Mechanism of the Histidine Kinase, PhoQ

Kathleen S. Molnar, Massimiliano Bonomi, Riccardo Pellarin, Graham D. Clinthorne, Gabriel Gonzalez, Shalom D. Goldberg, Mark Goulian, Andrej Sali, and William F. DeGrado

Supplemental Data (Figures)

Supplementary Figure 1 (Related to Figure 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**: 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**: 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).
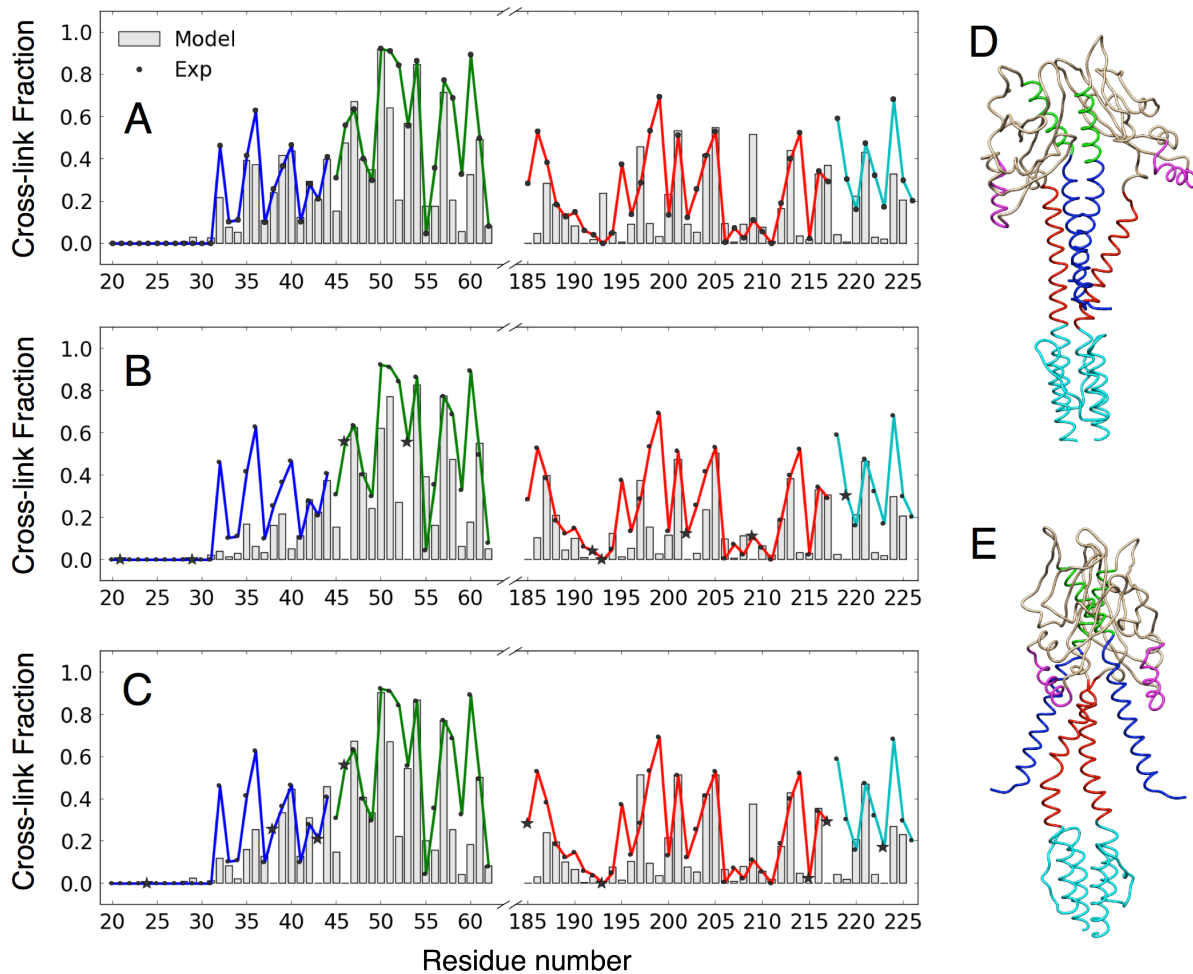
Supplementary Figure 2 (Related to Figure 3): Assessing the robustness of the two-state modeling Bayesian approach by data jackknifing. (A-C) Data predicted by the model (grey bars) with highest likelihood in the most populated cluster using the whole dataset of 85 cross-links (A) and randomly discarding 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.
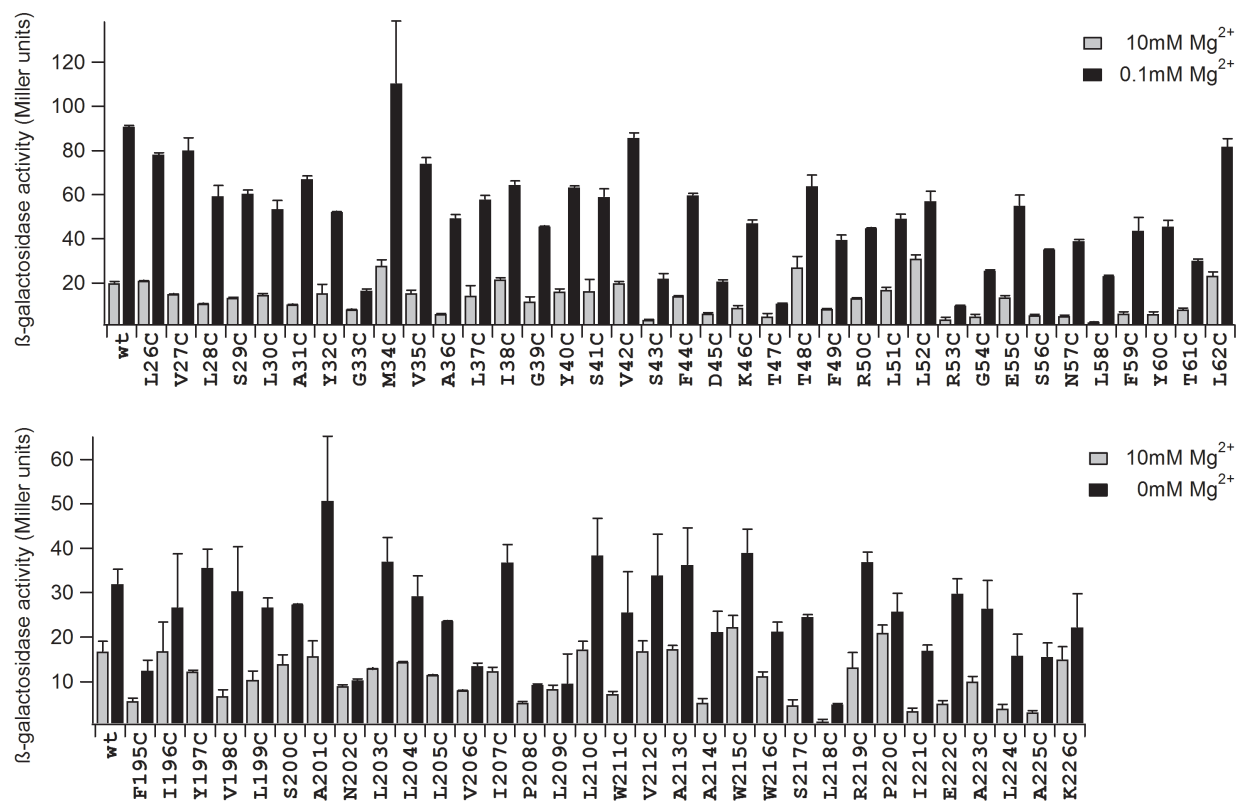
Supplementary Figure 3 (Related to Figure 4): Phenotypic changes in response to Cys mutations in PhoQ. We assessed the activity of Cys mutants by the Miller assay (Miller, 1972). TIM206 (*mgtA::lacZ ΔphoQ*) cells were transformed with a plasmid encoding PhoQ and a Cys mutation at a single position. Mutation 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+}.
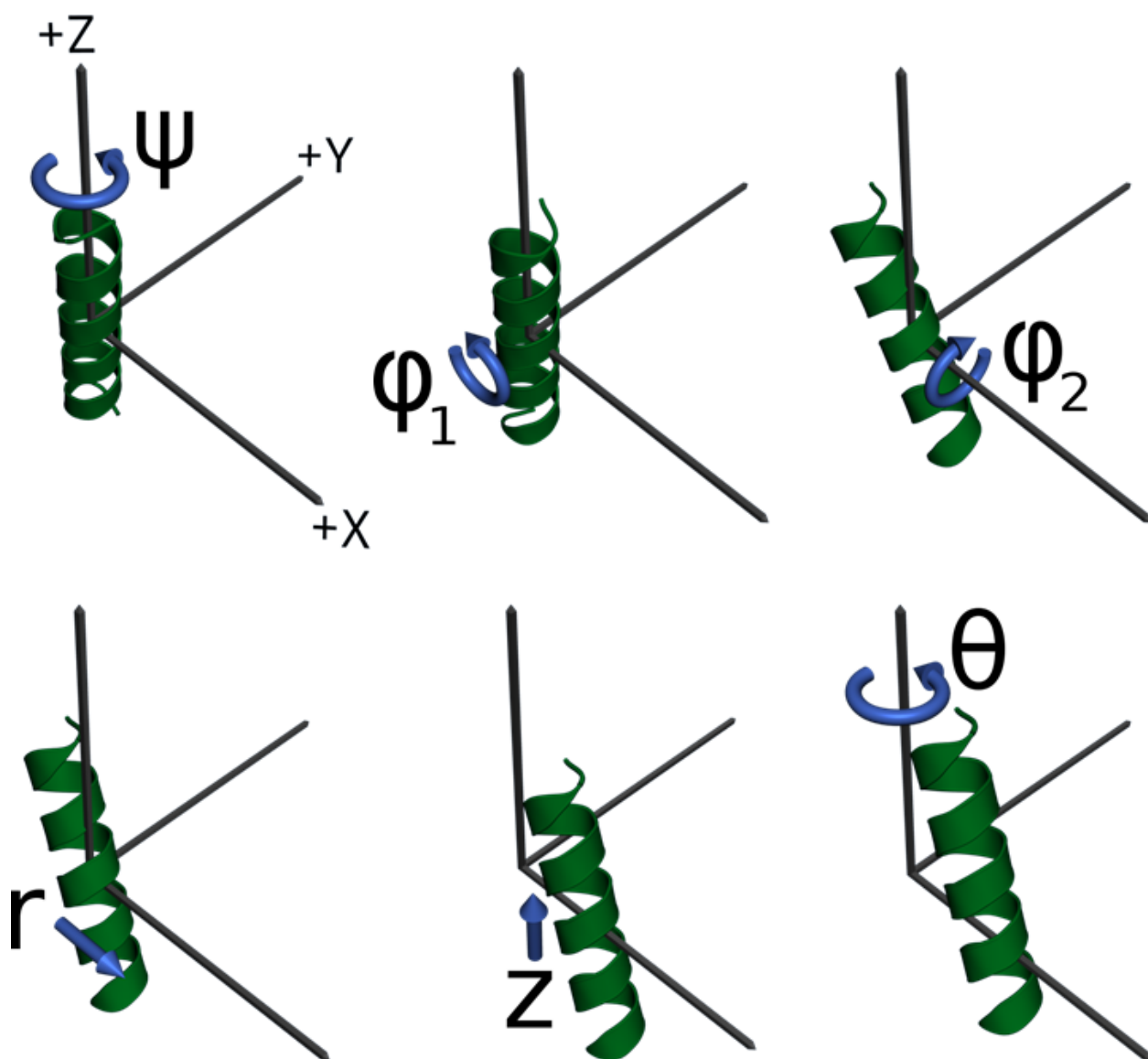
Supplementary Figure 4 (Related to Figure 5): The six degrees of motion in the order they are applied to fit any given helix: ψ: first rotation about Z axis; φ1: rotation about Y axis; φ2: rotation about X axis; r: translation along X axis; z: translation along Z axis; θ: second rotation about Z axis.

Sensor - CitA

HAMP - AF1503

DHp - DesK
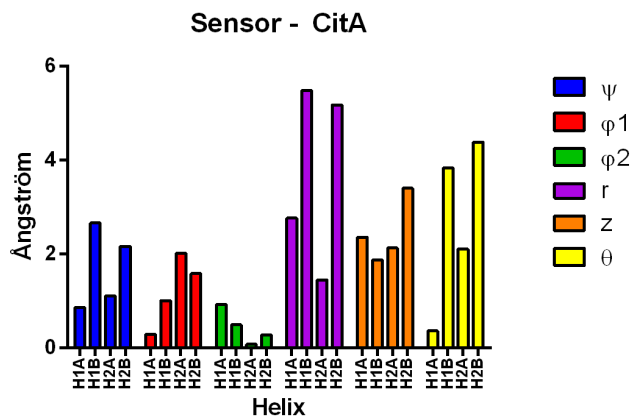
Supplementary Figure 5 (Related to Figure 8): Measured differences between equivalent helices in two component systems. Each chart corresponds to a single set of domains referenced by domain type (e.g. Sensor or HAMP) and protein (e.g. AF1503 or DesK). Each chart groups the differences by direction, 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.
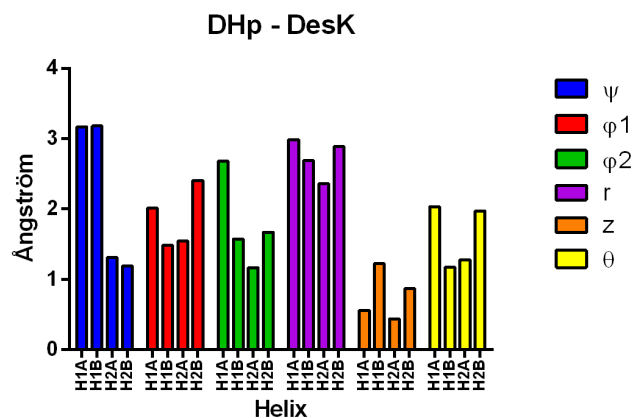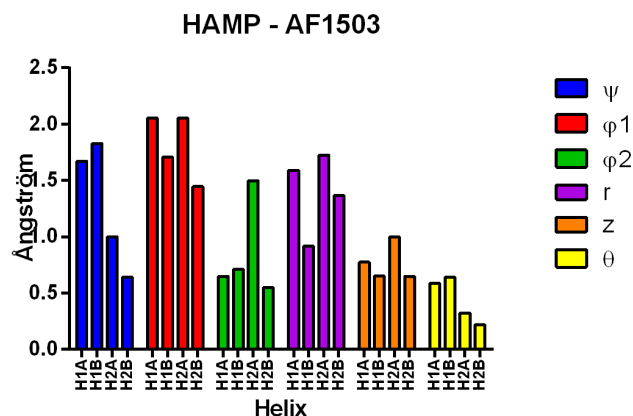
Supplemental Data (Tables)

Supplemental Table 1 (Related to Figure 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 χ^2	Center L	Best χ^2	Best 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 (Related to Figure 5). 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 3 (Related to Figure 5). 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 4**

Supplemental Data (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):

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

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 $p(\{X_i\}) \propto \exp\left(-\sum_i V(X_i)\right)$ 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 $V_{C_\alpha\ bonds}$, $V_{C_\alpha\ angles}$, 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 [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 probabilitys 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).

Supplemental References

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.

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.

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

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

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.

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

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.

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.

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