UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

A COMPUTATIONAL APPROACH FOR ACCESSING PHOSPHORYLATED RESPONSE REGULATOR CONFORMATIONS AND SIGNALING COMPLEXES INVOLVING THE FUNGAL PHOSPHORELAY PROTEIN YPD1

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

CLAY ALLEN FOSTER
Norman, Oklahoma
2016

A COMPUTATIONAL APPROACH FOR ACCESSING PHOSPHORYLATED RESPONSE REGULATOR CONFORMATIONS AND SIGNALING COMPLEXES INVOLVING THE FUNGAL PHOSPHORELAY PROTEIN YPD1

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF CHEMISTRY AND BIOCHEMISTRY

BY

_____
Dr. Ann H. West, Chair

_____
Dr. Elena Zgurskaya

_____
Dr. Robert Cichewicz

_____
Dr. Ari Berkowitz

_____
Dr. Anne Dunn

Acknowledgements

I would first like to acknowledge my major professor, Dr. Ann West, who has provided invaluable guidance and advice throughout my graduate career. Dr. West has supported my work continuously and enthusiastically, no matter the path each project took. This work would not have been possible without her, and she has my profound appreciation and gratitude. Additional thanks go to the members of my graduate committee, Dr. Elena Zgurskaya, Dr. Robert Cichewicz, Dr. Anne Dunn and Dr. Ari Berkowitz, for their continued guidance and feedback.

I would also like to thank my wife, Dr. Katie Branscum, for her love and support, as well as Dr. Paul Sims, Dr. Paul Cook, Dr. Len Thomas, Dr. Fares Najar and Dr. Eliza Ruben for their intellectual input. Dr. Len Thomas also assisted with protein crystallography and data collection. Dr. Katie Branscum, Dr. Smita Menon, Emily Kennedy, Jamie Sykes, Skyler Hebdon and Krutik Soni were involved collaboratively in various portions of these projects and provided insightful discussion in all areas of research.

Table of Contents

List of Tables

List of Figures

Abstract

Two-component signaling is the primary means by which bacteria, archaea and certain eukaryotes sense and respond to their environments. Signal transfer proceeds through sequential His-to-Asp phosphorylation of upstream histidine kinases and downstream response regulators. These systems share highly modular designs and have been incorporated into a myriad of cellular processes. The highly labile chemical natures of phosphoaspartate and phosphohistidine lead to relatively short experimental life-times, making study of the modified signaling proteins challenging. The focus of this research was to develop computational and experimental approaches for characterizing phosphorylated two-component signaling proteins. Following an introductory chapter, the first experimental section presents a computational technique for simulating the activation of individual response regulator proteins. This is accomplished using known experimental data on conserved active site chemistry to define a common set of restraints to drive each simulation. The protocol was verified on five genetically diverse response regulators with known experimental structures. The second section applies this principle to signaling complexes to study the effects of phosphorylation on protein-protein interactions within the *Saccharomyces cerevisiae* osmoregulatory signaling system. The third section describes the experimental characterization of a specific signaling complex from *Saccharomyces cerevisiae* between the response regulator Ssk1 and a point mutant (G68Q) of the histidine phosphotransfer protein Ypd1 using X-ray crystallography. This mutation occurs near the active site of both proteins and appears to interfere with phosphotransfer. Further *in silico* studies were performed to observe the role of G68 in catalysis of phosphotransfer.

Chapter 1: Introduction

1.1 Signal transduction

To survive, organisms must rapidly and continuously adapt to their environments. The process of detecting, internalizing, transmitting and responding to extracellular signals is known as signal transduction [1-3]. Cells are highly sensitive to their chemical environments and have developed numerous methods to detect minor changes in their surroundings. Most of these involve the use of membrane-bound receptors to bind specific ligands, though some non-polar molecules are able to diffuse directly through the membrane [1, 4-7]. These ligands, also known as primary messengers, are the first step in the cellular signaling process [1, 7].

Once a signal has been detected and internalized through a cellular receptor, it must then be converted into different chemical form(s) through a process known as transduction [1]. This generates another group of small signaling molecules within the cell, known as secondary messengers. These are involved in relaying signals from the receptors to the rest of the cellular machinery [8]. Several well-characterized examples of secondary messengers include calcium, cyclic AMP and cyclic GMP [9-11]. Secondary messengers can be used to both amplify a signal and activate further cellular components, such as enzymes or membrane channels. This entire signaling circuit is often referred to as a signal transduction cascade.

1.2 Protein phosphorylation in signaling cascades

Reversible, post-translational modifications are common regulatory mechanisms found in signal-transduction cascades [12]. One such modification involves the phosphorylation of a protein substrate. Phosphorylation is ubiquitous in both prokaryotic and eukaryotic organisms (reviewed in [13-15]). The first formal description of a phosphorylation-dependent signaling cascade was made over sixty years ago by two independent groups. Krebs and Fischer [16, 17] and Rall [18] demonstrated the interconversion of phosphorylase *a* and *b* by phosphorylase kinase. This discovery was the earliest documented example of reversible protein phosphorylation, a critical process eventually found in nearly every living cell [14]. The reversible process of phosphorylation/dephosphorylation is controlled by two families of enzymes: kinases and phosphatases [19]. Kinases are frequently activated by secondary messenger molecules and are responsible for binding ATP and transferring γ-phosphoryl groups to chemically appropriate acceptor residues [20]. Phosphatases are involved in the catalytic removal of these phosphoryl groups [21]. Eukaryotic kinases typically phosphorylate serine, threonine or tyrosine residues, often as part of a kinase cascade [14, 22-25]. Serine/threonine/tyrosine kinases are relatively stable in their phosphorylated forms and must be enzymatically dephosphorylated, usually by their phosphatase counterparts [26-28]. In prokaryotes, histidine and aspartate phosphorylation is the primary method of signal transduction (reviewed in [29,

30]). His-to-Asp phosphotransfer is fundamentally different from Ser/Thr/Tyr phosphorylation. Eukaryotic Ser/Thr/Tyr phosphorylation occurs as independent events, with each step requiring a molecule of ATP and a kinase [31, 32]. In contrast, prokaryotes utilize histidine and aspartate residues as relay points, shuttling the same phosphoryl group between multiple signaling components [31, 32]. This method takes advantage of the chemical labilities of phosphorylated histidine and aspartate residues to drive transduction [27, 28].

1.3 Mechanism of reversible phosphorylation

Due to its prevalence, phosphorylation is arguably the most highly-studied of all post-translational modifications. The covalent attachment of a negatively charged phosphoryl group onto an appropriate active site residue has been shown to cause dramatic structural and functional changes in target proteins [33-38]. This allosteric effect makes phosphorylation an ideal regulatory technique for cellular signaling pathways. The mechanism of phosphorylation begins with the universal phosphoryl donor, adenosine triphosphate (ATP). The energy found in ATP is stored in two high-energy phosphoanhydride bonds. Protein kinases bind ATP, catalyze the breaking of these linkages and facilitate the transfer of the terminal phosphate group (γ-PO_3^{2-}) to chemically appropriate amino acid residues (reviewed in [20, 39]). The large amount of energy released by the breaking of these bonds makes this reaction unidirectional [40].

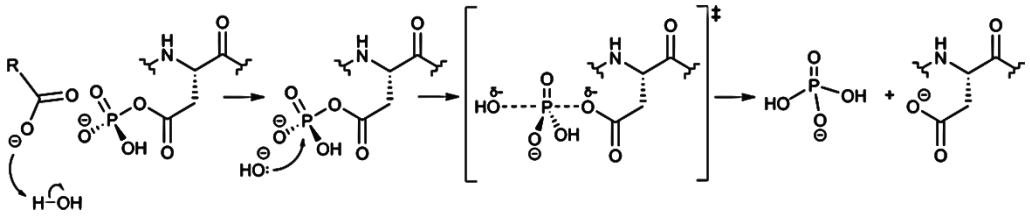

Figure 1. Hydrolytic dephosphorylation of phosphoaspartate.

An acidic side chain (Asp/Glu) is used to activate a nearby water molecule, which performs an in-line, nucleophilic attack at the phosphorus atom [28]. This results in the hydrolysis of the phosphoaspartate species and the generation of inorganic phosphate. The high energy acyl phosphate bond makes phosphoaspartate highly labile [28].

In eukaryotes, hydroxyl groups on serine, threonine and tyrosine side chains are activated by nearby bases to perform a nucleophilic attack on the phosphorus atom of the γ-PO_3^{2-} group. The process requires the presence of a divalent metal cation (e.g. Mg^{2+}) to neutralize the negatively-charged phosphate groups and facilitate the nucleophilic attack [20, 39]. In prokaryotes, the mechanism is similar, though it is usually the side chains of histidine and aspartate residues that act as nucleophiles [41]. The phosphoramidate and acyl phosphate bonds of phosphohistidine and phosphoaspartate are higher energy than their eukaryotic counterparts (see **Table 1**) [42, 43]. This explains their highly labile chemical natures and susceptibility to rapid hydrolytic dephosphorylation [44, 45] (see **Fig. 1**).

Table 1. Standard Gibbs free energy of hydrolysis

Phospho-amino acid	$\Delta G_{hydrolysis}$ (kcal/mol)	Reference(s)
Phosphohistidine	-12	[48]
Phosphoaspartate	-10.3*	[30, 45]
Phosphotyrosine	-9.5	[49, 50]
Phosphoserine	-6.5	[51]

Adenosine triphosphate	-7.3	[52]

* Approximation, based on phosphorylation by acetyl phosphate

1.4 Two-component signaling systems

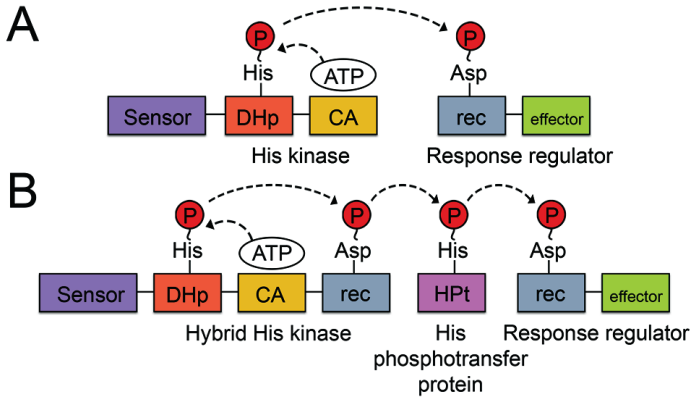

Figure 2. Architectures for a basic TCS system and a phosphorelay.

(A) Canonical two-component signaling pathway. Signals are sensed by the histidine kinase, which binds ATP in the catalytic (CA) domain, and autophosphorylates on a histidine residue in the dimerization and histidine phosphotransfer (DHp) domain. The phosphoryl is transferred to an aspartate on the receiver (rec) domain of a downstream response regulator (RR) protein. (B) Expanded multistep phosphorelay. Contains a hybrid HK, with its own receiver domain, and an intermediate known as a histidine-containing phosphotransfer (HPt) protein.

Two-component signaling (TCS) systems are the primary means by which all bacteria, as well as certain eukaryotes and archaea, detect and respond to their environments (extensively reviewed in [30] and [29]). TCS pathways utilize reversible protein phosphorylation and phosphotransfer as a means to couple external stimuli with a cellular response. A prototypical TCS system contains two signaling proteins: an upstream sensor histidine kinase (HK) and a downstream response regulator protein (RR) [29, 30]. The HK is responsible for detecting or receiving the extracellular stimulus, binding ATP

and a metal cation (Mg^{2+}), and autophosphorylating on a histidine residue. The RR receives this same phosphoryl from the HK on an aspartate residue within its receiver (rec) domain and triggers an appropriate cellular response, often through regulating the output of an attached effector domain [29, 30] (see **Fig. 2A**).

TCS systems are ubiquitous in bacteria, responsible for such diverse cellular functions as cellular communication, growth and propagation, stress response and pathogenesis [46-49]. Bacterial HKs are frequently encoded on the same operon as their cognate RRs, making bioinformatic census-taking relatively straightforward [50]. Genetic analysis suggests that the number of TCS genes for a given organism grows approximately as a square of genome size [51]. Most bacteria contain an average of 20-50 cognate HK-RR pairs, though numbers vary greatly, with some genomes encoding several hundred individual proteins [50, 51]. The number of encoded TCS proteins in a genome also appears to be related to the ecological niche filled by the organism [50-54]. Bacteria that inhabit relatively stable environments frequently contain few TCS genes, while the opposite is true for organisms that inhabit more volatile settings. The presence of so many paralogous signaling proteins within a single organism creates the possibility of cross-talk between different TCS systems [50, 55, 56]. While little sequence similarity exists between them, TCS proteins frequently share high structural and catalytic homology. Cellular viability depends on preventing the promiscuity of individual TCS

components. Though the mechanism(s) for regulating TCS proteins and preventing cross-talk is not fully understood, recent studies suggests that promiscuity within bacteria remains relatively low [46, 50, 56-58]. Most HKs display a marked kinetic preference for their cognate RR, as reviewed in [57]. Examples of this include EnvZ with OmpR (osmoregulation), CC1181 with CC1182 (function unknown), and RstB with RstA (biofilm formation and iron acquisition) [59-62]. While molecular recognition at an amino acid level is thought to generate the largest influence on the specificity of signaling interactions, recent work indicates that co-localization, transcriptional regulation and phosphatase activity all play parts in preventing heterologous activation [23, 56, 63-66]. In addition to preventing cross-talk between pathways, many of these same regulatory measures are required within single pathways, especially in complex eukaryotic systems that may contain multiple branch points.

1.5 Multi-step phosphorelays

Occurrence in prokaryotes

The modularity of two-component signaling pathways has allowed for the evolution of more complex system architectures [67, 68]. While the majority of prokaryotic TCS systems utilize a single HK and RR pair, expanded pathways that incorporate multiple phosphotransfer events have been observed [69, 70]. These systems are referred to as multi-step phosphorelays, in which the same phosphoryl group is shuttled from His-to-Asp-to-His-to-

Asp. Phosphorelays commonly utilize a hybrid HK, incorporating both an HK and RR rec domain in a single protein (**Fig. 2B**). Introduction of this hybrid protein coincides with the addition of an intermediary histidine phosphotransfer (HPt) domain. The *Escherichia coli* RcsC-YojN-RcsB signaling pathway, responsible for swarming behavior and capsular synthesis, is one example of this type of expanded system found in bacteria [71-73]. The kinase RcsC autophosphorylates on a histidine residue, transfers the phosphoryl to the HPt domain of YojN, which finally transfers to the cytosolic response regulator RcsB. Other phosphorelay architectures exist, such as the sporulation pathway in *Bacillus subtilis* [74]. This system includes separate proteins for every domain. Transduction begins with the autophosphorylation of a sensor kinase (KinA, KinB or KinC), which then transfers the phosphoryl to Spo0F. The intermediate regulator then phosphorylates Spo0B, which finally transfers the phosphoryl to Spo0A [75, 76]. A third type of expanded system also uses a hybrid HK, with the kinase, rec and HPt domains incorporated into the same protein. The *Bordetella pertussis* BvgS-BvgA system is one example of such a system, involving the transcriptional regulation of virulence factors [77, 78].

Occurrence in eukaryotes

For decades, Ser/Thr/Tyr phosphorylation was thought to be restricted to eukaryotic organisms, while His-to-Asp phosphotransfer was considered thoroughly prokaryotic. However, multiple studies published in 1993

overturned this belief with the discovery of bacterial Ser/Thr kinases and eukaryotic TCS proteins [79-81]. As genomic information became more widely available, two-component signaling was revealed to be surprisingly prevalent in eukaryotes, though overall TCS systems make up a relatively small percentage of their signaling pathways. Most eukaryotic TCS systems have been found in plants, slime mold and other fungi [82]. Genetic analysis of available prokaryotic and eukaryotic genomes reveals that hybrid HKs make up less than 20% of all bacterial TCS kinases, but more than 90% of eukaryotic histidine kinases [67]. This suggests that the majority of eukaryotic TCS systems are expanded phosphorelays, rather than canonical TCS pathways more commonly found in prokaryotes. To date, no TCS systems have been identified in animals [83]. Eukaryotic phosphorelays utilize the same principle for signal transduction as bacterial TCS systems, though they frequently include a greater number and diversity of signaling proteins. Almost every fungal and slime mold TCS pathway includes a hybrid HK, incorporating both a kinase and rec domain, and a stand-alone HPt protein [82, 84, 85] (**Fig. 2B**). The osmoregulatory Sln1 pathway found in *Saccharomyces cerevisiae* is one well-characterized example [86]. The membrane-bound hybrid sensor HK known as Sln1 binds ATP and autophosphorylates on a histidine residue under normal conditions. The phosphoryl group is then transferred to an attached rec domain, which binds and phosphorylates the intermediate HPt, Ypd1. The HPt can then phosphorylate two downstream RRs, Skn7 or Ssk1 [87-91]. Under

osmotic stress conditions, Sln1 ceases to autophosphorylate, allowing unphosphorylated downstream elements to accumulate and activating a cellular response. This branching architecture is a common feature of many eukaryotic TCS systems.

Evolutionary advantages of expanded phosphorelays

No definitive explanation exists for the development of phosphorelays or their prevalence in eukaryotes, though many have speculated on the evolutionary advantages that such systems might provide. The compartmentalization of the eukaryotic cell may explain the widespread adoption of the HPt protein [84]. A membrane-bound sensor kinase would likely be unable to transfer a phosphoryl to different compartments, such as the nucleus, where eukaryotic RR proteins frequently reside. A cytoplasmic HPt protein, such as Ypd1, would fulfill the role of an intermediate shuttle between the upstream and downstream components [92]. The presence of additional proteins provides more regulatory checkpoints within the pathway, as well as potential intersections with other signaling systems [69, 70]. Hybridization of HKs with their own downstream partners may also eliminate the danger of cross-talk between pathways. Fusion of two proteins effectively uses co-localization to prevent promiscuity [93]. Recent studies using systems biology approaches have suggested that expanded phosphorelays may provide multiple signal-processing advantages to organisms. Kinetic variation in transfer and hydrolysis rates can cause changes in the signal-response relationship of a

pathway, allowing either binary responses or responses that scale with the level of the incoming signal [94]. Response dynamics of phosphorelays also impart ultrasensitivity to the intermediate layers of the pathway above a certain threshold, but low responses below this threshold [95]. This creates a natural tolerance for cross-talk within the system, as well as a high signal-to-noise ratio at the final layer of the pathway.

1.6 Histidine kinases

Histidine kinase structure

Histidine kinases (HKs) are typically found at the apex of TCS pathways and are responsible for sensing, internalizing and transmitting the extracellular stimuli to the rest of the pathway [30]. They are frequently multi-domain proteins that exhibit significant architectural diversity. Activity is typically regulated by an attached sensor domain. Each HK is grouped into one of three different structural classes, based on the location of this sensory region [96]. Class I HKs are membrane-bound and utilize an extracellular sensor. Class II HKs are also membrane-bound, but their sensory region is thought to reside within the transmembrane domain. Class III HKs include a single cytoplasmic sensory domain. The large majority of HKs are Class I, coupling an extracellular sensor with a cytoplasmic, C-terminal catalytic region (reviewed extensively in [29, 97]). Though sensor domains are highly variable in structure and function, due to the wide range of environmental stimuli they must detect, the cytoplasmic kinase domain is far more conserved and contains

several sequence motifs that are characteristic of nearly all HKs. These motifs are known as the H, N, G1, G2 and F boxes, based on the conserved residues associated with the motif [98].

Canonical HKs are typically constitutive homodimers (**Fig. 3)**. Each monomeric subunit begins with an N-terminal, antiparallel helix, a periplasmic sensor domain and a second antiparallel helix [97]. Upon dimerization, the helices, also known as the transmembrane (TM) domain, form a four-helix bundle linking the extracellular sensor to the cytoplasmic kinase region. The TM helices are followed by an intracellular signal-transducing region, often incorporating HAMP, PAS, GAF and coiled coil domains [97]. Next is the dimerization and histidine phosphotransfer (DHp) domain, defined by two antiparallel helices connected through a hairpin loop. Upon dimerization, these helices form a conserved, homodimeric, four-helix bundle. The DHp domain contains the phosphorylatable histidine residue on the first helix, located within a region known as the H-box [97]. This area also forms a binding surface for downstream elements within the pathway. Finally, connected to the DHp domain by a flexible linker loop is the C-terminal catalytic and ATP-binding domain (CA). This region adopts a highly conserved α/β sandwich fold, formed by three α-helices and a five-stranded, mixed β-sheet. ATP binds between two helices within the N, G1, G2 and F box motifs and is locked in place through a flexible a loop known as the "ATP lid." The final helix within the CA domain, called the "Gripper" helix by Bhate, et al. (2015), interacts

with the DHp domain alongside the F box in response to nucleotide binding [97].

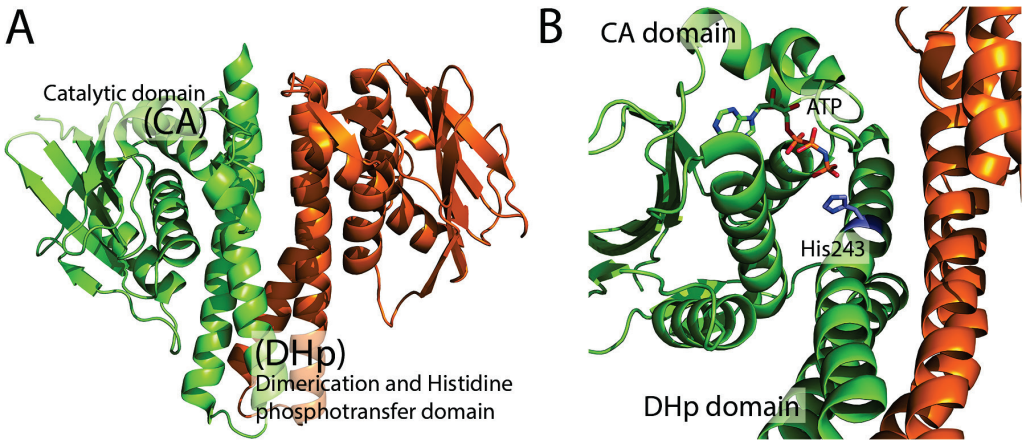

Figure 3. Ribbon representation of the dimeric histidine kinase, EnvZ.

(A) Chimeric homodimer of EnvZ (PDB 4KP4) from *Eschericia coli/Thermotoga maritima*. Shown are the dimerization and histidine phosphotransfer domain (DHp) and the catalytic domain (CA). Color corresponds to monomeric subunits. (B) Phosphorylatable histidine (purple) and bound ATP molecule (orange) shown in stick model.

Histidine kinase function

Kinase activity begins with the detection of an extracellular stimulus within the sensor domain [99]. Sensory regions vary greatly in both sequence and structure, producing significant effects on pathway function and architecture. A basic detection event involves direct interaction between the sensor domain and a cognate ligand, such as a small molecule or an additional binding protein. A well-studied example of this is the nitrate reductase (Nar) TCS system found within *E. coli* [100, 101]. In this phosphorelay, both NarX and NarQ fill the role of sensor kinsases, binding nitrate and nitrite groups independently. Detection activates the response regulators NarL and NarP,

which then modulate gene expression. In other instances, the sensor responds to mechanical or electrochemical stress, such as turgor pressure or an ionic gradient. The PhoQ/PhoP TCS system, responsible for pathogenesis in many Gram-negative bacteria, contains a sensor kinase (PhoQ) that is activated based cationic concentration, such as Mg^{2+} levels [102].

Activation of the sensor domain is believed to be transmitted through the periplasmic helix connecting the sensor to the TM domain [97]. This passes the signal through the signal-transducing domain to the cytoplasmic catalytic core, initiating nucleotide binding and autophosphorylation. ATP is bound within the CA domain, inducing a dramatic conformational change that reorients the CA and DHp domain through changes in the hydrophobic "Gripper" helix [97]. Autophosphorylation involves the transfer of a γ-phosphate from the nucleotide to the Nε-atom of the catalytic histidine, forming a high-energy phosphoramidate bond [26, 103]. While catalysis is influenced primarily through the CA domain, with numerous side chain interactions to the oxygen atoms of the γ-phosphate, the DHp domain also provides several conserved residues involved in the overall reaction [97, 104]. An acidic residue (Asp or Glu) located next to the catalytic histidine functions as a hydrogen bond acceptor, preparing the histidine for nucleophilic attack on the phosphate. A basic residue (Lys or Arg), located three positions from the histidine, is also believed to interact with or stabilize the acidic phosphate groups of the nucleotide [97, 104]. **Fig. 3** shows the cytoplasmic structure of a

chimeric construct of the kinase EnvZ from *Thermotoga maritima* [104]. The dimeric nature of the HK produces a structure reminiscent of a butterfly [29]. The autophosphorylation reaction can occur in both a *cis* and *trans* manner, depending on the handedness of the hairpin loop found between the DHp helices [97, 104-106]. If the loop turns left, the CA domain of one monomer ends up nearer to its own catalytic histidine and favors a *cis* reaction (same subunit). If the loop turns right, the opposite is true, and the CA domain of one subunit ends up nearer the catalytic histidine of the other subunit and favors a *trans* reaction (different subunits). Activation of the HK ultimately results in autophosphorylation and the binding of a downstream rec domain, either on a separate RR protein or its own attached domain in the case of a hybrid HK.

1.7 Histidine phosphotransfer proteins

Histidine phosphotransfer protein structure

Expanded TCS or multi-step His-to-Asp phosphorelay systems require the addition of intermediate phosphotransfer steps. This is accomplished through the histidine phosphotransfer (HPt) domain. HPt proteins are found in both prokaryotic and eukaryotic systems, though in prokaryotes they are frequently part of hybrid kinases, while eukaryotes typically utilize standalone domains [30, 84, 107, 108]. They are relatively small when compared to HKs, usually around 100-200 residues in length. Sequence conservation among HPts is quite low, though they do contain an invariant catalytic histidine residue surrounded by a short consensus motif that aids in phosphotransfer [109, 110].

HPt domains exhibit neither kinase nor phosphatase activity, making them well suited for their intermediate role in the relay [108, 109]. All HPts characterized to date adopt a universal, four-helix bundle structure shown in **Fig. 4.** This conserved tertiary motif is unusually reminiscent of the four-helix bundle created by canonical HKs, though HPts typically function as monomers and share little sequence homology with their kinase counterparts. **Fig. 4** shows the crystal structure of the *S. cerevisiae* HPt, Ypd1 [84, 111]. The core bundle is formed by four helices, αB-αC-αD-αG. Additionally, helix αA appears to stabilize the bundle by shielding hydrophobic residues on helices αB and αG and the conserved reverse turn connecting αC and αD. Helix αA is also important, as it has been shown to be involved in RR partner binding [112].

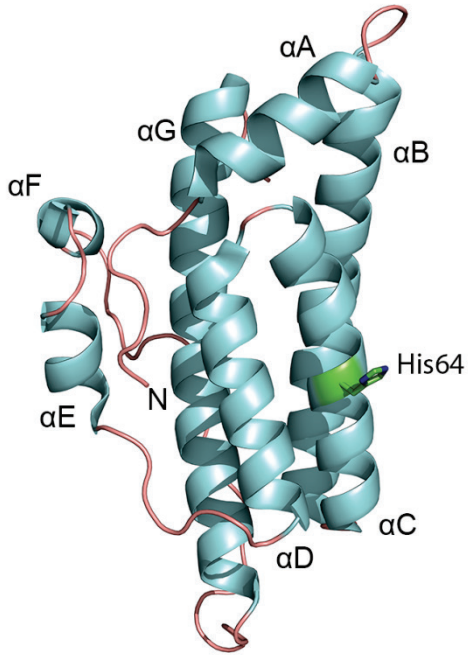

Figure 4. Ribbon representation of HPt protein Ypd1 from *S. cerevisiae.*

Crystal structure of the intermediate HPt protein, Ypd1, from the fungal Sln1 phosphorelay (PDB 1OXB:A). Shown is the conserved four-helix bundle fold adopted by all HPt domains. α-helices are labeled A-G. Phosphorylatable histidine is shown in stick model (green).

Histidine phosphotransfer protein function

In HPt domains, the invariant phosphorylatable histidine residue is located within the middle of the αC helix, and is almost completely exposed to solvent. The active sites of the few known HPt structures suggest that the surrounding residues are arranged to maintain the solvent accessibility of this histidine [112]. In Ypd1 (**Fig. 4**), a highly conserved glycine residue (Gly68) is located four positions downstream from the phosphorylatable His64 (H+4). This position is highly conserved in nearly every HPt protein. It has been hypothesized that this small residue may be necessary to allow access to the histidine side chain [113]. Recent findings also suggest that it may be required

for catalytic interaction with partner receiver domains (West lab, unpublished data). Additionally, the side chains of nearby Asn61 and Lys67 are pulled away from the imidazole ring of His64, maximizing accessibility and potentially stabilizing inter-helical structure through electrostatic interactions [112, 113]. The position occupied by HPts within phosphorelay systems indicates that they must be able to interact with multiple rec domain partners. High structural homology suggests that this occurs through a common interaction surface shared by all HPt domains [114, 115]. The structure of Ypd1 (**Fig. 4)** shows the phosphorylatable histidine residue (His64) in stick model located on the αC helix. This, along with helices αA, and αD, constitutes the conserved rec domain binding surface [116, 117]. Conserved side chains near the histidine are likely used to ensure proper orientation and accessibility of the imidazole ring for the donor/acceptor aspartate on a bound rec domain. A more in-depth analysis of Ypd1 can be found in Chapter 4.

1.8 Response regulator proteins

Response regulator structure

Response regulators (RRs) are typically found at the terminal end of TCS systems. They are responsible for regulating the output of the pathway, often at the level of transcription, ultimately resulting in a cellular response to the initial stimulus (reviewed in [29, 30, 118]). Many RRs are multi-domain proteins, composed of rec domains fused to secondary functional regions known as effector domains. While effector domains vary greatly in structure,

all RRs share a conserved receiver (rec) domain that acts as the phosphorylation-mediated switch for the rest of the protein. Sequence homology between RRs is typically <30%, but nearly all rec domains adopt a conserved $(\beta\alpha)_5$ fold, consisting of a five-stranded parallel β-sheet flanked by five amphipathic α-helices [118]. The C-terminal ends of the β-strands form a phosphorylatable active site, and the loops arrange into a phosphotransfer-competent binding surface. **Fig. 5A** shows this universal $(\beta\alpha)_5$ topology using the crystal structure of the well-characterized chemotaxis RR, CheY (PDB 3CHY) [119].

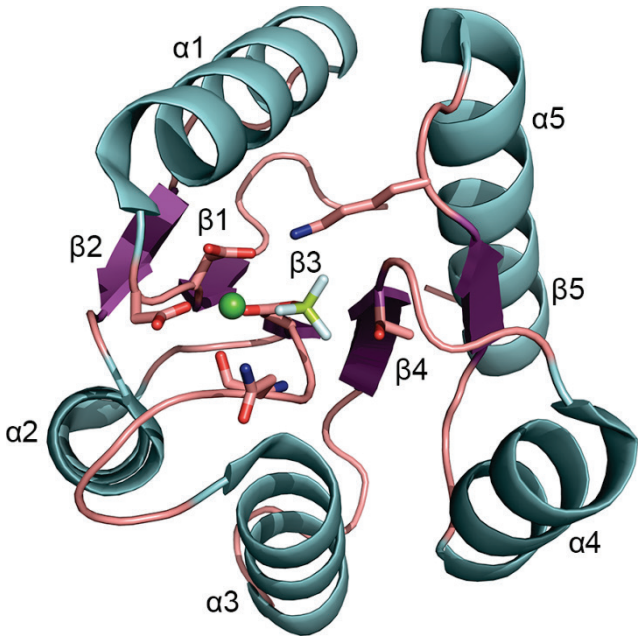

Figure 5. Ribbon representation of the rec domain for CheY from *E. coli.*

The single-domain chemotaxis RR, CheY bound with BeF_3^- (stick model; yellow). Shown is the prototypical $(\beta\alpha)_5$ topology adopted by nearly every rec domain. Shown in stick model are the conserved active site residues surrounding a divalent metal cation (Mn^{2+}; sphere; green).

Response regulator function

RR function is heavily dependent on the effector domain(s). Because of this, classification is normally based on effector domain sequence homology. The vast majority of effector domains have DNA-binding activity and are involved in transcriptional regulation [120]. Other categories include RNA-binding, protein binding and various enzymatic functions. Additionally, certain RRs lack an effector domain altogether and operate as stand-alone components within their signaling pathways, such as CheY from *E. coli* [121]. Because of this diversity, RRs are involved in an extremely wide range of cellular processes, including chemotaxis [121], sporulation [75], biofilm formation [122], and membrane-protein synthesis [123].

Rec domains fulfill two distinct chemical functions. Both involve a highly conserved phosphorylatable aspartate residue. In simple TCS systems, rec domains act as phosphoacceptors for the upstream phosphohistidine donor. In multi-step phosphorelays, rec domains can also function as phosphodonors to downstream histidine residues. Rec domain active sites feature a quintet of conserved, catalytic residues centered around interactions with a divalent metal cation and the phosphoryl group (reviewed in [118]; see **Fig. 5B**). The cation likely helps neutralize the highly negative nature of the active site pocket and is essential for the addition and removal of phosphoryl groups. Three conserved acidic residues participate in the coordination of this metal: the phosphorylatable aspartate, an aspartate near the β1-strand, and a third acidic

residue nearby (indirectly through an ordered water molecule). Metals such as Mg^{2+}, Mn^{2+}, Cu^{2+} and Ca^{2+} have all been observed in rec domain active sites [124, 125]. On the other side of the active site is a conserved lysine residue on the β5-strand and conserved Thr/Ser and Ala residues on the β4-strand/loop, both of which directly interact with the phosphoryl group. The organization of this active site allows rec domains to catalyze their phosphorylation, as evidenced by their ability to autophosphorylate *in vitro* using various small-molecule donors such as acetyl phosphate [126]. Autophosphorylation and/or phosphotransfer are initiated by the phosphorylatable aspartate, which performs an in-line, nucleophilic attack on the phosphorus atom and the donor bond (e.g., the phosphoramidate bond on phosphohistidine). This is thought to form a bipyramidal, pentavalent transition state, consisting of a planar PO_3^{2-} and axial attacking and leaving groups [118, 127]. The metal cation along with the conserved lysine and Thr/Ser residues are hypothesized to help stabilize this transition state by neutralizing the negative charges on the oxygen atoms of the phosphoryl group [128]. Autodephosphorylation likely occurs in a similar manner, with a hydrolytic water molecule assuming the nucleophilic attacking role (see **Fig. 1**). While the exact nature of this mechanism is poorly understood and difficult to verify experimentally, phosphotransfer mechanisms in proteins are typically classified by the tightness of the transition state, often described by their level of associative character [104, 129, 130].

Phosphorylation is known to induce several key conformational changes in rec domains. In addition to ordering the active site around the metal cation and the phosphoryl oxygen atoms, two key residues undergo allosteric rearrangement in response to phosphorylation (reviewed in [131]). In the few known active rec domain structures, a conserved Thr/Ser on the β4-strand repositions to hydrogen bond with a phosphoryl oxygen. This coincides with a shift in the β4α4 loop region. This change also typically results in the rotamerization of a conserved aromatic residue (Phe/Tyr/His) on the β5-strand towards an inward orientation, though this does not always occur. These two positions are known as "switch residues" and are traditional indicators for the conformational state of rec domains [131]. Ultimately, phosphorylation results in allosteric changes that propagate throughout the entire rec domain, leading to a diverse range of outputs, such as changes in dimerization [132] and partner binding [133].

1.9 Computational techniques provide complementary data to experimental methods

Protein structure and dynamics

Many experimental approaches exist for characterizing macromolecular structures, such as nuclear magnetic resonance (NMR), X-ray crystallography, small-angle X-ray scattering (SAXS) and electron microscopy (EM). Crystallography is often considered the gold standard of structural biology due to the high-resolution details it provides about protein structure. However, in

addition to the many challenges involved in X-ray crystallography, one of its biggest disadvantages is the static nature of the structures it reveals. An individual crystal structure essentially represents a conformationally averaged snapshot of a macromolecule. This can be misleading, as proteins are intrinsically dynamic and undergo constant and enormous conformational flux [134-138]. Large portions of the conformational space of a protein are hidden from experimental methods due to the low probability of observing high-energy conformers. Computational methods provide a link between protein structure and dynamics [139, 140]. With these methods, transient or high-energy sub-states of proteins can be observed, as well as events such as conformational transitions, allostery, ligand binding and protein folding (reviewed in [135, 141, 142]).

Molecular dynamics

Molecular dynamics (MD) is the science of simulating and observing the time-dependent behavior of a system of particles [143]. An MD simulation requires a three-dimensional structure or model as a starting point. An empirical force field is used to define the total potential energy surface for the particles within the structure, as a sum of potentials derived from simple physical forces such as bond-stretching and van der Waals contributions [139]. Newton's laws of motion can be used to determine the time-dependent evolution of the system [139]. Integration of Newton's equation of motion yields a trajectory that describes atomic positions, velocities and acceleration

as a function of time. This information can be used to predict the state of the system at any moment. With the virtual explosion in computing power and availability in the last decade, computational techniques like MD have become commonplace for large systems such as proteins, which can often contain hundreds of thousands of atoms. However, limitations still exist on the timescales that are accessible to these methods. Classical MD simulations must typically be sampled on similar timescales as the events they are studying, though certain biasing methods can be used to reduce this. Conventional techniques for simulating large, all-atom protein models are still impractical on the high microsecond to low millisecond timescales for most research laboratories. Ultimately, computational approaches like MD provide a powerful bridge between protein structure and protein dynamics, but they should be used in tandem with more traditional experiments. Data may be used to guide future laboratory studies and should always be supported by experimental evidence.

1.10 Sln1 pathway of *S. cerevisiae* as a model system

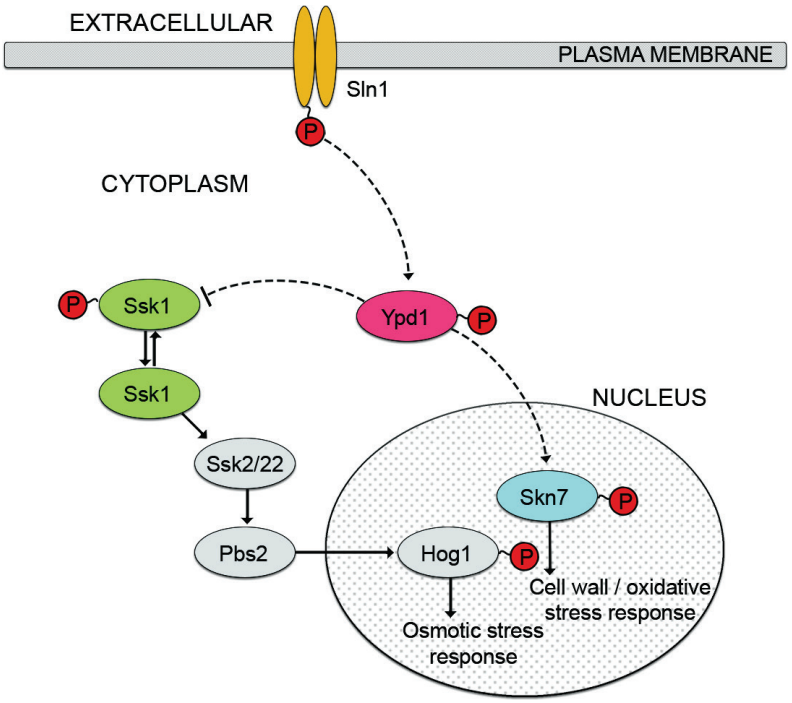

Figure 6. Diagram of the Sln1 phosphorelay from *S. cerevisiae*.

Expanded fungal phosphorelay system. Under non-stress conditions, Sln1 is constitutively autophosphorylated. It then transfers phosphoryl groups to Ypd1, which in turn relays phosphoryl groups to Ssk1. Phospho-Ssk1 is inhibited from interacting with Ssk2. Under hyperosmotic stress, the phosphorelay is attenuated, and unphosphorylated Ssk1 accumulates, activating the Ssk2-Pbs2-Hog1 system and triggering intracellular glycerol production to restore osmotic balance. Under cell wall stress, Ypd1 phosphorylates the transcription factor Skn7 in the nucleus, which regulates the expression of related genes in response.

S. cerevisiae has long been used as a model organism for studying two-component signaling, due to its genetic tractability and use of only a single TCS system (reviewed in [86]). The Sln1 pathway is an extended, branching phosphorelay with two roles. One branch detects and responds to osmotic stress [87-89]; the other branch responds to changes in cell wall integrity and oxidative stress [91]. **Fig. 6** shows a diagram of the pathway. Many pathogenic

fungi, such as *Cryptococcus neoformans* and *Candida albicans,* share homologous systems. Due to the absence of TCS systems in animals, these proteins have received much attention in recent years as potential therapeutic targets [144-147].

At the beginning of the pathway is a dimeric, trans-membrane hybrid HK known as Sln1 ([87], reviewed in [86]). Under normal environmental conditions, Sln1 constitutively binds ATP and autophosphorylates on a histidine residue. It then passes the phosphoryl to an aspartate on its attached rec domain [79]. The phosphoryl is then transferred to an intermediate HPt, Ypd1 [87]. In the absence of stress, Ypd1 passes the phosphoryl on to the downstream RR, Ssk1. While phosphorylated, Ssk1 is unable to activate the HOG pathway. Under hyperosmotic conditions, Sln1 autokinase activity is diminished, leading to an accumulation of unphosphorylated Ssk1, which then binds Ssk2 or Ssk22 [90, 148]. These MAP3Ks go on to activate the High Osmolarity Glycerol (HOG) response pathway, leading to increased intracellular glycerol production that restores osmotic balance [149, 150]. Upon cell wall perturbation, Ypd1 instead transfers the phosphoryl group to the nuclear RR/transcription factor, Skn7, altering expression of cell wall-related genes [151].

The mechanism by which Ypd1 is able to differentiate between three individual rec domains within its pathway is poorly understood. Each rec domain within the Sln1 pathway shares high structural homology and a

common hydrophobic binding site on the surface of Ypd1, while having low sequence homology and diverse molecular functions.. These facts indicate that the mechanism likely relies upon molecular specificity, or unique residues used for the individual interactions within the pathway that are outside the conserved common binding surface.

1.11 Research Focus

The chemical labilities of phosphohistidine and phosphoaspartate make experimental studies of modified phospho-proteins challenging. Of the hundreds of thousands of rec domain sequences available [152], a mere 150 non-redundant proteins have been structurally characterized with experimental methods [153]. Of these, only 20% have been obtained in a modified (i.e. phosphorylated) state, usually with the aid of phosphoryl analogs [153, 154]. This lack of structural information leaves our understanding of two-component signaling systems incomplete. In addition, the dynamic properties of the extremely diverse rec domain superfamily are almost completely unknown. By incorporating experimental data (structural, chemical, etc.) with MD, we can address these gaps in knowledge. This work is divided into three parts. Chapters 2 and 3 were written as manuscripts for peer review, though they have been formatted to better fit this dissertation. Chapter 4 is part of a collaborative effort with other members of the West laboratory.

For the first study, a method for simulating the “activation” of phosphorylated, monomeric rec domains is described. Kinetic data suggests

that phosphotransfer and associated conformational changes within rec domains occur on millisecond timescales [155-157]. Studying these timescales is impractical for standard research laboratories using classical MD simulations. To address this, conserved active site restraints are used to help drive the structure towards an active conformation, drastically reducing the required simulation time [158]. The accuracy of this method was first validated on five well-characterized rec domains. It was then applied to the rec domains found within the Sln1 pathway of *S. cerevisiae.* With this computational methodology, we can predict highly accurate phosphorylated models that closely resemble active crystal structures of rec domains. These models reveal unique structural features and physicochemical properties of their diverse RR proteins.

In the second study, the aforementioned computational approach is applied to heterodimeric signaling complexes within the fungal Sln1 pathway. This is done to explain differences in the interactions between the rec domains and Ypd1. In addition, the apo (unphosphorylated) proteins are simulated to detect transient structural features that may be hidden in their purely static crystal structures.

The third study contains an experimental and computational characterization of the Ypd1 mutant, G68Q. This residue, Gly68, is almost completely conserved in HPt domains. The mutation was previously determined to disrupt phosphotransfer from Sln1 to Ypd1 and from Ypd1 to

Ssk1 by a significant amount [113]. Along with its close proximity to the site of phosphorylation (His64) on Ypd1, these results suggest that it plays a key role in protein-protein binding and/or the catalysis of phosphotransfer with Ypd1. In an effort to explain this observation, crystallographic studies were carried out to co-crystallize the complex between Ypd1 G68Q and the downstream rec domain of Ssk1. *In silico* mutagenesis studies were also performed on the existing complex structure between Ypd1 G68Q and the upstream rec domain of Sln1.

References

1. JM Berg JT, L Stryer. Biochemistry. 5th ed. New York: W H Freeman; 2002.

2. Morris JC. The Touchstone of Life: Molecular Information, Cell Communication, and the Foundations of Life. Werner R. Loewenstein. The Quarterly Review of Biology. 2000;75(3):308-9.

3. Scott JD, Pawson T. Cell communication: the inside story. Scientific American. 2000;282(6):72-9.

4. Kiel C, Yus E, Serrano L. Engineering Signal Transduction Pathways. Cell. 2010;140(1):33-47.

5. Dumont JE, Dremier S, Pirson I, Maenhaut C. Cross signaling, cell specificity, and physiology. American journal of physiology Cell physiology. 2002;283(1):C2-28.

6. Signal transduction from the extracellular matrix. The Journal of cell biology. 1993;120(3):577-85.

7. Cooper G. The Cell: A Molecular Approach. 2nd ed. Sunderland (MA): Sinauer Associates; 2000.

8. Kaufman L, Rousseeuw P. Clustering by means of medoids: Faculty of Mathematics and Informatics; 1987.

9. Smith CM, Radzio-Andzelm E, Madhusudan, Akamine P, Taylor SS. The catalytic subunit of cAMP-dependent protein kinase: prototype for an extended network of communication. Progress in biophysics and molecular biology. 1999;71(3-4):313-41.

10. Dawson AP. Calcium signalling: how do IP3 receptors work? Current biology : CB. 1997;7(9):R544-7.

11. Lincoln TM, Dey N, Sellak H. Invited review: cGMP-dependent protein kinase signaling mechanisms in smooth muscle: from the regulation of tone to gene expression. Journal of applied physiology (Bethesda, Md : 1985). 2001;91(3):1421-30.

12. Walsh CT, Garneau-Tsodikova S, Gatto GJ. Protein Posttranslational Modifications: The Chemistry of Proteome Diversifications. Angewandte Chemie International Edition. 2005;44(45):7342-72.

13. Tarrant MK, Cole PA. The Chemical Biology of Protein Phosphorylation. Annual Rev Biochemistry. 2009;78(1):797-825.

14. Cohen P. The origins of protein phosphorylation. Nature cell biology. 2002;4(5):E127-30.

15. Humphrey SJ, James DE, Mann M. Protein Phosphorylation: A Major Switch Mechanism for Metabolic Regulation. Trends in Endocrinology & Metabolism. 26(12):676-87.

16. Fischer EH, Krebs EG. Conversion of phosphorylase b to phosphorylase a in muscle extracts. J Biol Chem. 1955;216(1):121-32.

17. Krebs EG, Fischer EH. Phosphorylase activity of skeletal muscle extracts. J Biol Chem. 1955;216(1):113-20.

18. Berthet J, Rall TW, Sutherland EW. The relationship of epinephrine and glucagon to liver phosphorylase. IV. Effect of epinephrine and glucagon on the reactivation of phosphorylase in liver homogenates. J Biol Chem. 1957;224(1):463-75.

19. Cheng H-C, Qi RZ, Paudel H, Zhu H-J. Regulation and Function of Protein Kinases and Phosphatases. Enzyme Research. 2011;2011:3.

20. Adams JA. Kinetic and catalytic mechanisms of protein kinases. Chemical reviews. 2001;101(8):2271-90.

21. Barford D, Das AK, Egloff MP. The structure and mechanism of protein phosphatases: insights into catalysis and regulation. Annual review of biophysics and biomolecular structure. 1998;27:133-64.

22. Kennelly PJ. Protein kinases and protein phosphatases in prokaryotes: a genomic perspective. FEMS Microbiology Letters. 2002;206(1):1-8.

23. Ubersax JA, Ferrell JE, Jr. Mechanisms of specificity in protein phosphorylation. Nature reviews Molecular cell biology. 2007;8(7):530-41.

24. Hanks SK, Quinn AM. Protein kinase catalytic domain sequence database: identification of conserved features of primary structure and classification of family members. Methods Enzymol. 1991;200:38-62.

25. Hanks SK, Hunter T. Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. FASEB J. 1995;9(8):576-96.

26. Klumpp S, Krieglstein J. Phosphorylation and dephosphorylation of histidine residues in proteins. European journal of biochemistry / FEBS. 2002;269(4):1067-71.

27. Attwood VP, Piggott JM, Zu LX, Besant GP. Focus on phosphohistidine. Amino Acids. 2007;32(1):145-56.

28. Attwood PV, Besant PG, Piggott MJ. Focus on phosphoaspartate and phosphoglutamate. Amino Acids. 2011;40(4):1035-51.

29. West AH, Stock AM. Histidine kinases and response regulator proteins in two-component signaling systems. Trends Biochem Sci. 2001;26(6):369-76.

30. Ann M. Stock, Victoria L. Robinson a, Goudreau PN. Two-component signal transduction. Annual Rev Biochemistry. 2000;69(1):183-215.

31. Cashin P, Goldsack L, Hall D, O'Toole R. Contrasting signal transduction mechanisms in bacterial and eukaryotic gene transcription. FEMS Microbiology Letters. 2006;261(2):155-64.

32. J B Stock, G S Lukat a, Stock AM. Bacterial Chemotaxis and the Molecular Logic of Intracellular Signal Transduction Networks. Annual Review of Biophysics and Biophysical Chemistry. 1991;20(1):109-36.

33. Barford D. Molecular mechanisms for the control of enzymic activity by protein phosphorylation. Biochimica et biophysica acta. 1991;1133(1):55-62.

34. Knowles JR. Enzyme-catalyzed phosphoryl transfer reactions. Annu Rev Biochem. 1980;49:877-919.

35. Laskowski RA, Gerick F, Thornton JM. The structural basis of allosteric regulation in proteins. FEBS Letters. 2009;583(11):1692-8.

36. Chebaro Y, Amal I, Rochel N, Rochette-Egly C, Stote RH, Dejaegere A. Phosphorylation of the Retinoic Acid Receptor Alpha Induces a Mechanical Allosteric Regulation and Changes in Internal Dynamics. PLoS Comput Biol. 2013;9(4):e1003012.

37. Edreira MM, Li S, Hochbaum D, Wong S, Gorfe AA, Ribeiro-Neto F, Woods VL, Jr., Altschuler DL. Phosphorylation-induced conformational changes in Rap1b: allosteric effects on switch domains and effector loop. J Biol Chem. 2009;284(40):27480-6.

38. Hofer HW, Schlatter S, Graefe M. Phosphorylation of phosphofructokinase by protein kinase C changes the allosteric properties of the enzyme. Biochem Biophys Res Commun. 1985;129(3):892-7.

39. Wang Z, Cole PA. Catalytic Mechanisms and Regulation of Protein Kinases. Methods Enzymol. 2014;548:1-21.

40. Lim W. Cell signaling : principles and mechanisms. In: Mayer B, Pawson T, editors.

41. Ciesla J, Fraczyk T, Rode W. Phosphorylation of basic amino acid residues in proteins: important but easily missed. Acta biochimica Polonica. 2011;58(2):137-48.

42. Stock JB, Stock AM, Mottonen JM. Signal transduction in bacteria. Nature. 1990;344(6265):395-400.

43. Pickart CM, Jencks WP. Energetics of the calcium-transporting ATPase. J Biol Chem. 1984;259(3):1629-43.

44. Koshland DE. Effect of Catalysts on the Hydrolysis of Acetyl Phosphate. Nucleophilic Displacement Mechanisms in Enzymatic Reactions1. J Am Chem Soc. 1952;74(9):2286-92.

45. Di Sabato G, Jencks WP. Mechanism and Catalysis of Reactions of Acyl Phosphates. I. Nucleophilic Reactions1. J Am Chem Soc. 1961;83(21):4393-400.

46. Skerker JM, Prasol MS, Perchuk BS, Biondi EG, Laub MT. Two-Component Signal Transduction Pathways Regulating Growth and Cell Cycle Progression in a Bacterium: A System-Level Analysis. PLoS Biol. 2005;3(10):e334.

47. Miller SI, Kukral AM, Mekalanos JJ. A two-component regulatory system (phoP phoQ) controls Salmonella typhimurium virulence. Proc Natl Acad Sci. 1989;86(13):5054-8.

48. Gooderham WJ, Hancock RE. Regulation of virulence and antibiotic resistance by two-component regulatory systems in Pseudomonas aeruginosa. FEMS Microbiol Rev. 2009;33(2):279-94.

49. Laub MT. The Role of Two-Component Signal Transduction Systems in Bacterial Stress Responses. Bacterial Stress Responses, Second Edition: American Society of Microbiology; 2011.

50. Capra EJ, Laub MT. Evolution of two-component signal transduction systems. Annual review of microbiology. 2012;66:325-47.

51. Galperin MY. A census of membrane-bound and intracellular signal transduction proteins in bacteria: Bacterial IQ, extroverts and introverts. BMC Microbiology. 2005;5(1):1-19.

52. Alm E, Huang K, Arkin A. The evolution of two-component systems in bacteria reveals different strategies for niche adaptation. PLoS Comput Biol. 2006;2(11):e143.

53. Galperin MY, Nikolskaya AN, Koonin EV. Novel domains of the prokaryotic two-component signal transduction systems. FEMS Microbiol Lett. 2001;203.

54. Koretke KK, Lupas AN, Warren PV, Rosenberg M, Brown JR. Evolution of two-component signal transduction. Mol Biol Evol. 2000;17.

55. Rowland MA, Deeds EJ. Crosstalk and the evolution of specificity in two-component signaling. Proc Natl Acad Sci. 2014;111(15):5550-5.

56. Laub MT, Goulian M. Specificity in two-component signal transduction pathways. Annual Rev Genetics. 2007;41(1):121-45.

57. Skerker JM, Perchuk BS, Siryaporn A, Lubin EA, Ashenberg O, Goulian M, Laub MT. Rewiring the Specificity of Two-Component Signal Transduction Systems. Cell. 2008;133(6):1043-54.

58. Yamamoto K, Hirao K, Oshima T, Aiba H, Utsumi R, Ishihama A. Functional characterization in vitro of all two-component signal transduction systems from Escherichia coli. J Biol Chem. 2005;280(2):1448-56.

59. Qin L, Cai S, Zhu Y, Inouye M. Cysteine-Scanning Analysis of the Dimerization Domain of EnvZ, an Osmosensing Histidine Kinase. J Bact. 2003;185(11):3429-35.

60. Aiba H, Mizuno T. Phosphorylation of a bacterial activator protein, OmpR, by a protein kinase, EnvZ, stimulates the transcription of the ompF and ompC genes in Escherichia coli. FEBS Letters. 1990;261(1):19-22.

61. Jeon J, Kim H, Yun J, Ryu S, Groisman EA, Shin D. RstA-Promoted Expression of the Ferrous Iron Transporter FeoB under Iron-Replete Conditions Enhances Fur Activity in Salmonella enterica. J Bact. 2008;190(22):7326-34.

62. Cabeza ML, Aguirre A, Soncini FC, Vescovi EG. Induction of RpoS degradation by the two-component system regulator RstA in Salmonella enterica. J Bact. 2007;189(20):7335-42.

63. Fisher SL, Jiang W, Wanner BL, Walsh CT. Cross-talk between the histidine protein kinase VanS and the response regulator PhoB. Characterization and identification of a VanS domain that inhibits activation of PhoB. J Biol Chem. 1995;270(39):23143-9.

64. Silva JC, Haldimann A, Prahalad MK, Walsh CT, Wanner BL. *In vivo* characterization of the type A and B vancomycin-resistant *enterococci* (VRE) VanRS two-component systems in *Escherichia coli*: A nonpathogenic model for studying the VRE signal transduction pathways. Proc Natl Acad Sci. 1998;95(20):11951-6.

65. Haldimann A, Fisher SL, Daniels LL, Walsh CT, Wanner BL. Transcriptional regulation of the *Enterococcus faecium* BM4147 vancomycin resistance gene cluster by the VanS-VanR two-component regulatory system in *Escherichia coli* K-12. J Bact. 1997;179(18):5903-13.

66. Schwartz MA, Madhani HD. Principles of MAP kinase signaling specificity in Saccharomyces cerevisiae. Annu Rev Genet. 2004;38:725-48.

67. Wuichet K, Cantwell BJ, Zhulin IB. Evolution and phyletic distribution of two-component signal transduction systems. Current Opinion in Microbiology. 2010;13(2):219-25.

68. Goulian M. Two-Component Signaling Circuit Structure and Properties. Current opinion in microbiology. 2010;13(2):184-9.

69. Zhang W, Shi L. Distribution and evolution of multiple-step phosphorelay in prokaryotes: lateral domain recruitment involved in the formation of hybrid-type histidine kinases. Microbiology. 2005;151(Pt 7):2159-73.

70. Appleby JL, Parkinson JS, Bourret RB. Signal Transduction via the Multi-Step Phosphorelay: Not Necessarily a Road Less Traveled. Cell. 1996;86(6):845-8.

71. Chen MH, Takeda S-i, Yamada H, Ishii Y, Yamashino T, Mizuno T. Characterization of the RcsC→YojN→RcsB Phosphorelay Signaling Pathway Involved in Capsular Synthesis in Escherichia coli. Bioscience, Biotechnology, and Biochemistry. 2001;65(10):2364-7.

72. Takeda S, Fujisawa Y, Matsubara M, Aiba H, Mizuno T. A novel feature of the multistep phosphorelay in Escherichia coli: a revised model of the RcsC --> YojN --> RcsB signalling pathway implicated in capsular synthesis and swarming behaviour. Mol Microbiol. 2001;40(2):440-50.

73. Ferrieres L, Clarke DJ. The RcsC sensor kinase is required for normal biofilm formation in Escherichia coli K-12 and controls the expression of a regulon in response to growth on a solid surface. Mol Microbiol. 2003;50(5):1665-82.

74. Piggot PJ, Hilbert DW. Sporulation of Bacillus subtilis. Curr Opin Microbiol. 2004;7(6):579-86.

75. Burbulys D, Trach KA, Hoch JA. Initiation of sporulation in B. subtilis is controlled by a multicomponent phosphorelay. Cell. 1991;64(3):545-52.

76. Jiang M, Shao W, Perego M, Hoch JA. Multiple histidine kinases regulate entry into stationary phase and sporulation in Bacillus subtilis. Mol Microbiol. 2000;38(3):535-42.

77. Uhl MA, Miller JF. Integration of multiple domains in a two-component sensor protein: the Bordetella pertussis BvgAS phosphorelay. EMBO J. 1996;15(5):1028-36.

78. Beier D, Gross R. The BvgS/BvgA phosphorelay system of pathogenic Bordetellae: structure, function and evolution. Advances in experimental medicine and biology. 2008;631:149-60.

79. Ota IM, Varshavsky A. A yeast protein similar to bacterial two-component regulators. Science. 1993;262(5133):566-9.

80. Chang C, Kwok SF, Bleecker AB, Meyerowitz EM. Arabidopsis ethylene-response gene ETR1: similarity of product to two-component regulators. Science. 1993;262(5133):539-44.

81. Munoz-Dorado J, Inouye S, Inouye M. Eukaryotic-like protein serine/threonine kinases in Myxococcus xanthus, a developmental bacterium exhibiting social behavior. Journal of cellular biochemistry. 1993;51(1):29-33.

82. Schaller GE, Shiu S-H, Armitage Judith P. Two-Component Systems and Their Co-Option for Eukaryotic Signal Transduction. Current Biology. 2011;21(9):R320-R30.

83. Srikantha T, Tsai L, Daniels K, Enger L, Highley K, Soll DR. The two-component hybrid kinase regulator CaNIK1 of Candida albicans. Microbiology. 1998;144 (Pt 10):2715-29.

84. Fassler JS, West AH. Histidine Phosphotransfer Proteins in Fungal Two-Component Signal Transduction Pathways. Eukaryotic Cell. 2013;12(8):1052-60.

85. Santos JL, Shiozaki K. Fungal histidine kinases. Science's STKE : signal transduction knowledge environment. 2001;2001(98):re1.

86. Fassler JS, West AH. Genetic and Biochemical Analysis of the SLN1 Pathway in Saccharomyces cerevisiae. Methods Enzymol. 2010;471:291-317.

87. Posas F, Wurgler-Murphy SM, Maeda T, Witten EA, Thai TC, Saito H. Yeast HOG1 MAP kinase cascade is regulated by a multistep phosphorelay mechanism in the SLN1-YPD1-SSK1 "two-component" osmosensor. Cell. 1996;86(6):865-75.

88. Maeda T, Wurgler-Murphy SM, Saito H. A two-component system that regulates an osmosensing MAP kinase cascade in yeast. Nature. 1994;369(6477):242-5.

89. Reiser V, Raitt DC, Saito H. Yeast osmosensor Sln1 and plant cytokinin receptor Cre1 respond to changes in turgor pressure. The Journal of cell biology. 2003;161(6):1035-40.

90. Posas F, Saito H. Osmotic activation of the HOG MAPK pathway via Ste11p MAPKKK: scaffold role of Pbs2p MAPKK. Science. 1997;276(5319):1702-5.

91. Shankarnarayan S, Malone CL, Deschenes RJ, Fassler JS. Modulation of yeast Sln1 kinase activity by the CCW12 cell wall protein. J Biol Chem. 2008;283(4):1962-73.

92. Lu JM, Deschenes RJ, Fassler JS. Saccharomyces cerevisiae histidine phosphotransferase Ypd1p shuttles between the nucleus and cytoplasm for SLN1-dependent phosphorylation of Ssk1p and Skn7p. Eukaryot Cell. 2003;2(6):1304-14.

93. Bijlsma JJE, Groisman EA. Making informed decisions: regulatory interactions between two-component systems. Trends in Microbiology. 2003;11(8):359-66.

94. Kothamachu VB, Feliu E, Wiuf C, Cardelli L, Soyer OS. Phosphorelays Provide Tunable Signal Processing Capabilities for the Cell. PLoS Comput Biol. 2013;9(11):e1003322.

95. Csikász-Nagy A, Cardelli L, Soyer OS. Response dynamics of phosphorelays suggest their potential utility in cell signalling. Journal of The Royal Society Interface. 2010.

96. Maslennikov I, Klammt C, Hwang E, Kefala G, Okamura M, Esquivies L, Mors K, Glaubitz C, Kwiatkowski W, Jeon YH, Choe S. Membrane domain structures of three classes of histidine kinase receptors by cell-free expression and rapid NMR analysis. Proc Natl Acad Sci. 2010;107(24):10902-7.

97. Bhate MP, Molnar KS, Goulian M, DeGrado WF. Signal transduction in histidine kinases: insights from new structures. Structure. 2015;23(6):981-94.

98. Kim D, Forst S. Genomic analysis of the histidine kinase family in bacteria and archaea. Microbiology. 2001;147(Pt 5):1197-212.

99. Mascher T, Helmann JD, Unden G. Stimulus Perception in Bacterial Signal-Transducing Histidine Kinases. Microbiology and Molecular Biology Reviews. 2006;70(4):910-38.

100. Stewart V, Bledsoe PJ. Synthetic lac operator substitutions for studying the nitrate- and nitrite-responsive NarX-NarL and NarQ-NarP two-component regulatory systems of Escherichia coli K-12. J Bact. 2003;185(7):2104-11.

101. Schroder I, Wolin CD, Cavicchioli R, Gunsalus RP. Phosphorylation and dephosphorylation of the NarQ, NarX, and NarL proteins of the nitrate-dependent two-component regulatory system of Escherichia coli. J Bact. 1994;176(16):4985-92.

102. Garcia Vescovi E, Soncini FC, Groisman EA. Mg2+ as an extracellular signal: environmental regulation of Salmonella virulence. Cell. 1996;84(1):165-74.

103. Robinson VL, Stock AM. High energy exchange: proteins that make or break phosphoramidate bonds. Structure. 1999;7(3):R47-53.

104. Casino P, Miguel-Romero L, Marina A. Visualizing autophosphorylation in histidine kinases. Nature communications. 2014;5:3258.

105. Casino P, Rubio V, Marina A. Structural Insight into Partner Specificity and Phosphoryl Transfer in Two-Component Signal Transduction. Cell. 2009;139(2):325-36.

106. Ashenberg O, Keating AE, Laub MT. Helix bundle loops determine whether histidine kinases autophosphorylate in cis or in trans. J Mol Biol. 2013;425(7):1198-209.

107. Alex LA, Simon MI. Protein histidine kinases and signal transduction in prokaryotes and eukaryotes. Trends Genet. 1994;10(4):133-8.

108. Tsuzuki M, Ishige K, Mizuno T. Phosphotransfer circuitry of the putative multi-signal transducer, ArcB, of Escherichia coli: in vitro studies with mutants. Mol Microbiol. 1995;18(5):953-62.

109. Mizuno T. Compilation of all genes encoding two-component phosphotransfer signal transducers in the genome of Escherichia coli. DNA Res. 1997;4.

110. Perraud AL, Kimmel B, Weiss V, Gross R. Specificity of the BvgAS and EvgAS phosphorelay is mediated by the C-terminal HPt domains of the sensor proteins. Mol Microbiol. 1998;27(5):875-87.

111. Xu Q, Porter SW, West AH. The yeast YPD1/SLN1 complex: Insights into molecular recognition in two-component signaling systems. Structure. 2003;11(12):1569-81.

112. Xu Q, West AH. Conservation of structure and function among histidine-containing phosphotransfer (HPt) domains as revealed by the crystal structure of YPD1. J Mol Biol. 1999;292(5):1039-50.

113. Janiak-Spens F, West AH. Functional roles of conserved amino acid residues surrounding the phosphorylatable histidine of the yeast phosphorelay protein YPD1. Mol Micro. 2000;37(1):136-44.

114. Kato M, Mizuno T, Shimizu T, Hakoshima T. Refined structure of the histidine-containing phosphotransfer (HPt) domain of the anaerobic sensor kinase ArcB from Escherichia coli at 1.57 A resolution. Acta

crystallographica Section D, Biological crystallography. 1999;55(Pt 11):1842-9.

115. Bauer J, Reiss K, Veerabagu M, Heunemann M, Harter K, Stehle T. Structure-function analysis of Arabidopsis thaliana histidine kinase AHK5 bound to its cognate phosphotransfer protein AHP1. Molecular plant. 2013;6(3):959-70.

116. Porter SW, West AH. A common docking site for response regulators on the yeast phosphorelay protein YPD1. Biochimica et biophysica acta. 2005;1748(2):138-45.

117. Porter SW, Xu Q, West AH. Ssk1p response regulator binding surface on histidine-containing phosphotransfer protein Ypd1p. Eukaryot Cell. 2003;2(1):27-33.

118. Bourret RB. Receiver domain structure and function in response regulator proteins. Current Opinion in Microbiology. 2010;13(2):142-9.

119. Volz K, Matsumura P. Crystal structure of *Escherichia coli* CheY refined at 1.7-A resolution. J Biol Chem. 1991;266(23):15511-9.

120. Galperin MY. Diversity of structure and function of response regulator output domains. Curr Opin Microbiol. 2010;13(2):150-9.

121. Hess JF, Oosawa K, Kaplan N, Simon MI. Phosphorylation of three proteins in the signaling pathway of bacterial chemotaxis. Cell. 1988;53(1):79-87.

122. Lehti TA, Heikkinen J, Korhonen TK, Westerlund-Wikström B. The Response Regulator RcsB Activates Expression of Mat Fimbriae in Meningitic Escherichia coli. J Bact. 2012;194(13):3475-85.

123. Overgaard M, Wegener-Feldbrügge S, Søgaard-Andersen L. The Orphan Response Regulator DigR Is Required for Synthesis of Extracellular Matrix Fibrils in Myxococcus xanthus. J Bact. 2006;188(12):4384-94.

124. Kojetin DJ, Thompson RJ, Benson LM, Naylor S, Waterman J, Davies KG, Opperman CH, Stephenson K, Hoch JA, Cavanagh J. Structural analysis of divalent metals binding to the Bacillus subtilis response regulator Spo0F: the possibility for in vitro metalloregulation in the initiation of sporulation. Biometals. 2005;18(5):449-66.

125. Bellsolell L, Prieto J, Serrano L, Coll M. Magnesium binding to the bacterial chemotaxis protein CheY results in large conformational changes involving its functional surface. J Mol Biol. 1994;238(4):489-95.

126. Lukat GS, McCleary WR, Stock AM, Stock JB. Phosphorylation of bacterial response regulator proteins by low molecular weight phospho-donors. Proc Natl Acad Sci. 1992;89(2):718-22.

127. Wang W, Cho HS, Kim R, Jancarik J, Yokota H, Nguyen HH, Grigoriev IV, Wemmer DE, Kim SH. Structural characterization of the reaction pathway in phosphoserine phosphatase: crystallographic "snapshots" of intermediate states. J Mol Biol. 2002;319(2):421-31.

128. Lukat GS, Stock AM, Stock JB. Divalent metal ion binding to the CheY protein and its significance to phosphotransfer in bacterial chemotaxis. Biochemistry. 1990;29(23):5436-42.

129. Varughese KI, Tsigelny I, Zhao H. The crystal structure of beryllofluoride Spo0F in complex with the phosphotransferase Spo0B represents a phosphotransfer pretransition state. J Bact. 2006;188(13):4970-7.

130. Mildvan AS. Mechanisms of signaling and related enzymes. Proteins. 1997;29(4):401-16.

131. Gao R, Stock AM. Molecular strategies for phosphorylation-mediated regulation of response regulator activity. Curr Opin Microbiol. 2010;13(2):160-7.

132. Da Re S, Schumacher J, Rousseau P, Fourment J, Ebel C, Kahn D. Phosphorylation-induced dimerization of the FixJ receiver domain. Mol Micro. 1999;34(3):504-11.

133. Lee SY, Cho HS, Pelton JG, Yan D, Henderson RK, King DS, Huang L, Kustu S, Berry EA, Wemmer DE. Crystal structure of an activated response regulator bound to its target. Nat Struct Biol. 2001;8(1):52-6.

134. Young RD, Fenimore PW. Coupling of protein and environment fluctuations. Biochimica et biophysica acta. 2011;1814(8):916-21.

135. Teilum K, Olsen JG, Kragelund BB. Protein stability, flexibility and function. Biochimica et biophysica acta. 2011;1814(8):969-76.

136. Frauenfelder H, Chen G, Berendzen J, Fenimore PW, Jansson H, McMahon BH, Stroe IR, Swenson J, Young RD. A unified model of protein dynamics. Proc Natl Acad Sci. 2009;106(13):5129-34.

137. Wang Y, Fisher JC, Mathew R, Ou L, Otieno S, Sublet J, Xiao L, Chen J, Roussel MF, Kriwacki RW. Intrinsic disorder mediates the diverse regulatory functions of the Cdk inhibitor p21. Nature chemical biology. 2011;7(4):214-21.

138. Bernado P, Blackledge M. Structural biology: Proteins in dynamic equilibrium. Nature. 2010;468(7327):1046-8.

139. Karplus M, Kuriyan J. Molecular dynamics and protein function. Proc Natl Acad Sci USA. 2005;102.

140. Adcock SA, McCammon JA. Molecular dynamics: survey of methods for simulating the activity of proteins. Chemical reviews. 2006;106(5):1589-615.

141. Peracchi A, Mozzarelli A. Exploring and exploiting allostery: Models, evolution, and drug targeting. Biochimica et biophysica acta. 2011;1814(8):922-33.

142. Buchner GS, Murphy RD, Buchete NV, Kubelka J. Dynamics of protein folding: probing the kinetic network of folding-unfolding transitions with experiment and theory. Biochimica et biophysica acta. 2011;1814(8):1001-20.

143. Karplus M, Petsko GA. Molecular dynamics simulations in biology. Nature. 1990;347(6294):631-9.

144. Shor E, Chauhan N. A Case for Two-Component Signaling Systems As Antifungal Drug Targets. PLoS Pathog. 2015;11(2):e1004632.

145. Barrett JF, Hoch JA. Two-Component Signal Transduction as a Target for Microbial Anti-Infective Therapy. Antimicrobial Agents and Chemotherapy. 1998;42(7):1529-36.

146. Watanabe T, Okada A, Gotoh Y, Utsumi R. Inhibitors Targeting Two-Component Signal Transduction. In: Utsumi R, editor. Bacterial Signal Transduction: Networks and Drug Targets. New York, NY: Springer New York; 2008. p. 229-36.

147. Worthington RJ, Blackledge MS, Melander C. Small-molecule inhibition of bacterial two-component systems to combat antibiotic resistance and virulence. Future Medicinal Chemistry. 2013;5(11):1265-84.

148. Maeda T, Takekawa M, Saito H. Activation of yeast PBS2 MAPKK by MAPKKKs or by binding of an SH3-containing osmosensor. Science. 1995;269(5223):554-8.

149. Albertyn J, Hohmann S, Thevelein JM, Prior BA. GPD1, which encodes glycerol-3-phosphate dehydrogenase, is essential for growth under osmotic stress in Saccharomyces cerevisiae, and its expression is regulated by the high-osmolarity glycerol response pathway. Molecular and cellular biology. 1994;14(6):4135-44.

150. Albertyn J, Hohmann S, Prior BA. Characterization of the osmotic-stress response in Saccharomyces cerevisiae: osmotic stress and glucose repression regulate glycerol-3-phosphate dehydrogenase independently. Current genetics. 1994;25(1):12-8.

151. Brown JL, Bussey H, Stewart RC. Yeast Skn7p functions in a eukaryotic two-component regulatory pathway. EMBO J. 1994;13(21):5186-94.

152. Tatusova T, Ciufo S, Fedorov B, O'Neill K, Tolstoy I. RefSeq microbial genomes database: New representation and annotation strategy. Nucleic Acids Res. 2014;42(Database issue):D553-9.

153. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids. 2000;28(1):235-42.

154. Cho H, Wang W, Kim R, Yokota H, Damo S, Kim SH, Wemmer D, Kustu S, Yan D. BeF_3^- acts as a phosphate analog in proteins phosphorylated on aspartate: structure of a BeF_3^- complex with phosphoserine phosphatase. Proc Natl Acad Sci. 2001;98(15):8525-30.

155. Bourret RB, Thomas SA, Page SC, Creager-Allen RL, Moore AM, Silversmith RE. Measurement of response regulator autodephosphorylation rates spanning six orders of magnitude. Methods Enzymol. 2010;471:89-114.

156. Feher VA, Cavanagh J. Millisecond-timescale motions contribute to the function of the bacterial response regulator protein Spo0F. Nature. 1999;400(6741):289-93.

157. Bobay BG, Hoch JA, Cavanagh J. Dynamics and activation in response regulators: The β4-α4 loop. Biomol Concepts. 2012;3(2):175-82.

158. Fiorin G, Klein ML, Hénin J. Using collective variables to drive molecular dynamics simulations. Mol Physics. 2013;111(22-23):3345-62.

Chapter 2: Simulating phosphorylated rec domains

2.1 Introduction

Two-component signaling (TCS) systems allow bacteria to respond to external environments using the reversible phosphorylation of conserved histidine and aspartate residues [1, 2]. Environmental changes stimulate a histidine kinase (HK) to autophosphorylate on a histidine residue. The phosphoryl group is then passed to an aspartate on the receiver (rec) domain of a downstream response regulator (RR), which modulates a cellular response (**Fig. 7A, top panel**). Some organisms, like plants, fungi and certain bacteria, have developed an expanded system known as a multistep His-Asp phosphorelay. This uses a hybrid HK that autophosphorylates on a His residue and passes the phosphoryl group to an Asp on an attached rec domain. The phosphoryl group is then transferred to a histidine-containing phosphotransfer (HPt) protein (**Fig. 7A, bottom panel**) [2]. In plants and fungi, HPt proteins often occupy branch points in the signaling pathway and can transfer phosphoryl groups to multiple downstream RRs. The osmosensing Sln1 pathway in *Saccharomyces cerevisiae* contains such a system. The HPt protein, Ypd1, acts as the relay point between the upstream hybrid sensor HK, Sln1, and the two downstream RRs, Ssk1 and Skn7 [3, 4]. To prevent cross-talk between multiple partners and pathways, protein-protein interactions within branched systems must be tightly regulated and/or highly specific [5-9].

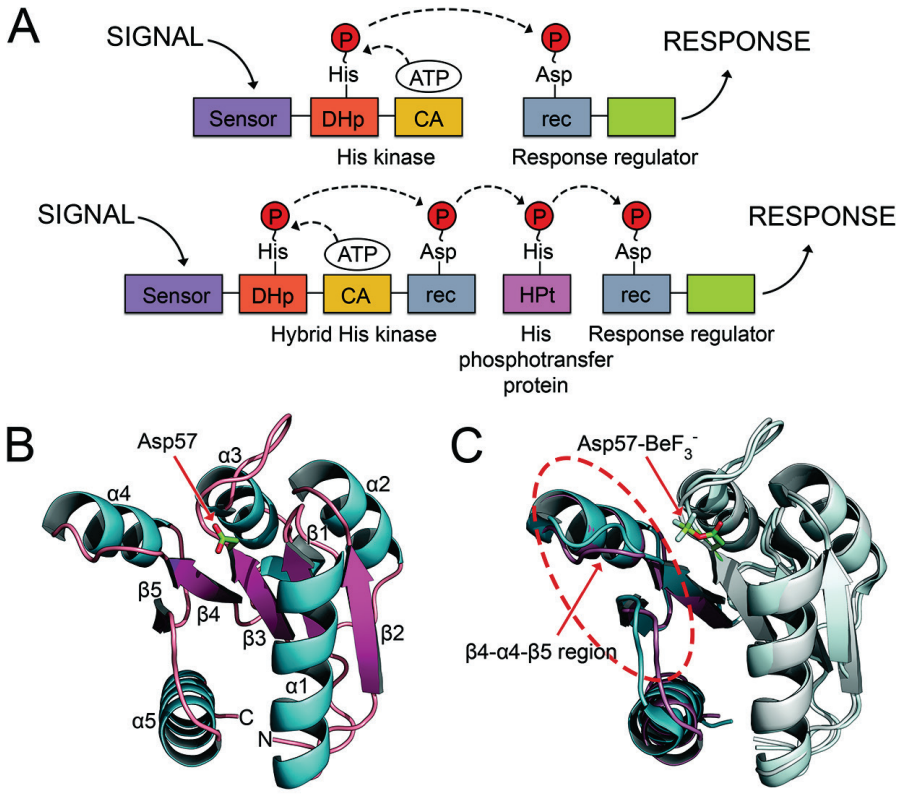

Figure 7. Domain arrangements for TCS systems; topology and induced conformational changes of RR rec domains.

(A) Canonical two-component signaling pathway (top) and expanded multistep phosphorelay (bottom). In a typical TCS pathway, the signal is sensed by a histidine kinase, which binds ATP in a catalytic domain (CA) and autophosphorylates a histidine residue within the dimerization histidine phosphotransfer domain (DHp). The phosphoryl group is transferred to an aspartate on the receiver domain (rec) of a downstream response regulator protein (RR). Activation of the rec domain elicits a cellular response. Expanded phosphorelay systems contain a hybrid HK, with its own receiver domain, and an intermediate known as a histidine-containing phosphotransfer protein (HPt). (B) Crystal structure of the representative RR, CheY (PDB 3CHY), in its unphosphorylated or inactive state shows the common $(\beta\alpha)_5$ topology. (C) Major conformational shifts upon phosphorylation occur within the β4-α4-β5 regions (circled). Inactive CheY (PDB 3CHY) in light cyan. BeF_3^- activated CheY (PDB 1FQW) in darker purple. Asp57-BeF_3^- is labeled and shown in stick model.

Nearly all rec domains adopt a common $(\beta\alpha)_5$ topology (**Fig. 7B**) [2], consisting of a five-stranded, parallel β-sheet surrounded by five amphipathic α-helices. Loops located at the C-terminal ends of β1, β3 and β5 form a docking surface for cognate HKs and HPt proteins. **Fig. 7B** shows a

representative structure of the chemotaxis RR, CheY [10]. Experimental structures show that rec domains use a highly conserved active site geometry that is essential for phosphotransfer. Rec domain active sites contain three conserved acidic residues (Asp or Glu) for coordinating a divalent metal cation, such as Mg^{2+}. The Asp at the C-terminal end of β3 serves as the site of phosphorylation [2, 11, 12]. Two additional residues complete the active site: a highly conserved lysine and a moderately conserved residue (often Gln, Asn or Lys) located two positions downstream from the phosphorylatable aspartate (D+2 position) [2, 11, 12].

The favored model of rec domain switching posits the existence of a dynamic equilibrium between apo and phosphorylated states. [13-19]. Phosphorylation is known to induce allosteric and functional changes in rec domains [11, 20-24]. This is thought to cause a population shift by stabilizing the modified state, thereby redistributing the equilibrium [17]. This shift is not fully understood, and additional intermediate states have been observed [13, 25-28]. The equilibrium shift model, also known as conformational selection, is in contrast to the more traditional idea of induced fit. In the induced fit scenario, phosphorylation is required to drive the conformational changes in the rec domain, which only occur when the ligand is bound (summarized in [29], specifically for rec domains in [13]). Despite more recent support for the conformational selection theory, controversy still surrounds the exact mode of rec domain switching.

Structural changes associated with phosphorylation are often subtle and difficult to identify. Structural alignments indicate average RMSDs between apo and phosphorylated conformations range from 1 to 3 Å, but deviations can be significantly greater (>7 Å) in functionally relevant areas. Experimental structures reveal several features common to most known modified rec domain structures [11, 23, 30-33]. **Fig. 7C** shows a rec domain alignment of CheY (PDB 3CHY [10] in its apo state; cyan) with CheY in its BeF_3^- bound state (PDB 1FQW [11]; purple). In CheY and the majority of other rec domains, large shifts consistently occur within the β4-α4-β5 region. Two conserved residues, a Ser/Thr on β4 and a Phe/Tyr/His on β5, adopt characteristically altered orientations in the crystal structures. Upon phosphorylation, the Ser/Thr rotamerizes and/or shifts with the β4α4 loop to form a hydrogen bond with the phosphoryl group, typically at a distance of 2.5-2.8 Å. The aromatic side chain of Phe/Tyr/His initially points away from the phosphoryl group, exposed to solvent in the apo state. Upon phosphorylation, it rotamerizes inwards to bury its side chain in a hydrophobic pocket recently vacated by the Ser/Thr on β4. These two conserved amino acids are referred to as "switch residues" and are used as traditional indicators of conformational state for rec domains [28]. Some studies have also implicated the aromatic residue (Phe/Tyr/His) as a factor in rec domain dimerization, due to its location in the center of a common dimerization interface [34, 35]. In CheY, these switch residues are Thr87 and Tyr106. More recently, a quartet of coupled residues have been implicated in

directing allosteric changes in CheY [16]. The model speculates that Trp58 (D+1) rotamerizes and shifts upwards, simultaneously increasing contact with Glu89 and decreasing contact with Met85. These changes are thought to be related to the rotamerization of Tyr106 and shifts in the β4α4 region. While these residues are only partially conserved, the relative phosphorylation-induced changes are visible in many experimental structures [30, 32, 33, 36, 37].

Even with highly conserved tertiary structure, genetic diversity between rec domains has led to divergent functional properties, making studies of individual rec domains necessary [11, 20, 21, 23, 24, 38-41]. Depending on their biological role, autodephosphorylation rates of rec domains can span several orders of magnitude, suggesting other structural elements may be affecting the stability of the phosphoaspartate [12, 42-44]. Many rec domains also oligomerize upon phosphorylation, usually through the α4-β5-α5 face [30, 34, 45-47]. Non-canonical interfaces utilizing the α1α5 helices have been observed as well [46, 48, 49]. Structural rearrangements associated with phosphorylation can cause changes in surface shape and physicochemical properties, affecting interactions with binding partners [47, 50-56]. While rec structures in their apo states provide basic structural information, they do not reveal the important allosteric changes that occur upon phosphorylation.

Of the >300,000 rec domain sequences found in the NCBI database, approximately 150 (non-redundant) have structures available in the RCSB

Protein Data Bank (PDB; http://www.rcsb.org) [57-59]. The majority of these structures are unphosphorylated, due to the difficulty of capturing the chemically labile phosphoaspartate species [60-64]. Most "phosphorylated" structures were obtained using phosphoryl analogs, like BeF_3^- [11, 23, 30-33]. This approach has met with limited success; fewer than 30 rec domains have been characterized in a conformationally modified state [57, 58]. Molecular dynamics (MD) can be used to circumvent the difficulties of characterizing these transient conformations. However, kinetic studies reveal that rec domain phosphoryation and conformational changes occur on high microsecond to low millisecond timescales [20, 21, 65]. Simulating these timescales at atomic resolution would be prohibitively expensive and impractical for most research laboratories. To make these studies more accessible, biasing terms can reduce the amount of sampling required [66-68]. The most conserved rec domain feature is the active site geometry required for phosphotransfer [11, 12, 33]. Restraints can be used to drive the formation of this active site configuration for any rec domain. This shifts the protein towards a phosphorylated state, leading to allosteric rearrangements throughout the structure. Though these phosphorylated conformations are theoretically observable without restraints, they would require millisecond-timescale simulations to reliably sample. The active site restraints drastically reduce the conformational search space and allow us to sample phosphorylated conformations in a fraction of the time [69].

With this approach, laboratories that lack the capability of millisecond timescale simulations are able to effectively simulate these events.

We have developed a methodology for accessing the phosphorylated conformation of a rec domain protein in a direct and expedient path using active site restraints. To validate this approach, we performed biased, all-atom simulations on five diverse rec domains with known modified structures. This produced models for the rec domains that are nearly identical to the experimental data, both in global conformation and in key local features. The success of this methodology suggests that while rec domains may utilize some form of conformational selection to achieve allosteric switching, the application of an induced fit model is sufficient to achieve an equivalent effect. Additionally, data on conformational dynamics obtained from each simulation were used to identify areas of major functional significance associated with phosphorylation. Finally, we applied the approach to two rec domains found within the osmoregulatory Sln1 pathway of *S. cerevisiae*, the upstream Sln1-rec and the downstream Ssk1-rec, to compare their structural and functional changes upon activation. The principle of using active site restraints to reduce conformational search space can be applied to other protein families individually, such as kinases and HPt proteins. In addition, this method can be expanded to study the effects of phosphorylation on rec domain interactions with related protein families.

2.2 Materials and Methods

Modeling and preparation

Initial structures. Crystal structures of single domain RRs (CheY and Spo0F), rec domains from multi-domain RRs (PhoP-rec and FixJ-rec), and the rec domain of a hybrid HK (Sln1-rec) were obtained from the PDB [57]. Modified rec structures contained both a metal cation and the modified aspartate residue, either with BeF_3^- or PO_3^{2-}. For a full list of the proteins used, see **Table 2**.

Table 2. Protein crystal structures used during simulations.

Protein (Organism)	*Inactive (PDB)*	*Active (PDB)*	*PO_3^{2-} analog*	*Ref.*
CheY *(E. coli)*	3CHY:A	1FQW:A	BeF_3^-	[10, 11]
PhoP-rec *(E. coli)*	2PKX:A	2PL1:A	BeF_3^-	[30]
Sln1-rec *(S. cerevisiae)*	1OXB:B	2R25:B	BeF_3^-	[4, 33]
FixJ-rec *(S. meliloti)*	1DCK:A	1D5W:A	PO_3^{2-}	[23, 37]
Spo0F *(B. subtilis)*	1NAT:A	2FTK:G	BeF_3^-	[32, 70]

Ssk1-rec structure prediction. The Protein Homology/analogY Recognition Engine v.2.0 (Phyre2) was used to generate a homology model of apo Ssk1-rec from *S. cerevisiae* [71]. The Ssk1-rec domain sequence (residues 505-580:603-651) was used as input for intensive modeling mode. Non-conserved α3β4 loop residues (581-602) were excluded due to poor modeling quality and their location on the surface opposite the active site. The initial

model was refined using approximately 5 ns of all-atom MD simulation. The resulting equilbrated structure was extracted and examined using Verify3D and RAMPAGE to assess model quality [72, 73]. The Ssk1-rec model was then prepared and simulated as described in subsequent sections to sample its phosphorylated state.

System preparation. Starting models were prepared from the apo crystal structures. Phosphoaspartate was modeled into the active sites with Coot [74]. A Mg^{2+} atom was also added based on available experimental coordinates. If no cation was present in the experimental structure, coordinates were copied from a homologous structure containing metal. While other metal types have been found in rec domain active sites, Mg^{2+} is sufficient for function of nearly all rec domains [75-77]. The initial positioning of the metal cation is not criticial in the context of the simulations. Each protein will ultimately adopt its phosphorylated conformation around the cation as long as it is present in the general vicinity of the active site. Initial structures were stripped of all other crystallographic waters, ligands and additional subunits prior to simulation.

Dowser was used to fill internal water cavities with a default probe radius of 0.2 Å [78, 79]. The SOLVATE program was then used to create a contoured solvent shell around each model with a shell thickness of 6.0 Å [67]. Systems were loaded into VMD and processed using the AutoPSF plug-in [80]. Finally, structures were immersed in a full orthorhombic water box with

15.0 Å padding on each side using a TIP3P water model with the VMD solvate plug-in [80]. Systems were neutralized and K^+/Cl^- ions were added to a final concentration of 100 mM using the VMD Autoionize plug-in [80].

Table 3. Partial charges assigned for phosphoaspartate.

Custom residue type - AST		
Atom name	***Atom type***	***Partial charge***
CB	CT2	-0.18
HB1	HA	0.09
HB2	HA	0.09
CG	CD	0.51
OD1	OB	-0.51
P	P	1.10
OD2	ON2	-0.40
H2	HN4	none
OP1	ON3	-0.90
OP2	ON4	-0.90
OP3	ON3	-0.90

Phosphoaspartate. Parameters for dianionic phosphoaspartate (AST) were defined for the CHARMM36 force field based on Damjanović et al. (2009) [81-83]. A tetrahedral phosphoryl group exists predominantly in the dianionic form at physiological pH (7.4) [63]. All simulations were run using the dianionic group instead of the monoanionic form. As described in [83], charges for the phosphoryl group along with various bond and angle terms were taken from dianionic methylphosphate. Additional dihedral angle terms were taken from phenol phosphate. The torsion angle (X-CD-ON2-X) and

partial charges (CG and OD1 atoms) were obtained from previous parameterization studies [83]. See **Table 3** for partial charges. Formatted entries for the CHARMM32 force field topology and parameter files are included below.

AST residue entry from CHARMM force field topology file.

```
RESI AST      -2.00
GROUP
ATOM          N       NH1 -0.47     !         |
ATOM          HN      H   0.31      !   HN-N
ATOM          CA      CT1  0.07     !         |  HB1      OD1
ATOM          HA      HB1  0.09     !         |     |        //
GROUP                               !          HA-CA--CB--CG
ATOM          CB      CT2  -0.18    !         |     |        \
ATOM          HB1     HA2  0.09     !         |  HB2       OD2
ATOM          HB2     HA2  0.09     !    O=C                 |
ATOM          CG      CD      0.51  !         |           O1P-P-O3P
ATOM          OD1     OB  -0.51     !                           |
ATOM          OD2     ON2  -0.40    !                        O2P
ATOM          P       P   1.10      !
ATOM          O1P     ON3  -0.90    !
ATOM          O2P     ON3  -0.90    !
ATOM          O3P     ON3  -0.90    !
GROUP
ATOM          C       C        0.51
ATOM          O       O        -0.51
BOND          CB      CA      CG      CB      OD2     CG
BOND          N       HN      N       CA      C       CA      C       +N
BOND          CA      HA      CB      HB1     CB      HB2
BOND          P       OD2     O1P     P       O3P     P       OT      P
DOUBLE        O       C       CG      OD1
IMPR  N       -C      CA      HN      C       CA      +N      O
IMPR  CG      CB      OD2     OD1
CMAP          -C      N       CA      C       N       CA      C       +N
DONOR         HN      N
ACCEPTOR      OD1     CG
ACCEPTOR      OD2     CG
ACCEPTOR      O       C
```

```
IC      -C      CA      *N      HN      1.3465 125.3100 180.0000 112.9400
0.9966
IC      -C      N       CA      C       1.3465 125.3100 180.0000 105.6300
1.5315
IC      N       CA      C       +N      1.4490 105.6300 180.0000 117.0600
1.3478
IC      +N      CA      *C      O       1.3478 117.0600 180.0000 120.7100
1.2330
IC      CA      C       +N      +CA     1.5315 117.0600 180.0000 125.3900
1.4484
IC      N       C       *CA     CB      1.4490 105.6300 122.3300 114.1000
1.5619
IC      N       C       *CA     HA      1.4490 105.6300 -116.4000 106.7700
1.0841
IC      N       CA      CB      CG      1.4490 111.1000 180.0000 112.6000
1.5218
IC      CG      CA      *CB     HB1     1.5218 112.6000 119.2200 109.2300
1.1086
IC      CG      CA      *CB     HB2     1.5218 112.6000 -121.6100 110.6400
1.1080
IC      CA      CB      CG      OD1     1.5619 112.6000 180.0000 117.9900
1.2565
IC      OD1     CB      *CG     OD2     1.2565 117.9900 -17
```

AST residue entry from CHARMM force field parameter file.

```
! Bond for phosphoaspartate
ON2     CD              230.00          1.4000
! Angles for phosphoaspartate
CD      ON2     P       20.0            120.0           35              2.33
ON2     CD      CT2     55.000          110.5000
ON2     CD      CT3     55.000          110.5000
ON2     CD      OB      50.000          123.00          210.00          2.26200
! Dihedrals for phosphoaspartate
X       CD      ON2     X       10.0            1       180.0
X       CD      ON2     X       2.25            2       180.00
X       CD      ON2     X       1.36            3       180.00
CD      ON2     P       ON3     0.1             3       0.00
```

Molecular dynamics

Simulations. Simulations were prepared and run using VMD and NAMD 2.9 [80, 84] with the CHARMM36 protein force field [81, 82]. A 12.0 Å cutoff was used for van der Waals interactions with a switching function distance of 10.0 Å. Long-range electrostatics were evaluated using the PME (Particle Mesh Ewald) method with a tolerance, interpolation order and grid spacing of $10e^{-6}$, 4.0 and 1.0 Å, respectively. A two fs integration step was used for all simulations. The SHAKE algorithm was applied to constrain covalent bonds involving hydrogen atoms throughout. Initial energy minimization was performed on non-backbone atoms for 100 ps of NVT MD. An all-atom minimization step was then done for an additional 100 ps of NVT MD with harmonic constraints of 0.5 kcal/mol applied to all Cα atoms. Following this, systems were heated to 310 K by increments of 1 K/ps over approximately 310 ps of NVT MD. Equilibration was then performed with Cα atoms restrained for 500 ps of NPT MD. A final, unrestrained equilibration step of 4 ns NPT MD was run to prepare systems for production. Proteins were equilibrated and then simulated in multiple 10 ns production runs to collect 100 ns of effective production data for each rec domain. Trajectories were combined for analysis. Parallel runs were performed to increase conformational sampling.

Simulation restraints. Biasing terms are frequently used to steer systems towards desired states for study. Parameters that regulate this are

known as collective variables (colvars) [69]. A colvar ξ is defined as a differentiable function of the vector of *3N* atomic Cartesian coordinates, *X*:

$$\xi(X) = \xi(x_1, x_2, \ldots x_N)$$

ξ(*X*) is generally a function with many fewer arguments than *3N*, or is a combination of these functions:

$$\xi(X) = \xi\left(z^{(1)}(X), z^{(2)}(X), \ldots z^{(\alpha)}(X) \ldots\right)$$

with the Cartesian gradient:

$$\nabla x \xi(X) = \left(\nabla_{x_1}\xi(X), \nabla_{x_2}\xi(X), \ldots\right)$$

where the number of functions $z^{(\alpha)}$ is much smaller than the number of atoms. The term $z^{(\alpha)}(X)$ describes a component and is a function of several atomic coordinates. A colvar is defined by a combination of one or more components. In the simplest scenario, the colvar ξ identifies a single component z, and often relates to properties such as atomic distances or dihedral angles.

Table 4. Colvars applied to drive active site formation in CheY.

Type	***Group1 atoms***	***Group2 atoms***	***Boundaries***
Distance	Asp57 OD1	Mg^{2+}	1.9-2.2 Å
CoordNum	Asp57 PO_3^{2-}	Mg^{2+}	1 group
Distance	Mg^{2+}	COM	0.0-0.1 Å
Distance	Ala88 N	Asp57 OP2	2.0-3.3 Å
CoordNum	Asp13 R	Mg^{2+}	1 group
Distance	Asp57 OP1	Mg^{2+}	0.0-2.2 Å
Distance	Asp13 OD1	Mg^{2+}	0.0-2.2 Å
Distance	Gln59 O	Mg^{2+}	0.0-2.2 Å

Using the colvar module in NAMD, we applied half-harmonic potential restraints to each rec domain to drive the formation of the appropriate active site geometry [84]. Restraints were added during equilibration and maintained for the full 100 ns production runs. For a full list of the restraints used in the CheY representative experiments, see **Table 4**. While this list was used on every system, minor changes were made if a given protein could not adopt a fully correct active site, often due to initial non-physiological conformations. A sample colvars list is provided below for the CheY simulations. It can be adapted for any target rec domain by changing the atomNumbers.

Sample NAMD colvars (.in) file for CheY.

```
colvarsTrajFrequency      5000
colvarsRestartFrequency   5000

colvar {
name asp57od1_to_mg               # distance of carbonyl to mg
width 0.1
lowerboundary 1.9
upperboundary 2.2
lowerWallConstant 100.0
upperWallConstant 100.0

distance {
group1 { atomNumbers 1986 }
group2 { atomNumbers 875 }
}
}

colvar {
name asp57phos_coord_to_mg        # coordination of phosphoryl to mg

lowerboundary 0.0
```

```
upperboundary 1.0
lowerWallConstant 100.0
upperWallConstant 100.0

coordNum {
group1 { atomNumbers 878 879 880 }
group2 { atomNumbers 1986}
cutoff 3.3
}
}

colvar {
name mgcom                    # anchor mg to its own starting coordinates
width 0.001

lowerboundary 0.0
upperboundary 0.1
lowerWallConstant 100.0
upperWallConstant 100.0

distance {
group1 { atomNumbers 1986 }
group2 { dummyAtom (0.069, 44.943, 123.420) }
}
}

colvar {
name ala88n_to_asp57ot              # distance of alanine to phosphoryl to
overcome initial clashes
width 0.001

lowerboundary 2.0
upperboundary 3.3
lowerWallConstant 100.0
upperWallConstant 100.0

distance {
group1 { atomNumbers 1345 }
group2 { atomNumbers 879 }
}
}

colvar {
```

```
name asp13_coord_to_mg                # coordination of active site aspartate to
mg to allow water

lowerboundary 0.0
upperboundary 1.0
lowerWallConstant 100.0
upperWallConstant 100.0

coordNum {
group1 { atomNumbers 193 194 195 }
group2 { atomNumbers 1986}
cutoff 3.3
}
}

colvar {
name asp57o1p_to_mg                   # distance of phosphoryl to mg
width 0.05

lowerboundary 2.2
upperboundary 2.3
lowerWallConstant 100.0
upperWallConstant 100.0

distance {
group1 { atomNumbers 1986 }
group2 { atomNumbers 878 }
}
}

colvar {
name asp13od1_to_mg                   # distance of active site aspartate to mg

lowerboundary 0.0
upperboundary 2.2
lowerWallConstant 100.0
upperWallConstant 100.0

distance {
group1 { atomNumbers 194 }
group2 { atomNumbers 1986 }
}
}
```

```
colvar {
name gln59co_to_mg            # distance of main chain carbonyl on D+2 to mg

lowerboundary 0.0
upperboundary 2.2
lowerWallConstant 100.0
upperWallConstant 100.0

distance {
group1 { atomNumbers 920 }
group2 { atomNumbers 1986 }
}
}
```

Data analysis

Trajectory analysis. Trajectories were analyzed using VMD and the Bio3D package in R [80, 85, 86]. Production runs were stripped of all non-protein atoms using VMD. Partial trajectories were then aligned by Cα atoms using the RMSDTT plug-in [80] and combined to form single trajectories for analysis.

Each simulation produced >20,000 conformations (snapshots) of its protein. Analysis of such high-dimensionality data is challenging. We performed principal component analysis (PCA) on the full combined trajectories to reduce the complexity of the data and analyze only the dominant conformations. PCA is a multivariate technique to simplify and visualize patterns within structural ensembles, including inter-conformer relationships [85, 87]. Typically it is performed using a reduced set of atoms, such as an Cα representation. The first step, after trajectory alignment, is construction of a

covariance matrix, *C*, for each trajectory. The elements, *Cij*, of the *3N* dimensional matrix are defined as:

$$Cij = \langle (xi - \langle xi \rangle)(xj - \langle xj \rangle) \rangle$$

where *xi* and *xj* are Cα coordinates, *i* and *j* represent all possible pairs of *3N* Cartesian coordinates (*N* = number of atoms being considered) and the brackets denote an ensemble. The covariance matrix is then diagonalized, with the columns of the transformation matrix becoming eigenvectors with corresponding eigenvalues. Each eigenvector represents a correlated displacement of groups of atoms through space, with the eigenvalues describing the variance of distribution along the corresponding eigenvector. Projection of the conformational distribution onto the subspace defined by the first few principal components (PCs; by eigenvalues) creates a lower-dimensional representation of the data, making analysis easier and faster [85, 87]. We then used a k-medoids approach to cluster the distribution in PC-space, grouping conformers based on structural similarity [88, 89]. The CLARA (Clustering for Large Applications) algorithm was used to handle the large data sets [89, 90].

We performed dynamical cross-correlation (DCC) analysis on the full 100 ns combined trajectories to detect collective motions between Cα atoms during each simulation. Trajectories were first superimposed against the initial frame. A cross-correlation coefficient, *Cij*, was calculated for the Cα atoms with the expression:

$$Cij = \frac{\langle \Delta r_i \cdot \Delta r_j \rangle}{\sqrt{(\langle \Delta r_i^2 \rangle \langle \Delta r_j^2 \rangle)}}$$

where Δr_i and Δr_j are the displacements from the mean position of the *i*th and *j*th atoms with respect to time [91]. *Cij* values were either positive, representing in-phase or positively correlated motion, or negative, representing out-of-phase or negatively correlated motion. These were collected in matrix form and mapped for analysis.

Structural analysis. Structural alignments and analyses were performed using UCSF Chimera [92]. Superpositioning was done using the internal β-sheet, as it showed minimal deviation upon activation in rec domain crystal structures. Per residue RMSDs were exported for graphing and comparison in R [86]. Electrostatic surface potential was calculated with Adaptive Poisson-Boltzmann Solver (APBS) package in 0.1 M KCl [93]. Surface cavities were calculated using the 3D-Surfer server with the Visgrid algorithm [94-96].

Software. MD simulations were performed with NAMD [84] on resources provided by the OU Supercomputing Center for Education and Research (OSCER) at the University of Oklahoma. System preparation, visualization and initial trajectory analyses were performed in VMD [80]. Models and figures were made in Coot and PyMOL [74, 97]. Structural analyses were performed in UCSF Chimera [92]. Trajectory analyses and plots were done with R using the Bio3D, fpc and cluster packages [85, 89, 98].

2.3 Results and Discussion

Restraints drive phosphorylation-induced conformational changes

Crystal structures of five apo rec domains were obtained from the PDB (**see Table 2**) for simulations. Phosphorylatable aspartate residues were replaced in the apo models with a custom dianionic phosphoaspartate (AST). A Mg^{2+} cation was modeled into the center of the active sites to act as a pivot point for each protein. Two options exist and were tested for the positioning of this cation in the apo model(s). The first was to use the same coordinates as the cation found in the modified (BeF_3^- bound) experimental structure for a given rec domain. The second was to align the apo model(s) to a known modified structure of a homologous rec domain containing a bound cation. The metal's coordinates were copied from the homologous structure and inserted into the unphosphorylated system(s). Results from the validation set showed that both approaches are effective, demonstrating how a rec domain with no known modified or metal-bound structure can be simulated by manually modeling the bound cation. Half-harmonic potential restraints were applied centered on this metal cation (**see Table 4**). Any restraints during a simulation can affect the dynamics of the system. Caution should be used when speculating on any mechanism of conformational transition. Nevertheless, we know that the active site is almost always arranged in the previously described geometry, as it is necessary for the chemistry of phosphotransfer. Sampling phosphorylated rec domain conformations without the active site restraints is possible, but would

require far longer (millisecond) timescales that are prohibitively expensive and impractical for standard research laboratories. We repeatedly failed to observe complete transitions from apo to phosphorylated states for the rec domains in the validation set without the correct combination of restraints. This suggests that on this timescale, we cannot reliably or adequately sample phosphorylated conformations without the biasing terms.

Biased MD simulations were performed in explicit solvent for each system. Structural comparisons of phosphorylated and apo rec domain crystal structures suggest that the shift towards a phosphorylated conformation is subtle when the entire domain is included (average RMSDs of 1-3 Å) [11, 33]. Cα deviations observed during the production simulations corroborated this, with most systems stabilizing within 5 ns and maintaining RMSDs of 1-2 Å throughout the runs (data not shown). This indicates that the simulated phosphorylation of these rec domains occurs on a similar magnitude as the experimental structures. The subtlety of these perturbations makes overall structural comparisons difficult for both experimental data and predicted models. In contrast, major changes are almost always seen along the β4-α4-β5 region, with some deviations occurring on magnitudes of >7 Å in known structures [30, 32, 33, 36, 37]. To explore if these changes were due to increased flexibility within this region, or more concerted intramolecular movements, we calculated the per residue RMSF for each protein during the production simulations. **Fig. 8** shows the plots for the rec domains in the

validation set. As expected, the loop regions generally possessed the highest levels of flexibility. The β4-α4-β5 region demonstrated no outstanding flexibility relative to the rest of the protein in the majority of the rec domains, suggesting that it is most likely involved in more concerted or specific changes associated with phosphorylation. FixJ-rec is a possible exception. Comparison of existing crystal structures reveals that the β4-α4-β5 region of FixJ-rec undergoes the most dramatic shift within the validation set. The high RMSF in this region suggests that the protein's intrinsic flexibility may contribute significantly to this deviation, in addition to the allosteric changes induced by phosphorylation.

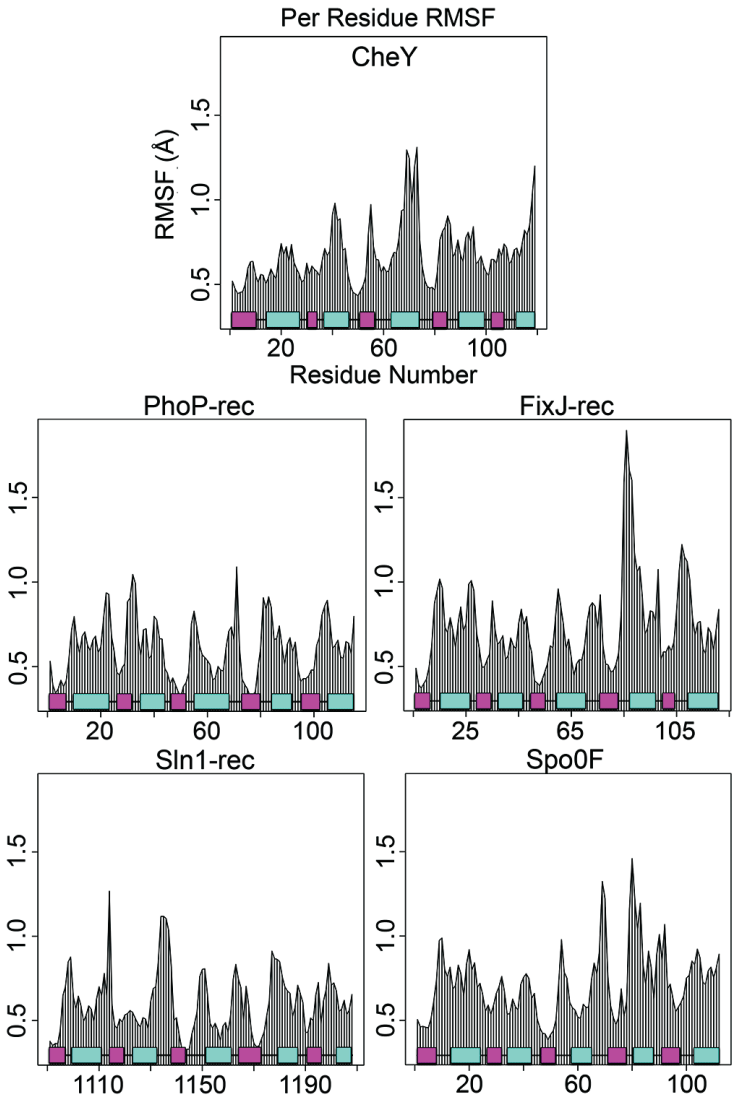

Figure 8. Residue mobilities during simulation for the five phosphorylated rec domains.

Comparison of average root mean square fluctuations (RMSFs) for combined MD trajectories shows regions of relative flexibility in each protein. β-strands (magenta); α-helices (cyan).

Analysis of collective motions reveals areas of functional relevance

Protein dynamics are intimately involved in enzyme activity and catalysis. Previous work has speculated that areas undergoing collective motions within proteins are likely distinct subdomains of functionally related

residues [91, 99-105]. To identify these regions, with the notion that they may prove relevant to the phosphorylated state and protein function, we generated dynamical cross-correlation maps (DCCM) for all five rec domains [85]. Cross-correlation plots show how various regions of a protein communicate by quantifying their relative motions. Motion can be positively correlated, meaning in the same direction, or negatively correlated (anticorrelated), meaning in the opposite direction. **Fig. 9** shows maps of correlated and anticorrelated motion within each protein calculated using the combined trajectories. These represent motions with the active site restraints already enforcing the conserved catalytic geometry. Correlation was normalized and scaled between -1.0 and 1.0. Collective motions between -0.25 and 0.25 were not considered significant (**Fig. 9 white**) [101, 106]. For CheY, the middle and upper right areas show positively correlated motions between residues 80-89 and 101-110 (**Fig. 9, yellow/green**). These correspond to the α3-β4-α4 and α4-β5-α5 regions, suggesting correlated motions between the β4α4 and β5α5 loops and throughout the β4 and β5 strands themselves. In previous studies on activated rec domains, most dramatic shifts were typically observed in these areas, particularly in the loops adjacent to the active site [11, 23, 30-33]. Residues 50-61 and 79-87 (α2-β3-α3 and α3-β4-α4 regions) also displayed positively correlated motions, again suggesting concerted loop movements and β-strand interactions. In the lower left, residues 4-14 and 31-38 (β1α1 and α1-β2-α2 regions) and residues 4-14 and 50-61 (β1α1 and α2-β3-α3 regions) also

showed similar positively correlated motions throughout every simulation. The areas of positive correlation suggest high overall stability of the central β-sheet (as expected) and its interactions with loops, but relatively little correlated motion between it and the flanking helices. Strong negatively correlated motion was found within the CheY molecule between the C-terminal portion of the α5 helix and the β5 strand and between regions of the α5 helix and most of the α4 helix. This indicates significant concerted movement within the α4-β5-α5 face of CheY. Other studies have shown that many rec domains dimerize through this surface [30, 34, 45-47]. While CheY does not form a homodimer, it does binds the downstream flagellar motor switch protein, FliM, through this surface [36]. Phosphorylation of CheY is known to increase its binding affinity for FliM 20-fold [107]. The trajectory data revealed significant reorientation of key residues (Ala90, Ile95, Tyr106, Lys119, Lys122; relative to the apo conformation) involved in, or potentially blocking, interaction with FliM [36]. Each of the five proteins produced remarkably similar patterns of positively correlated motions involving loop regions and the central β-strands. Due to the high similarity between profiles, this pattern of positively correlated motion can provide a check for the proper simulation of rec domains. Perhaps more important are the unique patterns of negatively correlated motion produced by each protein (**Fig. 9, blue/black**). These variations offer valuable insights related to differences in rec domain function and effects after allosteric modification. The results suggest that putative regions of functional

significance tied to phosphorylation-induced changes can be identified based on (anti)correlated motions. This provides potential targets for further experimental studies. However, it must be noted that DCCM alone cannot provide conclusive information on conformational transitions. Further study would be valuable for characterizing the extent of these changes, using complementary approaches such as normal mode analysis. The following sections briefly summarize the DCCM results for the four remaining rec domains in the validation set.

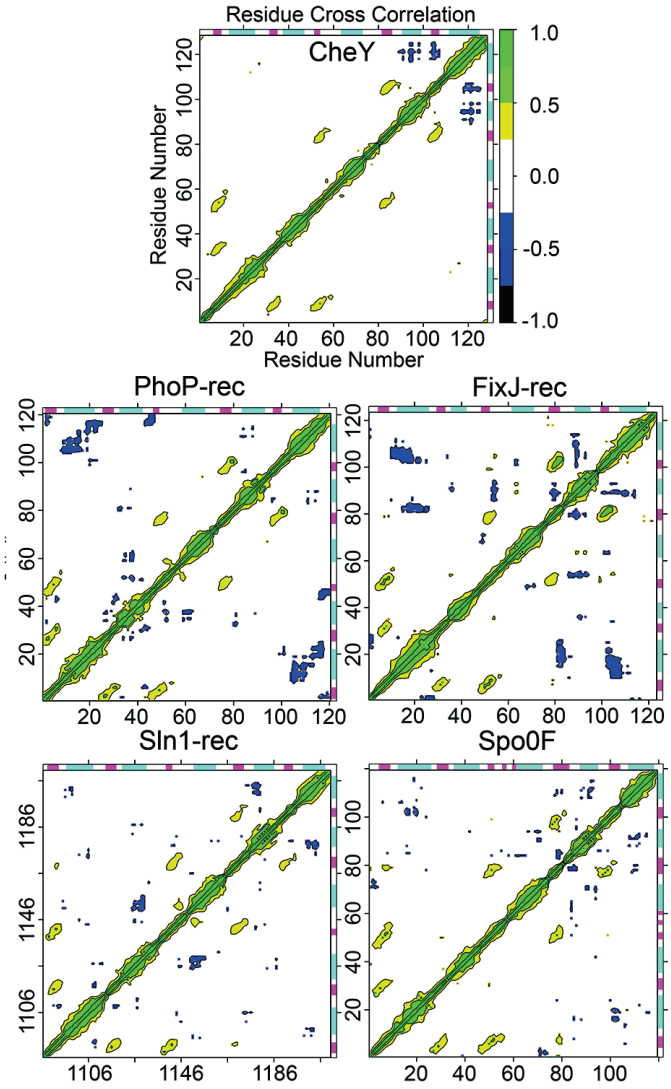

Figure 9. Dynamical cross-correlation maps (DCCM) reveal collective motions within the phosphorylated rec domains.

Positive correlation (yellow/green) suggests movement in phase, or in the same direction. Negative correlation (blue/black) suggests movement out of phase, or in along the opposite direction. β-strands (magenta); α-helices (cyan).

Major regions of FixJ-rec that exhibited significant negatively correlated motions were the α4-β5-α5 surface, including a large shift in the α4 helix, and the loop regions accessible on the active site surface. Phosphorylation-dependent FixJ dimerization and further downstream

signaling is believed to occur through the α4-β5-α5 interface (similar to CheY-FliM binding) [23, 53]. Key residues implicated in homodimerization that adopted altered orientations during simulation and were involved in negatively correlated motions include His84, Val87, Val91, Lys95 and the switch residue Phe101 [23]. Additionally, correlated perturbations seen in the α1α5 region are likely involved in the binding of the upstream partner, FixL. This is supported by the work of Saito, et al. (2003), which suggests that mutations at this surface (particularly on the α1 helix) interfere with phosphotransfer from FixL to FixJ [108]. As FixL also exhibits phosphatase activity, this likely affects the stability of phosphorylated FixJ-rec.

Regions of negatively correlated motion seen in PhoP-rec correspond to the α2α3 and α1α5 regions, similar to other rec domains. Also present is the characteristic movement of the α4-β5-α5 face, though to a much smaller degree than the other rec domains. PhoP dimerizes, even in an unphosphorylated form, through an asymmetric interface, but phosphorylation is required for downstream function (DNA binding) [30, 109]. Phosphorylation causes a shift in the α4 helix, potentially stabilizing the dimerization interface. Cross-correlation analysis also revealed significant correlated movements in the loop regions near the active site surface of PhoP-rec. All known rec domains must bind their upstream donors through this active site to facilitate phosphotransfer, suggesting that the correlated loop motions may be involved in interactions with the upstream partner, PhoQ.

Spo0F regions that demonstrated significant negatively correlated motion correspond to a characteristic rearrangement of the α4-β5-α5 face, along with more subtle motions at the α1α5 surface and the loops near the active site. The α4-β5-α5 surface of Spo0F is not involved in dimerization and does not directly bind its partner, Spo0B [32]. The exact purpose of this area is unknown, and these results suggest a target for additional studies. Collective motion observed between the α1 helix and the α4α5 region is most likely involved in protein-protein interactions with Spo0B, as phosphorylation is speculated to increase the binding affinity between the two proteins. The co-crystal structure of the Spo0F•Spo0B complex confirms this, showing direct intermolecular contacts located within this region [32]. In addition, previous studies indicated that mutations in α1 helix and the β4α4 region caused defects in binding with KinA [110].

Major regions in Sln1-rec that exhibited significant negatively correlated motion correspond to movements between active site loop regions and between the α2α3, α1α5 and α4-β5-α5 surfaces. Sln1 is a fungal hybrid histidine kinase believed to dimerize through its DHp domain [111]. The rec domain alone is believed to function as a monomer, though no full-length structure of the Sln1 hybrik HK exists. Cross-correlation analysis suggests that the α4-β5-α5 surface, and possibly the α1α5 and α2α3 helices, may have significant roles in the signaling pathway.These movements may involve interactions with the downstream partner HPt, Ypd1, or with its own hybrid

HK domain. Examination of the existing crystal structures suggest that the α2α3 and α1α5 helix regions are directly involved in binding Ypd1, though the purpose of the α4-β5-α5 region remains unknown [4, 33]. A more detailed analysis and comparison with another fungal rec domain, Ssk1-rec, is discussed in later sections.

Trajectory analysis and extraction of representative rec domain structures

Due to the dynamic nature of proteins, no single "correct" phosphorylated conformation exists for a rec domain. But based on the characteristic changes associated with rec domain phosphorylation found in existing structures, we can identify a representative structure that is most likely to resemble a phosphorylated conformation in a crystal structure. Determining which conformation(s) from within a trajectory best accomplishes this is indeed challenging. Each simulation produced a substantial amount of conformational data (>20,000 observations). We first had to simplify these complex datasets to identify which conformers were the most significant to this goal. To reduce the dimensionality of the data, we performed PCA on the combined trajectories for each rec domain (data not shown). Performing PCA prior to clustering can save significantly on computational cost and time, as well as filter out high-frequency conformational variance (noise) from the data [112].

Trajectories were then clustered in the resulting PC space to quantitatively identify the dominant conformations adopted by the proteins

during simulation. Due to the large size of each data set, an implementation of the CLARA algorithm was used. CLARA is an extension of the k-medoids clustering approach and can be used to extract the medoid for each identified partition [88, 89]. A medoid is an object within the cluster for which the dissimilarity to all other objects within the cluster is minimal. The CLARA algorithm requires an initial estimate of the desired number of clusters during input. To predict the ideal cluster number for each data set, we used the pamk function within the R package fpc. This estimates the ideal number of clusters for a data set by calculating the optimum average silhouette width for a range of clusters [90, 98, 113, 114]. Data were clustered using the top seven PCs to incorporate ~50% of the observed data variance. To produce a reasonable number of candidate structures from which to choose a final active structure, the minimum number of clusters was set to four, and the maximum was set to fifteen, with a corresponding medoid predicted for each identified group. For our purposes, the medoids best represent the dominant conformations observed during the activation of the rec domains. Among these models is the conformation(s) most likely to resemble the active crystal structure of the protein under study. **Figs. 10-14** show conformer plots produced for all five rec domain structures projected onto the top three PCs for simplicity (PC1-3). Each color represents an individual conformational cluster that was identified.

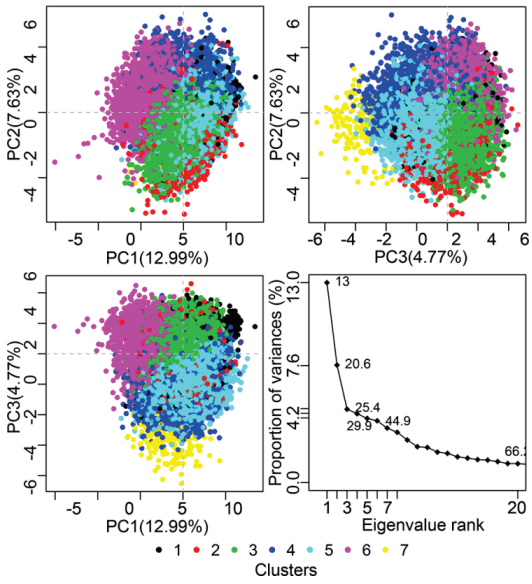

Figure 10. Projections of instantaneous CheY conformations onto the top three principal components.

Conformers were extracted from snapshots of each production run and plotted along the top three PCs (based on data variance). Shown are plots for the single domain RR, CheY, along with corresponding eigenvalue rankings. Each point represents one conformation adopted by CheY during MD simulations. Points were assigned colors based on groupings calculated using a k-medoids clustering approach.

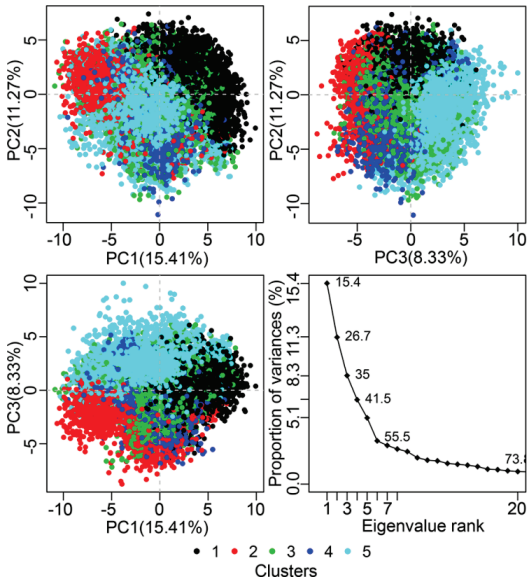

Figure 11. Projections of instantaneous PhoP-rec conformations onto the top three principal components.

Conformers were extracted from snapshots of each production run and plotted along the top three PCs (based on data variance). Shown are plots for PhoP-rec, along with corresponding

eigenvalue rankings. Each point represents one conformation adopted by PhoP-rec during MD simulations. Points were assigned colors based on groupings calculated using a k-medoids clustering approach.

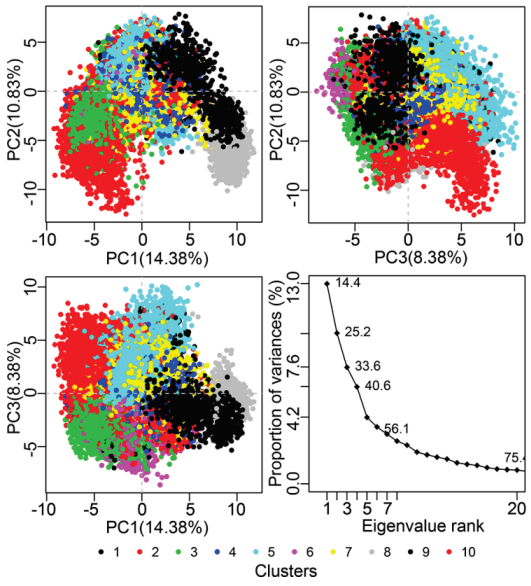

Figure 12. Projections of instantaneous FixJ-rec conformations onto the top three principal components.

Conformers were extracted from snapshots of each production run and plotted along the top three PCs (based on data variance). Shown are plots for FixJ-rec, along with corresponding eigenvalue rankings. Each point represents one conformation adopted by FixJ-rec during MD simulations. Points were assigned colors based on groupings calculated using a k-medoids clustering approach.

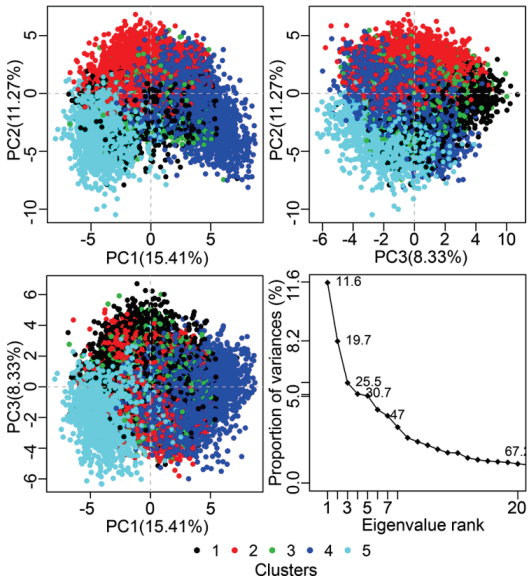

Figure 13. Projections of instantaneous Sln1-rec conformations onto the top three principal components.

Conformers were extracted from snapshots of each production run and plotted along the top three PCs (based on data variance). Shown are plots for Sln1-rec, along with corresponding eigenvalue rankings. Each point represents one conformation adopted by Sln1-rec during MD simulations. Points were assigned colors based on groupings calculated using a k-medoids clustering approach.

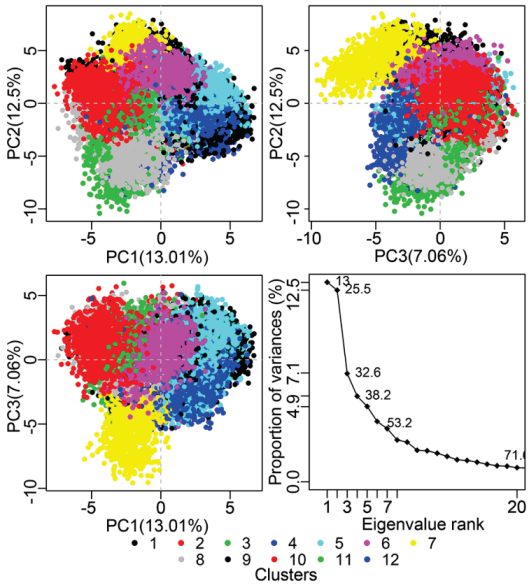

Figure 14. Projections of instantaneous Spo0F conformations onto the top three principal components.

Conformers were extracted from snapshots of each production run and plotted along the top three PCs (based on data variance). Shown are plots for Spo0F, along with corresponding eigenvalue rankings. Each point represents one conformation adopted by Spo0F during MD simulations. Points were assigned colors based on groupings calculated using a k-medoids clustering approach.

Each point in **Figs. 10-14** represents a single conformation extracted from the combined trajectory of that rec domain. Colors correspond to the ideal number of clusters identified using CLARA based on the optimum average silhouette width. Sln1-rec and PhoP-rec possessed the fewest distinct conformational clusters, with the other three proteins having seven or more groups.

Coordination of water to the divalent metal cation is an important indicator of correct RR active site geometry. Due to computational limitations, solvent molecules are typically stripped from trajectories prior to PCA and clustering methods. Using a k-medoids method of clustering allowed us to circumvent these limitations. After identifying medoids using the stripped trajectory data, the corresponding frames were extracted from the fully solvated partial trajectories. Through this process, a list of final candidate structures in explicit solvent was compiled for each of the rec domains that were simulated.

Identification and validation of representative predictions using experimental rec domain structures

Ideally, each candidate (medoid) structure should be analyzed individually for physiological relevance, as all are potentially valid

conformations of phosphorylated rec domains. No single scoring method exists to adequately identify the most representative structure of a phosphorylated rec domain. But to fulfill the ultimate goal of this study and for validation purposes, we developed a set of criteria to identify the top 1-2 candidate structures most likely to resemble the phosphorylated crystal structure of each rec domain. All factors must be examined collectively to make an informed decision. The process (depicted in the flowchart in **Fig. 15**) includes comparisons with the apo conformation to detect characteristic allosteric changes associated with phosphorylation, quantifying and ranking switch residue and quartet residue reorientations, eliminating inappropriate active site arrangements and ranking by conformer cluster population sizes.

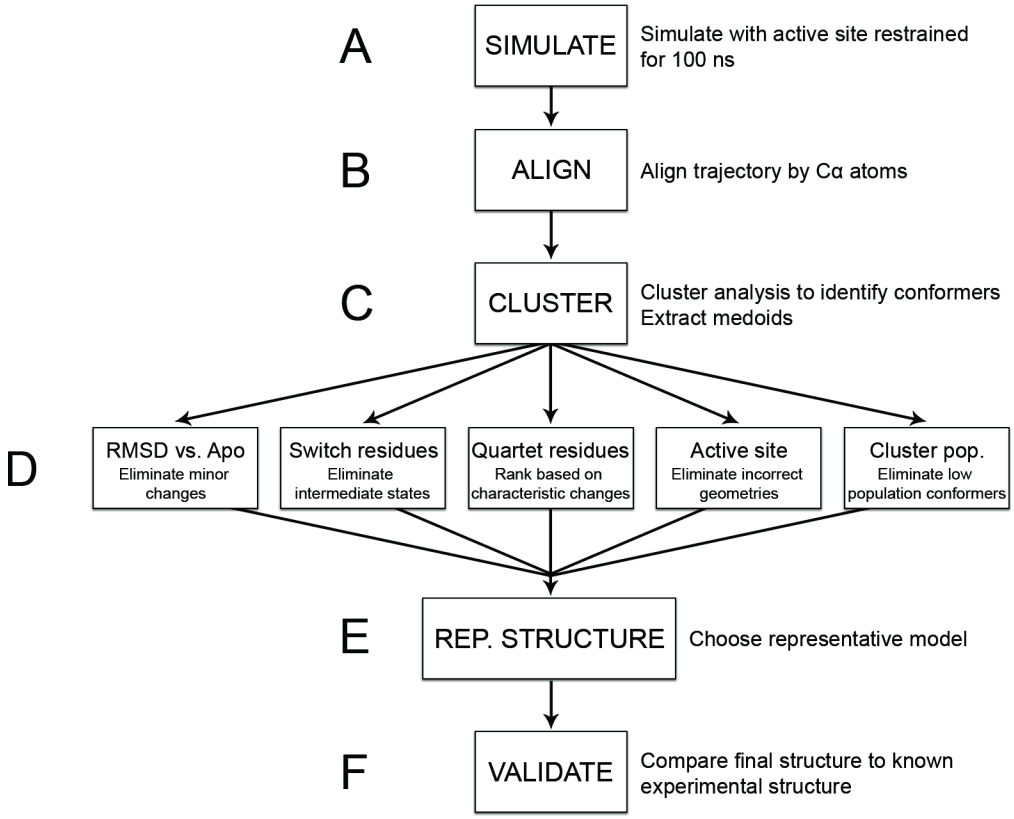

Figure 15. Flowchart describing the methodology.
(A) Initial apo structure is phosphorylated and simulated with active site restraints. (B) Resulting trajectory is aligned to remove translational and rotational fluctuation. (C) Conformers are clustered together based on similarity using k-medoids based algorithm. Candidate structures (medoids) are extracted for each cluster. (D) Each medoid is analyzed to identify candidate most likely to resemble a phosphorylated crystal structure. Candidates are aligned to the apo structure to eliminate medoids that show little to no conformational shift. Switch residue and quartet residue positions are measured. Active site geometry is checked. Medoids from transient clusters are eliminated. (E) Final representative structure is identified. (F) Representative structure is validated against existing crystal structure(s).

Tabulated results and summaries from the protocol are provided for each of the rec domains below (**Tables 5-9**). The process is then described in detail for CheY in the subsequent sections.

Table 5. Results for identification of representative structure for CheY.

Protein or	[1]Pop.	[2]Max RMSD	T87	Y106 χ1	[3]W58-T87	[4]W58-M85	W58 χ2	M85 χ1	[5]E89-W58	Active	[6]PO_3^2 geometry

Medoid		(Å)		(degrees)	(Å)	(Å)	(degrees)	(degrees)	(Å)	Site	(degrees)
Apo	**N/A**	N/A	No	-25	4.2	4.3	78	70	9.1	No	N/A
BeF_3^-	**N/A**	**~5**	**Yes**	**76**	**3.4**	**6.6**	**136**	**-178**	**4.8**	**Yes**	**-168**
1	1575	**~5**	**Yes**	**79**	**3.3**	5.8	79	136	5.3	No	**-167**
2	1139	~3	**Yes**	60	**3.4**	6.1	90	**174**	5.5	No	-158
3	**6468**	**~4**	**Yes**	**78**	3.5	6.3	95	-165	6.9	No	-157
4	**3019**	**~5**	**Yes**	**68**	**3.3**	**6.9**	**109**	**174**	**4.9**	**Yes**	**-174**
5	**5494**	**~4**	**Yes**	65	3.7	**6.5**	**104**	**176**	5.7	No	-144
6	**3187**	**~4**	**Yes**	**78**	**3.2**	**7.0**	90	**-171**	**4.5**	No	-161
7	182	**~4.5**	**Yes**	65	3.5	5.9	93	-162	5.7	No	-152

[1] conformer cluster population; [2] in the β4-α4-β5 region; [3] Cβ-Oε; [4] Cβ-Cβ; [5] Cβ-Cβ; [6] dihedral angle, phosphoanhydride bond

Summary:

Cells are shaded grey when they meet a specific criteria.

Medoids 1, 2 and 7 were placed at a lower priority due to their low conformer cluster populations (Pop.)

Phosphorylation is known to induce significant changes in the β4-α4-β5 region. Structural alignment reveals changes of up to 5 Å within the area in the crystal structures. All medoids exhibited similar changes, except medoid 2, which was placed at a lower priority.

Upon phosphorylation, T87 is known to shift to within 2.5 Å of the phosphoryl group, forming a hydrogen bond. Medoids found with this shift were assigned "Yes" in this column. T87 was found switched in 100% of the trajectory, meaning all medoids were retained. Upon phosphorylation, Y106 rotamerizes inwards, quantified by its χ1 torsion angle changing from -25° (apo) to 76° (BeF_3^-). This corresponds to a fully buried orientation. Medoids 2, 5 and 7 showed lower degrees of rotamerization and were placed at a lower priority, meaning only ~68% of the trajectory showed completely rotamerized orientations.

Phosphorylation consistently causes W58 to shift closer to T87, going from 4.2 Å to 3.4 Å (Cβ-Oε atoms). This is shown in all medoids, though medoids 3, 5 and 7 exhibited smaller shifts and were placed at a lower priority. Simultaneously, W58 shifts away from M85 upon phosphorylation. This is seen in all medoids, though medoids 1, 2, 3 and 7 showed much smaller shifts and were placed at a lower priority. W58 is known to rotamerize upon phosphorylation, measured by the χ2 side chain torsion angle (78° to 136°). This was found to a significant degree in only medoids 4 and 5, representing ~40% of the trajectory. M85 also rotamerizes upon phosphorylation, going from a χ1 torsion angle of 70° to -178°. Rotamerization was observed in all medoids, though medoids 1, 3 and 7 exhibited smaller shifts and were placed at a lower priority. E89 is also known to shift down upon phosphorylation, increasing its contact with W58. This shift was found in all medoids, though strong interactions were only observed in medoids 4 and 6, representing ~29% of the trajectory.

Analysis of the active site geometry in each medoid revealed that all candidates displayed distorted phosphoryl group geometry (quantified by the dihedral angle of the phosphoanhydride bond) and/or K109 arrangement, except medoid 4, representing only ~14% of the trajectory.

Comparing each row, medoid 4 (dark) shows the greatest number of shaded cells and was chosen as the most representative structure for phosphorylated CheY, representing only ~14% of the trajectory.

Table 6. Results for identification of representative structure for PhoP.

Protein or Medoid	[1]Pop.	[2]Max RMSD (Å)	T79	Y98 χ1 (degrees)	[3]L52-T79 (Å)	[4]L52-V77 (Å)	L52 χ2 (degrees)	V77 χ1 (degrees)	[5]R81-L52 (Å)	Active Site	[6]PO_3^{2-} geometry (degrees)
Apo	**N/A**	N/A	No	-59	4.7	6.3	N/A	N/A	7.4	No	N/A
BeF_3^-	**N/A**	**~3.5**	**Yes**	**71**	**3.7**	**6.3**	**N/A**	**N/A**	**5.3**	**Yes**	**-165**
1	**4979**	**~4**	No*	-59	**3.8**	**6.8**	**N/A**	**N/A**	**6.1**	No	-155
2	**2303**	**~3.5**	No*	-63	**3.8**	**6.8**	**N/A**	**N/A**	**5.4**	No	**-163**
3	**6998**	**~3.5**	**Yes**	-72	**3.7**	**6.4**	**N/A**	**N/A**	**6.0**	**Yes**	**-161**
4	**1961**	**~3.5**	**Yes**	-62	**3.9**	**6.9**	**N/A**	**N/A**	**5.6**	**Yes**	-155
5	**4823**	**~3.5**	No*	-76	**4.0**	**6.8**	**N/A**	**N/A**	**6.0**	No	-154

[1] conformer cluster population; [2] in the β4-α4-β5 region; [3] Cβ-Oε; [4] Cβ-Cβ; [5] Cβ-Cβ; [6] dihedral angle, phosphoanhydride bond
*Refers to distance between A80 amine and phosphoryl group, similar to T79

Summary:

Cells are shaded grey when they meet a specific criteria.

All medoids were retained based on conformer cluster population (Pop.)

Phosphorylation is known to induce significant changes in the β4-α4-β5 region. Structural alignment reveals unusually small changes of up to 3.5 Å within the area in the crystal structures. All medoids exhibited similar levels of change and were retained.

Upon phosphorylation, T79 is known to shift to within 2.8 Å of the phosphoryl group, forming a hydrogen bond. Medoids found with this shift were assigned "Yes" in this column. This was observed in all medoids. However, the closely related A80 was found to be positioned farther away than normal in medoids 1, 2 and 5 (3.1-3.4 Å). These were placed at lower priority. Upon phosphorylation, Y98 rotamerizes inwards, quantified by its χ1 torsion angle changing from -59° (apo) to 71° (BeF_3^-). This corresponds to a fully buried orientation. No medoids were found in a fully buried orientation. A possible reason for this is addressed in the corresponding Results subsection within the text.

Phosphorylation consistently causes L52 to shift closer to T79, going from 4.7 Å to 3.7 Å (Cβ-Oε atoms), similar to the analogous residues in CheY (W58-T87). This shift is found in all medoids, meaning 100% of the trajectory. Normally, this would suggest that L52 should also shift away from V77 (analogous to M85 in CheY).

Though this is not observed in the crystal structures (where the distance is maintained at 6.3 Å), every medoid showed a slight increase in this distance. L52 and V77 were not examined for rotamerization. However, R81 (analogous to E89 in CheY) should shift to increase contact with L52. This is seen in the crystal structures (going from 7.4 Å to 5.3 Å upon modification) as well as in every medoid structure.

Analysis of the active site geometry in each medoid revealed that only medoids 3 and 4 formed proper catalytic geometry. Medoids 2 and 5 were unable to form a complete active site due to A80. Medoid 1 had a weaker orientation of the conserved K101 residue. Medoid 4 introduced significant distortion to the phosphoryl group geometry.

Comparing each row, medoid 3 (dark) shows the greatest number of shaded cells and was chosen as the most representative structure for phosphorylated PhoP-rec, representing ~33% of the trajectory.

Table 7. Results for identification of representative structure for FixJ.

Protein or Medoid	[1]Pop.	[2]Max RMSD (Å)	T82	F101 χ1 (degrees)	[3]L55-T82 (Å)	[4]L55-V80 (Å)	L55 χ2 (degrees)	V80 χ1 (degrees)	[5]H84-L55 (Å)	Active Site	[6]PO_3^{2-} geometry (degrees)
Apo	**N/A**	N/A	No	-41	6.6	5.3	-158	N/A	11.5	No	N/A
BeF_3^-	**N/A**	**~7**	**Yes**	**70**	**3.3**	**5.5**	**N/A**	**N/A**	**6.5**	**Yes**	**-177**
1	749	~1.5	No	-58	5.8	**6.0**	**N/A**	**N/A**	11.2	No	**-138**
2	**3551**	~3.5	**Yes**	**68**	**3.3**	5.8	**N/A**	**N/A**	**10.5**	**Yes**	-120
3	**2052**	**~4.5**	No	61	**3.6**	5.4	**N/A**	**N/A**	**10.1**	No	**-153**
4	**3084**	**~4**	**Yes**	62	**3.1**	**6.2**	**N/A**	**N/A**	**10.3**	**Yes**	**-140**
5	**2495**	**~4**	**Yes**	**72**	**3.2**	5.9	**N/A**	**N/A**	**10.6**	No	**-130**
6	**1385**	**~6.5**	**Yes**	62	3.9	**6.6**	**N/A**	**N/A**	**7.6**	**Yes**	**-131**
7	**3870**	**~4.5**	**Yes**	63	**3.5**	5.8	**N/A**	**N/A**	11.0	**Yes**	-113
8	**1311**	**~4**	**Yes**	**74**	**3.4**	**6.2**	**N/A**	**N/A**	**10.7**	**Yes**	**-123**
9	904	~3.5	**Yes**	55	**3.3**	**6.2**	**N/A**	**N/A**	**9.5**	No	**-132**
10	**1663**	**~4**	No	58	4.2	**7.4**	**N/A**	**N/A**	11.0	No	**-148**

[1] conformer cluster population; [2] in the β4-α4-β5 region; [3] Cβ-Oε; [4] Cβ-Cβ; [5] Cβ-Cβ; [6] dihedral angle, phosphoanhydride bond

Summary:

Cells are shaded grey when they meet a specific criteria.

Because of the high amount of conformational variance observed for FixJ-rec, along with conflicting criteria results, the population cutoff was lowered from 1500 to 1000. Medoids 1 and 9 were placed at a lower priority due to their low conformer cluster populations (Pop.)

Phosphorylation is known to induce significant changes in the β4-α4-β5 region. Structural alignment reveals changes of up to 7 Å within the area in the crystal

structures. Medoids 1, 2 and 9 exhibited unusually small deviations in this region (< 3.5 Å) and were placed at a lower priority.

Upon phosphorylation, T82 is known to shift to within 2.7 Å of the phosphoryl group, forming a hydrogen bond. Medoids found with this shift were assigned "Yes" in this column. T82 was found switched in ~79% of the trajectory. Medoids 1, 3 and 10 were placed at lower priority due to lack of T82 switching. Upon phosphorylation, F101 rotamerizes inwards, quantified by its χ1 torsion angle changing from -41° (apo) to 70° (BeF_3^-). This corresponds to a fully buried orientation. Only medoids 2, 5 and 7 showed fully buried side chain orientaitons for F101, representing ~35% of the trajectory.

Phosphorylation consistently causes L55 to shift closer to T82, going from 6.6 Å to 3.3 Å (Cβ-Oε atoms), similar to the analogous residues in CheY (W58-T87). This shift is found in all medoids except 1, and is weaker in medoids 6 and 10. These were placed at a lower priority. Normally, this would suggest that L55 should also shift away from V80 (analogous to M85 in CheY). A slight increase is observed in the crystal structures (5.3 to 5.5 Å), though larger changes were exhibited in every medoid. To simulate a predictive situation, medoids with little to no increases were placed at a lower priority, mimicking CheY. L55 and V80 were not examined for rotamerization. However, H84 (analogous to E89 in CheY) should shift to increase contact with L55. This is seen in the crystal structures (going from 11.5 Å to 6.5 Å upon modification). However, H84 is flanked on by glycine residues and exhibits extremely high flexibility during simulation. It is likely that this relationship is less homologous to the CheY E89-W58 shift. Every medoid showed a decrease in this distance, though only medoid 6 showed a large change. Medoid 6 was also eliminated based on its partially exposed aromatic switch residue orientation. Because of this, any medoid exhibiting an H84-L55 distance of < 11 Å was retained.

Analysis of the active site geometry in each medoid revealed that only medoids 4, 6, 7 and 8 formed a proper catalytic geometry. The other medoids (representing ~37% of the trajectory) do not resemble the apo conformation, but their side chain arrangement is non-optimal. Significantly more phosphoryl group geometry distortion was observed for FixJ-rec, also seen in the experimental crystal structures. However, medoids 2 and 7 exhibited reverse orientation of the phosphorylatable aspartate residue and were placed at a lower priority.

Comparing each row, medoid 8 (dark) shows the greatest number of shaded cells and was chosen as the most representative structure for phosphorylated FixJ-rec, representing an unusually low ~6% of the trajectory.

Table 8. Results for identification of representative structure for Sln1-rec.

Protein or Medoid	[1]Pop	[2]Max RMSD (Å)	T1173	F1192 χ1 (degrees)	[3]V1145-T1173 (Å)	[4]V1145-A1171 (Å)	V1145 χ2 (degrees)	A1171 χ1 (degrees)	[5]F1175-V1145 (Å)	Active Site	[6]PO_3^{2-} geometry (degrees)
Apo	**N/A**	N/A	No	-41	4.4	5.2	N/A	N/A	7.5	No	N/A
BeF_3^-	**N/A**	**~3**	**Yes**	**63**	**3.9**	**6.4**	**N/A**	**N/A**	**5.8**	**Yes**	**-174**
1	**6307**	**~3**	**Yes**	47	**3.8**	5.4	**N/A**	**N/A**	6.6	No	**-151**
2	**5482**	~2.5	**Yes**	53	**3.9**	5.9	**N/A**	**N/A**	6.3	**Yes**	**-153**
3	**5246**	**~3**	**Yes**	**64**	**3.9**	**6.4**	**N/A**	**N/A**	8.6	**Yes**	**-152**

4	**2533**	**~3**	**Yes**	**60**	**3.7**	**6.3**	**N/A**	**N/A**	6.6	No	**-174**
5	1496	**~4**	**Yes**	38	**3.8**	5.9	**N/A**	**N/A**	6.7	No	**-161**

[1] conformer cluster population; [2] in the β4-α4-β5 region; [3] Cβ-Oε; [4] Cβ-Cβ; [5] Cβ-Cβ; [6] dihedral angle, phosphoanhydride bond

Summary:

Cells are shaded grey when they meet a specific criteria.

Medoid 5 was placed at a lower priority due to its smaller cluster population (Pop.)

Phosphorylation is known to induce significant changes in the β4-α4-β5 region. Structural alignment reveals unusually small changes of up to ~3 Å within the area in the crystal structures. All medoids exhibited similar levels of change and were retained, except medoid 2.

Upon phosphorylation, T1173 is known to shift to within 2.6 Å of the phosphoryl group, forming a hydrogen bond. Medoids found with this shift were assigned "Yes" in this column. This was observed in all medoids. Upon phosphorylation, F1192 rotamerizes inwards, quantified by its χ1 torsion angle changing from -41° (apo) to 63° (BeF_3^-). This corresponds to a fully buried orientation. Medoids 3 and 4 also exhibited this fully buried orientation. Medoids 1, 2 and 5 showed only partial rotamerization.

Phosphorylation consistently causes V1145 to shift closer to T1173, going from 4.4 Å to 3.9 Å (Cβ-Oε atoms), similar to the analogous residues in CheY (W58-T87). This shift is found in all medoids, meaning 100% of the trajectory. This suggests that V1145 should also shift away from A1171 (analogous to M85 in CheY). This is seen in the crystal structures, going from 5.2 Å to 6.4 Å. An increase is observed in every medoid, though medoids 1, 2 and 5 showed much smaller shifts and were placed at a lower priority. V1145 and A1171 were not examined for rotamerization. F1175 (analogous to E89 in CheY) was expected to shift towards V1145, as seen in the crystal structures. Minor shifts were seen in all medoids except medoid 3. However, since medoid 3 was the only candidate that met the other criteria, it was retained.

Analysis of the active site geometry in each medoid revealed that only medoids 2 and 3 formed a proper catalytic geometry. In medoids 1, 4 and 5, residues E1094 and D1095 were unable to coordinate the divalent Mg^{2+}.

Comparing each row, medoid 3 (dark) shows the greatest number of shaded cells and was chosen as the most representative structure for phosphorylated Sln1-rec, representing ~25% of the trajectory.

Table 9. Results for identification of representative structure for Spo0F.

Protein or Medoid	[1]Pop.	[2]Max RMSD (Å)	T82	H101 χ1 (degrees)	[3]M55-T82 (Å)	[4]M55-I80 (Å)	M55 χ2 (degrees)	I80 χ1 (degrees)	[5]E89-M55 (Å)	Active Site	[6]PO_3^{2-} geometry (degrees)

Apo	N/A	N/A	No	-39	4.5	5.6	-54	-67	7.3	No	N/A
BeF_3^-	**N/A**	**~3.5**	**Yes**	**-61**	**3.2**	**6.0**	**-49**	**169**	**5.8**	**Yes**	**-143**
1	1180	~2.5	No	**-48**	4.2	6.0	**-71**	-73	7.8	No	**178**
2	**1831**	**~3.5**	**Yes**	**-65**	**3.8**	**7.0**	**-81**	**191**	**6.7**	**Yes**	**-160**
3	1434	~2.5	**Yes**	53	4.0	6.8	**-67**	**195**	**6.7**	**Yes**	-159
4	**1740**	**~3.5**	**Yes**	43	4.1	**7.1**	**-77**	-84	7.0	**Yes**	-158
5	**4567**	**~4**	**Yes**	43	**3.8**	6.7	**-59**	-79	8.1	**Yes**	**-179**
6	**3228**	**~4**	**Yes**	33	4.1	6.6	**-79**	-67	**6.4**	No	**-175**
7	1135	**~9**	**Yes**	38	4.2	6.3	-176	**182**	9.0	**Yes**	**-173**
8	**2009**	~3	**Yes**	42	**3.7**	6.9	**-71**	-81	8.0	**Yes**	**-166**
9	**2000**	**~4**	**Yes**	49	4.0	**7.3**	**-80**	-77	7.0	**Yes**	**-170**
10	1053	~2.5	**Yes**	51	3.9	6.7	**-56**	**192**	**6.2**	**Yes**	**173**
11	511	**~5**	**Yes**	55	**3.7**	**7.1**	**-73**	-71	**6.4**	**Yes**	**179**
12	376	**~4**	**Yes**	59	4.2	6.4	**-67**	-67	7.5	**Yes**	**-179**

[1] conformer cluster population; [2] in the β4-α4-β5 region; [3] Cβ-Oε; [4] Cβ-Cβ; [5] Cβ-Cβ; [6] dihedral angle, phosphoanhydride bond

Summary:

Cells are shaded grey when they meet a specific criteria.

Medoids 1, 3, 7, 10, 11 and 12 were placed at a lower priority due to their smaller cluster populations (Pop.)

Phosphorylation is known to induce significant changes in the β4-α4-β5 region. Structural alignment reveals changes of up to ~4 Å within the area in the crystal structures. Medoids 4, 5, 6, 7, 9, 11 and 12 exhibited changes from 4-9 Å and were retained. Medoids 1, 3, 8 and 10 were placed at a lower priority because of their smaller deviations.

Upon phosphorylation, T82 is known to shift to within 2.6 Å of the phosphoryl group, forming a hydrogen bond. Medoids found with this shift were assigned "Yes" in this column. This was observed in all candidates except medoid 1. Upon phosphorylation, H101 was found to remain fully solvent exposed, closely resembling the apo orientation (-39° and -61°). This was found only in medoids 1 and 2. Medoids 3-12 exhibited partial rotamerization, adopting intermediate orientations and representing ~86% of the trajectory. These medoids were placed at a lower priority.

Phosphorylation consistently causes M55 to shift closer to T82, going from 4.5 Å to 3.2 Å (Cβ-Oε atoms), similar to the analogous residues in CheY (W58-T87). Significant shifts between these residues were only observed in medoids 2, 5, 8 and 11. This suggests that M55 should also shift away from I80 (analogous to M85 in CheY). This is seen in the crystal structures, going from 5.6 Å to 6.0 Å. An increase is observed in every medoid, though medoids with the largest changes were placed at a higher priority. M55 was also examined for rotamerization (analogous to W58 in CheY). In the crystal structures, no significant rotamerization was observed, unlike in CheY. Only medoid 7 showed a significant rotamerization and was placed at a lower priority. I80 rotamerization was also examined. In the crystal structures, we see a

change in the χ1 torsion angle from -67° (apo) to 169° (BeF_3^-). Similar significant rotamerizations are only found in medoids 2, 3, 7 and 10. E86 (likely analogous to E89 in CheY) was expected to shift towards M55. This is observed in the crystal structures, shifting from 7.3 Å (apo) to 5.8 Å (BeF_3^-). Medoids 2, 3, 6, 10 and 11 showed similar significant shifts. Medoids 1, 4, 5, 7, 8, 9 and 12 were placed at a lower priority because they lacked this change.

Analysis of the active site geometry in each medoid revealed that all formed correct active site geometry arrangements except medoids 1 and 6, which showed distortion of the K104 residue.

Comparing each row, medoid 2 (dark) shows the greatest number of shaded cells and was chosen as the most representative structure for phosphorylated Spo0F, representing only ~9% of the trajectory.

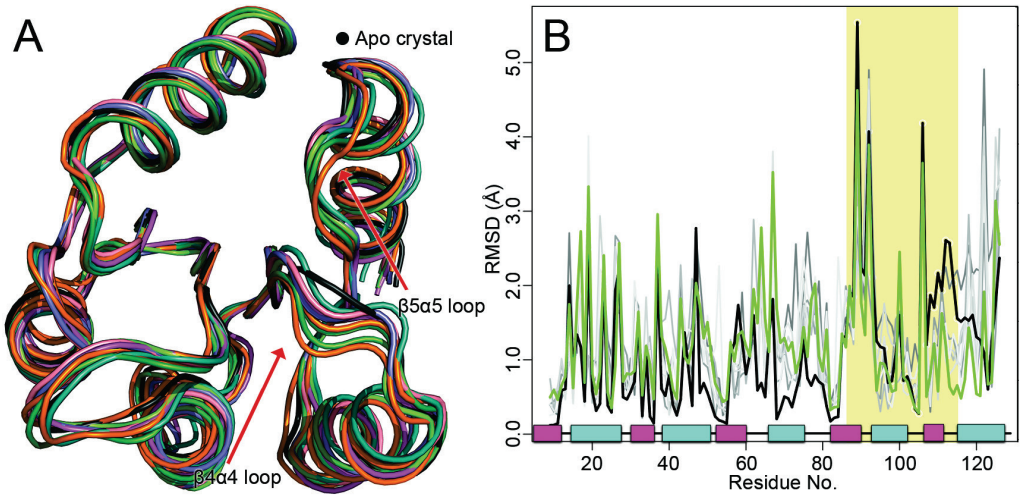

Figure 16. Alignments of phosphorylated candidates to the apo structure of CheY.

(A) Superposition of apo crystal structure (PDB 3CHY, black) with candidate (medoid) structures identified during cluster analysis. (B) Per residue, all-atom RMSD for phosphorylated structures vs. the apo crystal structure. Shaded area corresponds to β4-α4-β5 regions. Trace color corresponds to chosen structures in alignment: example candidate structure vs. apo crystal (green); BeF_3^- crystal vs. apo crystal (black). Traces for additional candidate structures appear in gray. α-helix (cyan); β-strand (magenta).

Structural alignments were performed with each set of candidate structures and their corresponding apo crystal structure. **Fig. 16A** shows a superposition of all seven candidate structures extracted from the CheY

trajectories aligned to the apo CheY structure (PDB 3CHY; black ribbon) using the invariant β-sheet core [10]. Most of the conformational variance we observed between clusters was small (~1.5 Å) and unsurprisingly located in the flexible loop regions. Qualitatively, this can be seen in the CheY alignment in **Fig. 16A,** where the majority of the conformers show little deviation, especially throughout the helix and strand elements. Other rec domains (such as Spo0F and FixJ-rec) exhibited far more diversity among the medoid conformers. The largest changes induced by phosphorylation occur in the β4-α4-β5 region of CheY. This area has consistently shown high deviations in nearly all modified rec structures [11, 23, 31-33]. The shift from apo to phosphorylated state is quantified in **Fig. 16B** for CheY, which shows the per residue all-atom RMSD after alignment between the **apo** (PDB 3CHY) and **BeF_3^-** bound crystal (PDB 1FQW) in **black trace**. The **grey traces** in **Fig. 16B** shows the same for each **medoid** of CheY aligned to the **apo** crystal. The **colored trace** shows the RMSD between the apo crystal structure and a sample medoid structure, demonstrating a significant conformational shift characteristic of phosphorylation. By examining the deviations in the β4-α4-β5 region (**Fig. 16B, shaded area**), we were able to eliminate any candidate structures that showed unusually low relative conformational changes associated with activation, indicating a closer resemblance to the apo conformation. For CheY, typical max deviations observed among the medoids were ~4.5-5 Å within this area (see **Table 5**). A single medoid was found

(representing ~5% of the trajectory) with significantly smaller changes (~2.5-3 Å) . This medoid was assigned lower priority based on these RMSD calculations. The remaining candidates displayed more characteristic changes when aligned to the apo structure, and comparison between the colored and black traces suggest a phosphorylation-induced conformational shift on a similar magnitude. Similar analyses were performed on the other protein systems (**Figs. 17-20**).

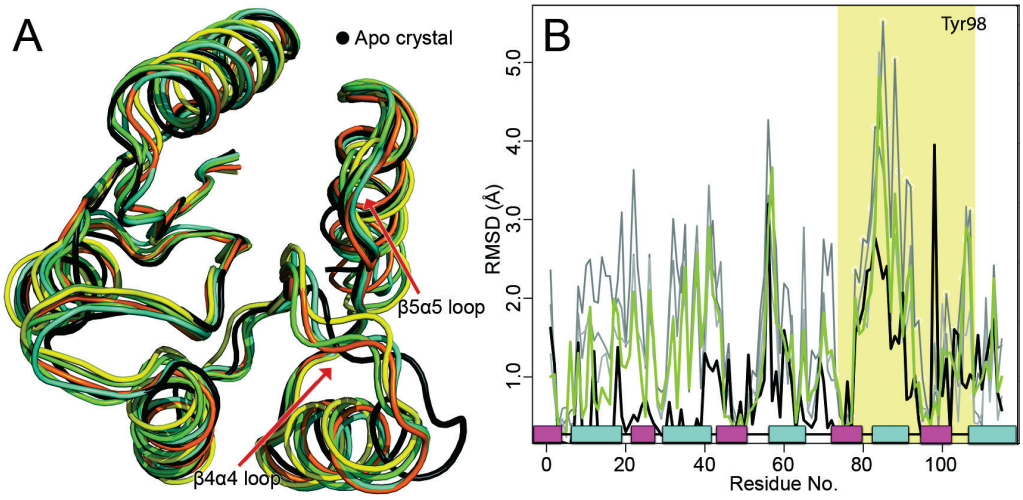

Figure 17. Alignments of phosphorylated candidates to the apo structure of PhoP-rec.

(A) Superposition of apo crystal structure (PDB 2PKX:A; black) with candidate (medoid) structures identified during cluster analysis. (B) Per residue all-atom RMSD for phosphorylated structures vs. the apo crystal structure. Shaded area corresponds to β4-α4-β5 regions. Trace color corresponds to chosen structures in alignment: example candidate structure vs. apo crystal (lime); BeF_3^- crystal vs. apo crystal (black). Traces for additional candidate structures appear in gray. α-helix (cyan); β-strand (magenta).

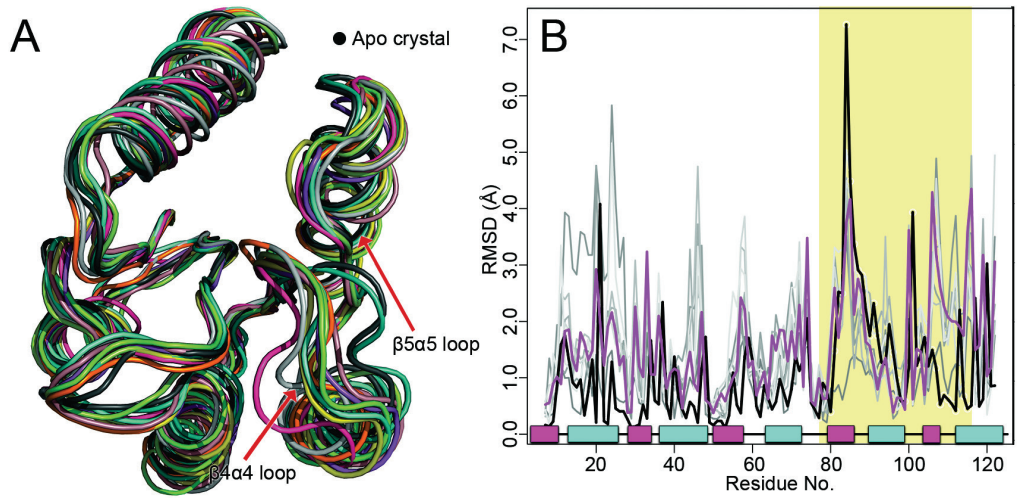

Figure 18. Alignments of phosphorylated candidates to the apo structure of FixJ-rec.

(A) Superposition of apo crystal structure (PDB 1DCK:A; black) with candidate (medoid) structures identified through cluster analysis. (B) Per residue all-atom RMSD for phosphorylated structures vs. the apo crystal structure. Shaded area corresponds to β4-α4-β5regions. Trace color corresponds to chosen structures in alignment: example candidate structure vs. apo crystal (purple); phosphorylated crystal vs. apo crystal (black). Traces for additional candidate structures appear in gray. α-helix (cyan); β-strand (magenta).

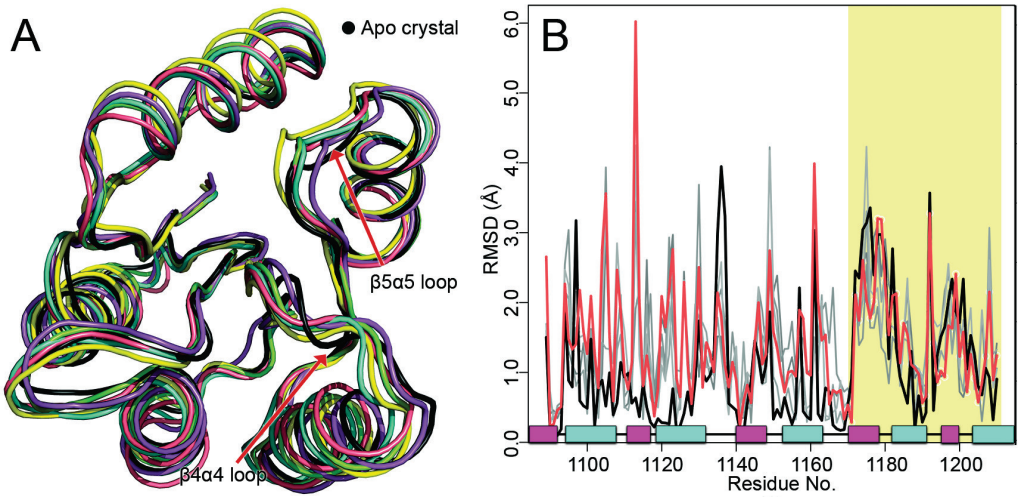

Figure 19. Alignments of phosphorylated candidates to the apo structure of Sln1-rec.

(A) Superposition of apo crystal structure (PDB 1OXB:B; black) with candidate (medoid) structures identified through cluster analysis. (B) Per residue all-atom RMSD for phosphorylated structures vs. the apo crystal structure. Shaded area corresponds to β4-α4-β5regions. Trace color corresponds to chosen structures in alignment: example candidate structure vs. apo crystal (salmon); BeF_3^- crystal vs. apo crystal (black). Traces for additional candidate structures appear in gray. α-helix (cyan); β-strand (magenta).

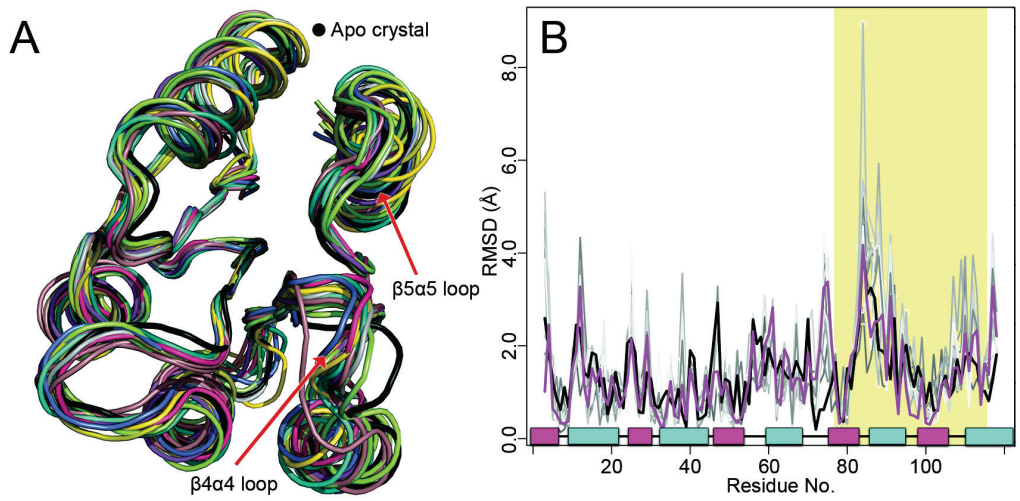

Figure 20. Alignments of phosphorylated candidates to the apo structure of Spo0F.

(A) Superposition of apo crystal structure (PDB 1NAT; black) with candidate (medoid) structures identified through cluster analysis. (B) Per residue all-atom RMSD for phosphorylated structures vs. the apo crystal structure. Shaded area corresponds to β4-α4-β5regions. Trace color corresponds to chosen structures in alignment: example candidate structure vs. apo crystal (purple); BeF_3^- crystal vs. apo crystal (black). Traces for additional candidate structures appear in gray. α-helix (cyan); β-strand (magenta).

Switch residue orientation was also examined in each of the extracted candidate structures. **Fig. 21** shows a close-up comparison of the switch residues found in CheY. In **Fig. 21A**, the apo crystal structure of CheY shows the unmodified positions of the switch residues. The side chain of Tyr106 adopts a torsion angle (χ_1) of -25^0 and Thr87 is positioned 6.1 Å away from the phosphorylatable Asp57. **Fig. 21B** shows how Thr87 shifts to within 2.5 Å of Asp57, and Tyr109 adopts a torsion angle (χ_1) of 76^0 in the BeF_3^- bound crystal structure. **Fig. 21C** demonstrates the same "switched" orientations in a sample medoid structure predicted for CheY. A similar shift of Thr87 was

observed in 100% of frames for CheY, indicating an almost complete and instantaneous switch upon adoption of the catalytic active site geometry. However, medoids representing ~32% of frames exhibited a smaller rotamerization of Tyr109 (torsion angle χ_1 of ~60^0). These medoids were assigned lower priority. Medoids representing the other ~68% of the trajectory adopted rotamers corresponding to more buried orientations characteristic of phosphorylated conformations (torsion angle χ_1 of ~75-80^0). In CheY these differences were subtle, but in other rec domains (Sln1-rec, Spo0F) the switch residues showed far greater deviations. Candidates that did not exhibit proper switch residue reorientations were eliminated as possible representative structures (see **Tables 5-9** for detailed measurements). Figures of switch residues for the remaining rec domains are shown below (**Figs. 22-24**).

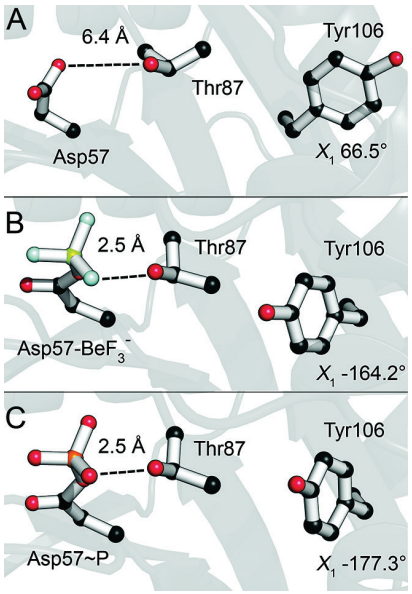

Figure 21. Characteristic switch residues shift in CheY.

(A) Apo orientation (PDB 3CHY). (B) BeF_3^- bound orientation (PDB 1FQW). Thr87 has shifted to form a hydrogen bond with Asp57. Tyr106 has rotamerized inwards into the hydrophobic pocket. (C) A sample candidate structure of phosphorylated CheY demonstrates near-identical orientations to the BeF_3^- bound crystal.

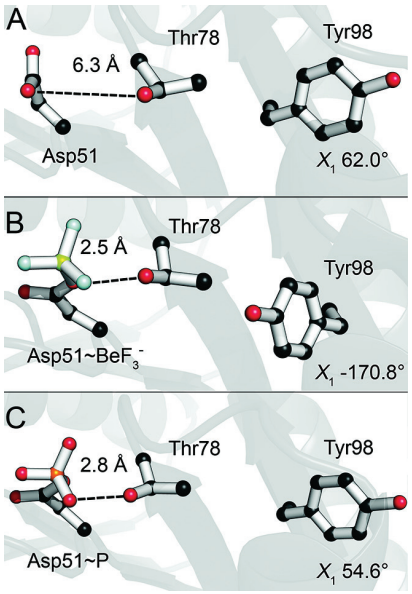

Figure 22. Characteristic switch residues shift in PhoP-rec.

(A) Apo orientation (PDB 2PKX:A). (B) Phosphorylated orientation (PDB 2PL1). Thr78 has shifted to form a hydrogen bond with Asp51. Tyr98 has rotamerized inwards into the hydrophobic pocket. (C) A sample candidate structure of phosphorylated PhoP-rec shows Thr78 orientation is identical to the BeF_3^- bound crystal. Tyr98 remains in the solvent-exposed orientation. Examination of the BeF_3^- bound structure of PhoP-rec reveals that the receiver domain crystallizes as a homodimer. Tyr98 is located in the central region of the dimerization inteface, suggesting that its side chain orientation plays an important role in dimerization and vice versa.

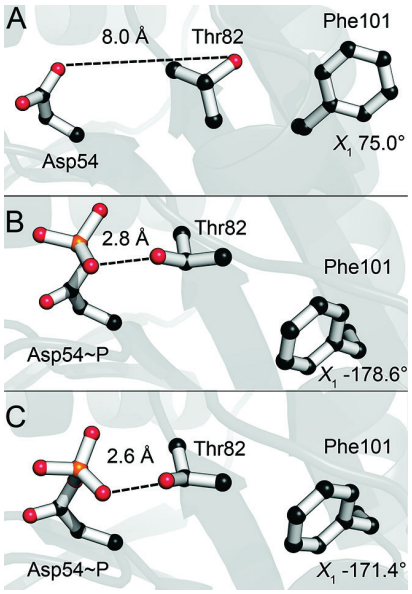

Figure 23. Characteristic switch residues shift in FixJ-rec.

(A) Apo orientation (PDB 1DCK:A). (B) Phosphorylated orientation (PDB 1D5W:A). Thr82 has shifted to form a hydrogen bond with Asp54. Phe101 has rotamerized inwards into the hydrophobic pocket. (C) A sample candidate structure of phosphorylated FixJ-rec demonstrates near-identical orientations to the phosphorylated crystal.

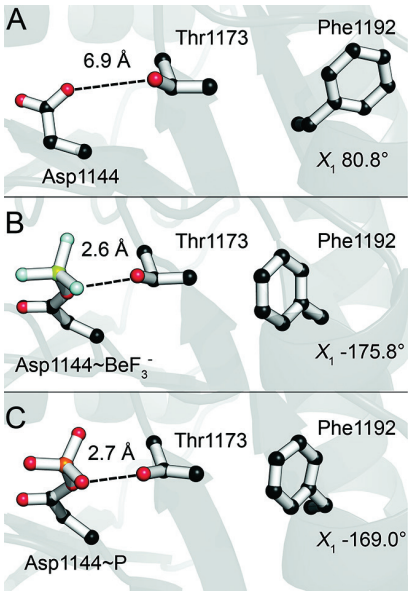

Figure 24. Characteristic switch residues shift in Sln1-rec.

(A) Apo orientation (PDB 1OXB:B). (B) Phosphorylated orientation (PDB 2R25:B). Thr1173 has shifted to form a hydrogen bond with Asp1144. Phe1192 has rotamerized inwards into the

hydrophobic pocket. (C) A sample candidate structure of phosphorylated Sln1-rec demonstrates near-identical orientations to the BeF_3^- bound crystal.

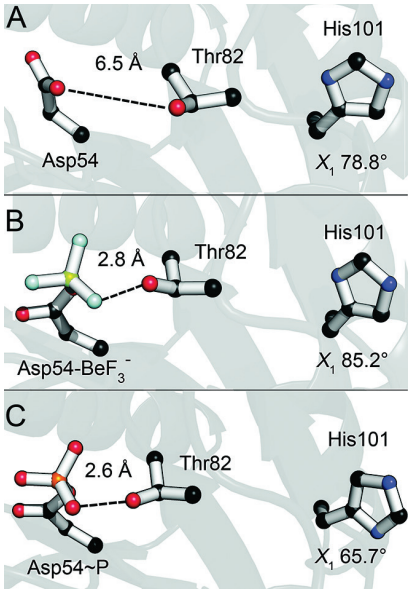

Figure 25. Characteristic switch residues shift in Spo0F.

(A) Apo orientation (PDB 1NAT:A). Thr82 is too far to form a hydrogen bond with the phosphorylatable aspartate, and His101 is solvent exposed, pointing away from the active site. (B) Phosphorylated orientation (PDB 2FTK:G). Thr82 has shifted to form a hydrogen bond with Asp54. His101 remains solvent-exposed. (C) A sample candidate structure of phosphorylated Spo0F demonstrates near-identical orientations to the BeF_3^- bound crystal. His101 remains solvent-exposed.

Each of the five rec domains in the validation set produced at least one candidate structure with switch residue orientations matching the modified crystal structures, except for PhoP-rec, where Tyr98 remained completely in a solvent-exposed orientation in 100% of the trajectory. Further examination of the PhoP-rec structure reveals that the rec domain crystallized as a homodimer, with Tyr98 located in the center of the dimerization interface [30]. Previous studies have implicated the aromatic switch residue in certain dimeric RR interactions [34, 35]. Comparison of the modified crystal structure and the

predicted model reveals similar environments around the Tyr98 side chain, with adequate volume to allow rotamerization in both. In *E. coli* PhoP, the pocket created by the shift in the β4α4 loop is less hydrophobic in nature than the other rec domains in the validation set, likely making rotamerization less favorable. We speculate that by using only monomeric PhoP-rec during the simulations, the rotamerization was not inducible. Also of significant note is the aromatic switch residue His101 in Spo0F. In the crystal structures of Spo0F, the side chain of His101 adopts multiple conformations, but remains solvent-exposed regardless of phosphorylation state [32]. This residue and region have no known role in dimerization or partner binding for Spo0F, which may be why it fails to rotamerize [15, 32, 115, 116]. Most of the extracted Spo0F candidate structures exhibited various degrees of rotamerization at His101. The majority adopted a partially buried orientation (medoids representing ~91% of the population), likely resembling an intermediate conformation. A single medoid (~9%) was identified that remained fully solvent-exposed. No candidates were observed to adopt a completely buried orientation. This supports the existence of an intermediate state(s) in solution, though medoids with partially exposed rotamers were eliminated as possible representative structures. More importantly, it indicates that this computational approach can capture unique allosteric changes associated with phosphorylation that may differ between rec domain proteins.

Quartet residue positions were also analyzed within the medoid structures. These residues have been hypothesized to form an allosteric network, able to switch between apo and phosphorylated-like conformational states [16]. Comparisons rely on the relative changes in interatomic distances between residues upon transitioning from apo to phosphorylated conformation. While most quartet residues share moderate to low sequence conservation, we found that their relationships are relatively maintained in the rec domains used in the study and likely in most rec domains. In the CheY candidates (**Table 5**), we examined the distance between Trp58 (D+1 position) and the switch Thr87, which are consistently found in closer proximity in modified crystal structures. All medoids showed a decrease in the atomic distance between these side chains, though only the most dramatic shifts were considered for potential representative structures. We also examined the distance between Trp58 and Met85, which is believed to increase upon phosphorylation. Again, all medoids showed an increase in relative distance upon phosphorylation, but higher priority was placed on the medoids that exhibited the largest shifts. Interaction between Trp58 and Glu89 is also thought to increase as a result of phosphorylation. Each medoid showed various interatomic distances, but only two exhibited strong interaction between these residues, representing ~29% of the trajectory. Also analyzed were the rotamerization of Trp58, related to the interaction with Glu89, and the rotamerization of Met85. Only two medoids (representing ~40%) showed sufficiently large rotamerization at Trp58, going

from 75^0 to ~110^0 upon phosphorylation. The χ_1 torsion angle of Met85 shifts from 70^0 (apo crystal) to -178^0 (BeF_3^- crystal). A similar rotamerization is observed in four different CheY medoids (representing ~61% of the trajectory). Interestingly, the complete "phosphorylated" organization of quartet residues is observed in only two candidates (representing ~31% of the trajectory). Analysis of the final quartet residue, Tyr109, can be found in the previous section. By ranking the changes in interatomic distance between the quartet residues in relation to the apo state, we were able to narrow down which medoids were more likely to resemble the phosphorylated crystal structures for each rec domain. Tabulated results for the other systems (M55, I80, E89 in Spo0F; V1145, A1171, F1175 in Sln1-rec; L55, V80, H84 in FixJ-rec; L52, V77, R81 in PhoP-rec) can be found in **Tables 6-9**.

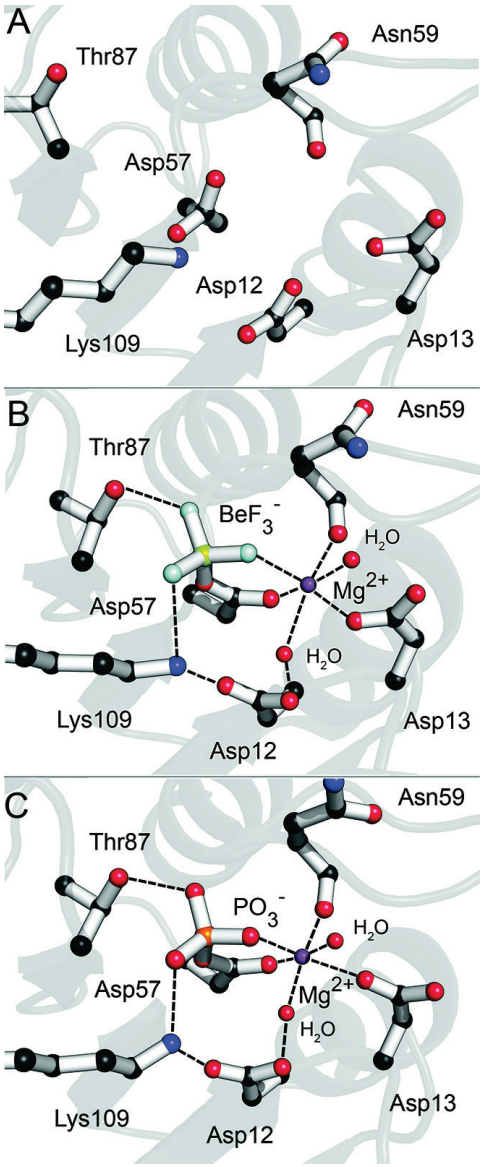

Figure 26. Highly conserved active site geometry of active CheY.

(A) Disordered active site side chains of apo CheY (PDB 3CHY). (B) Active site side chains of CheY bound to BeF_3^- (PDB 1FQW). The addition of Mg^{2+} orders the active site, generating an octahedral coordination geometry with the recruitment of two additional water molecules. Switch residue Thr87 has shifted to hydrogen bond with Asp57. (C) Active site side chains in a sample candidate structure of phosphorylated CheY demonstrates near-identical geometry to the BeF_3^- bound crystal.

We next examined the phosphorylated active sites of each candidate structure. Proper active site geometry is essential for the phosphotransfer

reaction to occur in all rec domains, and a divalent metal cation is required for catalysis. **Fig. 26** shows this geometry for CheY. In **Fig. 26A**, the active site lacks the metal cation, leading to a disordered side chain arrangement. With the addition of a Mg^{2+} cation and the phosphoryl analog BeF_3^-, the active site side chains arrange in an octahedral configuration centered around the metal ion (**Fig. 26B**). This is also demonstrated in a sample medoid for CheY (**Fig. 26C**). Despite the restraints, we frequently observed incomplete or incorrect active sites in the extracted medoids. If a candidate structure failed to fully adopt the conserved active site arrangement, we eliminated it as a possible representative structure. One of the most frequently seen deviations was a failure to recruit the two water molecules needed to complete the octahedral coordination around the cation. This was usually due to one of the conserved acidic residues forming multiple points of coordination between their side chains and the metal, rather than a single point of coordination from each. Those candidate structures with a warped phosphoryl group geometry were placed assigned lower priority as well. Among the CheY candidates, a single medoid structure adopted the ideal catalytic active site, representing ~14% of the trajectory. This medoid also had the closest to ideal linear phosphoryl geometry as well (measured by dihedral angle of the phosphoaspartate bond). The remaining candidates failed to properly coordinate the metal cation and/or the phosphoryl group. Active site results for the other rec domains used in this study were analyzed and tabulated (see **Tables 6-9**; **Figs. 27-30**).

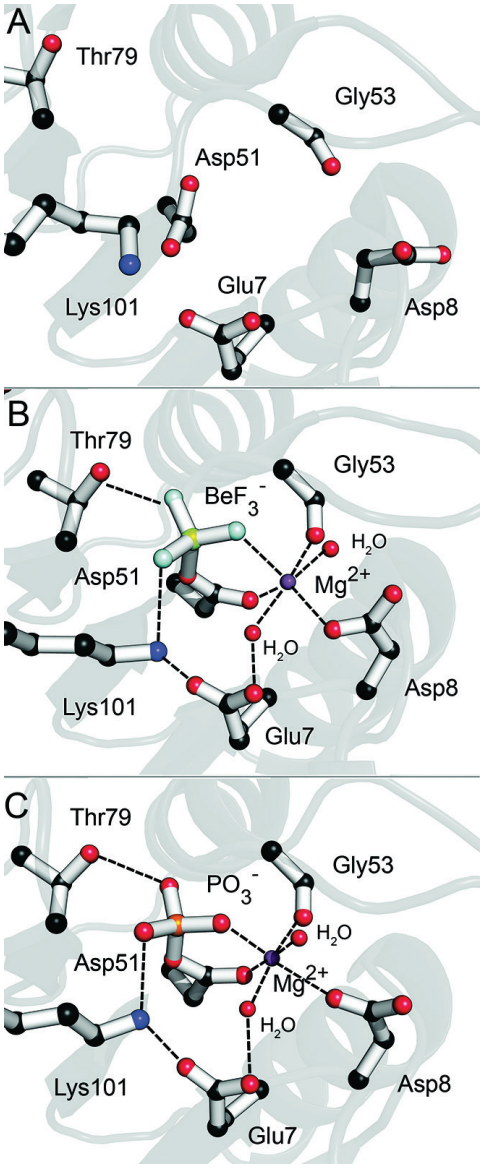

Figure 27. Highly conserved active site geometry of PhoP-rec.

(A) Disordered active site side chains of apo PhoP-rec (PDB 2PKX:A). (B) Active site side chains of PhoP-rec bound to BeF_3^- (PDB 2PL1). The addition of Mg^{2+} orders the active site, generating an octahedral coordination geometry with the recruitment of two additional water molecules. Switch residue Thr79 has shifted to hydrogen bond with Asp51. (C) Active site side chains in a sample candidate structure of phosphorylated PhoP-rec demonstrates near-identical geometry to the BeF_3^- bound crystal

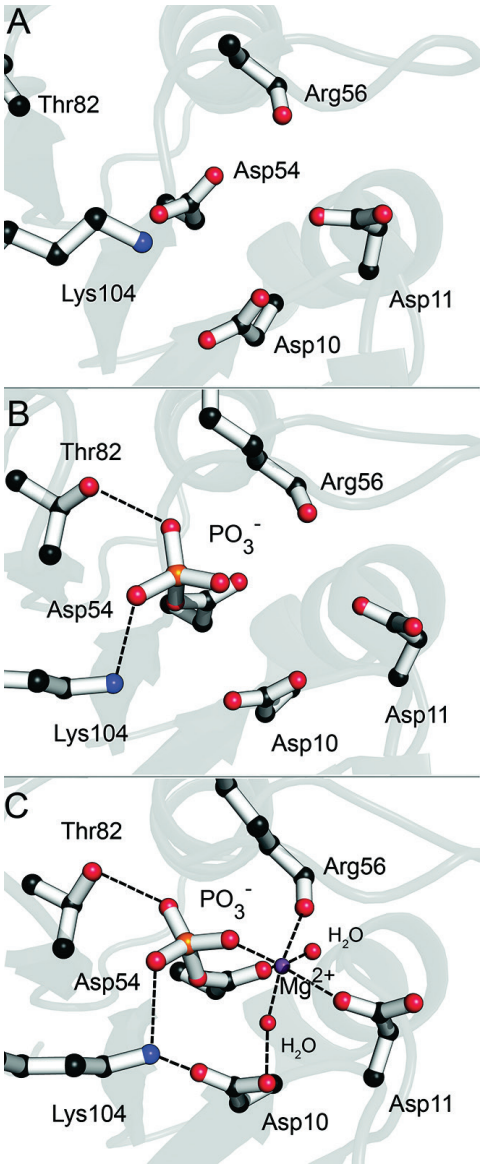

Figure 28. Highly conserved active site geometry of FixJ-rec.

(A) Disordered active site side chains of apo FixJ-rec (PDB 1DCK:A). (B) Active site side chains of phosphorylated FixJ-rec (PDB 1D5W:A). Crystal structure lacks defined waters and the metal cation. Switch residue Thr82 has shifted to hydrogen bond with Asp54. (C) Active site side chains in a sample candidate structure of phosphorylated FixJ-rec demonstrates near-identical geometry to the phosphorylated crystal.

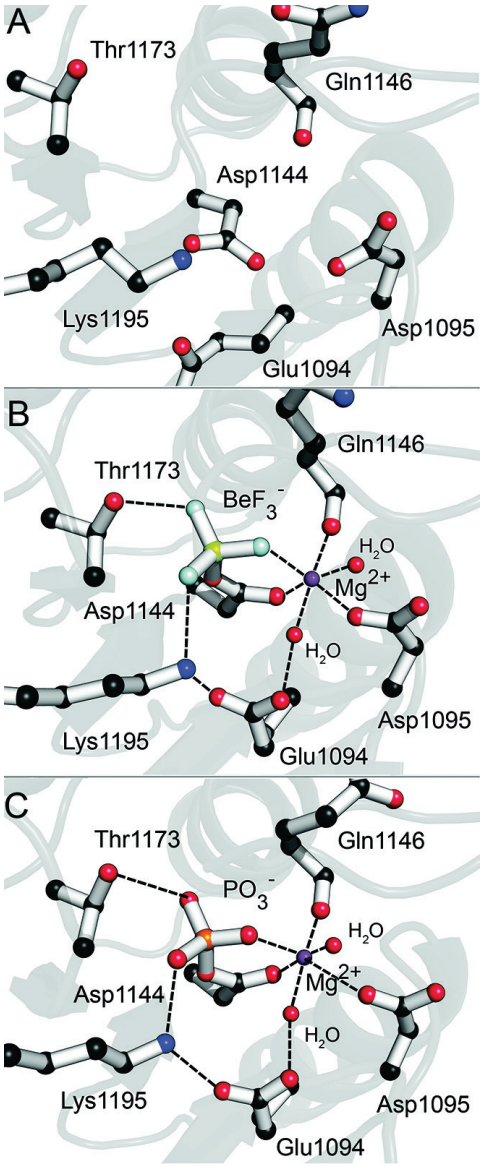

Figure 29. Highly conserved active site geometry of Sln1-rec.

(A) Disordered active site side chains of apo Sln1-rec (PDB 1OXB:B). (B) Active site side chains of Sln1-rec bound to BeF_3^- (PDB 2R25:B). The addition of Mg^{2+} orders the active site, generating an octahedral geometry with the recruitment of two additional water molecules. Switch residue Thr1173 has shifted to hydrogen bond with Asp1144. (C) Active site side chains in a sample candidate structure of phosphorylated Sln1-rec demonstrates near-identical geometry to the BeF_3^- bound crystal.

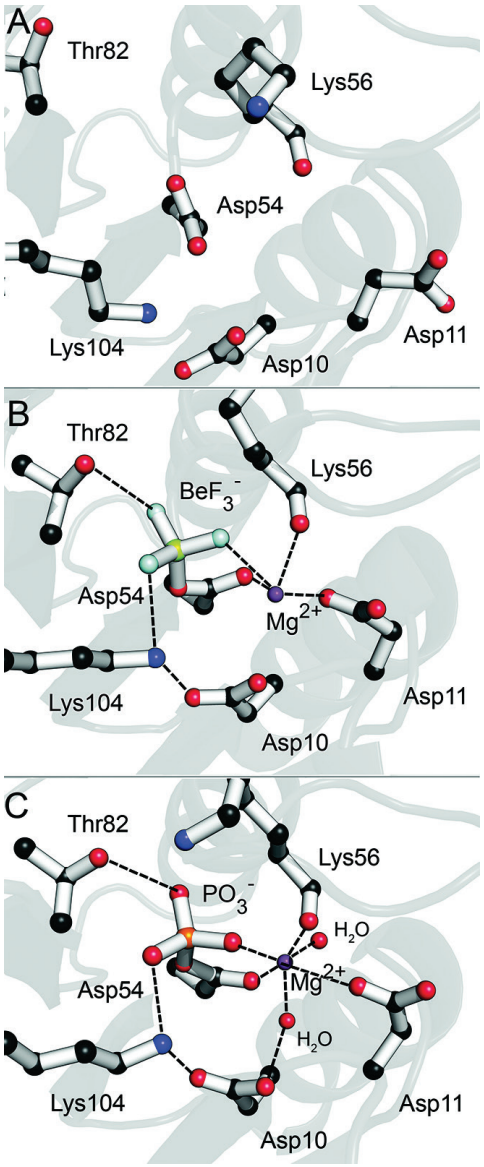

Figure 30. Highly conserved active site geometry of Spo0F.

(A) Disordered active site side chains of apo Spo0F (PDB 1NAT:A). (B) Active site side chains of Spo0F bound to BeF_3^- (PDB 2FTK:G). The addition of Mg^{2+} orders the active site, generating an octahedral geometry (note: water molecules were not defined in the crystal structure). Switch residue Thr82 has shifted to hydrogen bond with Asp54. (C) Active site side chains in a sample candidate structure of phosphorylated Spo0F demonstrates near-identical geometry to the BeF_3^- bound crystal.

Finally, the population sizes of the clusters were calculated. When performing geometrical clustering on MD simulations, we assume that

conformations grouped into the same cluster are structurally similar, and therefore lie within the same basin on the free-energy landscape of the protein [117]. Clusters are typically different sizes, due to conformational sampling progressing according to a Boltzmann distribution during simulation. Sparsely populated clusters are more likely to represent transient and/or higher energy states. When trying to identify which conformations will best resemble a crystal structure, which we normally assume is at an energetic minimum, neither of these effects is desirable. For this reason, conformers extracted from minor clusters (< ~1500) were typically eliminated, unless otherwise described. Additionally, if multiple medoids were identified as likely representative structures, the candidate from the larger cluster was retained.

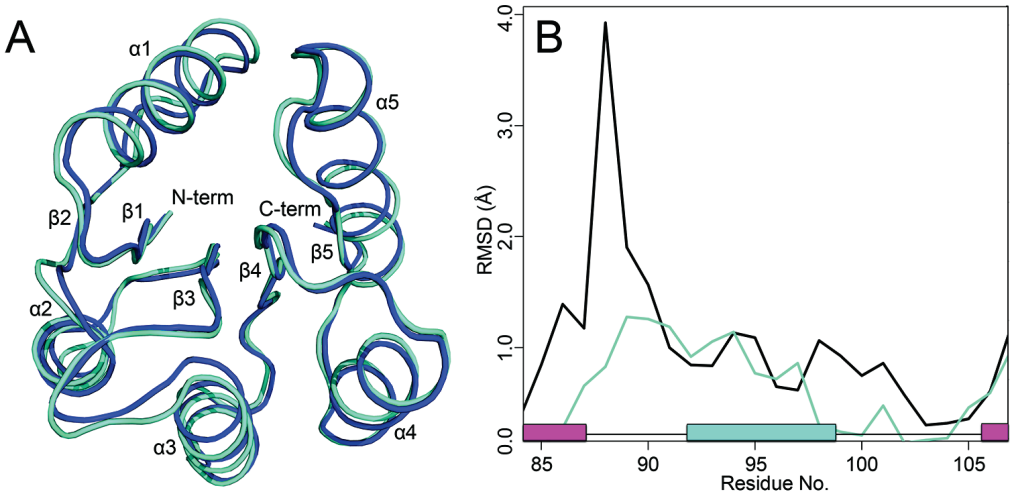

Figure 31. Global structural alignment of the top predicted model to the active CheY crystal structure and comparisons of the β4-α4-β5 region.

(A) Final predicted model (light cyan). Active crystal structure of CheY with BeF_3^- bound (PDB 1FQW, dark blue). (B) Per residue Cα RMSD for inactive vs. active crystals (black), and for predicted model vs. active crystal (cyan). The black trace shows the differences between

the active and inactive conformations. The cyan trace shows how closely the prediction matches the true active crystal structure. α-helix (cyan); β-strand (magenta).

Taking all of these criteria into consideration, a process of elimination allowed for the convenient identification of the top representative structure most likely to resemble a phosphorylated structure for each rec domain. This is apparent in the tabulated results (**Tables 5-9**). Table cells retained for analysis are shown in grey. Medoids with solid grey in every column were found to be most representative of a phosphorylated conformation. Each of these representative structures was then compared to their modified crystal structure counterpart to test the predictions. We first performed overall structural alignments to compare global protein structure. **Fig. 31A** shows the final CheY representative model (representing ~14% of the trajectory) superimposed to the BeF_3^- bound crystal structure (PDB 1FQW, dark blue ribbon). Structural alignments were also performed on each of the other rec domains (see **Figs. 32-35**). Overall average RMSDs for the predicted models ranged from 0.5 to 1.0 Å when compared to the existing modified crystal structures obtained from the PDB. While these measurements suggest that the models are nearly identical to their experimental counterparts, RMSD averages are often misleading, especially considering that the majority of the important conformational changes occur in flexible loop areas. Global RMSDs comparing apo and modified crystal structures are typically higher, but only moderately so. In addition, both our simulations and structures of identical proteins obtained from different experimental sources (crystallography, NMR)

suggest that the same rec domain in a common phosphorylated state may deviate by up to 2 Å, presumably through the protein's inherent flexibility. This likely generates considerable "noise," making it challenging to identify significant conformational differences. A more prudent analysis should focus on those areas known to change significantly and directly in response to phosphorylation. To this end, we generated per residue Cα RMSD traces for the β4-α4-β5 region, to compare the deviations between both the apo vs. BeF_3^- bound crystal structures and the predicted models vs. BeF_3^- bound crystal structures. Structures were aligned by the invariant β-sheet core, as this was found to exhibit the least amount of deviation upon phosphorylation. **Fig. 31B** contains two RMSD traces for CheY. The **black** trace represents the deviation between the apo crystal (PDB 3CHY) and BeF_3^- bound crystal (PDB 1FQW) structures, showing the degree to which the experimental conformations change. The **colored** trace represents the deviation between the top predicted model and the BeF_3^- bound crystal (PDB 1FQW) structure, showing how closely the prediction matches the experimental conformation. While natural fluctuations were present in most, the top representative models were still consistently and significantly closer to the experimental modified structures in every example, except for PhoP-rec, which showed only moderate improvement (see **Fig 32B**). In experimental structures, PhoP-rec experiences only minor changes upon phosphorylation, suggesting that visualizing the significant differences using simple RMSD measurements may be ineffective.

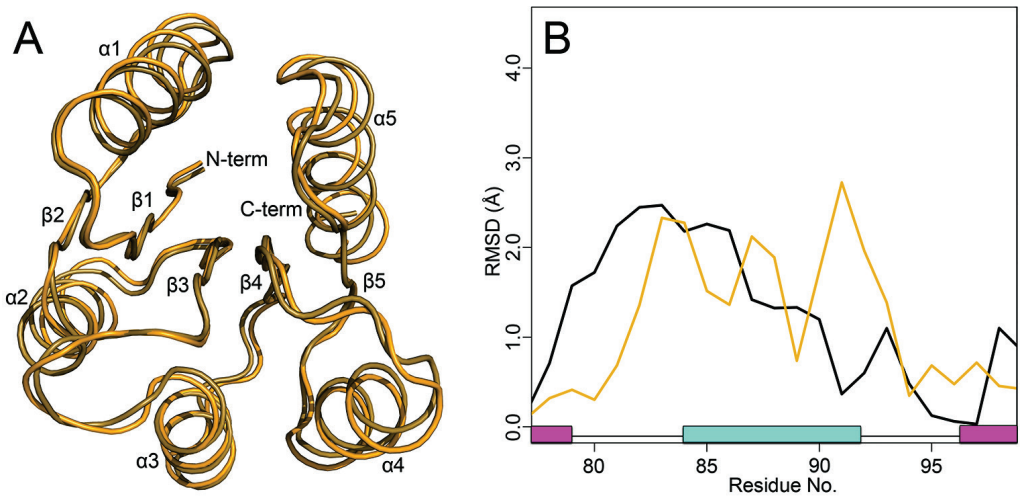

Figure 32. Global structural alignment of the top predicted model to the active PhoP-rec crystal and comparisons of the β4-α4-β5 region.

(A) Final predicted model (gold). Active crystal structure of PhoP-rec with BeF_3^- bound (PDB 2PL1, brown). (B) Per-residue RMSD for inactive vs. active crystals (black), and for predicted model vs. active crystal (gold). The black trace shows the differences between the active and inactive conformation. The gold trace shows how closely the prediction matches the true experimental structure. α-helix (cyan); β-strand (magenta).

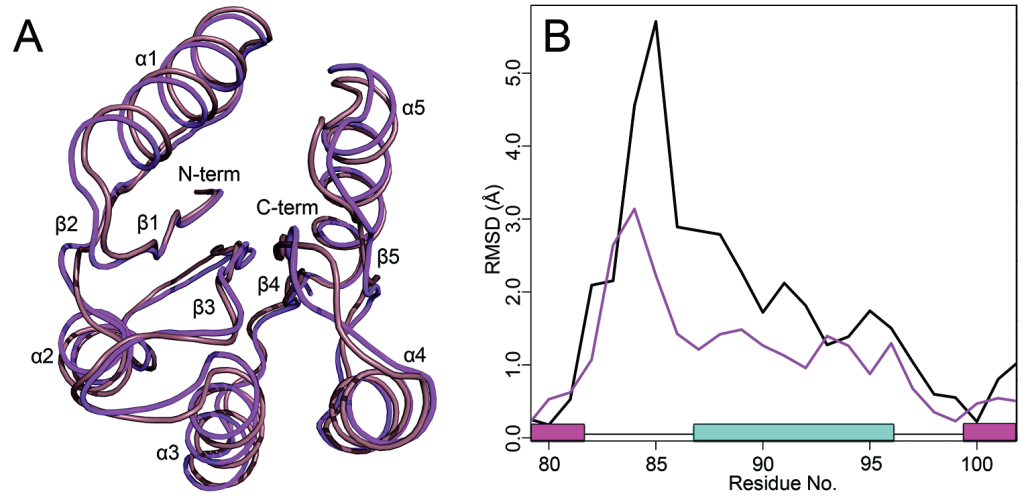

Figure 33. Global structural alignment of the top predicted model to the active FixJ-rec crystal and comparisons of the β4-α4-β5 region.

(A) Final predicted model (magenta). Active crystal structure of FixJ-rec with BeF_3^- bound (PDB 1D5W:A, dark purple). (B) Per-residue RMSD for inactive vs. active crystals (black),

and for predicted model vs. active crystal (magenta). The black trace shows the differences between the active and inactive conformation. The magenta trace shows how closely the prediction matches the true experimental structure. α-helix (cyan); β-strand (magenta).

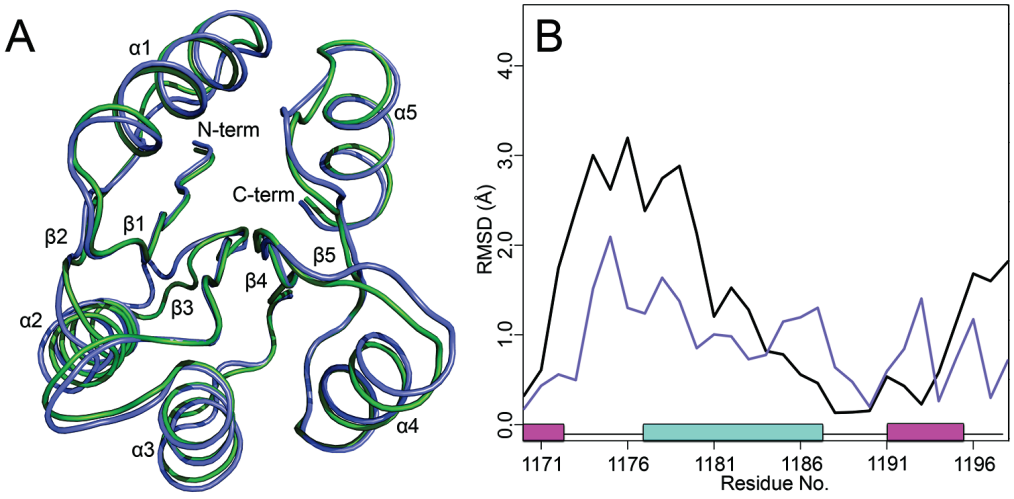

Figure 34. Global structural alignment of the top predicted model to the active Sln1-rec crystal and comparisons of the β4-α4-β5 region.

(A) Final predicted model (purple). Active crystal structure of Sln1-rec with BeF_3^- bound (PDB 2R25:B, green). (B) Per-residue RMSD for inactive vs. active crystals (black), and for predicted model vs. active crystal (purple). The black trace shows the differences between the active and inactive conformation. The purple trace shows how closely the prediction matches the true experimental structure. α-helix (cyan); β-strand (magenta).

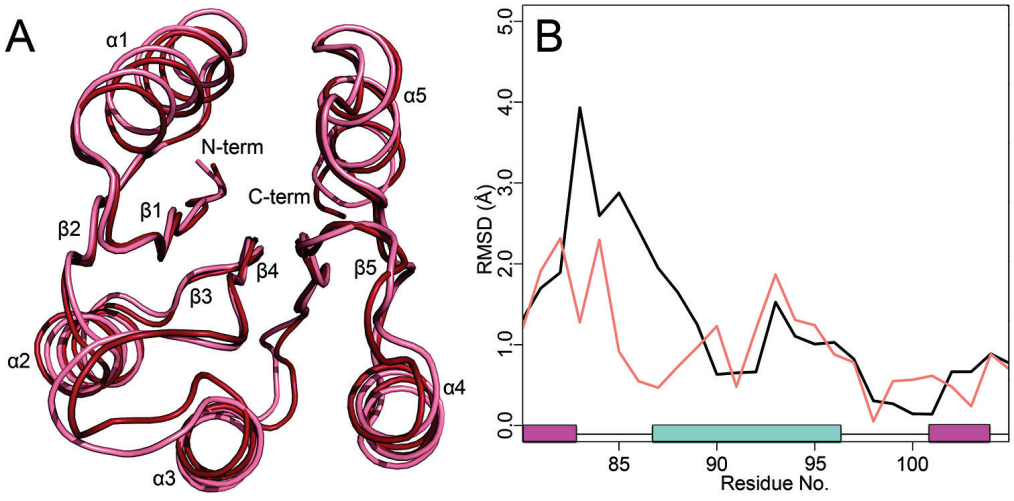

Figure 35. Global structural alignment of the top predicted model to the active Spo0F crystal and comparisons of the β4-α4-β5 region.

(A) Final predicted model (pink). Active crystal structure of Spo0F with BeF_3^- bound (PDB 2FTK:G, red). (B) Per-residue RMSD for inactive vs. active crystals (black), and for predicted model vs. active crystal (pink). The black trace shows the differences between the active and inactive conformation. The pink trace shows how closely the prediction matches the true experimental structure. α-helix (cyan); β-strand (magenta).

To further analyze the conformational changes in the β4-α4-β5 region, we calculated the pseudodihedral angle formed by consecutive Cα atoms in the β4α4 loop for each of the rec domains. This loop is related to the reorientation of the conserved switch Thr/Ser residue and is intimately involved in the transition between apo and phosphorylated conformations. Representing flexible regions in proteins using pseudodihedral angles is an effective way of comparing local conformations without relying on structural alignment methods and RMSD calculations [118]. **Table 10** contains the angles observed in the modified crystal structures, the top predicted models and the apo crystal structures for each of the proteins within the validation set.

Table 10. Pseudodihedral angles defined by Cα atoms in the β4α4 loop.

Protein	***Active Crystal***	***Predicted***	***Inactive Crystal***	***Residues***
CheY	160.36	161.12	-125.10	Ala88-Lys91
PhoP-rec	172.45	167.78	-168.23	Ala80-Ser83
Sln1-rec	-169.75	-177.81	160.29	A1174-D1177
FixJ-rec	132.20	127.22	-7.51	Gly83-Asp86
Spo0F	52.56, 109.61*	65.09	56.46	Ala83-Glu86

* Multiple conformations in the active crystal structure.

All models match their corresponding modified crystal structure closely, except for the seemingly incorrect Spo0F. Futher investigation revealed the existence of multiple conformations for the β4α4 loop in the BeF_3^- bound crystal structure of Spo0F (PDB 2FTK), suggesting that the

phosphorylated rec domain can adopt at least two separate forms. In addition, this BeF_3^- bound Spo0F structure was crystallized bound to its cognate phosphotransferase, Spo0B, which interacts extensively with the β4α4 loop. In the absence of its binding partner, the β4α4 loop likely adopts an altered conformation. Analysis of the NMR structural ensemble of BeF_3^- bound Spo0F (PDB 1PUX) shows pseudodihedral angles for the β4α4 loop ranging from 70 to 95^0. These results suggest that our predicted structure likely resembles a phosphorylated crystal structure of monomeric Spo0F.

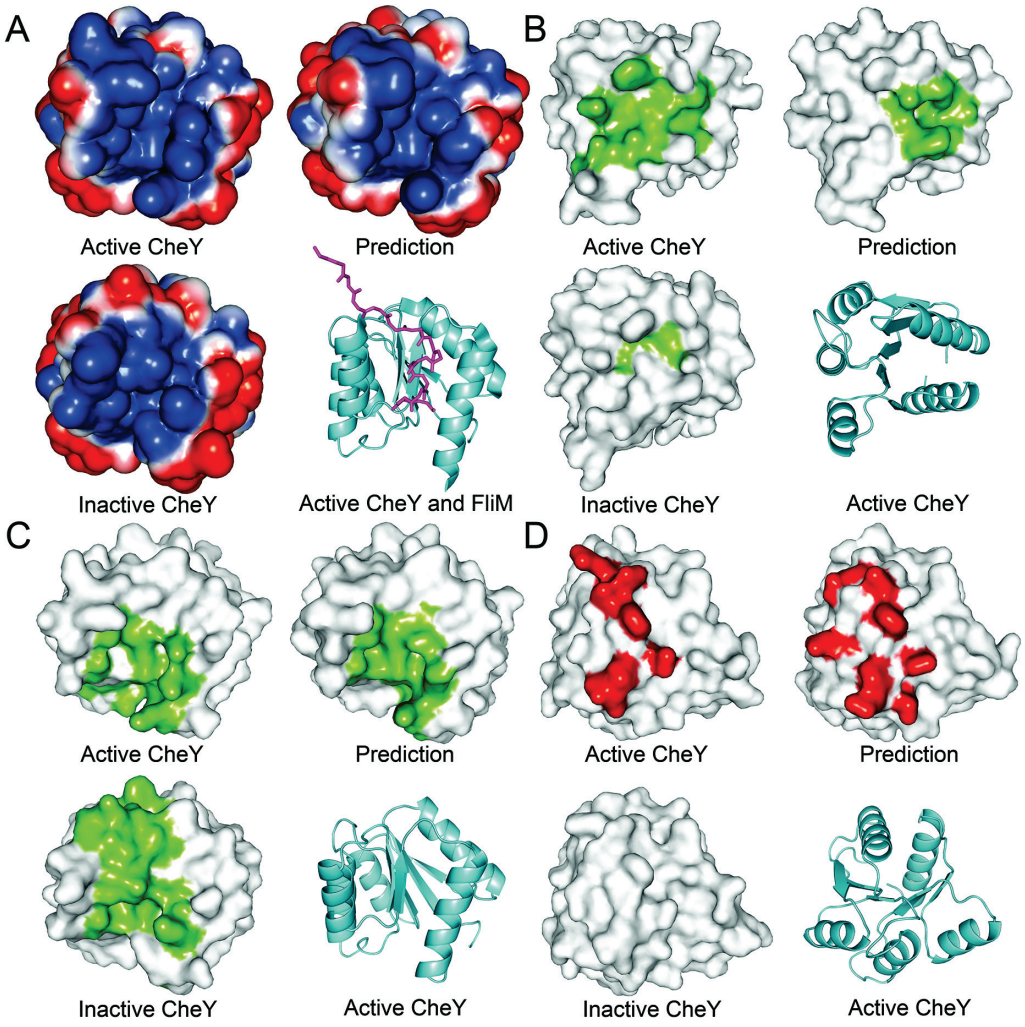

Figure 36. Changes in physicochemical properties of CheY.

Structures were aligned and partitioned into quadrants. View is of the same region on every structure. (A) Electrostatic surface potential calculated for the CheY crystal structures and the representative model. Shown is the β4-α4-β5 face. A ribbon representation of the BeF_3^- structure of CheY bound to FliM (PDB 1F4V) is shown for reference. Negative potential is blue; positive potential is red; FliM is magenta. (B) Detected molecular cavity located near the active site of CheY. (C) Detected molecular cavity located on the β4-α4-β5 surface. (D) Detected molecular protrusion located near the active site of CheY.

As a final validation step, we compared the physicochemical characteristics of the existing crystal structures with the predicted representative models. Properties such as electrostatic potential, hydrophobicity and geometric surface

features provide insight into the biological functions of proteins. By identifying differences between the phosphorylation states of the rec domain crystal structures, and testing if those differences are also present in the representative models, we can more accurately determine the quality of the models as predictive tools. Any visible changes that occur upon phosphorylation can be used to identify functionally relevant areas and targets for future experimental studies**. Fig. 36** shows the most significant changes found in the CheY structures. Structures were aligned and then arranged side-by-side to compare common regions. We first calculated the electrostatic surface potential for each of the structures (apo crystal, BeF_3^- bound crystal, predicted model) using the Adaptive Poisson-Boltzmann Solver (APBS) package [93]. **Fig. 36A** shows the results for CheY when viewed along the β4-α4-β5 surface. Upon phosphorylation, characteristic shifts facilitated by the movement of the β4α4 loop in CheY lead to a 20-fold increase in binding affinity for its partner, FliM [107]. This is accompanied by a change in the electrostatic potential of the α4-β5-α5 face, caused by residue shifts near the α4 and α5 helices and the β4α4 loop (**Fig. 36A)**. The structure of CheY bound to FliM (PDB 1F4V) reveals that the areas of altered potential, particularly the β4α4 loop, interact directly with the N-terminal region of FliM (**Fig. 36A)** [36]. These changes are nearly identical in the BeF_3^- bound crystal structure and predicted model, but are absent in the apo structure. An implementation of the VisGrid algorithm on the 3D-SURFER server was then used to characterize the local geometric features

of the proteins [94, 96]. VisGrid classifies the shape of a molecule's surface based on a visibility criterion, essentially the fraction of open or visible directions from a given target position. Cavities and protrusions are defined as clusters of positions based on the level of their visibility. VisGrid provides a simple and robust method of comparing the relative geometric shapes and regions of potential functional significance for different protein structures. **Fig. 36B** shows the first pocket detected in CheY. The BeF_3^- bound structure and model both show a cavity near the center of the active site surface, but this cavity is largely absent in the apo structure. **Fig. 36C** shows a second pocket found on the β4-α4-β5 surface. While the pocket is detected in all three structures, it has a significantly altered shape and location in the apo structure. **Fig. 36D** shows a protrusion detected on the active site surface. This protrusion is formed by the loops adjacent to the active site. No large or continuous protrusion was detected at this location in the apo crystal structure. Similar comparisons were performed for each of the other rec domains from this study (data not shown). The predicted models closely matched the modified crystal structures for all members of the validation set. Nearly every unique structural difference identified between the modified and apo crystal structures was accounted for in the models, including regions of known functional significance.

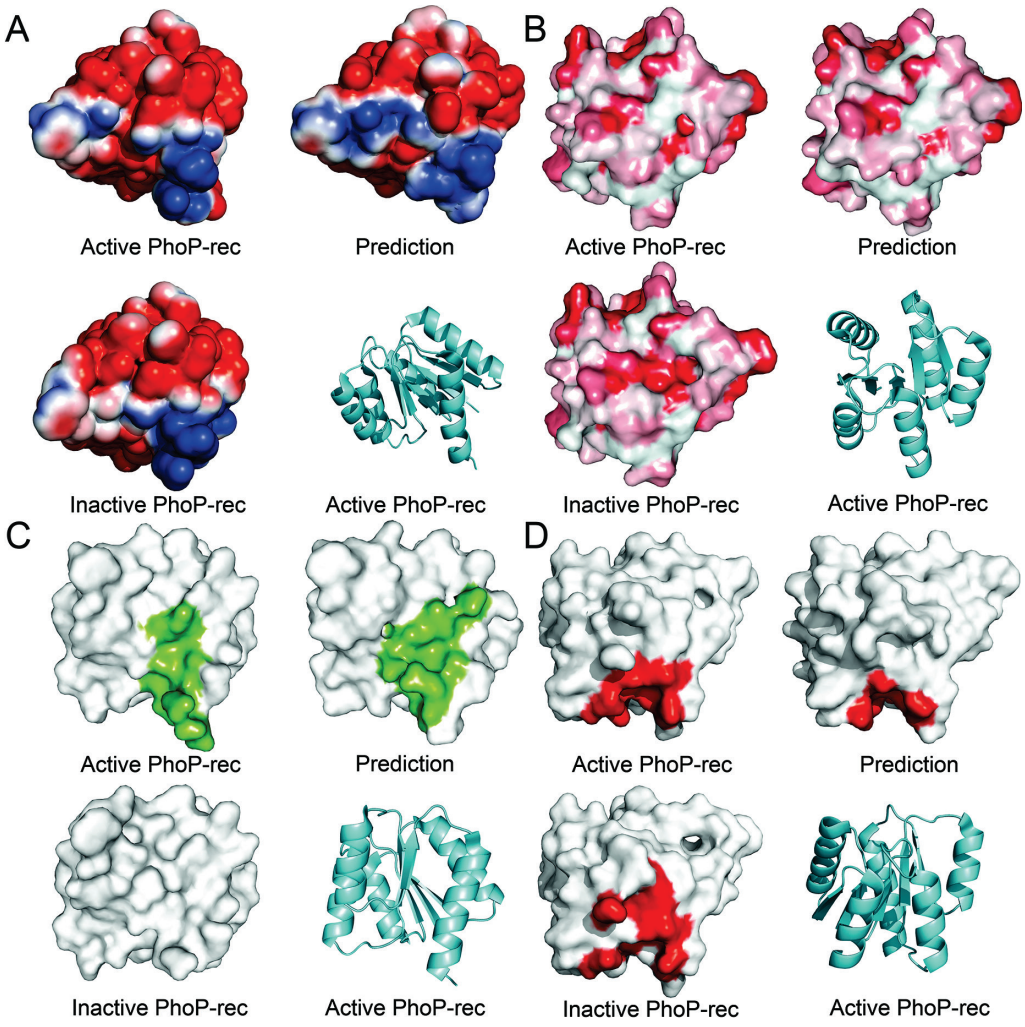

Figure 37. Changes in physicochemical properties of PhoP-rec.

Structures were aligned and partitioned into quadrants. View is of the same region on every structure. (A) Electrostatic surface potential calculated for the PhoP-rec crystal structures and the representative model. Shown is the β4-α4-β5 face. A ribbon representation of the BeF_3^- bound structure of PhoP-rec (PDB 2PL1) is shown for reference. Negative potential is blue; positive potential is red. (B) Calculated hydrophobicity of the PhoP-rec active site surface. (C) Detected molecular cavity located on the β4-α4-β5 surface. This cavity is absent in the apo structure (PDB 2PKX:A). (D) Detected molecular cavity located near α5- α1.

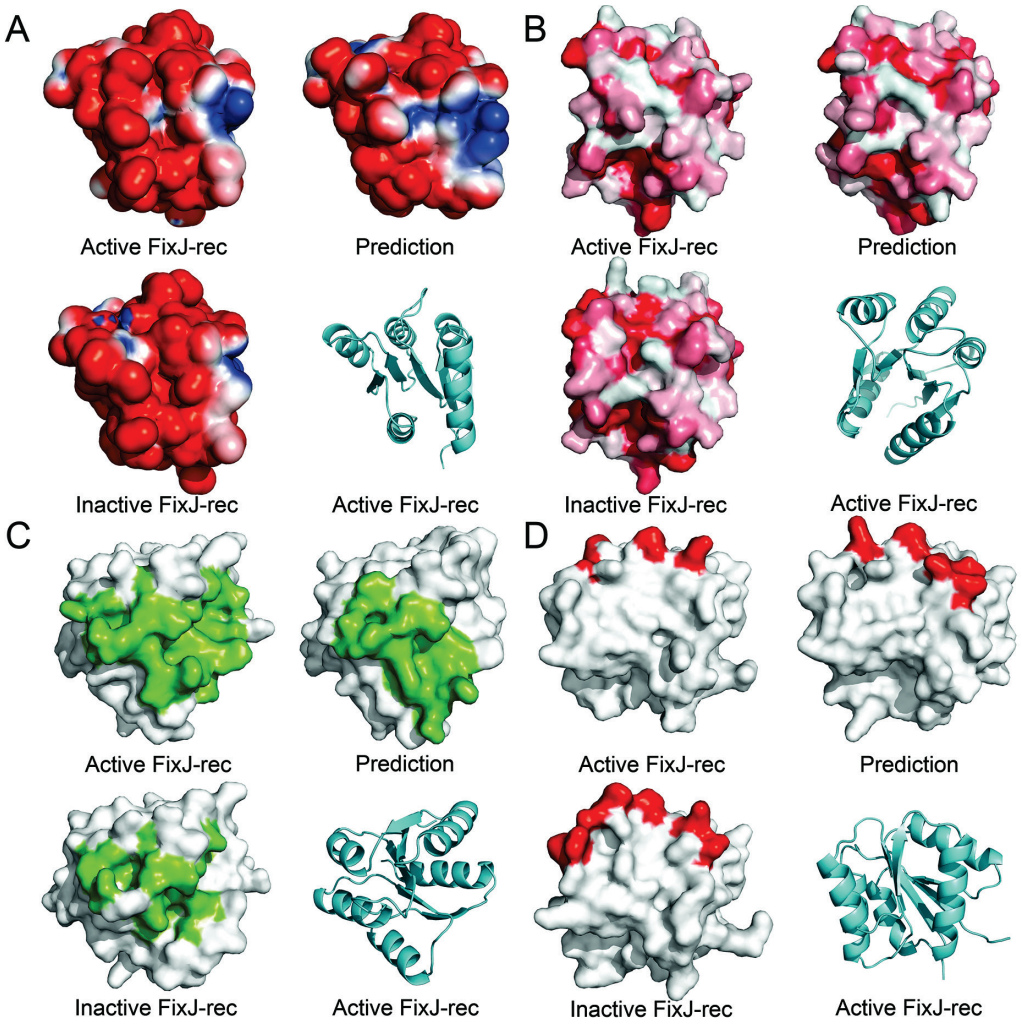

Figure 38. Changes in physicochemical properties of FixJ-rec.

Structures were aligned and partitioned into quadrants. View is of the same region on every structure. (A) Electrostatic surface potential calculated for the FixJ-rec crystal structures and the chosen representative model. Shown is the active site, and along the α5-α1 face. A ribbon representation of the phosphorylated structure of FixJ-rec (PDB 1D5W:A) is shown for reference. Negative potential is blue; positive potential is red. (B) Calculated hydrophobicity of the FixJ-rec active site surface, highlighting the drastic shape change upon phosphorylation. (C) Detected molecular cavity located at the surface of the active site. (D) Detected molecular protrusions located along the ridge of the β4-α4-β5 face. Positioning of the residues located near the β4-α4 loop (upper left) is critical for dimerization and function. (apo crystal PDB 1DCK:A).

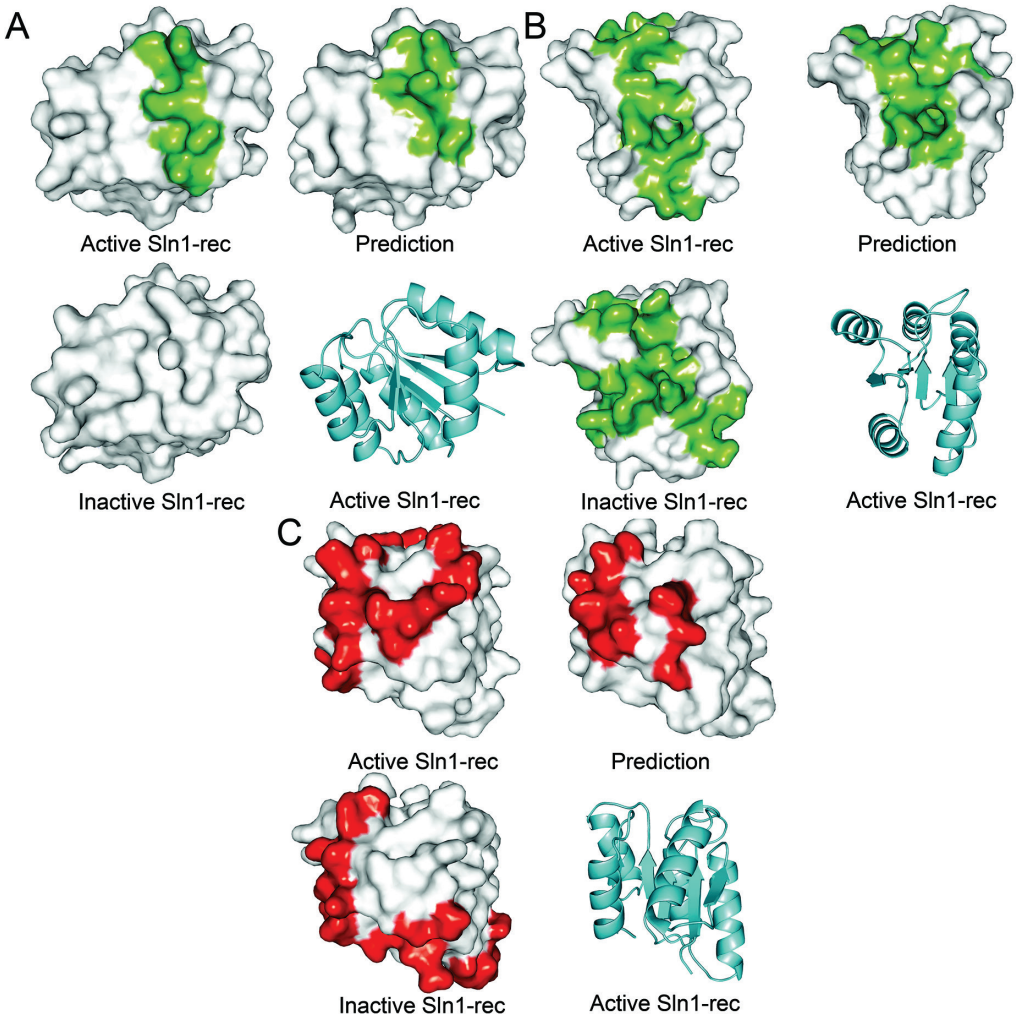

Figure 39. Changes in physicochemical properties of Sln1-rec.

Structures were aligned and partitioned into quadrants. View is of the same region on every structure. No electrostatic potential differences were detected between the apo and BeF_3^- bound crystal structures of Sln1-rec. (A) Detected molecular cavity located at the β4-α4-β5 face, near the α5 helix. This cavity is absent in the apo crystal structure (PDB 1OXB:B). A ribbon representation of the active structure of Sln1-rec (PDB 2R25:B) is shown for reference. (B) Detected molecular cavity located on the active site surface. (C) Detected molecular protrusions located along the α5- α1 face.

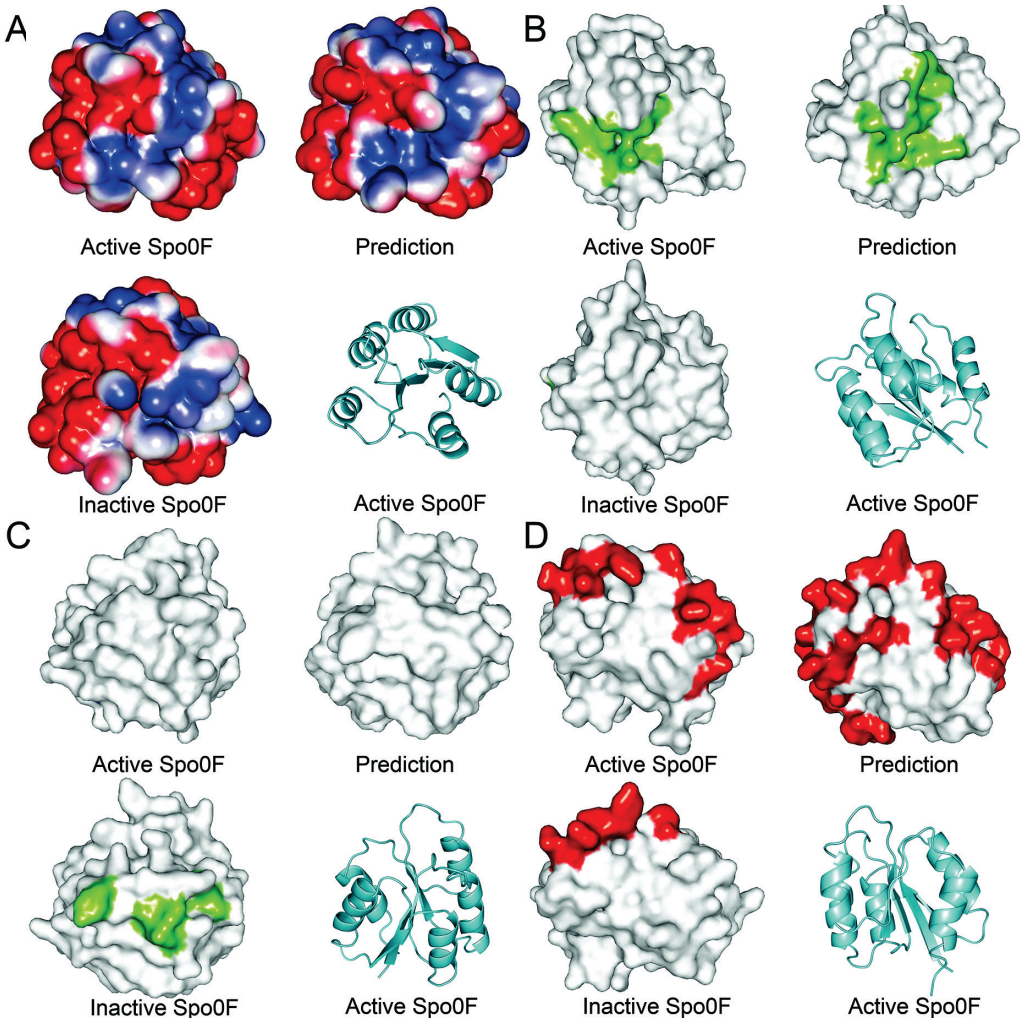

Figure 40. Changes in physicochemical properties of Spo0F.

Structures were aligned and partitioned into quadrants. View is of the same region on every structure. (A) Electrostatic surface potential calculated for the Spo0F crystal structures and the chosen representative model. Shown is the active site surface. A ribbon representation of the BeF_3^- bound structure of Spo0F (PDB 2FTK:G) is shown for reference. Negative potential is blue; positive potential is red. (B) Detected molecular cavity located at the α4 helix. (C) Detected molecular cavity located at the α2-α3 surface. (D) Detected molecular protrusions located along the ridge of the β4-α4-β5 face.

By combining comparisons of global structure, functionally relevant areas, physicochemical characteristics and local geometric features, we have demonstrated that the models predicted using the previously described

approach provide accurate representations of phosphorylated rec domain structures. Though changes are challenging to detect, the predictions are able to capture the key structural features associated with the phosphorylation of rec domains. The fact that these features are not detectable in the apo crystal structures shows the value that such models hold for guiding experimental studies.

Phosphorylation of Ssk1-rec and comparison with the upstream Sln1-rec

Having validated the new methodology, we then compared the rec domains of the upstream hybrid HK, Sln1, with the downstream RR, Ssk1, from the *S. cerevisiae* osmotic stress response pathway. Ssk1-rec is slightly unusual in that the phosphorylated conformation is the resting state, and activation occurs only upon dephosphorylation. While crystal structures of apo and BeF_3^- bound Sln1-rec exist [4, 33], no structure is currently available for the Ssk1-rec domain. A homology model for apo Ssk1-rec was generated using the Phyre2 server as an alternative [71]. Though they share little overall sequence similarity, Ssk1-rec possesses a canonical rec domain at residues 506-651 that aligns well with Sln1-rec, except for a short, 22-residue loop (residues 580-603) between α3 and β4 on the surface distal to the active site. This loop consistently modeled poorly and was excluded from the final structure prediction.

We subjected the apo Ssk1-rec model to energy minimization and equilibration to alleviate any non-physiological structural features or model

bias produced by homology modeling. The resulting structure was validated with the Verify3D server [72]. Ramachandran plots generated with RAMPAGE indicated that ~99% of residues fell within favored or allowed regions [73]. This "relaxed" model was phosphorylated and used for restrained MD. Trajectories were analyzed and clustered to extract candidate structures for phosphorylated Ssk1-rec. A final representative medoid was chosen based on the previously described criteria. We then compared this model and the dynamic information obtained from the simulations with the Sln1-rec data, to provide insight into the components of the fungal signaling pathway.

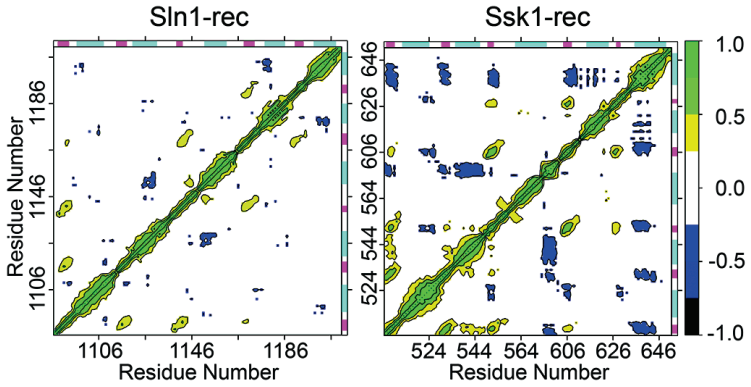

Figure 41. Dynamical cross-correlation maps for Sln1-rec and Ssk1-rec.

Positive correlation (yellow/green) suggests movement in phase, or in the same direction. Negative correlation (blue/black) suggests movement out of phase, or along the opposite direction. α-helix (cyan); β-strand (magenta).

The characteristic pattern of positively correlated motion found in all rec domains can be seen in the DCCM for Ssk1-rec **(Fig. 41)**. Additional patches were also observed between the β1 strand with the α2 helix, and the β3 strand with the α4β5 region. None of the other rec domains in this study displayed positively correlated motion in these areas. Ssk1-rec also

demonstrated a large network of negatively correlated movements, far more extensive than in Sln1-rec. These areas of negative correlation correspond to the β1 strand with the α5 helix, the β1 strand with the α3 helix/loop, the α1 helix/loop with the β3 strand, the α1 helix/loop with the β4 loop/strand, the α1 helix/loop with the β5 strand, the entire β2-α2-β3 region with the α3β4 region, the entire β4-α4-β5 region with the α5 helix, and the α5 helix with the β1, β2 and β3 strands. The results suggest that Ssk1-rec undergoes a multitude of significant collective movements in its phosphorylated state, in addition to the characteristic perturbations at the β4-α4-β5 surface. Ssk1-rec and Sln1-rec appear to share negatively correlated motions between their α2 and α3 regions and their α4 and α5 regions, though in both cases the affected areas are larger in Ssk1-rec.

The differences seen in the collective motions of Ssk1-rec and Sln1-rec are likely attributable to the distinct functional roles of each protein. Sln1-rec, as part of a hybrid HK, interacts with its upstream donor (its own catalytic kinase domain) and downstream acceptor (the HPt protein, Ypd1) through a common binding surface formed by the loops adjacent to the active site and portions of the α1 and α5 helices. The interface between the rec and HK domains of Sln1 is likely similar, though no structure currently exists for the complex. The concentrated regions of negatively correlated motion seen in Sln1-rec are localized to these relatively small areas. We speculate that this is due to the presence of only the single binding surface. In contrast, Ssk1-rec

must bind the upstream Ypd1, the downstream Ssk2 and possibly its own extensive N-terminal domain. The Ypd1 binding interface is known to be homologous to Sln1-rec, formed at the active site surface by adjacent loop regions [4]. However, mutational studies by Horie et al. (2008) suggest that the Ssk2 binding site is located elsewhere on the rec domain [119]. The same researchers also determined that Ssk1 forms a homodimer in solution, though whether the dimerization interface is a part of the rec domain or the N-terminal region is unknown. These multiple regions of functional significance likely correspond to the far more complex network of collective motions detected in phosphorylated Ssk1-rec.

Most significant structural changes induced by phosphorylation in Ssk1-rec were centered around the β4-α4-β5 region, with the β4α4 loop and N-terminal portion of the α4 helix undergoing a dramatic shift away from the α4-β5-α5 face. Large changes also occurred in the α1, α2 and C-terminal half of α5 helices. Smaller shifts were observed in the N-terminal portion of the α3 helix and the β1α1, β3α3 and β5α5 loops. **Fig. 42** shows a structural alignment of the top representative structure and the apo Ssk1-rec model. Visually, the shift in the α4 helix and loop is far more dramatic than in Sln1-rec. The α1, α2 and α5 helices of Ssk1-rec may also be functionally important, as their shifts correspond to strong, broad collective motions not seen in Sln1-rec. Surface analysis shows that phosphorylated Ssk1-rec forms additional protrusions on the β4α4 loop and helix, the β5α5 loop and down the α5 helix (data not

shown). Additionally, the α4-β5-α5 surface cavity undergoes significant shape changes (data not shown).

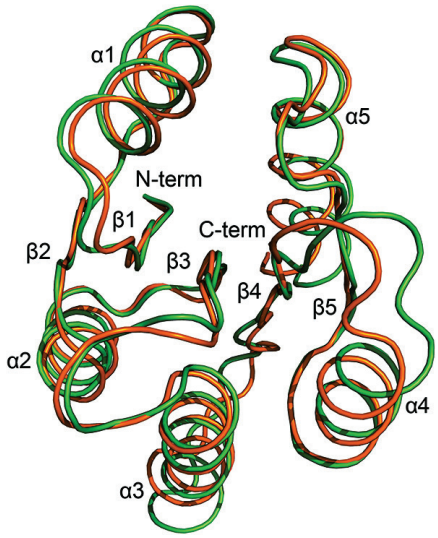

Figure 42. Structural alignment of predicted model and apo Ssk1-rec.
Final predicted model (orange). Homology model of apo Ssk1-rec (green).

Horie et al. (2008) discovered a mutation (D628G) on Ssk1-rec that abolished binding with Ssk2 and inhibited the activation of Ssk2 by wild-type Ssk1-rec through dimerization [119]. This suggests that the likely binding surface for Ssk2 is located on the α4-β5-α5 surface of Ssk1-rec. The D628G mutation occurs next to the aromatic switch residue, Tyr629. The authors hypothesized that replacing this aspartate with a glycine locks the protein into a conformation that is unable to bind Ssk2. They also speculated that phosphorylation induces a similar shift, possibly through the characteristic rotamerization of the aromatic switch residue. Our results may shed further light on this phenomenon.

Surprisingly, Tyr629 did not rotamerize in any of the phosphorylated simulations, remaining completely solvent-exposed in both phosphorylated and apo conformations. We speculate that this is caused by the severe shift of the β4α4 loop and α4 helix upon phosphorylation. Unlike most of the other rec domains, the pocket created by the shift of the Thr/Ser switch residue in Ssk1-rec is significantly less hydrophobic due to Lys618. In the apo Ssk1-rec structure, the side chains Lys618 and Tyr629 are hydrogen-bonded and solvent-exposed, pointing outwards. Upon phosphorylation, the conformational change causes Lys618 to shift up and in, partially occupying the pocket that would normally be filled by the rotamerized Tyr629 and forming hydrogen bonds with residues on the α4 helixand β4α4 loop. PhoP-rec contains a similar organization with residues Lys87 and Tyr98, though it lacks the dramatic shift in the α4 helix upon phosphorylation. Because of this, Lys87 never enters the pocket, and Tyr98 is able to rotamerize and satisfy an additional hydrogen bond with Arg81. These results suggest that unlike the majority of rec domains, Tyr629 is incapable of rotamerizing while Ssk1-rec is phosphorylated. This may be related to Ssk1-rec requiring dephosphorylation to activate further downstream binding.

Although Asp628 does not appear to affect the rotamerization of Tyr629, it may play a role in the overall conformational change that occurs in the α4-β5-α5 surface. Comparison of the Ssk1-rec models reveals that phosphorylation causes the α4 and α5 helices to shift away from the β5 strand

in an opening motion. The DCCM also shows negatively correlated motion between several areas on the α5 helix and the β5 strand (**Fig. 41)**, including Asp628. In the representative phosphorylated model, the side chain of Asp628 forms an hydrogen bond with Trp646 on the α5 helix. Lys642 is also near, potentially creating an additional salt bridge. We compared brief trajectories of phosphorylated and unphosphorylated Ssk1-rec to investigate the role of Asp628 (data not shown) . When phosphorylated, Asp628 forms a salt bridge with Lys642 in ~67% of frames, and a hydrogen bond with Trp646 in ~32% of frames. When unphosphorylated, Asp628 forms a salt bridge with Lys642 in ~85% of frames, and a hydrogen bond with Trp646 in ~84% of frames. These preliminary numbers suggest a significant change in interaction stability corresponding to the α5 helix shifting closer to the β5 strand upon dephosphorylation. We speculate that mutating Asp628 to glycine abolishes this effect, preventing binding to Ssk2 through the α4-β4-α5 surface.

2.4 Conclusions

We have presented a computational methodology that utilizes biased molecular dynamics to predict the phosphorylated conformations of response regulator rec domains. This approach uses multiple half-harmonic restraints to drive formation of the conserved rec domain active site. This rearrangement pulls the rest of the protein into its phosphorylated state, allowing sampling of modified conformations on a nanosecond timescale. The biasing terms make these studies accessible to research laboratories that may lack the specialized

equipment and experience required to observe this process on the millisecond timescale on which it normally occurs. We validated this technique on five diverse, well-characterized rec domains: CheY, PhoP-rec, FixJ-rec, Sln1-rec and Spo0F. Cross-correlation analysis revealed areas of correlated motion that have been implicated in phosphorylation-mediated behavior for the five rec domains. This information can be used to guide further experimental studies towards areas of functional interest. We extracted representative models for each rec domain and compared them to the existing crystal structures using global structure, switch residue and quartet residue orientations, active site geometry, loop conformations and physicochemical surface properties to demonstrate the accuracy of the predictions. The models closely matched the experimentally-determined structures of the conformationally modified rec domains in all aspects, validating the approach and suggesting that an induced fit model of conformational switching is sufficient to sample these transient phosphorylated conformations. Researchers studying two-component signaling pathways can use this approach to gain structural insight into specific rec domain characteristics, such as changes in surface properties and binding surface accessibility. We demonstrated this by comparing the known and predicted structures of the *S. cerevisiae* signaling proteins, Sln1-rec and Ssk1-rec, respectively. This information allows for more complete analysis of the functional differences in these two rec domains in the absence of experimental structural data. We predict that this approach can also be applied to signaling

complexes, to study changes in protein-protein interactions within two-component signaling systems as rec domains transition from apo to phosphorylated conformations.

References

1. Ann M. Stock, Victoria L. Robinson a, Goudreau PN. Two-component signal transduction. Annual Rev Biochemistry. 2000;69(1):183-215.

2. West AH, Stock AM. Histidine kinases and response regulator proteins in two-component signaling systems. Trends Biochem Sci. 2001;26(6):369-76.

3. Posas F, Wurgler-Murphy SM, Maeda T, Witten EA, Thai TC, Saito H. Yeast HOG1 MAP kinase cascade is regulated by a multistep phosphorelay mechanism in the SLN1-YPD1-SSK1 "two-component" osmosensor. Cell. 1996;86(6):865-75.

4. Xu Q, Porter SW, West AH. The yeast YPD1/SLN1 complex: Insights into molecular recognition in two-component signaling systems. Structure. 2003;11(12):1569-81.

5. Fisher SL, Jiang W, Wanner BL, Walsh CT. Cross-talk between the histidine protein kinase VanS and the response regulator PhoB. Characterization and identification of a VanS domain that inhibits activation of PhoB. J Biol Chem. 1995;270(39):23143-9.

6. Laub MT, Goulian M. Specificity in two-component signal transduction pathways. Annual Rev Genetics. 2007;41(1):121-45.

7. Rowland MA, Deeds EJ. Crosstalk and the evolution of specificity in two-component signaling. Proc Natl Acad Sci. 2014;111(15):5550-5.

8. Silva JC, Haldimann A, Prahalad MK, Walsh CT, Wanner BL. *In vivo* characterization of the type A and B vancomycin-resistant *enterococci* (VRE) VanRS two-component systems in *Escherichia coli*: A nonpathogenic model for studying the VRE signal transduction pathways. Proc Natl Acad Sci. 1998;95(20):11951-6.

9. Willett JW, Tiwari N, Müller S, Hummels KR, Houtman JCD, Fuentes EJ, Kirby JR. Specificity residues determine binding affinity for two-component signal transduction systems. mBio. 2013;4(6).

10. Volz K, Matsumura P. Crystal structure of *Escherichia coli* CheY refined at 1.7-A resolution. J Biol Chem. 1991;266(23):15511-9.

11. Lee SY, Cho HS, Pelton JG, Yan D, Berry EA, Wemmer DE. Crystal structure of activated CheY. Comparison with other activated receiver domains. J Biol Chem. 2001;276(19):16425-31.

12. Thomas SA, Brewster JA, Bourret RB. Two variable active site residues modulate response regulator phosphoryl group stability. Mol Micro. 2008;69(2):453-65.

13. Dyer CM, Dahlquist FW. Switched or not: The structure of unphosphorylated CheY bound to the N terminus of FliM. J Bact. 2006;188(21):7354-63.

14. Gao R, Mukhopadhyay A, Fang F, Lynn DG. Constitutive activation of two-component response regulators: Characterization of VirG activation in *Agrobacterium tumefaciens*. J Bact. 2006;188(14):5204-11.

15. Gardino AK, Volkman BF, Cho HS, Lee SY, Wemmer DE, Kern D. The NMR solution structure of BeF_3^--activated Spo0F reveals the conformational switch in a phosphorelay system. J Mol Biol. 2003;331(1):245-54.

16. McDonald Leanna R, Boyer Joshua A, Lee Andrew L. Segmental notions, not a two-state concerted switch, underlie allostery in CheY. Structure. 2012;20(8):1363-73.

17. Volkman BF, Lipson D, Wemmer DE, Kern D. Two-state allosteric behavior in a single-domain signaling protein. Science. 2001;291(5512):2429-33.

18. Gardino AK, Villali J, Kivenson A, Lei M, Liu CF, Steindel P, Eisenmesser EZ, Labeikovsky W, Wolf-Watz M, Clarkson MW, Kern D. Transient non-native hydrogen bonds promote activation of a signaling protein. Cell. 2009;139(6):1109-18.

19. Gao R, Mack TR, Stock AM. Bacterial response regulators: Versatile regulatory strategies from common domains. Trends Biochem Sci. 2007;32(5):225-34.

20. Bourret RB, Thomas SA, Page SC, Creager-Allen RL, Moore AM, Silversmith RE. Measurement of response regulator autodephosphorylation rates spanning six orders of magnitude. Methods Enzymol. 2010;471:89-114.

21. Bobay BG, Hoch JA, Cavanagh J. Dynamics and activation in response regulators: The β4-α4 loop. Biomol Concepts. 2012;3(2):175-82.

22. Groban ES, Narayanan A, Jacobson MP. Conformational changes in protein loops and helices induced by post-translational phosphorylation. PLoS Comput Biol. 2006;2(4):e32.

23. Birck C, Mourey L, Gouet P, Fabry B, Schumacher J, Rousseau P, Kahn D, Samama J-P. Conformational changes induced by phosphorylation of the FixJ receiver domain. Structure. 1999;7(12):1505-15.

24. Kaserer AO, Andi B, Cook PF, West AH. Effects of osmolytes on the SLN1-YPD1-SSK1 phosphorelay system from *Saccharomyces cerevisiae*. Biochemistry. 2009;48(33):8044-50.

25. Re SD, Tolstykh T, Wolanin PM, Stock JB. Genetic analysis of response regulator activation in bacterial chemotaxis suggests an intermolecular mechanism. Protein Sci. 2002;11(11):2644-54.

26. Guhaniyogi J, Robinson VL, Stock AM. Crystal structures of beryllium fluoride-free and beryllium fluoride-bound CheY in complex with the conserved C-terminal peptide of CheZ reveal dual binding modes specific to CheY conformation. J Mol Biol. 2006;359(3):624-45.

27. Simonovic M, Volz K. A distinct meta-active conformation in the 1.1-A resolution structure of wild-type apo CheY. J Biol Chem. 2001;276(31):28637-40.

28. Stock AM, Guhaniyogi J. A new perspective on response regulator activation. J Bact. 2006;188(21):7328-30.

29. Csermely P, Palotai R, Nussinov R. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. Trends Biochem Sci. 2010;35(10):539-46.

30. Bachhawat P, Stock AM. Crystal structures of the receiver domain of the response regulator PhoP from *Escherichia coli* in the absence and presence of the phosphoryl analog beryllofluoride. J Bact. 2007;189(16):5987-95.

31. Cho H, Wang W, Kim R, Yokota H, Damo S, Kim SH, Wemmer D, Kustu S, Yan D. BeF_3^- acts as a phosphate analog in proteins phosphorylated on aspartate: structure of a BeF_3^- complex with phosphoserine phosphatase. Proc Natl Acad Sci. 2001;98(15):8525-30.

32. Varughese KI, Tsigelny I, Zhao H. The crystal structure of beryllofluoride Spo0F in complex with the phosphotransferase Spo0B represents a phosphotransfer pretransition state. J Bact. 2006;188(13):4970-7.

33. Zhao X, Copeland DM, Soares AS, West AH. Crystal structure of a complex between the phosphorelay protein YPD1 and the response regulator domain of SLN1 bound to a phosphoryl analog. J Mol Biol. 2008;375(4):1141-51.

34. Toro-Roman A, Wu T, Stock AM. A common dimerization interface in bacterial response regulators KdpE and TorR. Protein Sci. 2005;14(12):3077-88.

35. Hickey JM, Lovell S, Battaile KP, Hu L, Middaugh CR, Hefty PS. The atypical response regulator protein ChxR has structural characteristics and dimer interface interactions that are unique within the OmpR/PhoB subfamily. J Biol Chem. 2011;286(37):32606-16.

36. Lee SY, Cho HS, Pelton JG, Yan D, Henderson RK, King DS, Huang L, Kustu S, Berry EA, Wemmer DE. Crystal structure of an activated response regulator bound to its target. Nat Struct Biol. 2001;8(1):52-6.

37. Gouet P, Fabry B, Guillet V, Birck C, Mourey L, Kahn D, Samama JP. Structural transitions in the FixJ receiver domain. Structure. 1999;7(12):1517-26.

38. Bourret RB. Receiver domain structure and function in response regulator proteins. Current Opinion in Microbiology. 2010;13(2):142-9.

39. Hess JF, Oosawa K, Kaplan N, Simon MI. Phosphorylation of three proteins in the signaling pathway of bacterial chemotaxis. Cell. 1988;53(1):79-87.

40. Janiak-Spens F, Sparling JM, Gurfinkel M, West AH. Differential stabilities of phosphorylated response regulator domains reflect functional roles of the yeast osmoregulatory SLN1 and SSK1 proteins. J Bact. 1999;181(2):411-7.

41. Weiss V, Magasanik B. Phosphorylation of nitrogen regulator I (NRI) of *Escherichia coli*. Proc Natl Acad Sci. 1988;85(23):8919-23.

42. Pazy Y, Wollish AC, Thomas SA, Miller PJ, Collins EJ, Bourret RB, Silversmith RE. Matching biochemical reaction kinetics to the timescales of life: Structural determinants that influence the autodephosphorylation rate of response regulator proteins. J Mol Biol. 2009;392(5):1205-20.

43. Goudreau PN, Lee PJ, Stock AM. Stabilization of the phospho-aspartyl residue in a two-component signal transduction system in *Thermotoga maritima*. Biochemistry. 1998;37(41):14575-84.

44. Jagadeesan S, Mann P, Schink CW, Higgs PI. A novel "four-component" two-component signal transduction mechanism regulates developmental progression in *Myxococcus xanthus*. J Biol Chem. 2009;284(32):21435-45.

45. Menon S, Wang S. Structure of the response regulator PhoP from *Mycobacterium tuberculosis* reveals a dimer through the receiver domain. Biochemistry. 2011;50(26):5948-57.

46. Bachhawat P, Swapna GV, Montelione GT, Stock AM. Mechanism of activation for transcription factor PhoB suggested by different modes of dimerization in the inactive and active states. Structure. 2005;13(9):1353-63.

47. Fiedler U, Weiss V. A common switch in activation of the response regulators NtrC and PhoB: phosphorylation induces dimerization of the receiver modules. EMBO J. 1995;14(15):3696-705.

48. Sola M, Gomis-Ruth FX, Serrano L, Gonzalez A, Coll M. Three-dimensional crystal structure of the transcription factor PhoB receiver domain. J Mol Biol. 1999;285(2):675-87.

49. Trajtenberg F, Albanesi D, Ruétalo N, Botti H, Mechaly AE, Nieves M, Aguilar PS, Cybulski L, Larrieux N, de Mendoza D, Buschiazzo A. Allosteric activation of bacterial response regulators: The role of the cognate histidine kinase beyond phosphorylation. mBio. 2014;5(6).

50. McCleary WR. The activation of PhoB by acetylphosphate. Mol Micro. 1996;20(6):1155-63.

51. Lee SY, De La Torre A, Yan D, Kustu S, Nixon BT, Wemmer DE. Regulation of the transcriptional activator NtrC1: Structural studies of the regulatory and AAA+ ATPase domains. Genes Dev. 2003;17(20):2552-63.

52. Wassmann P, Chan C, Paul R, Beck A, Heerklotz H, Jenal U, Schirmer T. Structure of BeF_3^--modified response regulator PleD: Implications for diguanylate cyclase activation, catalysis, and feedback inhibition. Structure. 2007;15(8):915-27.

53. Da Re S, Schumacher J, Rousseau P, Fourment J, Ebel C, Kahn D. Phosphorylation-induced dimerization of the FixJ receiver domain. Mol Micro. 1999;34(3):504-11.

54. Park S, Meyer M, Jones AD, Yennawar HP, Yennawar NH, Nixon BT. Two-component signaling in the AAA + ATPase DctD: Binding Mg^{2+} and BeF_3^- selects between alternate dimeric states of the receiver domain. FASEB J. 2002;16(14):1964-6.

55. Sheftic SR, White E, Gage DJ, Alexandrescu AT. NMR structure of the HWE kinase associated response regulator Sma0114 in its activated state. Biochemistry. 2014;53(2):311-22.

56. Park AK, Moon JH, Lee KS, Chi YM. Crystal structure of receiver domain of putative NarL family response regulator spr1814 from *Streptococcus pneumoniae* in the absence and presence of the phosphoryl analog beryllofluoride. Biochem Biophys Res Commun. 2012;421(2):403-7.

57. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids. 2000;28(1):235-42.

58. Bateman A, Birney E, Cerruti L, Durbin R, Etwiller L, Eddy Sean R, Griffiths-Jones S, Howe KL, Marshall M, Sonnhammer ELL. The Pfam Protein Families Database. Nucleic Acids. 2002;30(1):276-80.

59. Tatusova T, Ciufo S, Fedorov B, O'Neill K, Tolstoy I. RefSeq microbial genomes database: New representation and annotation strategy. Nucleic Acids Res. 2014;42(Database issue):D553-9.

60. Attwood PV, Besant PG, Piggott MJ. Focus on phosphoaspartate and phosphoglutamate. Amino Acids. 2011;40(4):1035-51.

61. Carpenter FH. The free energy change in hydrolytic reactions: The non-ionized compound convention. J Am Chem Soc. 1960;82(5):1111-22.

62. Hayes DM, Kenyon GL, Kollman PA. Theoretical calculations of the hydrolysis energies of some "high-energy" molecules. A survey of some biologically important hydrolytic reactions. J Am Chem Soc. 1978;100(14):4331-40.

63. Koshland DE. Effect of catalysts on the hydrolysis of acetyl phosphate. Nucleophilic displacement mechanisms in enzymatic reactions. J Am Chem Soc. 1952;74(9):2286-92.

64. Ruben EA, Plumley JA, Chapman MS, Evanseck JD. Anomeric effect in "high energy" phosphate bonds. Selective destabilization of the scissile bond and modulation of the exothermicity of hydrolysis. J Am Chem Soc. 2008;130(11):3349-58.

65. Feher VA, Cavanagh J. Millisecond-timescale motions contribute to the function of the bacterial response regulator protein Spo0F. Nature. 1999;400(6741):289-93.

66. Carter EA, Ciccotti G, Hynes JT, Kapral R. Constrained reaction coordinate dynamics for the simulation of rare events. Chemical Physics Letters. 1989;156(5):472-7.

67. Grubmüller H. Predicting slow structural transitions in macromolecular systems: Conformational flooding. Phys Rev E. 1995;52(3):2893-906.

68. Huber T, Torda A, van Gunsteren W. Local elevation: A method for improving the searching properties of molecular dynamics simulation. J Computer-Aided Mol Des. 1994;8(6):695-708.

69. Fiorin G, Klein ML, Hénin J. Using collective variables to drive molecular dynamics simulations. Mol Physics. 2013;111(22-23):3345-62.

70. Madhusudan M, Zapf J, Hoch JA, Whiteley JM, Xuong NH, Varughese KI. A response regulatory protein with the site of phosphorylation blocked by an arginine interaction: Crystal structure of Spo0F from *Bacillus subtilis*. Biochemistry. 1997;36(42):12739-45.

71. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protocols. 2015;10(6):845-58.

72. Eisenberg D, Luthy R, Bowie JU. VERIFY3D: assessment of protein models with three-dimensional profiles. Methods Enzymol. 1997;277:396-404.

73. Lovell SC, Davis IW, Arendall WB, 3rd, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC. Structure validation by Calpha geometry: phi,psi and Cbeta deviation. Proteins. 2003;50(3):437-50.

74. Emsley PL, Bernhard; Scott, W.G.; Cowtan, Kevan. Features and development of Coot. Acta Cryst D. 2010;66:15.

75. Kojetin DJ, Thompson RJ, Benson LM, Naylor S, Waterman J, Davies KG, Opperman CH, Stephenson K, Hoch JA, Cavanagh J. Structural analysis of divalent metals binding to the *Bacillus subtilis* response regulator Spo0F: The possibility for in vitro metalloregulation in the initiation of sporulation. Biometals. 2005;18(5):449-66.

76. Lukat GS, Stock AM, Stock JB. Divalent metal ion binding to the CheY protein and its significance to phosphotransfer in bacterial chemotaxis. Biochemistry. 1990;29(23):5436-42.

77. Stock AM, Martinez-Hackert E, Rasmussen BF, West AH, Stock JB, Ringe D, Petsko GA. Structure of the Mg^{2+}-bound form of CheY and mechanism of phosphoryl transfer in bacterial chemotaxis. Biochemistry. 1993;32(49):13375-80.

78. Gumbart J, Trabuco LG, Schreiner E, Villa E, Schulten K. Regulation of the protein-conducting channel by a bound ribosome. Structure. 2009;17(11):1453-64.

79. Zhang L, Hermans J. Hydrophilicity of cavities in proteins. Proteins. 1996;24(4):433-8.

80. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. J Mol Graphics. 1996;14(1):33-8.

81. Brooks BR, Brooks CL, 3rd, Mackerell AD, Jr., Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M. CHARMM: The biomolecular simulation program. J Comput Chem. 2009;30(10):1545-614.

82. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B. 1998;102(18):3586-616.

83. Damjanović A, García-Moreno E B, Brooks BR. Self-guided Langevin dynamics study of regulatory interactions in NtrC. Proteins. 2009;76(4):1007-19.

84. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. Scalable molecular dynamics with NAMD. J Comput Chem. 2005;26(16):1781-802.

85. Grant BJ, Rodrigues APC, ElSawy KM, McCammon JA, Caves LSD. Bio3d: an R package for the comparative analysis of protein structures. Bioinformatics. 2006;22(21):2695-6.

86. Team RDC. R: A language and environment for statistical computing. R Foundation. 2013.

87. Skjaerven L, Martinez A, Reuter N. Principal component and normal mode analysis of proteins; A quantitative comparison using the GroEL subunit. Proteins. 2011;79(1):232-43.

88. Kaufman L, Rousseeuw P. Clustering by means of medoids: Faculty of Mathematics and Informatics; 1987.

89. Maechler MR, P.; Struyf, A.; Hubert, M.; Hornik, K. Cluster: cluster analysis basics and extensions. R Package version 1.14.4. 2013.

90. Kaufman L, Rousseeuw PJ. Clustering Large Applications (Program CLARA). Finding Groups in Data: John Wiley & Sons, Inc.; 2008. p. 126-63.

91. Ichiye T, Karplus M. Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. Proteins. 1991;11(3):205-17.

92. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem. 2004;25(13):1605-12.

93. Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, Klebe G, Baker NA. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. Nucleic Acids Res. 2007;35(Web Server issue):W522-5.

94. La D, Esquivel-Rodriguez J, Venkatraman V, Li B, Sael L, Ueng S, Ahrendt S, Kihara D. 3D-SURFER: software for high-throughput protein surface comparison and analysis. Bioinformatics. 2009;25(21):2843-4.

95. Sael L, Li B, La D, Fang Y, Ramani K, Rustamov R, Kihara D. Fast protein tertiary structure retrieval based on global surface shape similarity. Proteins. 2008;72(4):1259-73.

96. Li B, Turuvekere S, Agrawal M, La D, Ramani K, Kihara D. Characterization of local geometry of protein surfaces with the visibility criterion. Proteins. 2008;71(2):670-83.

97. Schrodinger L. The PyMOL molecular graphics system, version 1.8. 2015.

98. Hennig C. Fpc: Flexible procedures for clustering. R package version 2.1-10. 2015.

99. Swaminathan S, Harte WE, Beveridge DL. Investigation of domain structure in proteins via molecular dynamics simulation: application to HIV-1 protease dimer. J Am Chem Soc. 1991;113(7):2717-21.

100. Pieniazek SN, Hingorani MM, Beveridge DL. Dynamical allosterism in the mechanism of action of DNA mismatch repair protein MutS. Biophys J. 2011;101(7):1730-9.

101. VanSchouwen B, Akimoto M, Sayadi M, Fogolari F, Melacini G. Role of Dynamics in the Autoinhibition and Activation of the Hyperpolarization-activated Cyclic Nucleotide-modulated (HCN) Ion Channels. J Biol Chem. 2015;290(29):17642-54.

102. Kormos BL, Baranger AM, Beveridge DL. A study of collective atomic fluctuations and cooperativity in the U1A-RNA complex based on molecular dynamics simulations. Journal of structural biology. 2007;157(3):500-13.

103. Wan H, Hu JP, Tian XH, Chang S. Molecular dynamics simulations of wild-type and mutants of human complement receptor 2 complexed with C3d. Physical chemistry chemical physics : PCCP. 2013;15(4):1241-51.

104. Xu L, Kong R, Zhu J, Sun H, Chang S. Unraveling the conformational determinants of LARP7 and 7SK small nuclear RNA by theoretical approaches. Molecular BioSystems. 2016;12(8):2613-21.

105. Bhattacharyya M, Vishveshwara S. Elucidation of the conformational free energy landscape in H.pylori LuxS and its implications to catalysis. BMC Structural Biology. 2010;10(1):1-15.

106. Lukman S, Grant BJ, Gorfe AA, Grant GH, McCammon JA. The Distinct Conformational Dynamics of K-Ras and H-Ras A59G. PLoS Comput Biol. 2010;6(9):e1000922.

107. Bren A, Eisenbach M. The N terminus of the flagellar switch protein, FliM, is the binding domain for the chemotactic response regulator, CheY. J Mol Biol. 1998;278(3):507-14.

108. Saito K, Ito E, Hosono K, Nakamura K, Imai K, Iizuka T, Shiro Y, Nakamura H. The uncoupling of oxygen sensing, phosphorylation signalling and transcriptional activation in oxygen sensor FixL and FixJ mutants. Mol Microbiol. 2003;48(2):373-83.

109. Kato A, Tanabe H, Utsumi R. Molecular characterization of the PhoP-PhoQ two-component system in *Escherichia coli* K-12: identification of extracellular Mg^{2+}-responsive promoters. J Bact. 1999;181(17):5516-20.

110. Tzeng YL, Hoch JA. Molecular recognition in signal transduction: the interaction surfaces of the Spo0F response regulator with its cognate phosphorelay proteins revealed by alanine scanning mutagenesis. J Mol Biol. 1997;272(2):200-12.

111. Maeda T, Wurgler-Murphy SM, Saito H. A two-component system that regulates an osmosensing MAP kinase cascade in yeast. Nature. 1994;369(6477):242-5.

112. Wolf A, Kirschner KN. Principal component and clustering analysis on molecular dynamics data of the ribosomal L11.23S subdomain. Journal of molecular modeling. 2013;19(2):539-49.

113. Duda RO, Hart PE. Pattern classification and scene analysis: Wiley; 1973.

114. Caliński T, Harabasz J. A dendrite method for cluster analysis. Stats Comms. 1974;3(1):1-27.

115. Varughese KI. Conformational Changes of Spo0F along the Phosphotransfer Pathway. J Bact. 2005;187(24):8221-7.

116. Zapf J, Sen U, Madhusudan, Hoch JA, Varughese KI. A transient interaction between two phosphorelay proteins trapped in a crystal

lattice reveals the mechanism of molecular recognition and phosphotransfer in signal transduction. Structure. 2000;8(8):851-62.

117. Fraccalvieri D, Pandini A, Stella F, Bonati L. Conformational and functional analysis of molecular dynamics trajectories by Self-Organising Maps. BMC Bioinformatics. 2011;12(1):1-19.

118. DeWitte RS, Shakhnovich EI. Pseudodihedrals: simplified protein backbone representation with knowledge-based energy. Protein Sci. 1994;3(9):1570-81.

119. Horie T, Tatebayashi K, Yamada R, Saito H. Phosphorylated Ssk1 prevents unphosphorylated Ssk1 from activating the Ssk2 mitogen-activated protein kinase kinase kinase in the yeast high-osmolarity glycerol osmoregulatory pathway. Molecular and cellular biology. 2008;28(17):5172-83.

Chapter 3: Simulating complexes from the Sln1 pathway

3.1 Introduction

In bacteria and certain eukaryotes, two-component signaling (TCS) is the most prevalent strategy for sensing and adapting to environmental stimuli [1, 2]. Signal transduction occurs through serial phosphorylation, facilitated by an upstream histidine kinase (HK) and a downstream response regulator (RR) protein (reviewed in [1]). The HK detects a specific stimulus, internalizes the signal, and responds by binding ATP and autophosphorylating on a histidine residue. The phosphoryl group is then passed to an aspartate on the receiver (rec) domain of an RR, which modulates the cellular response. Some organisms, like plants, fungi and certain bacteria, use an expanded version of

this system called a multi-step phosphorelay. These systems incorporate multiple phosphotransfer events into a single pathway [1, 2]. This more complex architecture coincides with the introduction of an intermediate signaling module known as a histidine phosphotransfer (HPt) protein [3, 4]. The HPt is a non-enzymatic domain containing a phosphorylatable histidine residue that acts as a relay point between the upstream and downstream components. In plants and fungi, HPt proteins often occupy branch positions in signaling pathways, transferring phosphoryl groups to multiple downstream RRs [5-8]. To prevent cross-talk between multiple partners and individual pathways, these interactions must be tightly regulated and/or highly specific [5-9]. The widespread use and importance of two-component signaling in the microbial world has drawn considerable attention in recent years. The absence of any known TCS systems in animals makes them ideal therapeutic targets against both pathogenic bacteria, such as *Bacillus anthracis* [9, 10], *Clostridium difficile* [11] and *Staphylococcus aureus* [12], as well as pathogenic fungi, such as *Cryptococcus neoformans* [13] and *Candida albicans* [14-16].

Saccharomyces cerevisiae has long been used as a model organism for research on two-component signaling, due to both its genetic tractability and its use of a single TCS system [17, 18]. Discovered in the 1990s, the Sln1 pathway is a branched, multi-step phosphorelay involved in the fungal osmoregulatory response [19-21]. At the head of the pathway is a membrane-

bound, hybrid sensor HK known as Sln1. The intermediate HPt protein, Ypd1, occupies a branch position, acting as a relay between the upstream Sln1 and two different downstream RR proteins, Ssk1 and Skn7 [3, 4]. How Ypd1 is able to differentiate between these multiple partners is of significant interest to researchers. Each rec domain is highly homologous structurally, though sequence similarity is only ~20-30%. Yeast two-hybrid data suggest that each rec domain in the Sln1 pathway shares a common hydrophobic docking site on Ypd1, formed by regions of the αA, αB and αC helices [22]. This appears to be a conserved feature of most HPt domains, as evidenced by studies on *E. coli* ArcB [23] and CheA-P1 [24, 25]. Reconstitution of the Sln1 pathway *in vitro* has allowed for limited biochemical and kinetic characterization of the proteins involved. The following provides a brief overview on the two best-characterized rec domains, Sln1-rec and Ssk1-rec, and their interactions with Ypd1.

Sln1 is a prototypical homodimeric, hybrid HK found at the head of the pathway [20]. The N-terminal portion contains an extracellular sensor domain linked to the C-terminal catalytic kinase and rec domains through a helical trans-membrane region. Under normal conditions, the kinase actively binds ATP and constitutively autophosphorylates on His576 within its catalytic region [18, 20]. As a hybrid HK, it then transfers the phosphoryl group to the phospho-accepting Asp1144 within its own C-terminal rec domain [21]. The ability of many rec domains to utilize small molecule phospho-donors, like

acetyl phosphate, suggests that this final step is catalyzed by the rec domain itself [26]. The phosphorylated rec domain then dissociates from the HK and binds to the downstream HPt, Ypd1 [27]. In the presence of hyperosmotic stress, the Sln1 kinase activity is diminished, allowing the accumulation of unphosphorylated downstream elements (reviewed in [28]).

Occupying an intermediate position in the pathway, the standalone Ypd1 is a relatively small (167 residue) protein that adopts the canonical four-helix bundle characteristic of most HPt domains. The phospho-accepting His64 is located in a surface-exposed region upon the αC (third) helix [29]. The adjacent surface forms the bulk of the binding site for the three rec domains within the pathway [22]. Phosphotransfer between Sln1-rec and Ypd1 occurs at a maximum rate of $29 \pm 3\ s^{-1}$ during the forward reaction [30]. While reverse transfer is possible and observable, based on *in vitro* experiments (maximum rate of $230 \pm 130\ s^{-1}$), the rapid dissociation of Sln1-rec and Ypd1 (~30x faster than the reverse transfer rate) ensures that 97% of Ypd1 remains phosphorylated [30]. The experimentally determined binding affinity of Ypd1 and Sln1-rec (K_d = 1.4 μM) classifies the complex as a weak-transient interaction, typical of a two-component signaling pathway ([30], West lab unpublished data). The combination of these factors results in a net forward direction of phosphotransfer within the upstream portion of the relay, moving from kinase to rec domain to HPt protein (Sln1-HK → Sln1-rec → Ypd1).

Located at the tail end of the Sln1 pathway, Ssk1 is a cytoplasmic, multi-domain RR protein (712 residues) that incorporates a large, N-terminal region of unknown function with a conserved, C-terminal rec domain [20]. The C-terminal region adopts the canonical $(\beta\alpha)_5$ topology shared by most rec domains (see Chapter 2), formed by a five-stranded β-sheet flanked by five amphipathic α-helices. The N-terminal regions of β1, β3 and β5 contain the conserved phosphorylatable aspartate residue (Asp554) and form the bulk of the Ypd1 binding surface. Under normal osmotic conditions, Ssk1-rec is constitutively phosphorylated, suppressing its ability to interact with the downstream MAP3Ks, Ssk2 and Ssk22 [20]. In response to osmotic shock, the accumulation of unphosphorylated Ssk1-rec activates Ssk2/22, which in turn activate the HOG pathway, resulting in the production of intracellular glycerol and the restoration of osmotic balance [20, 28].

Rapid-quench phosphotransfer experiments indicate that forward transfer from Ypd1 to Ssk1-rec occurs at a maximum rate of $160 \pm 70\ s^{-1}$ [30]. Interestingly, no reverse transfer was observable under the same conditions [30]. Unlike the upstream components, the experimentally determined binding affinity of Ypd1 and Ssk1-rec (K_d = 25 nM) is much stronger, indicative of a strong-transient interaction ([31], Katie Branscum, unpublished data). The kinetic data suggest that Ypd1 and Ssk1-rec spend a significant percentage of time in complex with each other. Overall, these characteristics propagate a strong forward direction of phosphotransfer within the pathway, from HPt to

downstream rec domain (Ypd1 → Ssk1-rec). *In vitro* radiolabeled phosphotransfer assays have shown that Ypd1 exhibits a significant kinetic preference for Ssk1-rec over both Sln1-rec and the pathway's other RR protein, Skn7-rec [30].

In addition to this unidirectional kinetic preference, phosphorylated Ssk1-rec exhibits an altered stability under certain conditions [32]. *In vitro* experiments suggest that the Sln1-rec, Ssk1-rec and Skn7-rec domains individually exhibit comparable phosphorylated half-lives of approximately 10-15 minutes. This is likely due to the chemical lability of the phosphoaspartate species, which is prone to rapid hydrolytic autodephosphorylation in solution [33]. However, the presence of Ypd1 dramatically extends the half-life of phosphorylated Ssk1-rec to over 40 hours (likely longer, but not observable due to degradation). This effect was not observed for either Sln1-rec or Skn7-rec [32].

The strong kinetic preference and stabilization of phosphorylated Ssk1-rec by Ypd1 are understandable from a physiological standpoint. As described above, Ssk1-rec must remain in a phosphorylated state to prevent unnecessary activation of the HOG pathway, as uncontrolled glycerol production would ultimately lead to cellular death [20, 28]. The high binding affinity and phosphostabilizing effect between Ypd1 and Ssk1-rec suggest that the proteins form a relatively stable complex, effectively shielding the phosphoaspartate from hydrolysis and sequestering the rec domain in an "inactive" state. When

cells are exposed to hyperosmotic conditions, the rapid efflux of water alters intracellular osmolyte concentrations and suppresses Sln1 kinase activity [28]. Without the upstream source of phosphoryl groups, unphosphorylated Ssk1-rec accumulates. In addition, increased concentrations of salt and glycerol are known to negatively affect the interaction between Ssk1-rec and Ypd1, removing the stabilizing effect on the phosphoaspartate [34]. The accumulation of unphosphorylated Ssk1-rec ultimately triggers the HOG pathway, resulting in the activation of genes involved in intracellular glycerol production and a return to osmotic balance. At higher concentrations (~1.0 M salt/glycerol), the previously destabilizing osmolytes were found to actually stabilize the interaction of Ssk1-rec and Ypd1, suggesting a regulatory mechanism for returning the pathway to homeostasis [34].

The co-crystal structures between the upstream Sln1-rec and Ypd1 provided valuable insights into the mechanisms governing the Sln1 pathway [29, 35]. However, little was revealed that might explain key differences between the interactions of Ypd1 with Sln1-rec and Ssk1-rec. Recently, a beneficial point mutation was discovered on Ssk1-rec (W638A) that allowed for the successful co-crystallization and structural characterization of Ssk1-rec W638A and Ypd1 (Katie Branscum, Smita Menon, unpublished). This complex structure explained much about the unique interface features between Ypd1 and its downstream partner, confirming and elaborating on the significant role that electrostatics play in their interaction. The Ssk1-rec

W638A mutant was characterized as a pseudo-wild-type variant, due to its comparable binding affinity and phosphotransfer ability. In addition, phosphorylated Ssk1-rec W638A is similarly stabilized by the presence of Ypd1. However, *in vitro* radiolabeled phosphotransfer assays indicate that reverse transfer from Ssk1-rec W638A to Ypd1 is possible, though to what extent is unknown.

Our understanding of the differences between Sln1-rec and Ssk1-rec function remains incomplete. Comparisons between the co-crystal complexes have shown structural features unique to each rec domain that help explain the differences in binding affinity and responses to osmolyte concentration. However, the static nature of the structures does not reveal how these features affect the interactions of the active sites. Outstanding questions that remain to be answered include how Ypd1 is able to stabilize the phosphorylated state of Ssk1-rec, as well as what causes the unidirectional nature of phosphotransfer within the pathway. In addition, the full effect of the W638A mutation is unknown. These gaps in knowledge highlight one of the most significant disadvantages of X-ray crystallography, that it provides only a snapshot of the protein in question. Dynamic structural information would provide valuable insight into these phenomena, allowing for a more complete characterization of the signaling pathway.

To the best of our knowledge, of the ~300,000 rec domain sequences within the NCBI database [36], approximately a dozen have been solved in co-

crystal structures with their cognate partner (either a phosphodonor or acceptor) [37]. Of these, only three also possess a co-crystal structure with the rec domain in a "phosphorylated" state [35, 38, 39]. This lack of structural information, particularly of allosterically modified conformations, is likely due to the difficulties of successfully co-crystallizing these signaling complexes. The chemical lability of phosphoaspartate also makes capturing a true "phosphorylated" rec domain challenging [33]. A previously proposed methodology (described and validated in Chapter 2) can be used to overcome this difficulty. Briefly, biased molecular dynamics can be used to directly sample rec domain phosphorylated conformations [40-42]. Phosphotransfer and rec domain activation occurs on high microsecond to low millisecond time scales [43-45]. Simulating such time scales is still largely impractical for the majority of the scientific community. Through the application of conserved active site restraints on the rec domain, we can drive the protein towards its phosphorylated conformation, enhancing sampling and reducing the conformational search space [35, 46, 47]. This same principle can be used to observe how phosphorylation and activation affect interaction between the rec domain and its cognate partners. We must stress that the application of the active site restraints is simply a tool to enhance and direct sampling towards the active conformations. These same conformations are theoretically accessible in unrestrained simulation, assuming proper phosphoaspartate parameters are used. However, the nature of rec domain activation would

require microsecond if not millisecond time scale simulations to adequately sample. Previous work indicates that applying even a few active site restraints reduces the required sampling time to within 100 ns. It also suggests that fully active conformations likely cannot be reliably accessed on this time scale without the full set of restraints.

The purpose of this work is to more fully characterize the regulatory interactions between Ypd1 and its upstream and downstream partners, Sln1-rec and Ssk1-rec. Combining biased MD and experimentally determined structural data, we simulated the activation of three artificially phosphorylated rec domains while in complex with Ypd1 (Sln1-rec, Ssk1-rec, and Ssk1-rec W638A). In addition, unrestrained simulations were performed on the apo complexes for comparative purposes. The dynamic data were then used to provide putative explanations for the unidirectional nature of the Sln1 multi-step phosphorelay, the role of Ypd1 in the stabilization of phosphorylated Ssk1-rec, and the mechanistic effects of the W638A mutation found in the recently obtained co-crystal structure. This information provides useful insights into the regulatory mechanisms of the Sln1 pathway, potential targets for future experimental studies and a further proof of concept for applying the biased MD methodology to signaling complexes.

3.2 Materials and Methods

Modeling and preparation

Initial crystal structures. Co-crystal structures of Sln1-rec•Ypd1 and Ssk1-rec W638A•Ypd1 were obtained from the PDB [37]. Sln1-rec has been crystallized in both an apo and BeF_3^- bound "active" form [29, 35]. Coordinates for the divalent metal cation (Mg^{2+}) during the Sln1-rec simulations were copied from the existing BeF_3^- bound structure. The Ssk1-rec W638A structure is only available in the apo form. To obtain the wild-type Ssk1-rec structure, the Ala638 in the crystal structure was mutated back to a tryptophan residue manually. The existing crystal structure of Ssk1-rec W638A•Ypd1 does not contain electron density for residues ~110-130 on Ypd1 and ~580-600 on Ssk1-rec. Both of these regions make up highly flexible loops located on surfaces distal to the binding interface and active sites. These segments were omitted from the Ssk1-rec•Ypd1 simulations. Coordinates for the metal cation in the Ssk1-rec models were obtained by aligning the rec domain structures with the BeF_3^- bound Sln1-rec and simply copying the metal atom [35].

Protein-protein docking. To demonstrate the viability of applying this methodology to a previously unknown dimeric structure, and to act as a control for the Sln1-rec•Ypd1 simulations, the High Ambiguity Driven protein-protein DOCKing (HADDOCK) server was used to dock known monomers of Sln1-rec and Ypd1, taken from the crystal structure (PDB 1OXB) [29, 48]. Likely

intermolecular contacts were identified as highly conserved positions found within all three rec domains (Sln1-rec, Ssk1-rec and Skn7-rec) that are known to be located at the interface with Ypd1, based on existing crystal structure. These residues were used as ambiguous interaction restraints (AIRs) to guide the HADDOCK complex docking protocol. For Ypd1, AIRs were identified using yeast two-hybrid data showing critical residues required for interaction with Sln1-rec *in vivo* [22]. The following lists the residues used as active restraints. Passive restraints were automatically determined as adjacent residues.

Sln1-rec residues used as active restraints:

Asp1095, Asn1096, Asn1099, Asp1144, Pro1148, Lys1195, Pro1196

Ypd1 residues used as active restraints:

Ile13, Glu16, Met20, Asp21, Asp23, Asp24, Glu27, Leu31, Gln38, Asp60, Phe65, Lys67, Gly68, Ser69, Ser70, Leu73, Gln76, Trp80, Glu83

To demonstrate the utility of this method in sampling a true active complex, even when the starting structure is an imperfectly docked model, the inactive subunit of Sln1-rec in the Sln1-rec•Ypd1 HADDOCK model was then replaced with a "pre-activated" Sln1-rec chosen from the top representative structures predicted in Chapter 2. This generated an imperfect co-complex, with the

phosphodonor and acceptor atoms out of alignment. The model was then simulated as described below.

System preparation. Starting models were prepared from inactive crystal structures unless otherwise described. Phosphoaspartate was modeled into the active sites with Coot [49]. A Mg^{2+} atom was also added based on available experimental coordinates from the co-crystal structure of Sln1-rec and Ypd1 bound to BeF_3^- (PDB 2R25)[35]. As shown in Chapter 2, precise positioning of the metal cation is not critical for the activation of the protein as long as the metal is present near the active site. Structures were stripped of all other crystallographic waters, ligands and subunits prior to preparation and simulation.

Dowser was used to fill internal water cavities with a probe radius of 0.2 Å [50, 51]. The SOLVATE program was then used to create a contoured solvent shell around each model with a shell thickness of 6.0 Å calculated using 6 gaussians [41]. Systems were loaded into VMD and processed using the AutoPSF plug-in [52]. Finally, structures were immersed in a full orthorhombic water box with 15.0 Å padding on each side using a TIP3P water model with the VMD solvate plug-in [52]. Systems were neutralized and K^+/Cl^- ions were added to a final concentration of 100 mM using the VMD Autoionize plug-in [52].

Phosphoaspartate. Parameters for dianionic phosphoaspartate (AST) were defined for the CHARMM36 force field based on Damjanović et al.

(2009) [53-55]. For a full list of partial charges used for AST, see **Table 3** in Chapter 2.

Molecular dynamics

Simulations. Simulations were prepared and run using VMD and NAMD 2.9 [52, 56] with the CHARMM36 protein force field [53, 54]. A 12.0 Å cutoff was used for van der Waals interactions with a switching function distance of 10.0 Å. Long-range electrostatics were evaluated using the PME (Particle Mesh Ewald) method with a tolerance, interpolation order and grid spacing of $10e^{-6}$, 4.0 and 1.0 Å, respectively. A two fs integration step was used for all simulations. The SHAKE algorithm was applied to constrain covalent bonds involving hydrogen atoms throughout. Initial energy minimization was performed on non-backbone atoms for 100 ps of NVT MD. An all-atom minimization step was then done for an additional 100 ps of NVT MD with harmonic constraints of 0.5 kcal/mol applied to all Cα atoms. Following this, systems were heated to 310 K by increments of 1 K/ps over approximately 310 ps of NVT MD. Equilibration was then performed with Cα atoms restrained for 500 ps of NPT MD. A final, unrestrained equilibration step of 4 ns NPT MD was run to prepare systems for production. A total of 50-100 ns of effective production data were then collected following equilibration in 10 ns partial runs.

The following table lists the multiple setups for each simulation. Systems were simulated under identical conditions unless otherwise specified.

Table 11. Simulations variants performed in this study.

Proteins used	*Initial model source*	*Ligands present*
Sln1-rec•Ypd1	Crystal (PDB 1OXB)	Mg^{2+}, PO_3^{2-}
Sln1-rec•Ypd1	Crystal (PDB 1OXB)	N/A
Sln1-rec•Ypd1	HADDOCK* (model)	Mg^{2+}, PO_3^{2-}
Ssk1-rec•Ypd1	Crystal (unpublished)	Mg^{2+}, PO_3^{2-}
Ssk1-rec•Ypd1	Crystal (unpublished)	N/A
Ssk1-rec W638A•Ypd1	Crystal (unpublished)	Mg^{2+}, PO_3^{2-}
Ssk1-rec W638A•Ypd1	Crystal (unpublished)	N/A

*An "pre-activated" model of Sln1-rec was first inserted into a docked complex between Ypd1 and Sln1-rec predicted using experimentally determined restraints. The complex was them simulated as previously described.

Simulation restraints. Biasing terms were applied during the simulations to enrich sampling of active rec domain conformations. Restraints are frequently used to steer systems towards desired states for study. Kinetic data suggest that phosphotransfer and rec domain conformational transitions occur on millisecond time scales [41, 45]. Simulating such events at all-atomic resolutions is impractical for most research laboratories. The application of restraints can reduce the amount of time required by decreasing the conformational search space. In NAMD, parameters that regulate this are known as collective variables (colvars) [57]. Using the colvars module, we applied half-harmonic potential restraints to Sln1-rec and Ssk1-rec during simulation to drive the formation of the appropriate active site geometry. **Table 12** provides a brief summary of these restraints. Formatted colvars input files (.in) are also provided for both Sln1-rec and Ssk1-rec. These colvars

parameters were chosen to successfully achieve a phosphotransfer-competent active site geometry.

Table 12. Summary of colvars applied to drive active site formation.

Type	*Group1 atoms**	*Group2 atoms*	*Boundaries*
Distance	Asp1144 OD1	Mg^{2+}	1.9-2.2 Å
CoordNum	Asp1144 PO_3^{2-}	Mg^{2+}	1 group
Distance	Mg^{2+}	COM	0.0-0.1 Å
Distance	Ala1174 N	Asp57 OP2	2.0-3.3 Å
CoordNum	Asp1095 R	Mg^{2+}	1 group
Distance	Asp1095 OD1	Mg^{2+}	0.0-2.2 Å
Distance	Asp1144 OP1	Mg^{2+}	0.0-2.2 Å
Distance	Asp1094 OE1	Mg^{2+}	0.0-2.2 Å
Distance	Asp1146 O	Mg^{2+}	0.0-2.2 Å

*Residue numbers based on Sln1-rec

Formatted NAMD colvars (.in) file for Sln1-rec.

```
colvarsTrajFrequency        5000
colvarsRestartFrequency     5000

colvar {
name asp1144od1_to_mg       #distance of carbonyl to mg
width 0.1

lowerboundary 1.9
upperboundary 2.2
lowerWallConstant 100.0
upperWallConstant 100.0

distance {
group1 { atomNumbers 3595 }
group2 { atomNumbers 2673 }
}
}
```

```
colvar {
name asp1144phos_coord_to_mg      #coordination of phosphoryl to mg

lowerboundary 0.0
upperboundary 1.0
lowerWallConstant 100.0
upperWallConstant 100.0
coordNum {
group1 { atomNumbers 3598 3599 3600 }
group2 { atomNumbers 2673 }
cutoff 3.3
}
}

colvar {
name mgcom                                #anchor mg to its own starting
coordinates
width 0.001

lowerboundary 0.0
upperboundary 0.1
lowerWallConstant 1000.0
upperWallConstant 1000.0

distance {
group1 { atomNumbers 2673 }
group2 { dummyAtom (-6.600, 5.708, 12.497) }
}
}

colvar {
name ala1174n_to_asp1144ot                #distance of ala to phosphoryl ot
width 0.001

lowerboundary 2.6
upperboundary 2.8
lowerWallConstant 100.0
upperWallConstant 100.0

distance {
group1 { atomNumbers 4069 }
group2 { atomNumbers 3599 }
}
}
```

```
colvar {
name asp1095_coord_to_mg            #coordination of asp to mg

lowerboundary 0.0
upperboundary 1.0
lowerWallConstant 100.0
upperWallConstant 100.0
coordNum {
group1 { atomNumbers 2817 2818 2819 }
group2 { atomNumbers 2673}
cutoff 3.3
}
}

colvar {
name asp1144o1p_to_mg               #distance of asp to mg
width 0.05

lowerboundary 2.2
upperboundary 2.3
lowerWallConstant 100.0
upperWallConstant 100.0

distance {
group1 { atomNumbers 3598 }
group2 { atomNumbers 2673 }
}
}

colvar {
name asp1095od1_to_mg               #distance of asp to mg

lowerboundary 0.0
upperboundary 2.2
lowerWallConstant 1000.0
upperWallConstant 1000.0

distance {
group1 { atomNumbers 2818 }
group2 { atomNumbers 2673 }
}
}
```

```
colvar {
name gln1146co_to_mg        #distance of mainchain carbonyl (d+2) to mg

lowerboundary 0.0
upperboundary 2.2
lowerWallConstant 100.0
upperWallConstant 100.0

distance {
group1 { atomNumbers 3635 }
group2 { atomNumbers 2673 }
}
}

colvar {
name glu1094oe1_to_mg               #distance of glu to mg

lowerboundary 4.0
upperboundary 4.5
lowerWallConstant 100.0
upperWallConstant 100.0

distance {
group1 { atomNumbers 2806 }
group2 { atomNumbers 2673 }
}
}
```

Formatted NAMD colvars (.in) file for Ssk1-rec.

```
colvarsTrajFrequency        5000
colvarsRestartFrequency     5000

colvar {
name asp554od1_to_mg               #distance of asp to mg
width 0.1

lowerboundary 1.9
upperboundary 2.2
lowerWallConstant 100.0
upperWallConstant 100.0

distance {
group1 { atomNumbers 4884 }
```

```
group2 { atomNumbers 9 }
}
}

colvar {
name asp554phos_coord_to_mg       #coordination of phosphoryl to mg

lowerboundary 0.0
upperboundary 1.0
lowerWallConstant 100.0
upperWallConstant 100.0
coordNum {
group1 { atomNumbers 12 13 14 }
group2 { atomNumbers 4884 }
cutoff 3.3
}
}

colvar {
name mgcom                        #anchor mg to its own starting
coordinates
width 0.001

lowerboundary 0.0
upperboundary 0.1
lowerWallConstant 1000.0
upperWallConstant 1000.0

distance {
group1 { atomNumbers 4884 }
group2 { dummyAtom (29.561, 74.060, 50.744) }
}
}

colvar {
name ala611n_to_asp554ot          #distance of ala to phosphoryl
width 0.001

lowerboundary 2.6
upperboundary 2.8
lowerWallConstant 100.0
upperWallConstant 100.0

distance {
```

```
group1 { atomNumbers 4093 }
group2 { atomNumbers 13 }
}
}

colvar {
name asp511_coord_to_mg              #coordination of asp to mg

lowerboundary 0.0
upperboundary 1.0
lowerWallConstant 100.0
upperWallConstant 100.0
coordNum {
group1 { atomNumbers 2746 2747 2748 }
group2 { atomNumbers 4884}
cutoff 3.3
}
}

colvar {
name asp554o1p_to_mg                 #distance of asp to mg
width 0.05

lowerboundary 2.2
upperboundary 2.3
lowerWallConstant 100.0
upperWallConstant 100.0

distance {
group1 { atomNumbers 4884 }
group2 { atomNumbers 12 }
}
}

colvar {
name asp511od1_to_mg                 #distance of asp to mg

lowerboundary 0.0
upperboundary 2.2
lowerWallConstant 1000.0
upperWallConstant 1000.0

distance {
group1 { atomNumbers 4884 }
```

```
group2 { atomNumbers 2747 }
}
}

colvar {
name gln556co_to_mg          #distance of mainchain carbonyl (d+2) to mg

lowerboundary 0.0
upperboundary 2.2
lowerWallConstant 100.0
upperWallConstant 100.0

distance {
group1 { atomNumbers 4884 }
group2 { atomNumbers 3478 }
}
}

colvar {
name glu510oe1_to_mg                  #distance of glu to mg

lowerboundary 4.0
upperboundary 4.5
lowerWallConstant 100.0
upperWallConstant 100.0

distance {
group1 { atomNumbers 4884 }
group2 { atomNumbers 2735 }
}
}
```

In addition to the regular restraints, an extra distance restraint was used for the simulation involving the "pre-activated" Sln1-rec•Ypd1 model. During the equilibration stage, a catalytic pre-transition state geometry was enforced upon the donor phosphoaspartate (Asp1144) and acceptor histidine (His64) residues. A distance restraint of 2.2-3.5 Å was applied between the Nε2 atom on the histidine residue and the phosphoryl group, mimicking the distances observed

in known BeF_3^- structures. This effectively reorients the imperfectly docked complex by pulling the donor and acceptor together. The restraint was removed after equilibration and was not enforced during production runs.

Data analysis

Trajectory preparation. Trajectories were analyzed using VMD and the Bio3D package in R [52, 58, 59]. Production runs were stripped of all non-protein atoms using VMD. Partial trajectories were aligned by Cα atoms using the RMSDTT plug-in and combined to form single trajectories for analysis [52].

Consensus maps. MDCons was used to calculate and score intermolecular residue contacts (ICs) during the simulations [60]. Two residues were considered intermolecular contacts if at least two heavy atoms were detected within 5.0 Å of each other during simulation. Stripped production runs were first loaded into VMD. Residues on Ypd1 were designated chain A and residues on the rec domain were designated chain B. Individual frames (snapshots) were exported individually at 50 ps intervals, using the following Tcl script:

```
for {set i 1} {$i < 2000} {incr i} {
[atomselect top all frame $i] writepdb frame_$i.pdb }
```

The resulting 2000 snapshots were then used to calculate a consensus score for each intermolecular contact that was detected. Percentages correlate with the frequency at which each contact is detected. Scores were output to MATLAB for plotting (MATLAB v2016a, MathWorks Inc., © 2016).

Intermolecular contacts (ICs) with conservation scores <15% were eliminated. Scores >70% were considered stable. Available crystal structures (PDB 1OXB; 2R25; unpublished Ssk1-rec W638A) were similarly processed. Crystallographic contacts were considered stable for comparison purposes.

Buried solvent accessible surface area (SASA) calculations. The interfacial buried surface area between the rec domain and Ypd1 in each complex was determined using the following formula:

$$Buried\ Interface\ Surface\ Area = (SASA_A + SASA_B) - SASA_{AB}$$

Calculations were done using the following Tcl script template:

```
set fp [open "SASA-output_complex.dat" w]
set nf [molinfo top get numframes]
set complex [atomselect top "protein and noh"]
for {set i 0} {$i < $nf} {incr i} {
$complex frame $i
Set sasa_complex [measure sasa 1.40 $complex –restrict]
puts $fp "$i $sasa_complex"
}
close $fp
```

MATLAB was used to plot the buried interface surface area for each trajectory frame. Mean interfacial SASA was calculated by averaging the calculated SASA over all frames in the trajectories.

Hydrogen bond detection. The H-bond plugin within VMD was used to calculate the number of potential hydrogen bonds formed between specified atom selections during each simulation [52]. Hydrogen bonding occupancies were also calculated between the catalytic His64 on Ypd1 and the phosphoryl group and/or phosphorylatable aspartate residue on the rec domains. A

hydrogen bond distance cutoff of 4.0 Å and an angle cutoff of 35^0 between heavy atom donor and acceptor were used for each calculation. The number of hydrogen bonds per trajectory frame was tracked and plotted using MATLAB.

Distance calculations. The distance between the His64 Nε2 atom on Ypd1 and the phosphorus atom on the rec domain was plotted as a function of time using VMD's native atom selection tools. For apo structures, the distance between the His64 Nε2 atom on Ypd1 and the phosphorylatable aspartate CG atom on the rec was used. Mean distances were calculated by averaging over all frames in the combined trajectories.

RMSD calculations. Full trajectories were aligned using Cα atoms. Cα RMSD values were calculated for each frame and graphed using MATLAB. Mean RMSDs were calculated by averaging over all frames in the combined trajectories.

Active site hydration analysis. Water accessibility of the active sites during simulation was analyzed using several techniques. First, the number of intact water molecules within 5.0 Å of the phosphorylatable aspartate residues was tracked for each frame within the trajectories using the following Tcl script (using Sln1-rec residue numbering):

```
set mol [molinfo top]
set out [open "water_within_5A_of_resid1144.txt" w]
set sel [atomselect $mol "(water and oxygen within 5 of (protein and
resid 1144 and name P))"]
set frames [molinfo $mol get numframes]
for {set i 0} {$i < $frames} {incr i} {
$sel frame $i
set n [$sel num]
```

```
puts $out "$i $n"
}
close $out
```

The mean number of water molecules near the phosphorylatable residue was calculated by averaging the total number of waters detected over all frames in the combined trajectories. To more quantitatively represent the amount of water around and within the active site during the simulations, the average volume occupied by water molecules (density using water oxygen atoms) within the active site was calculated using the VolMap plugin within VMD [52]. Fully solvated trajectories were first combined at 50 ps frame intervals (~2000 snapshots) and aligned to a Cα trace. Water densities were calculated with a grid spacing of 1.0 Å, producing volumetric maps containing the average density of water oxygen atoms over a matrix of cubic voxels (1.0 $Å^3$ each) observed during the simulation. Higher density correlates with a greater the number of water molecules. Water occupancies were also calculated using the VolMap plugin, showing the percent occupancies of water molecules averaged over each trajectory. Occupancy indicates how frequently the space is occupied by water molecules. Each volumetric data set was visualized in VMD using a three-dimensional isosurface positioned above the phosphorylatable aspartate (drawn at 15% and 50% for density and occupancy, respectively).

Electrostatic potential maps. Electrostatic potential for each trajectory was computed using the PME method within NAMD [56, 61, 62]. Potential maps were obtained using the PMEpot plugin with VMD using a grid

resolution of 1 and an Ewald factor of 0.25. Maps were visualized using a two-dimensional slice along the XY-plane, showing a heat map of the active site centered on the phosphoryl group, and a molecular surface view with each atom colored by potential. Potential values were scaled between -15 and +30 *kT/e*.

Software. Structural alignments and analyses were performed using UCSF Chimera [63]. Figures were generated in both PyMOL (PyMOL, v1.8. 2015, Schrodinger LLC, © 2016) and VMD [52]. Data plotting was performed in MATLAB (MATLAB v2016a, MathWorks Inc., © 2016) and R [58]. MD simulations were performed with NAMD [56] on resources provided by the OU Supercomputing Center for Education and Research (OSCER) at the University of Oklahoma. System preparation, visualization and initial trajectory analyses were performed in VMD [52].

3.3 Results and Discussion

Effects of phosphorylation and activation on complex stability

Overall effects of phosphorylation on the stability and structure of the signaling complexes were examined using the following comparisons: phosphorylated vs. apo Sln1-rec•Ypd1; phosphorylated vs. pre-activated Sln1-rec•Ypd1; phosphorylated vs. apo wild-type Ssk1-rec•Ypd1; phosphorylated vs. apo Ssk1-rec W638A•Ypd1. Effective production simulations (100 ns) were collected for each regular system. In each phosphorylated scenario, harmonic restraints were applied centered around the Mg^{2+} atom bound to the

rec domain active site. Apo complexes were left completely unrestrained. Standard structural checks revealed proper rec domain active site alignment in each of the systems, as well as characteristic switch residue and conformational changes upon activation (typically in the β4α4-β5α5 regions) during the simulations. The only exception to this was the orientation of Phe1192 on Sln1-rec. In the active crystal structure, the conserved aromatic ring rotamerizes inward, to point towards the phosphoryl group. In previous monomeric simulations, Phe1192 readily rotamerizes, but only rarely was this observed in the dimeric simulations. These findings suggests that Phe1192 likely rotamerizes prior to binding with Ypd1. While this appears to have no effect on the overall conformation of active Sln1-rec, the pre-active Sln1-rec•Ypd1 HADDOCK model was used as a control for comparison. As generating the model produced an imperfect catalytic active site, an extra distance restraint was applied between His64 on Ypd1 and the phosphoryl group on Sln1-rec to enforce ideal phosphotransfer geometry. This restraint was only present during the initial equilibration period. 50 ns of effective production data were collected using the normal restraints described above. The resulting data set was then compared to the normal Sln1-rec•Ypd1 simulations.

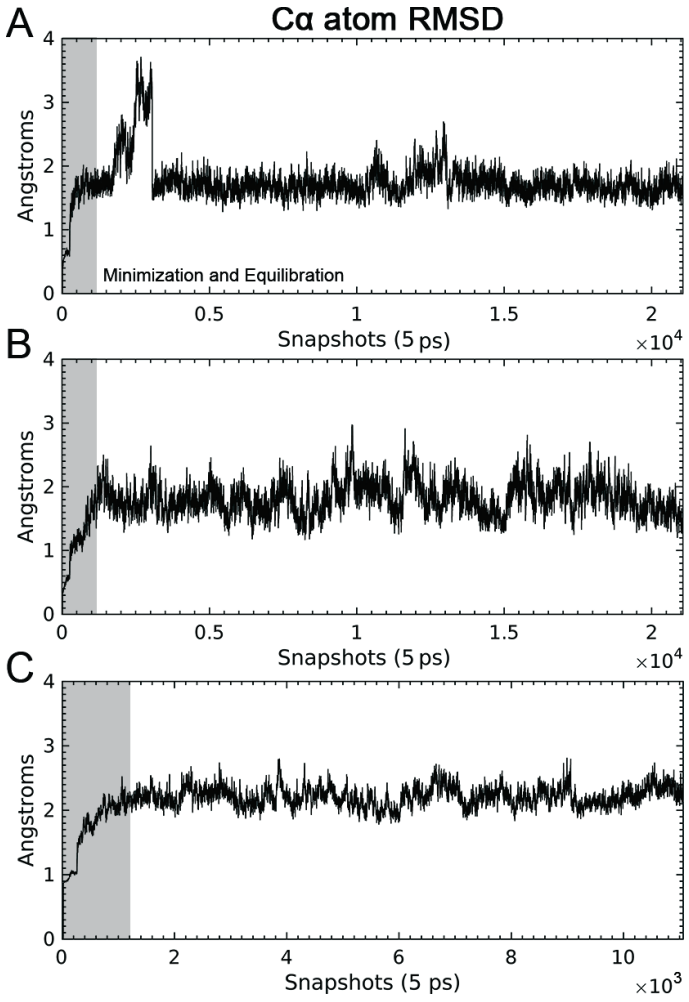

Figure 43. RMSD traces from Sln1-rec•Ypd1 simulations.

Cα deviations for combined trajectories. (A) Phosphorylated Sln1-rec•Ypd1 (100 ns). (B) Apo Sln1-rec•Ypd1 (100 ns). (C) Pre-activated phosphorylated Sln1-rec•Ypd1 (50 ns). Shaded regions represent minimization/equilibration segments.

Convergence and initial stabilization upon phosphorylation were monitored using total Cα RMSD. Each trajectory was aligned to its Cα trace using the RMSDTT plugin. Corresponding inactive X-ray crystal structures were used as reference points. RMSD measurements typically reached a plateau within the first 500-1000 snapshots (2.5-5.0 ns) for both apo and phosphorylated systems. **Fig. 43** shows RMSD traces per frame for

phosphorylated (**A**) and apo (**B**) Sln1-rec•Ypd1 and the pre-activated phosphorylated (**C**) Sln1-rec•Ypd1 complex. Curiously, barring the anomalous segment within the phosphorylated data set between frames ~2000 and 3000, the traces exhibit noticeably different profiles. The apo complex (**B**) appears to fluctuate slightly but significantly more than the phosphorylated system (**A**). This effect may be partially attributable to the application of the active site restraints, as would be expected in this scenario. The trend is maintained in the pre-activated complex system (**C**) as well. However, examination of the phosphorylated Ssk1-rec•Ypd1 and apo Ssk1-rec•Ypd1 traces (**Fig. 44A-B**) shows a reversal of this phenomenon, with the apo structure exhibiting less fluctuation and an overall lower level of deviation.

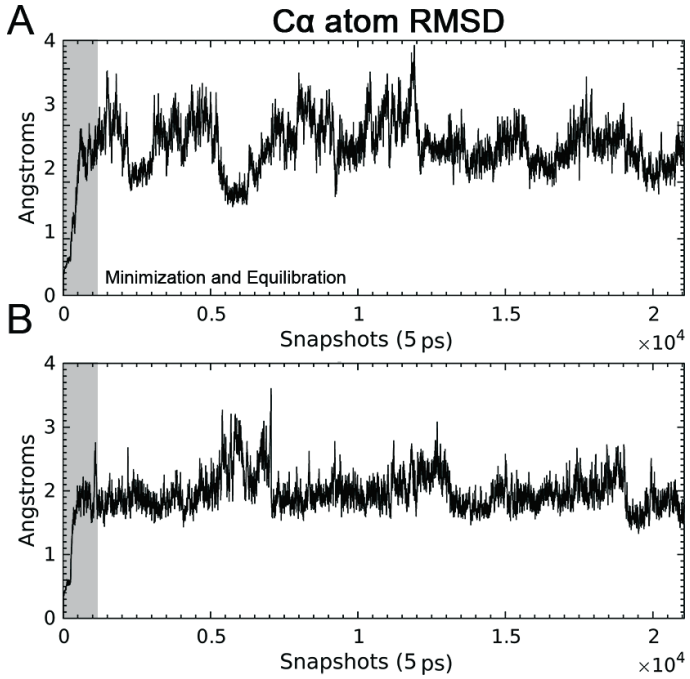

Figure 44. RMSD traces from wild-type Ssk1-rec•Ypd1 simulations.

Cα deviations for combined trajectories. (A) Phosphorylated Ssk1-rec•Ypd1 (100 ns). (B) Apo Ssk1-rec•Ypd1 (100 ns). Shaded regions represent minimization/equilibration segments.

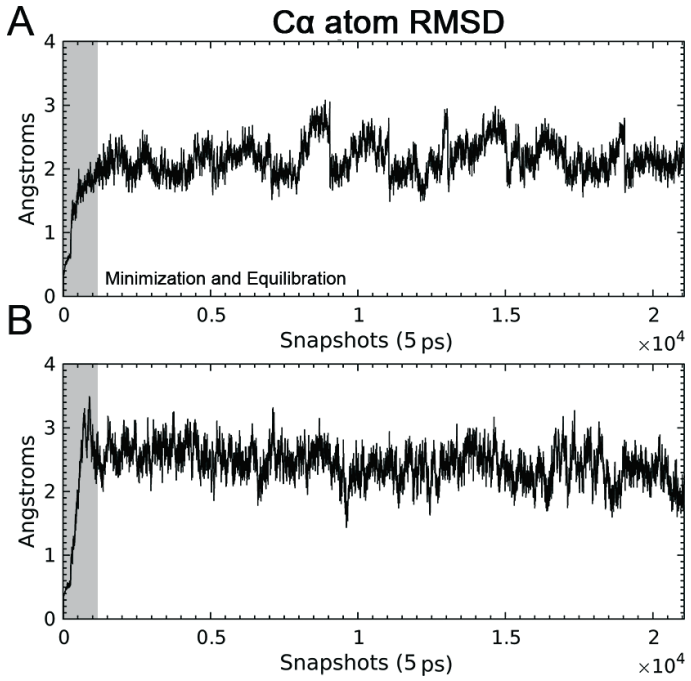

Figure 45. RMSD traces from the Ssk1-rec W638A•Ypd1 simulations.

Cα deviations for combined trajectories. (A) Phosphorylated Ssk1-rec W638A•Ypd1 (100 ns). (B) Apo Ssk1-rec W638A•Ypd1 (100 ns). Shaded regions represent minimization/equilibration segments.

Ssk1-rec W638A•Ypd1 simulations show RMSD profiles more similar to the Sln1-rec•Ypd1 data, though the disparity between apo and phosphorylated systems is slightly greater (**Fig. 45A-B**). **Table 13** lists average Cα RMSD values and standard deviations calculated over the combined production trajectories.

Table 13. Cα RMSDs for heterodimeric complexes (production only)

Proteins	*Average Cα RMSD*	*Standard deviation*
P~Sln1-rec•Ypd1	1.77 Å	0.31 Å
P~Sln1-rec•Ypd1*	N/A	N/A

Sln1-rec•Ypd1	1.81 Å	0.25 Å
P~Ssk1-rec•Ypd1	2.66 Å	0.42 Å
Ssk1-rec•Ypd1	1.98 Å	0.28 Å
P~Ssk1-rec W638A•Ypd1	2.16 Å	0.25 Å
Ssk1-rec W638A•Ypd1	2.43 Å	0.27 Å

*Pre-activated phosphorylated Sln1-rec docked to Ypd1, 50 ns trajectory.

To compare the relative stabilities of the observed protein-protein interfaces, we measured the buried interfacial surface area for each complex. The interfacial surface area was defined as the sum of the solvent accessible surface area (SASA) for both monomers minus the SASA of the total protein complex. ~2000 snapshots were produced for each trajectory at 50 ps intervals (25 ps for pre-activated model). The interfacial SA was computed for each frame and plotted as a function of time to observe changes in interfacial SASA.

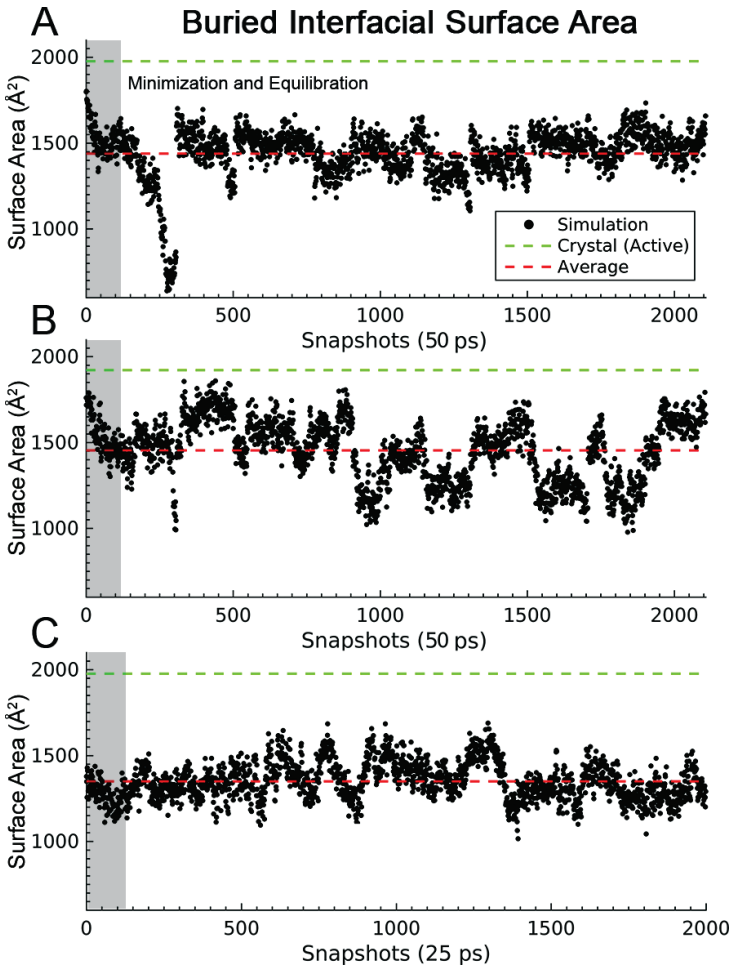

Figure 46. Interfacial surface area for Sln1-rec•Ypd1.

The change in buried interfacial surface area ([$SASA_{Ypd1}$ + $SASA_{Sln1\text{-}rec}$] -$SASA_{Complex}$) plotted for each trajectory frame. (A) Interfacial surface area for phosphorylated Sln1-rec•Ypd1. (B) Interfacial surface area for apo Sln1-rec•Ypd1. (C) Interfacial surface area for pre-activated phosphorylated Sln1-rec•Ypd1.

Fig. 46 shows a comparison of the relative interface stabilities for phosphorylated and apo Sln1-rec•Ypd1. It is immediately apparent that the phosphorylated complex forms a more stabilized protein-protein interface in terms of surface area (again, barring the anomalous event within frames ~200-300), with the apo form exhibiting significantly larger fluctuation. Stabilization of the binding interface and active site has been hypothesized as a feature of

Sln1-rec to aid in the formation of a phosphotransfer-competent alignment with His64 on Ypd1. The average surface area for both apo and phosphorylated complexes remained comparable, though both systems showed large, almost immediate decreases in interfacial surface area relative to their corresponding crystal structures. The significance of this effect is unknown at this time, though it is likely a natural result of molecular "breathing" and directly affected by interface flexibility. Crystallization conditions and crystal packing may also affect the "tightness" of the interface, factors that are obviously absent in the *in silico* studies. Comparison to the pre-activated phosphorylated Sln1-rec•Ypd1 system (**Fig. 46C**) maintains the trend, suggesting increased stability as a result of phosphorylation.

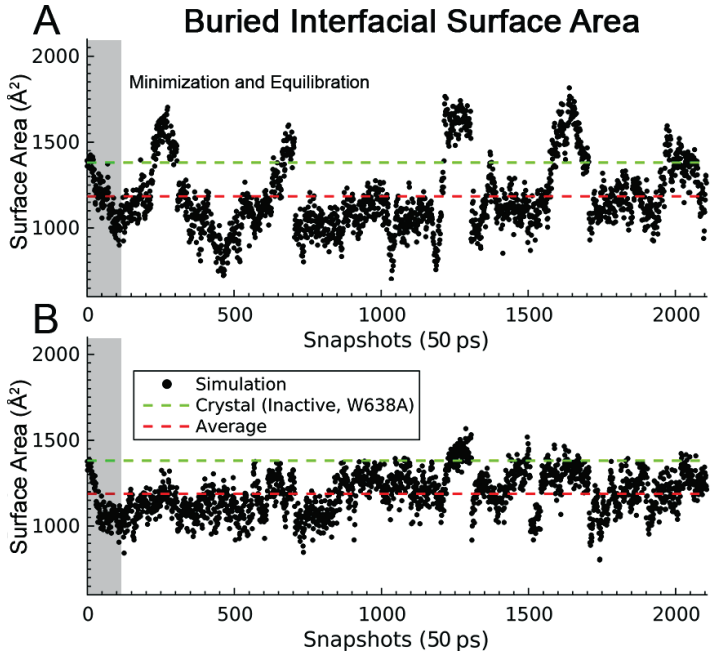

Figure 47. Interfacial surface area for wild-type Ssk1-rec•Ypd1.

The change in buried interfacial surface area ([$SASA_{Ypd1}$ + $SASA_{Ssk1\text{-}rec}$] -$SASA_{Complex}$) plotted as a function of trajectory frame. (A) Interfacial surface area for phosphorylated wild-type Ssk1-rec•Ypd1. (B) Interfacial surface area for apo wild-type Ssk1-rec•Ypd1.

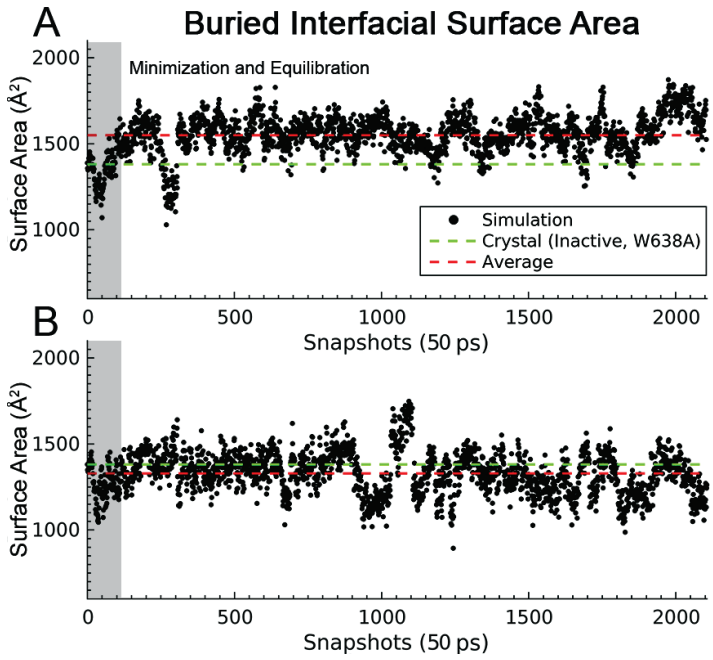

Figure 48. Interfacial surface area for Ssk1-rec W638A•Ypd1.

The change in buried interfacial surface area ([$SASA_{Ypd1}$ + $SASA_{Ssk1\text{-}rec}$] -$SASA_{Complex}$) plotted as a function of trajectory frame. (A) Interfacial surface area for phosphorylated Ssk1-rec W638A•Ypd1. (B) Interfacial surface area for apo Ssk1-rec W638A•Ypd1.

As before, the complex structure of wild-type Ssk1-rec•Ypd1 tells a different story (**Fig. 47**). For these systems, phosphorylation appears to "weaken" the protein-protein interface (relative to the co-crystal structure), increasing surface area fluctuation. The apo and phosphorylated Ssk1-rec W638A•Ypd1 mutant complexes (**Fig. 48**) produced results more similar to Sln1-rec•Ypd1. In the mutant system, phosphorylation appears to simultaneously stabilize and increase the size of the interface, causing a 300 $Å^2$ jump in the buried surface area relative to the co-crystal structure. Due to the

time scales used for these simulations (50-100 ns), it is unlikely that the larger decreases in interfacial surface area were the result of true dissociation events. Each complex remained relatively intact for the duration of the simulations. **Table 14** lists average interfacial surface areas and standard deviations calculated for each trajectory.

Table 14. Average buried interfacial surface area.

Proteins	*Data set*	*Interfacial SA*	*Std.*
P~Sln1-rec•Ypd1	Simulation	1438.4 (1460.3) Å^2	147.22 (102.10) Å^2
P~Sln1-rec•Ypd1	HADDOCK Model	N/A	N/A
Sln1-rec•Ypd1	Simulation	1454.0 Å^2	179.0 Å^2
Sln1-rec•Ypd1	Crystal (1OXB)	1921.2 Å^2	N/A
BeF_3^- Sln1-rec•Ypd1	Crystal (2R25)	1976.4 Å^2	N/A
P~Ssk1-rec•Ypd1	Simulation	1183.9 Å^2	207.2 Å^2
Ssk1-rec•Ypd1	Simulation	1187.6 Å^2	120.4 Å^2
P~Ssk1-rec W638A•Ypd1	Simulation	1549.2 (1570.4) Å^2	117.1 (95.4) Å^2
Ssk1-rec W638A•Ypd1	Simulation	1328.4 Å^2	121.8 Å^2
Ssk1-rec W638A•Ypd1	Crystal (unpublished)	1380.7 Å^2	N/A

(Values in parentheses were calculated by excluding anomalous frames ~200-300)

Intermolecular interactions related to phosphorylation

The program MDCons was used to construct consensus contact maps for each of the complexes [60]. This involved ranking detected intermolecular contacts (ICs), defined as any two inter-subunit atoms of distinct residues within 5.0 Å of one another, based on the frequency of occurrence within a trajectory. Consensus maps provide a level of dynamic detail on specific intermolecular interactions that may be hidden or misleading in static crystal structures. They are a relatively simple yet robust method for detecting patterns

within protein-protein interactions that may be affected by events such as ligand binding or post-translational modification and related conformational shifts. They also allow quantification of interface conservation, as a measure of similarity between different snapshots. Using these measurements, we can infer properties about protein-protein complexes such as relative stability and interface flexibility. Additionally, areas of interest can be easily identified as targets for future experimental work.

2000 snapshots were extracted (at 50 ps intervals) from the production trajectories, representing 100 ns of effective simulation for each system. These were processed for formatting and used as input to the MDCons program. The resulting ICs and conservation rates (frequencies) were then imported into MATLAB for further analysis. For each simulation, a set of visual consensus maps was generated and compared to a corresponding set of co-crystal structures. Crystallographic contact maps were calculated and each contact was arbitrarily set to 100% conservation. ICs conserved in <15% of frames were discarded. The remaining residue pairs were plotted and colored according to conservation rate. This provided a filtered overview regarding the intermolecular interactions of each complex, useful for detecting changes in overall patterns. Contacts conserved in ≥70% of frames were classified as "stable" interactions. Stable ICs were plotted separately and examined for overlap with the crystallographic contacts. Comparisons between apo and phosphorylated forms were also made using the stable IC lists. Remaining

contacts between 15 and 70% conservation were classified as "transient." These were plotted to examine overlap with the remaining crystallographic pairs that was not detected among the stable IC list. MDCons was used to calculate C_{90}, C_{70} and C_{50} scores, representing the fraction of maintained ICs in 90%, 70% and 50% of frames, respectively. Consistently higher values correspond approximately to higher interfacial stability. **Table 15** shows these three parameters computed for each of the simulated systems. The scores again suggest that phosphorylation stabilizes the interfaces within Sln1-rec•Ypd1 and Ssk1-rec W638A•Ypd1, but strongly destabilizes the interface within wild-type Ssk1-rec•Ypd1.

Table 15. Overall IC conservation scores.

Proteins	C_{90}	C_{70}	C_{50}
Sln1-rec•Ypd1	57 %	76%	100%
P~Sln1-rec•Ypd1	63%	79%	92%
Ssk1-rec•Ypd1	61%	78%	89%
P~Ssk1-rec•Ypd1	50%	64%	87%
Ssk1-rec W638A•Ypd1	62%	73%	96%
P~Ssk1-rec W638A•Ypd1	62%	80%	96%

Fig. 49 shows a comparison between the filtered consensus maps generated from the phosphorylated and apo simulations of Sln1-rec•Ypd1. Seven main IC groupings are visible, encompassing the following areas on Ypd1: the majority of the αA and the N-terminal half of the αB helix; the majority of the αC helix and its C-terminal linker; the C-terminal region of the

αD helix. Overall, more stable contacts are seen on the phosphorylated map (yellow), specifically within groups 1, 2 and 4. Each of these groups involves interactions with the α4-β5-α5 region on Sln1-rec. Crystallographic data from other rec domains suggest this region is at the center of the allosteric changes induced by phosphorylation, particularly involving the conserved Thr/Ser switch residue found on β4-strand and the β4α4 loop. The increase in the number of stable ICs corroborates the previously observed trend in buried interfacial surface area, suggesting that phosphorylation helps to stabilize the interaction between Sln1-rec and Ypd1.

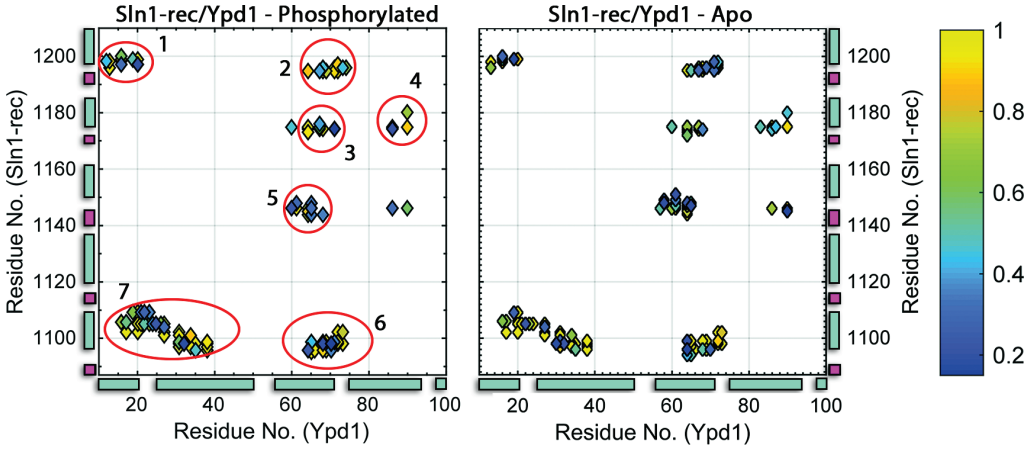

Figure 49. Filtered consensus intermolecular contact maps for Sln1-rec•Ypd1 simulations.

Combined ICs with conservation scores ≥15% were graphed and colored based on conservation. (Left panel) Filtered ICs for phosphorylated Sln1-rec•Ypd1. (Right panel) Filtered ICs for apo Sln1-rec•Ypd1. A conservation score of 1.0 corresponds to detection in 100% of frames. β-strands (magenta); α-helices (cyan).

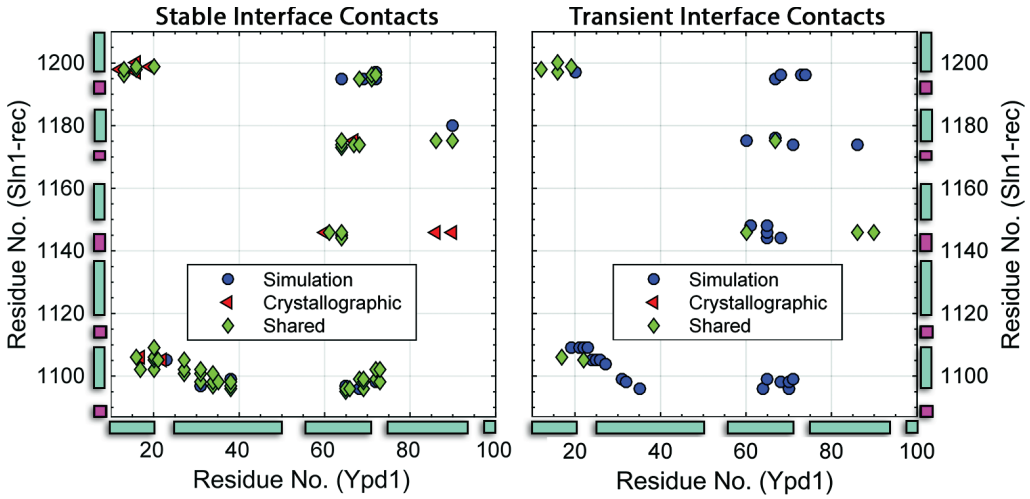

Figure 50. Comparisons between simulated and crystallographic ICs for phosphorylated Sln1-rec•Ypd1.

ICs with conservation scores ≥70% were classified as stable. Remaining ICs with scores between 15% and 70% were classified as transient. (Left panel) Stable ICs were compared with crystallographic contacts. Unique ICs seen only in the simulation are shown in blue. Crystallographic ICs not found as stable contacts in the trajectory are shown in red. Overlapping contacts are shown in green. (Right panel) Transient ICs were compared with the remaining crystallographic contacts not found in the left panel. Unique transient ICs observed only in the simulation are shown in blue. Truly unique crystallographic contact pairs found in neither stable nor transient classes are shown in red. Overlapping contact pairs are shown in green. β-strands (magenta); α-helices (cyan).

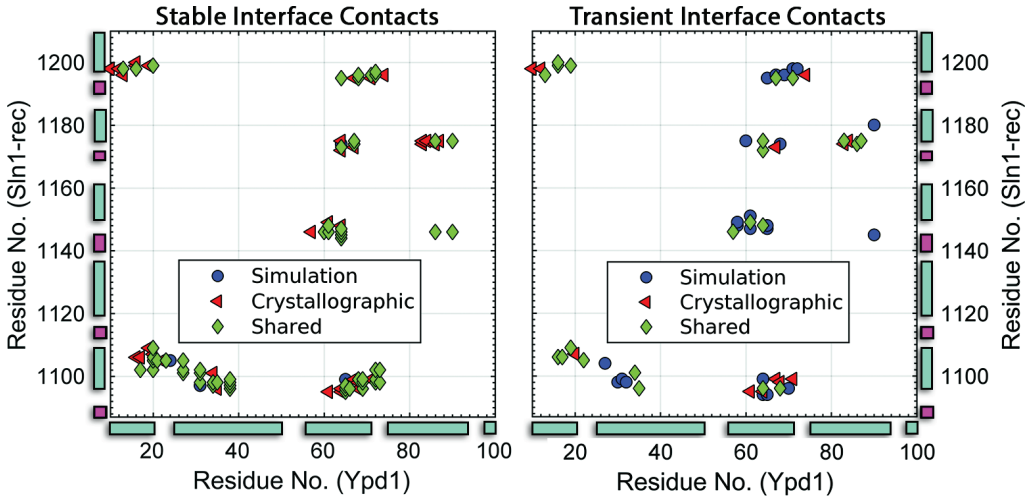

Figure 51. Comparisons between simulated and crystallographic ICs for apo Sln1-rec•Ypd1.

ICs with conservation scores ≥70% were classified as stable. Remaining ICs with scores between 15% and 70% were classified as transient. (Left panel) Stable ICs were compared with crystallographic contacts. Unique ICs seen only in the simulation are shown in blue. Crystallographic ICs not found as a stable contacts in the trajectory are shown in red. Overlapping contacts are shown in green. (Right panel) Transient ICs were compared with the remaining crystallographic contacts not found in the left panel. Unique transient ICs observed only in the simulation are shown in blue. Truly unique crystallographic contact pairs found in neither stable nor transient classes are shown in red. Overlapping contact pairs are shown in green. β-strands (magenta); α-helices (cyan).

Fig. 50 shows a comparison between the simulated ICs and the BeF_3^- bound crystal structure for Sln1-rec•Ypd1 (PDB 2R25). Most crystallographic contacts are detected as stable IC pairs from the simulation, with the few remaining contacts being accounted for by the detected transient ICs. Every crystallographic inter-residue pair was identified within the simulation data. In addition, numerous unique IC pairs were observed that were not seen in the crystal complex. These novel residue pairs are listed in **Table 16.** The results demonstrate that not only is the significant interaction data provided by the co-crystal structure replicated by the simulation data, but the novel ICs provide prime targets for future mutagenesis studies.

Table 16. Novel ICs observed in phosphorylated Sln1-rec•Ypd1.

Ypd1 - Residue	***Sln1-rec - Residue***
Ala72	Ile1197
Ala72	Lys1195
Ala72	Val1098
Arg90	Asn1180
Asp23	Arg1105
Gln34	His1097
Gln38	Asn1099
Gln38	His1097

Gly68	Asn1096
Leu31	His1097
Phe65	His1097
Ser69	Lys1195

Many novel pairs were also observed within the transient group, though at relatively low conservation levels, suggesting they are likely a result of flexibility around the binding interface and not filling specific functional roles. Similar comparisons were performed on the apo complex against the inactive co-crystal structure (PDB 1OXB). **Fig. 51** shows a greater number of crystallographic contacts not observed in the stable residue pairs list. The vast majority of these contacts were observed as transient ICs during the simulation, but several crystallographic contacts were absent from this list (see **Table 17**). Each of these ICs were detected in the simulation with conservation scores below the 15% cutoff, except for Met20-Leu1107, which was completely absent.

Table 17. Crystallographic ICs not observed in >15% of frames of the apo Sln1-rec•Ypd1 simulation.

Ypd1 - Residue	***Sln1-rec - Residue***
Ala71	Asn1099
Arg84	Phe1175
Asn10	Lys1198
Asn61	Asp1095
Glu83	Ala1174
Gly68	Val1098
Gly74	Pro1196

His64	Asp1095
Lys67	Asn1099
Lys67	Thr1173
*Met20	*Leu1107
Thr12	Lys1198

*All pairs were detected below the 15% cutoff except for Met20-Leu1107, which was not observed during simulation.

Stable ICs for the apo and phosphorylated Sln1-rec•Ypd1 trajectories were compared to identify interface changes as a result of phosphorylation (**Fig. 52**). A large increase in unique stable ICs associated with phosphorylation was observed in areas near the β4α4-β5α5 region of Sln1-rec, corresponding to regions of the αA, αC and parts of the αD helices on Ypd1. In the apo simulation, a similar effect is visible at residues adjacent to the Sln1-rec active site and involving parts of the αC and αD helices on Ypd1. The loss of these stable contacts upon phosphorylation is likely a response to the additional volume introduced by the phosphoryl group in the active site cavity. **Table 18** lists ICs identified as stable contacts (>70%) only in the phosphorylated Sln1-rec•Ypd1 simulation. Key interactions of significant interest include Arg90-Asn1180 and Gly68-Ala1174. Arg90 on Ypd1 has been speculated to aid in stabilizing the αC-D helices, as well as contributing to the formation of the active site geometry through Gln1146 on Sln1-rec. No prior studies have suggested its interaction with Asn1180, located on the α4 helix that undergoes a large conformational shift upon phosphorylation of Sln1-rec. Disruption of this interaction may have unforeseen consequences on the

conformational shift upon activation of the rec domain. The highly conserved residues Gly68 on Ypd1 and Ala1174 (also called the T+1 position) on Sln1-rec have both been implicated in affecting the catalysis of phosphotransfer (see Chapter 4). **Table 19** contains a similar list found only in the apo Sln1-rec•Ypd1 simulation. Key interactions include Arg90-Gln1146, Asp60-Gln1146, Gln86-Gln1146 and Lys67-Phe1175, all of which have been postulated to be involved in the formation of a phosphotransfer-competent active site.

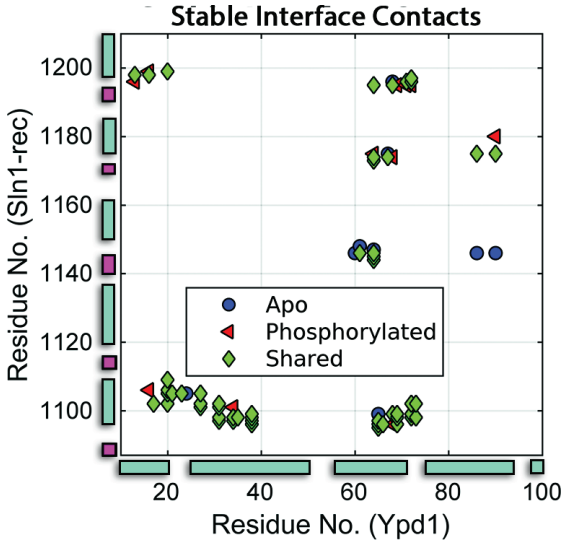

Figure 52. Comparison of stable ICs calculated from apo and phosphorylated Sln1-rec•Ypd1 simulations.

Stable residue pairs extracted from the apo and phosphorylated trajectories were overlaid to identify interactions novel to each form. ICs only observed in the apo simulation are in blue. ICs only observed in the phosphorylated simulation are in red. Overlapping ICs found in both simulations are in green. β-strands (magenta); α-helices (cyan).

Table 18. Stable ICs unique to phosphorylated Sln1-rec•Ypd1.

Ypd1 - Residue	*Sln1-rec - Residue*
Ala71	Lys1195

Ala72	Lys1195
Arg90	Asn1180
Gln34	Glu1101
Glu16	Arg1199
Glu16	Met1106
Gly68	Ala1174
Gly68	Asn1096
His64	Phe1175
Ile13	Pro1196
Ser69	Lys1195

Table 19. Stable ICs unique to apo Sln1-rec•Ypd1.

Ypd1 - Residue	***Sln1-rec - Residue***
Arg90	Gln1146
Asn61	Pro1148
Asp24	Arg1105
Asp60	Gln1146
Gln86	Gln1146
Gly68	Pro1196
His64	Met1147
Lys67	Phe1175
Phe65	Asn1099

Fig. 53 shows the stable interface maps of the normal phosphorylated Sln1-rec•Ypd1 system overlaid with the pre-activated HADDOCK model of Sln1-rec•Ypd1. Due to its nature as a docked model, the pre-activated complex almost certainly exhibits more transient contacts overall than the crystallographic-based simulation. However, the two trajectories share

impressive similarity. All but two of the stable ICs detected in the pre-activated complex are also detected in the normal simulation. These two ICs are present in the more transient IC list for the regular simulation, suggesting they were simply lower than the 70% cutoff used for the stable ICs. Taken with the previous data, the results indicate that the information provided by the normal activated complex and the pre-activated HADDOCK model is comparable, with both revealing highly similar characteristics. The significance of these findings is not limited to their role as a control for this study. They also suggest that an active complex can be generated even from an imperfect protein-protein interaction, simply by temporarily enforcing alignment of the catalytic donor-acceptor pair. The pre-activated control demonstrates how a protein-protein complex of previously unknown structure can be predicted and simulated to extract novel and useful information.

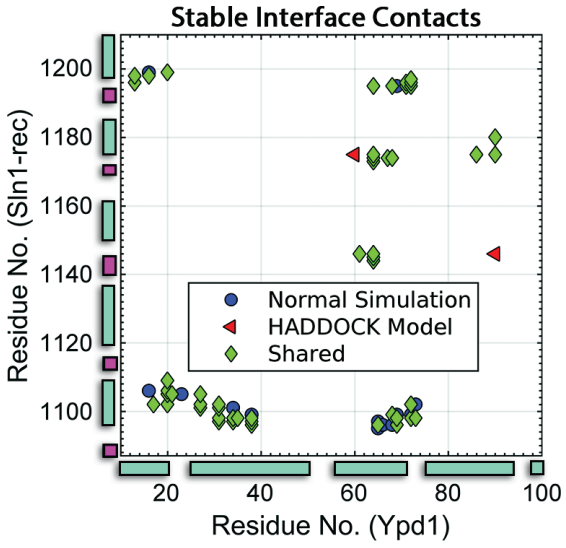

Figure 53. Comparison of stable ICs calculated from the original phosphorylated Sln1-rec•Ypd1 and the HADDOCK model simulations.

Stable residue pairs extracted from the phosphorylated trajectories were overlaid to identify interactions novel to each form. ICs only observed in the original phosphorylated simulation are in blue. ICs only observed in the HADDOCK model simulation are in red. Overlapping ICs found in both simulations are in green. β-strands (magenta); α-helices (cyan).

Fig. 54 shows a comparison between the filtered consensus maps calculated for the phosphorylated and apo wild-type Ssk1-rec•Ypd1 simulations. Seven major IC groupings are visible, similar to those detected in the Sln1-rec•Ypd1 interactions. Immediate differences are visible between the phosphorylated and apo maps. In the apo simulations, group 2, (αC helix on Ypd1 and β5α5 region on Ssk1-rec) contains an increased number of stable interactions. In addition, groups 3-4 (αC and αD helices on Ypd1 and β4α4 region of Ssk1-rec) disappear almost completely in the unphosphorylated complex. This is different from the Sln1-rec•Ypd1 complexes, where these ICs are detectable in both apo and phosphorylated forms. Overall, a general decrease in IC stability was detected upon phosphorylation. This result again corroborates the previously observed trend in buried interfacial surface area, further supporting the hypothesis that phosphorylation actually destabilizes the interface between wild-type Ssk1-rec and Ypd1.

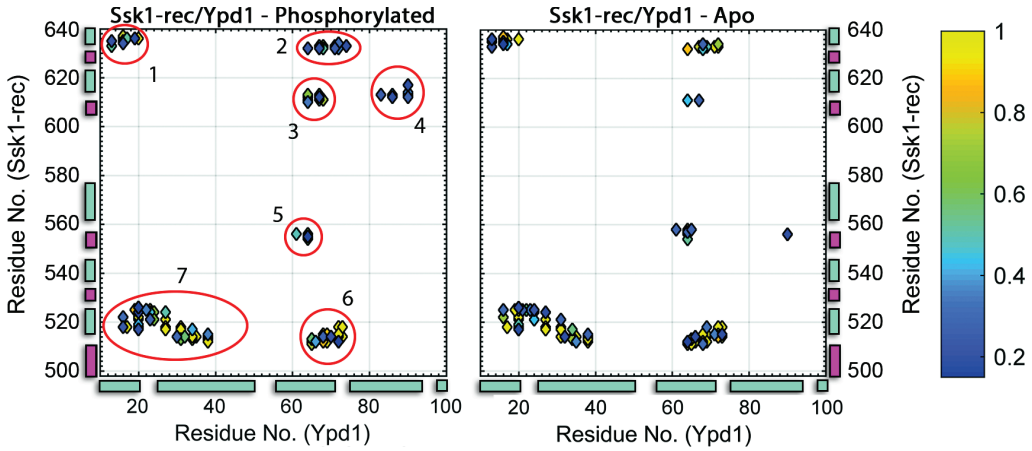

Figure 54. Filtered consensus intermolecular contact maps for wild-type Ssk1-rec•Ypd1 simulations.

Combined ICs with conservation scores ≥15% were graphed and colored based on conservation. (Left panel) Filtered ICs for phosphorylated Ssk1-rec•Ypd1. (Right panel) Filtered ICs for apo Ssk1-rec•Ypd1. A conservation score of 1.0 corresponds to detection in 100% of frames. β-strands (magenta); α-helices (cyan).

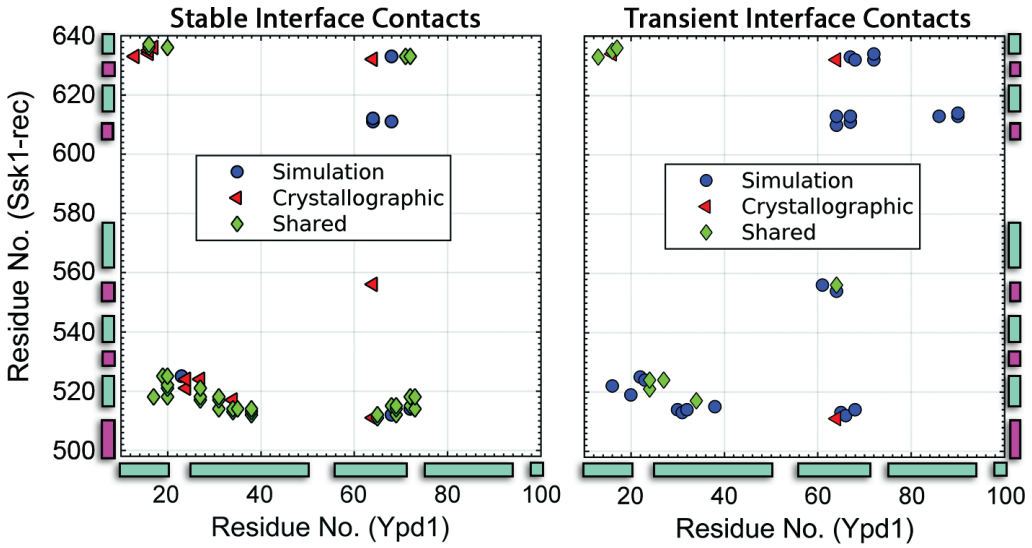

Figure 55. Comparisons between simulated and crystallographic ICs for phosphorylated wild-type Ssk1-rec•Ypd1.

ICs with conservation scores ≥70% were classified as stable. Remaining ICs with scores between 15% and 70% were classified as transient. (Left panel) Stable ICs were compared with crystallographic contacts. Unique ICs seen only in the simulation are shown in blue. Crystallographic ICs not found as a stable contact in the trajectory are shown in red. Overlapping contacts are shown in green. (Right panel) Transient ICs were compared with the

remaining crystallographic contacts not found in the left panel. Unique transient ICs observed only in the simulation are shown in blue. Truly unique crystallographic contact pairs found in neither stable nor transient classes are shown in red. Overlapping contact pairs are shown in green. β-strands (magenta); α-helices (cyan).

Fig. 55 shows a comparison between the simulated ICs for wild-type phosphorylated Ssk1-rec•Ypd1 and the inactive co-crystal structure for Ssk1-rec W638A•Ypd1 (no other crystal structure is currently available). Nearly all crystallographic contacts were detected within the simulation data, either as stable or transient ICs, with only His64-Asp511 being absent. This is likely a result of phosphorylation, as this pair is detected within the apo simulation data set. Multiple stable IC pairs were observed that are not seen in the co-crystal structure. **Table 20** lists those novel ICs with conservation scores ≥70%. Residues of interest include Gly68-Ala611, His64-Ala611 and His64-Ser612. Gly68 and Ala611 have been implicated in affecting the catalysis of phosphotransfer. Additionally, the interactions His64-Ala611 and His64-Ser612 may be involved in the reversibility of the phosphotransfer event.

Table 20. Novel ICs observed in phosphorylated wild-type Ssk1-rec•Ypd1.

Ypd1 - Residue	***Ssk1-rec - Residue***
Ala72	Ile514
Asp23	Lys525
Glu16	His637
Gly68	Ala611
Gly68	Asn512
Gly68	Pro633
His64	Ala611
His64	Ser612

Fig. 56 shows the same comparison for the apo wild-type Ssk1-rec•Ypd1 simulation and the apo Ssk1-rec W638A•Ypd1 co-crystal structure. Again, every crystallographic contact is accounted for within the filtered IC list (>15%), and multiple novel interactions were detected. **Table 21** lists ICs observed in the apo Ssk1-rec•Ypd1 simulation that are not present in the co-crystal structure of Ssk1-rec W638A•Ypd1.

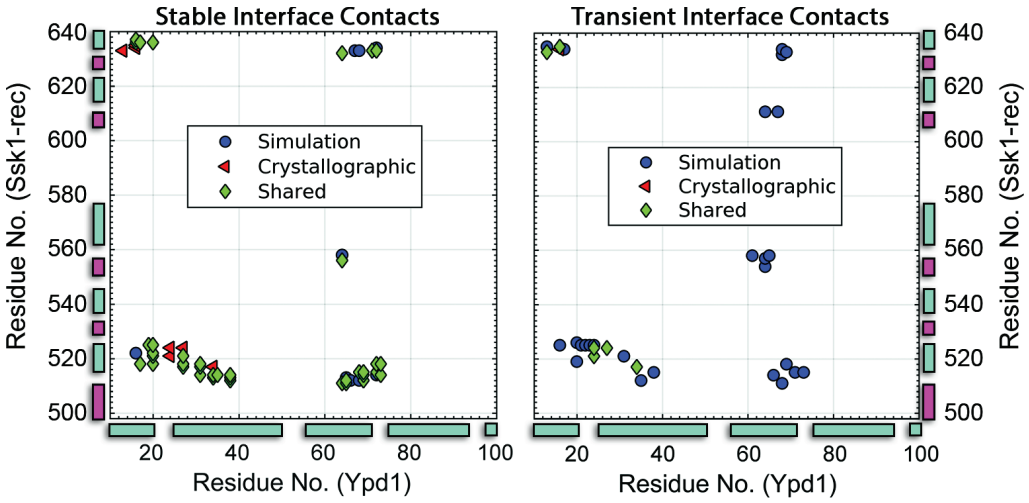

Figure 56. Comparison between simulated and crystallographic ICs for apo wild-type Ssk1-rec•Ypd1.

ICs with conservation scores ≥70% were classified as stable. Remaining ICs with scores between 15% and 70% were classified as transient. (Left panel) Stable ICs were compared with crystallographic contacts. Unique ICs seen only in the simulation are shown in blue. Crystallographic ICs not found as a stable contact in the trajectory are shown in red. Overlapping contacts are shown in green. (Right panel) Transient ICs were compared with the remaining crystallographic contacts not found in the left panel. Unique transient ICs observed only in the simulation are shown in blue. Truly unique crystallographic contact pairs found in neither stable nor transient classes are shown in red. Overlapping contact pairs are shown in green. β-strands (magenta); α-helices (cyan).

Table 21. Novel ICs observed in apo wild-type Ssk1-rec•Ypd1.

Ypd1 - Residue	*Ssk1-rec - Residue*

Ala72	Ile514
Ala72	Val634
Glu16	Phe522
Gly68	Asn512
Gly68	Pro633
His64	Pro558
Leu66	Asn512
Lys67	Pro633
Phe65	Val513

Stable contact pairs detected during the apo and phosphorylated Ssk1-rec•Ypd1 trajectories were compared to identify the effects of phosphorylation on the interface (**Fig. 57**). The apo contact map shows strong concentrations of stable ICs involving the αA and αC helices on Ypd1 and the β5α5 region on Ssk1-rec. These ICs are largely absent in the phosphorylated simulation, suggesting a significant shift in the orientation of the complex. Similarly, a small but highly conserved concentration of ICs was seen in the phosphorylated simulation involving the αC helix of Ypd1 and the β4α4 region of Ssk1-rec. These contacts appear as poorly conserved, transient ICs in the apo simulation and are absent in the co-crystal structure. Additionally, a large group of transient ICs found in the phosphorylated simulation (shown in **Fig. 54 group 4**) are completely absent in both the apo simulation and crystal complex. **Table 22** lists ICs identified as stable contacts (>70%) only in the phosphorylated wild-type Ssk1-rec•Ypd1 simulation. Key interactions of interest include those involving Ala611 and Ser612, both of which have been

implicated in affecting the catalysis of phosphotransfer. **Table 23** contains a similar list found only in the apo Ssk1-rec•Ypd1 simulation.

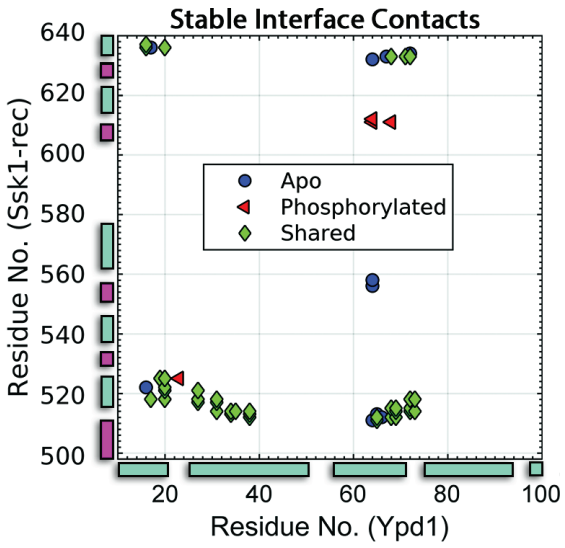

Figure 57. Comparison of stable ICs calculated for apo and phosphorylated wild-type Ssk1-rec•Ypd1.

Stable residue pairs extracted from the apo and phosphorylated trajectories overlaid to identify interactions novel to each form. ICs only observed in the apo simulation are in blue. ICs only observed in the phosphorylated simulation are in red. Overlapping ICs found in both simulations are in green. β-strands (magenta); α-helices (cyan).

Table 22. Stable ICs unique to phosphorylated wild-type Ssk1-rec•Ypd1.

Ypd1 - Residue	*Ssk1-rec - Residue*
Asp23	Lys525
Gly68	Ala611
His64	Ala611
His64	Ser612

Table 23. Stable ICs unique to apo wild-type Ssk1-rec•Ypd1.

Ypd1 - Residue	*Ssk1-rec - Residue*
Ala72	Val634

Glu16	Phe522
His64	Asp511
His64	Gln556
His64	Lys632
His64	Pro558
Ile17	Leu 636
Leu66	Asn512
Lys67	Pro633
Phe65	Val513

Fig. 57 shows a comparison between the filtered consensus maps calculated for the phosphorylated and apo Ssk1-rec W638A•Ypd1 simulations. Seven major IC groupings are again present, like those found in the wild-type Ssk1-rec•Ypd1 simulation. However, phosphorylation of the mutant complex appears to increase the stability of the IC pairs, specifically in areas involving the αC-D helices on Ypd1 and the β3α3, β4α4 and β5α5 regions on Ssk1-rec W638A. Phosphorylation stabilizes the contacts within groups 2, 3 and 4. Additionally, groups 1 and 5 show an increase in the overall number of ICs. These results suggest divergent characteristics between the wild-type and mutant Ssk1-rec•Ypd1 complexes that may help explain their kinetic differences.

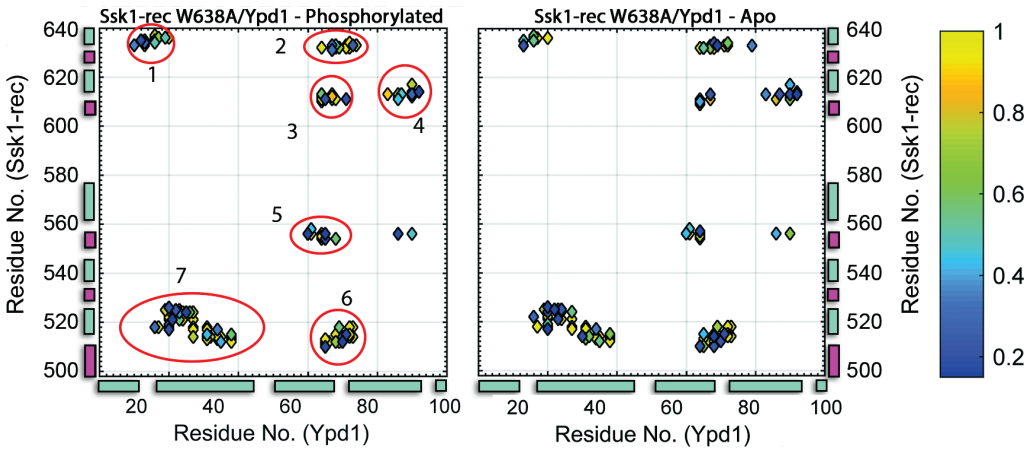

Figure 58. Filtered consensus intermolecular contact maps for Ssk1-rec W638A•Ypd1 simulations.

Detected ICs with conservation scores lower than 15% (corresponding to 15% frequency of trajectory frames) were discarded. Remaining ICs were graphed and colored based on conservation scores. (Left panel) Filtered ICs for phosphorylated Ssk1-rec W638A•Ypd1. (Right panel) Filtered ICs for apo Ssk1-rec W638A•Ypd1. A conservation score of 1.0 corresponds to detection in 100% of frames. β-strands (magenta); α-helices (cyan).

Fig. 58 shows a comparison between the simulated ICs for phosphorylated Ssk1-rec W638A•Ypd1 and the inactive co-crystal structure for Ssk1-rec W638A•Ypd1. Almost every crystallographic contact was detected during the simulation, with most being identified as stable ICs. Overall, only His64-Asp511 was lost during simulation. Numerous stable IC pairs were observed during the simulation that are absent within the crystal complex. **Table 24** contains novel residue pairs observed in the phosphorylated simulation with conservation scores ≥70%. Interactions of interest include Arg90-Asp617 and Arg90-Ser614, which may be involved in the conformational shift upon phosphorylation of Ssk1-rec W638A. Additionally,

Gly68-Ala611 and Gly68-Lys632 may serve roles in the catalysis of phosphotransfer, and Gln86-Ser612, Glu83-Asn613, His64-Ser612, His64-Ala611, Lys67-Ala611, Lys67-Asn613 and Lys67-Ser612 may include residues involved in formation of a phosphotransfer-competent active site and/or influence the directionality of the phosphotransfer reaction.

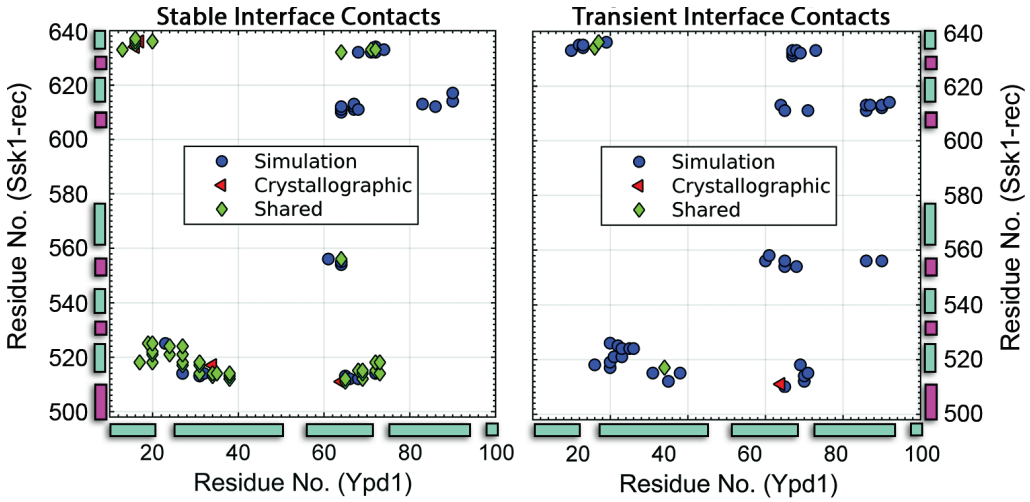

Figure 59. Comparisons between simulated and crystallographic ICs for phosphorylated Ssk1-rec W638A•Ypd1.

ICs with conservation scores ≥70% were classified as stable. Remaining ICs with scores between 15% and 70% were classified as transient. (Left panel) Stable ICs were compared with crystallographic contacts. Unique ICs seen only in the simulation are shown in blue. Crystallographic ICs not found as a stable contact in the trajectory are shown in red. Overlapping contacts are shown in green. (Right panel) Transient ICs were compared with the remaining crystallographic contacts not found in the left panel. Unique transient ICs observed only in the simulation are shown in blue. Truly unique crystallographic contact pairs found in neither stable nor transient classes are shown in red. Overlapping contact pairs are shown in green. β-strands (magenta); α-helices (cyan).

Table 24. Novel ICs observed in phosphorylated Ssk1-rec W638A•Ypd1.

Ypd1 - Residue	*Ssk1-rec - Residue*
Ala71	Lys632
Ala72	Ile514
Ala72	Lys632

Ala72	Val634
Arg90	Asp617
Arg90	Ser614
Asn61	Gln556
Asp23	Lys525
Gln86	Ser612
Glu83	Asn613
Gly68	Ala611
Gly68	Asn512
Gly68	Lys632
Gly74	Pro633
His64	Ala611
His64	Asp554
His64	Leu555
His64	Ser612
His64	Thr610
Ile32	Ile514
Leu31	Val513
Leu66	Asn512
Lys67	Ala611
Lys67	Asn613
Lys67	Ser612
Phe27	Ile514
Phe65	Val513

Fig. 60 shows the same comparison for the apo Ssk1-rec W638A•Ypd1 simulation and the apo Ssk1-rec W638A•Ypd1 crystal complex. Like the phosphorylated simulation, most crystallographic contacts were also observed

in the filtered list of ICs (>15% conservation), and multiple stable contacts were detected that are not present in the crystal structure (see **Table 25**).

Table 25. Novel ICs observed in apo Ssk1-rec W638A•Ypd1

Ypd1 - Residue	***Ssk1-rec - Residue***
Ala72	Ile514
Arg90	Ser612
Asn61	Gln556
Asp23	Lys525
Gln86	Ala611
Gly68	Asn512
Gly68	Lys632
Gly68	Pro633
His64	Ala611
His64	Asp554
His64	Leu555
Leu66	Asn512
Lys67	Ala611
Lys67	Pro633
Phe65	Val513
Ser69	Ile518

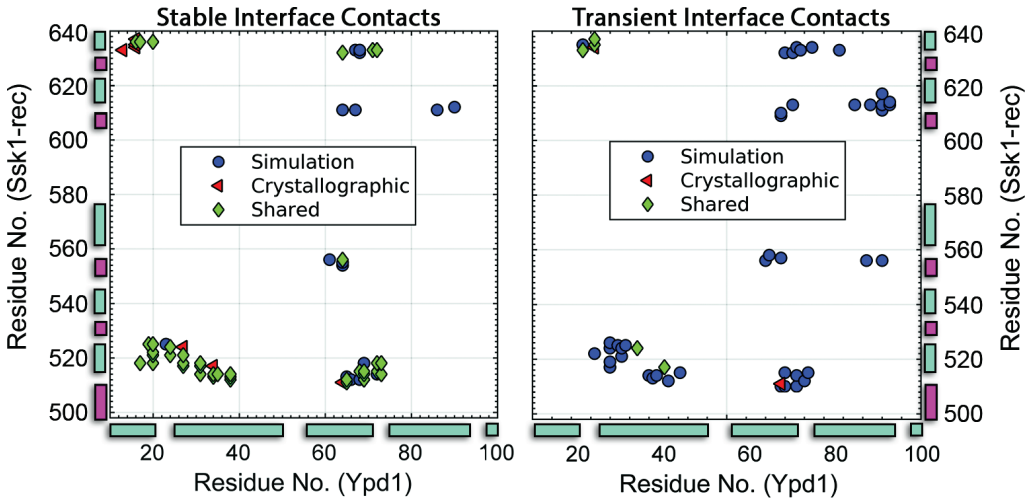

Figure 60. Comparisons between simulated and crystallographic ICs for apo Ssk1-rec W638A•Ypd1 simulation.

ICs with conservation scores ≥70% were classified as stable. Remaining ICs with scores between 15% and 70% were classified as transient. (Left panel) Stable ICs were compared with crystallographic contacts. Unique ICs seen only in the simulation are shown in blue. Crystallographic ICs not found as a stable contact in the trajectory are shown in red. Overlapping contacts are shown in green. (Right panel) Transient ICs were compared with the remaining crystallographic contacts not found in the left panel. Unique transient ICs observed only in the simulation are shown in blue. Truly unique crystallographic contact pairs found in neither stable nor transient classes are shown in red. Overlapping contact pairs are shown in green. β-strands (magenta); α-helices (cyan).

The lists of stable ICs seen during the apo and phosphorylated Ssk1-rec•Ypd1 trajectories were compared to identify interface changes related to phosphorylation (**Fig. 61**). The phosphorylated consensus map shows a trend reminiscent of the Sln1-rec•Ypd1 changes upon phosphorylation, with large concentrations of stable ICs forming between residues of the β4α4 and β5α5 regions on Ssk1-rec W638A and the αA, αC and αD helices of Ypd1. However, Ssk1-rec W638A appears to form a greater number of ICs within these regions than both Sln1-rec and wild-type Ssk1-rec. **Table 26** lists ICs identified as stable contacts (>70%) only in the phosphorylated Ssk1-rec

W638A•Ypd1 simulation. Residue pairs of interest include Arg90-Asp617, Arg90-Ser614, Gln86-Ser612, Glu83-Asn613, Lys67-Asn613 and Lys67-Ser612, all of which may be involved in the conformational change induced by the phosphorylation of Ssk1-rec W638A. **Table 27** contains a similar list found only in the apo Ssk1-rec W638A•Ypd1 simulation.

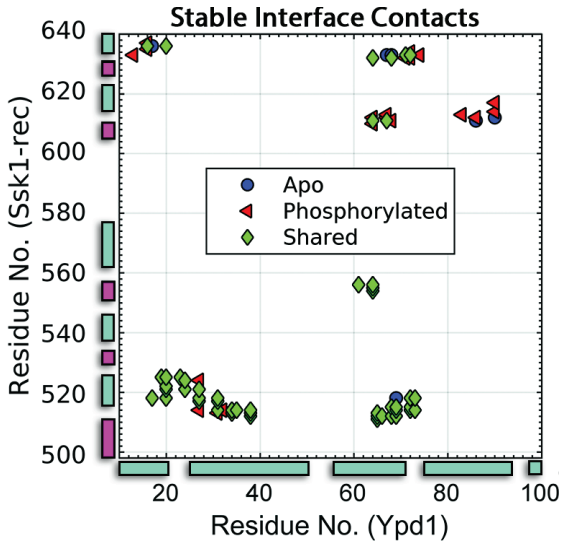

Figure 61. Comparison of stable ICs calculated for apo and phosphorylated Ssk1-rec W638A•Ypd1.

Stable residue pairs extracted from the apo and phosphorylated trajectories overlaid to identify interactions novel to each form. ICs only observed in the apo simulation are in blue. ICs only observed in the phosphorylated simulation are in red. Overlapping ICs found in both simulations are in green.

Table 26. Unique stable ICs in phosphorylated Ssk1-rec W638A•Ypd1.

Ypd1 - Residue	*Ssk1-rec - Residue*
Ala71	Lys632
Ala72	Lys632
Ala72	Val634
Arg90	Asp617
Arg90	Ser614

Gln86	Ser612
Glu16	Asn635
Glu16	His637
Glu83	Asn613
Gly68	Ala611
Gly74	Pro633
His64	Ser612
His64	Thr610
Ile13	Pro633
Ile32	Ile514
Leu31	Val513
Lys67	Asn613
Lys67	Ser612
Phe27	Arg524
Phe27	Ile514

Table 27. Unique stable ICs in apo Ssk1-rec W638A•Ypd1.

Ypd1 - Residue	***Ssk1-rec - Residue***
Arg90	Ser612
Gln86	Ala611
Gly68	Pro633
Ile17	Leu636
Lys67	Pro633
Ser69	Ile518

In summary, phosphorylation and its related structural changes were found to stabilize the interface between Sln1-rec•Ypd1 and Ssk1-rec W638A•Ypd1, but appeared to destabilize wild-type Ssk1-rec•Ypd1. It should

be noted that the (de)stabilization of the interface does not necessarily equate to affecting the binding of the proteins, in kinetic terms. Rather, the flexibility of the protein-protein interface is affected once bound. This may ultimately affect the binding affinity of the partners, as is the case with Sln1-rec and Ypd1 [30]. The largest changes in intermolecular contact stability involve regions of the αC and αD helices on Ypd1, and the β4α4 and β5α5 surfaces on the rec domains. Ssk1-rec W638A forms the most numerous and stable ICs within this region amongst the three rec domains, while wild-type Ssk1-rec forms the fewest. It is perhaps no coincidence that the W638A mutation is located near the center of this region. Removal of the large tryptophan side chain has a dramatic effect on the flexibility and/or stability of the interface and appears to cause a shift in the orientation of the phosphotransfer complex.

Active site geometry affects the reversibility of the phosphotransfer event

Radiolabeled phosphotransfer experiments have demonstrated the apparent irreversibility of transfer from Ypd1 to wild-type Ssk1-rec [30]. No plausible explanation has yet been presented that can account for this unusual characteristic, as chemically the transfer should be reversible. In contrast, the recently characterized Ssk1-rec W638A shows at least some level of reverse transfer when exposed to unphosphorylated Ypd1 (Katie Branscum, unpublished). It was previously postulated that replacing the large Trp638 side chain with a smaller amino acid, like alanine, might shift the orientation of the rec domain and/or Ypd1 interface (Katie Branscum, unpublished). This shift

could alter interatomic distances within the active site, bringing donor and acceptor residues (Asp554 and His64, respectively) into closer proximity.

In an attempt to further explain this observation, the simulations were used to examine the number of potential hydrogen bonds formed between the side chain of His64 on Ypd1 (acceptor, Nε2-atom) and the conserved, phosphorylatable aspartate (Asp1144 in Sln1-rec, Asp554 in Ssk1-rec) on the bound rec domain. For phosphotransfer to occur, donor and acceptor must be at least partially accessible geometrically and in relative proximity to one another. In the BeF_3^- bound structure of Sln1-rec•Ypd1 (PDB 2R25), the distance between the donor (Asp1144 Oδ1) and acceptor (His64 Nε2) residues is 4.9 Å, with the nitrogen and hydrogen atoms forming a hydrogen bond with the tetrahedral BeF_3^- group in a nearly linear geometry appropriate for phosphotransfer. A similar arrangement is seen in the BeF_3^- bound structure of Spo0F/Spo0B. Presumably these structures represent pre-transition state transfer complexes, suggesting that these are the approximate distances for catalytic alignment. To detect if similar geometries and interactions occur within the simulated active sites, the number of potential hydrogen bonds formed between His64 on Ypd1 and the phosphoryl group or phosphorylatable aspartate on each rec domain was calculated and tracked over the trajectories. Average interatomic distance between donor and acceptor atoms was also determined for each simulation. **Fig. 62** shows hydrogen-bond calculations for both the phosphorylated (A) and apo (B) simulations of Sln1-rec•Ypd1. During

the phosphorylated simulation, His64 remained in relatively close proximity to the phosphoryl group, forming an average of 0.96 hydrogen bonds per frame and maintaining an average distance between the Nε2 and phosphorus atoms of 3.93 Å. This corresponds to a hydrogen-bond occupancy (fraction of time that a hydrogen bond is formed) of ~96%. The apo simulation suggests that these levels are maintained regardless of phosphorylation state, with a hydrogen-bond occupancy of ~100%. Similar results were obtained for the pre-activated phosphorylated Sln1-rec•Ypd1 simulation, with His64 and the phosphoryl group forming an average of 0.96 hydrogen bonds per frame and maintaining an average distance between the Nε2 and phosphorus atoms of 3.42 Å. This corresponds to a hydrogen-bond occupancy of ~96%.

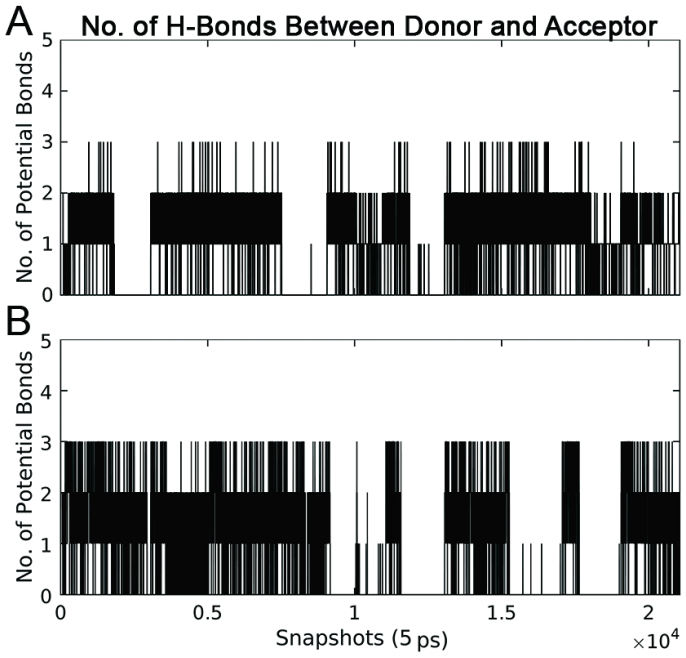

Figure 62. Number of hydrogen bonds between His64 and Asp1144.

(A) The number of potential hydrogen bonds formed in each frame between the side chain of His64 and the phosphoryl group during the phosphorylated Sln1-rec•Ypd1 simulation. (B) The

number of potential hydrogen bonds formed in each frame between the side chain of His64 and Asp1144 during the apo Sln1-rec•Ypd1 simulation.

Fig. 63 shows a dramatic change in the wild-type Ssk1-rec•Ypd1 simulations. The average number of hydrogen bonds formed between His64 on Ypd1 and Asp554 on Ssk1-rec drops to 0.12, and the average distance between the Nε2 and phosphorus atoms rises to 6.30 Å for the phosphorylated simulation (0.0 hydrogen bonds and 7.1 Å in the apo simulation). These correspond to a hydrogen-bond occupancy of ~12%. The almost complete abolishment of hydrogen-bonding capability strongly suggests that donor and acceptor are no longer accessible to one another. His64 was repeatedly observed interacting with residues on the β4α4 loop, including Ala611 and Ser612.

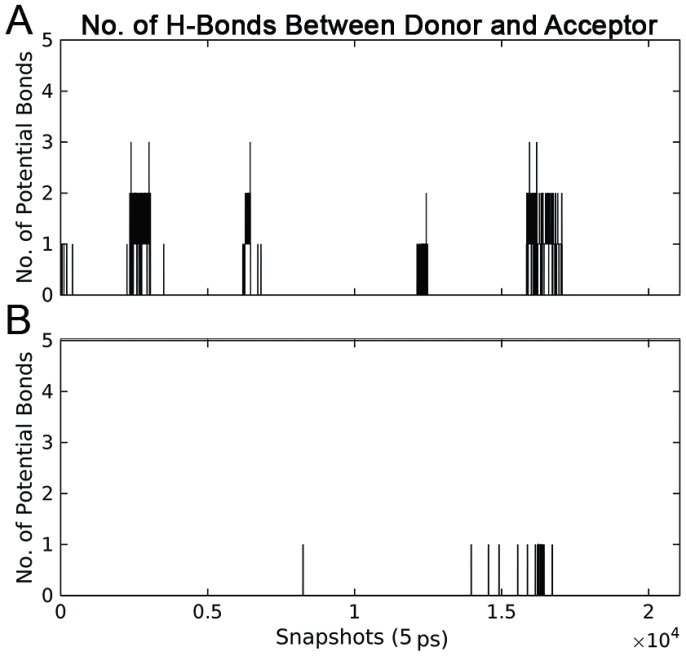

Figure 63. Number of hydrogen bonds between His64 and Asp554.

(A) The number of potential hydrogen bonds formed in each frame between the side chain of His64 and the phosphoryl group during the phosphorylated wild-type Ssk1-rec•Ypd1 simulation. (B) The number of potential hydrogen bonds formed in each frame between the side chain of His64 and As554 during the apo wild-type Ssk1-rec•Ypd1 simulation.

Fig. 64 shows the same analysis for the mutant Ssk1-rec W638A•Ypd1. In the phosphorylated simulation, His64 on Ypd1 remains near the phosphoryl group for the majority of the trajectory, forming an average of 0.86 hydrogen bonds per frame and maintaining an average distance of 3.7 Å to the phosphorus atom (0.69 hydrogen bonds and 4.36 Å in the apo simulation). These correspond to a hydrogen-bond occupancy of ~86%.

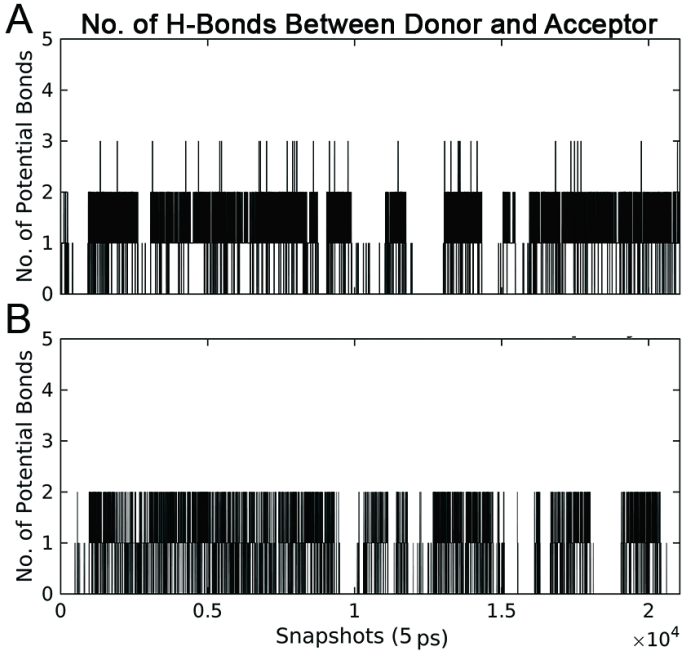

Figure 64. Number of hydrogen bonds between His64 and Asp554.

(A) The number of potential hydrogen bonds formed in each frame between the side chain of His64 and the phosphoryl group during the phosphorylated Ssk1-rec W638A•Ypd1 simulation. (B) The number of potential hydrogen bonds formed in each frame between the side chain of His64 and Asp554 during the apo Ssk1-rec W638A•Ypd1 simulation.

The significant differences between the systems suggest that under normal conditions, the phosphorylatable His64 is either unable to hydrogen bond with the phosphoryl or is preoccupied with additional interactions that make the hydrogen bonding less favorable in wild-type Ssk1-rec. The lack of interaction and overall increase in the interatomic distance between donor and acceptor makes phosphotransfer unlikely. This may explain the lack of observable reverse transfer from Ssk1-rec to Ypd1. Taking into account the hydrogen bonding tendency of the donor and acceptor in Sln1-rec•Ypd1 and Ssk1-rec W638A•Ypd1, as well as the stabilization effect observed in both of their interfaces upon phosphorylation, it is reasonable to speculate that the infrequent interaction between donor and acceptor atoms in wild-type Ssk1-rec•Ypd1 is due to the instability of its interface contacts. Mutation of Trp638 to an alanine stabilizes these contacts, thereby stabilizing the interaction between donor and acceptor residues and allowing for reverse transfer to occur.

Complex formation occludes hydrolytic water from the active site and alters the local chemical environment

Despite utilizing highly conserved active site geometrics, rec domain autodephosphorylation rates are extremely variable, spanning nearly six orders of magnitude [64]. These rates typically correlate with the biological processes controlled by the rec domain [65]. How rec domains are able to customize their own dephosphorylation kinetics is unknown. The high genetic diversity found

within the rec domain family indicates that the answer is likely different for each protein. Autodephosphorylation (hydrolytic) is thought to occur through an in-line attack at the phosphorus atom by a nucleophilic water molecule positioned directly above the phosphoryl group [39]. The reaction then proceeds through a trigonal bipyramidal transition state, similar to the normal associative phosphotransfer mechanism [66]. The metal cation and conserved lysine residue are essential for the reaction, likely for neutralizing and stabilizing the highly charged transition state [67].

Alone, both Ssk1-rec and Sln1-rec possess comparable half-lives of approximately 10-15 minutes [32]. However, in the presence of Ypd1, the phosphorylated half-life of Ssk1-rec (both wild-type and W638A) is stabilized to over 40 hours [32]. This same effect is not observed with Sln1-rec or the third rec domain of the pathway, Skn7-rec. Several possibilities exist that may explain the stabilization of phosphorylated Ssk1-rec by Ypd1 (reviewed in [65]). Because the hydrolysis reaction must proceed along an in-line route at the phosphorus atom, lowering the accessibility of the phosphoryl group to solvent may protect the phosphorylated state. Ypd1 may aid in occluding water from the proximity of the phosphoryl group, or may facilitate the same effect through another residue on Ssk1-rec. If dephosphorylation is linked to the conformational change seen in Ssk1-rec upon activation, sequestering the rec domain in an active conformation may also affect hydrolysis rates. Finally, Ypd1 may alter the chemical environment of the active site, making it less

favorable energetically for the reaction to occur or making the phosphorus a less attractive site for nucleophilic attack.

To examine the solvent accessibility of the phosphoryl group, average water densities and occupancies were calculated for each of the simulated systems. Fully solvated snapshots were taken from each production trajectory at 50 ps intervals (~2000 snapshots total). Frames were first aligned by Cα atoms using their initial frame as a reference. The average number of waters near the phosphorylatable aspartate of each rec domain was then calculated and is shown in **Table 28**.

Table 28. Water molecules within 5 Å of the catalytic aspartate residue.

Proteins	***Waters***
P~Sln1-rec•Ypd1	4
Sln1-rec•Ypd1	5
P~Ssk1-rec•Ypd1	5
Ssk1-rec•Ypd1	5
P~Ssk1-rec W638A•Ypd1	5
Ssk1-rec W638A•Ypd1	4

To more accurately quantify the relative positions and concentrations of active site water molecules, the VMD VolMap plugin was then used to calculate the average atomic densities and occupancies contributed by water (oxygen) atoms within each simulation. Data were mapped as isosurfaces, showing the three-dimensional locations and relative densities of water molecules. **Fig. 65** shows these maps calculated from the phosphorylated Sln1-

rec•Ypd1simulation. While each figure shows relatively high active site accessibility to bulk solvent, no waters were observed directly above the phosphoryl group, where an in-line attack would originate. **Fig. 66** shows the same view of apo Sln1-rec•Ypd1 for reference. Here, the lack of an organized active site appears to inhibit the ordered arrangement of active site waters as seen in **Fig. 65**.

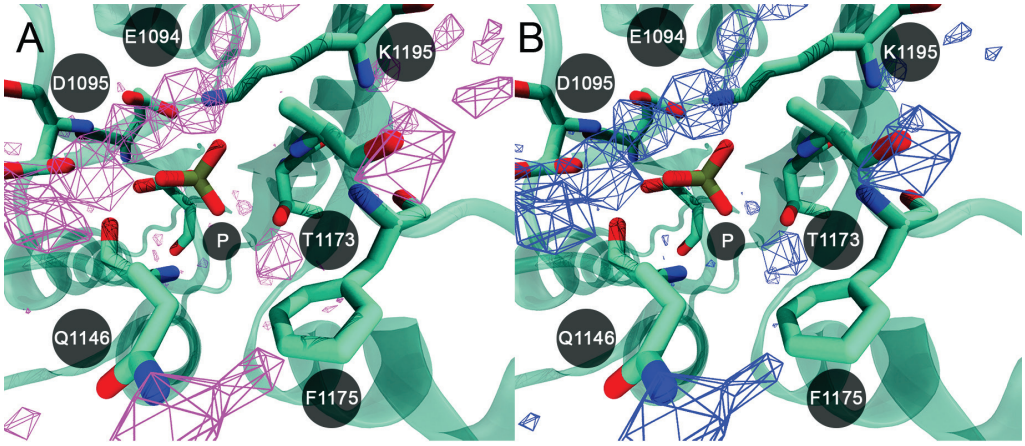

Figure 65. Time-averaged water density and occupancy maps calculated for phosphorylated Sln1-rec•Ypd1.

(A) Atomic water density (oxygen atoms) averaged over 100 ns (pink wire-frame). Density isosurface is shown at 0.05 (on a range of 0-0.15) (B) Atomic water occupancy (oxygen atoms) averaged over 100 ns (green wire-frame). Occupancy isosurface is shown at 50%.

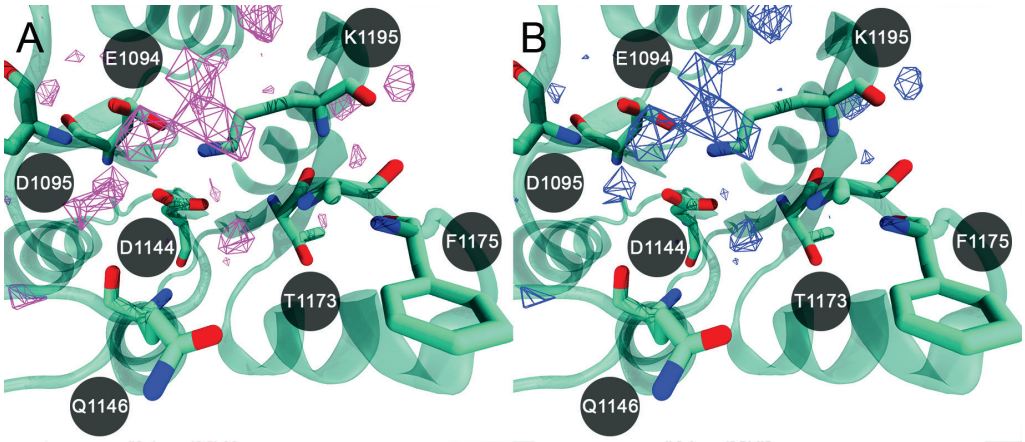

Figure 66. Time-averaged water density and occupancy maps calculated for apo Sln1-rec•Ypd1.

(A) Atomic water density (oxygen atoms) averaged over 100 ns (pink wire-frame). Density isosurface is shown at 0.05 (on a range of 0-0.15) (B) Atomic water occupancy (oxygen atoms) averaged over 100 ns (green wire-frame). Occupancy isosurface is shown at 50%.

Comparison with volumetric maps for wild-type Ssk1-rec•Ypd1 (**Fig. 67 phosphorylated; Fig. 68 apo**) indicates that this effect is common to both rec domain complexes and is likely a result of Ypd1 partitioning the active site away from the bulk solvent. Overall active site water density and occupancy were lower in the wild-type Ssk1-rec•Ypd1 simulation than in the Sln1-rec•Ypd1 complex. Originally it was hypothesized that this may also contribute to the more protective effect Ypd1 shares with Ssk1-rec. However, analysis of the mutant Ssk1-rec W638A•Ypd1 simulations suggests that this is not the case (**Figs. 69-70**). These showed high active site water density and occupancy in locations similar to the Sln1-rec•Ypd1 complex (**Fig. 65**), though no water was detected directly above the phosphoryl group.

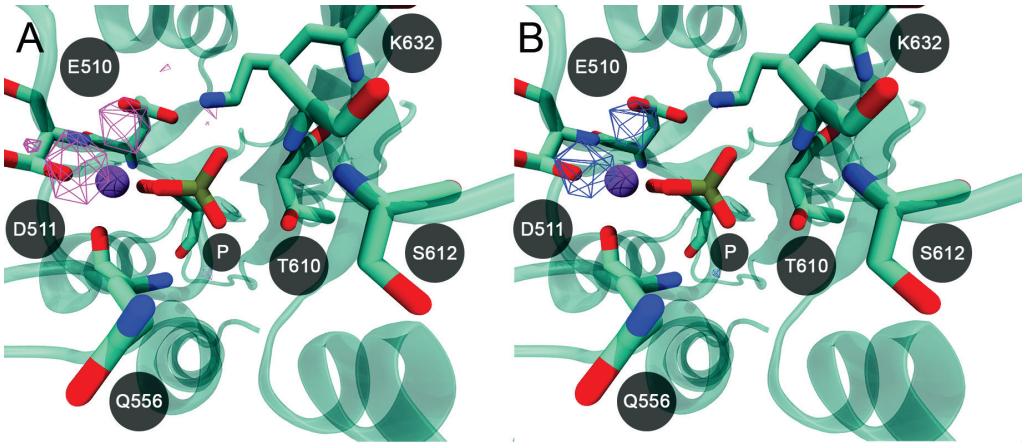

Figure 67. Time-averaged water density and occupancy maps calculated for phosphorylated wild-type Ssk1-rec•Ypd1.

(A) Atomic water density (oxygen atoms) averaged over 100 ns (pink wire-frame). Density isosurface is shown at 0.05 (on a range of 0-0.15) (B) Atomic water occupancy (oxygen atoms) averaged over 100 ns (green wire-frame). Occupancy isosurface is shown at 50%.

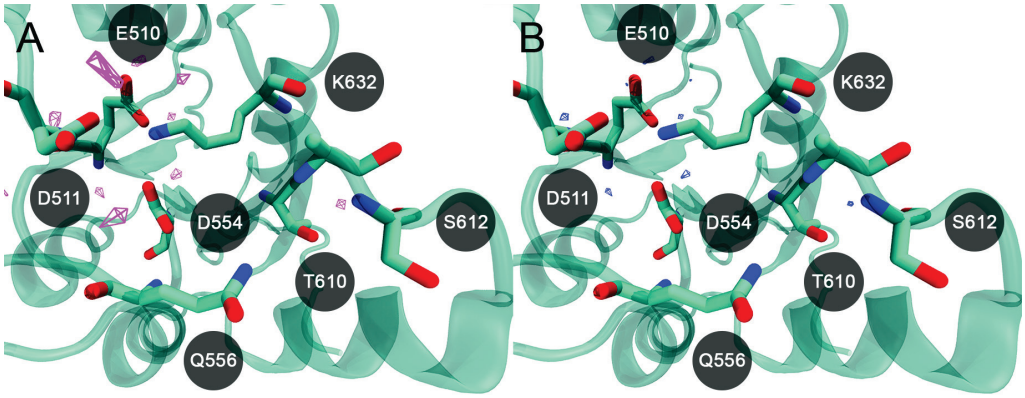

Figure 68. Time-averaged water density and occupancy maps calculated for apo wild-type Ssk1-rec•Ypd1.

(A) Atomic water density (oxygen atoms) averaged over 100 ns (pink wire-frame). Density isosurface is shown at 0.05 (on a range of 0-0.15) (B) Atomic water occupancy (oxygen atoms) averaged over 100 ns (green wire-frame). Occupancy isosurface is shown at 50%.

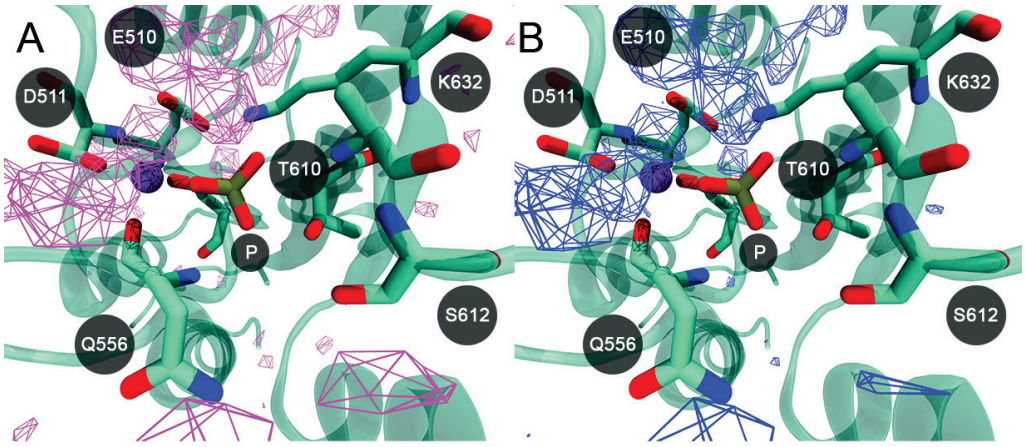

Figure 69. Time-averaged water density and occupancy maps calculated for phosphorylated Ssk1-rec W638A•Ypd1.

(A) Atomic water density (oxygen atoms) averaged over 100 ns (pink wire-frame). Density isosurface is shown at 0.05 (on a range of 0-0.15) (B) Atomic water occupancy (oxygen atoms) averaged over 100 ns (green wire-frame). Occupancy isosurface is shown at 50%.

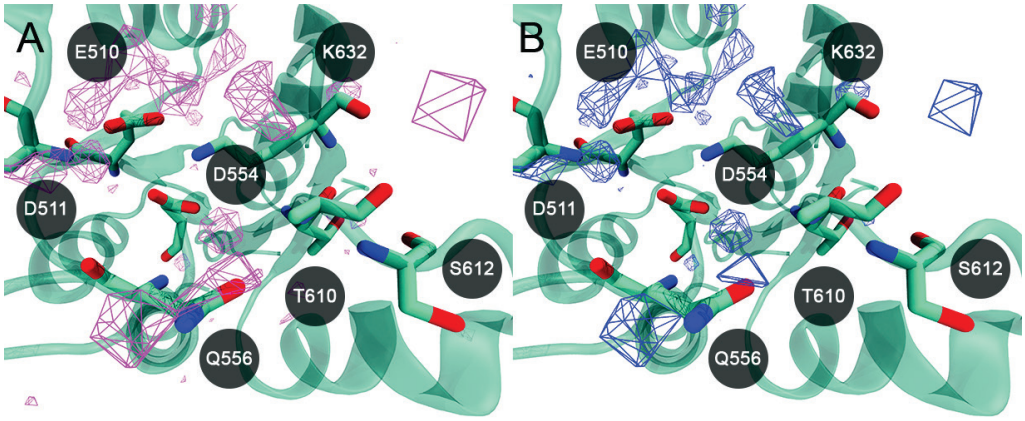

Figure 70. Time-averaged water density and occupancy maps calculated for apo Ssk1-rec W638A•Ypd1.

(A) Atomic water density (oxygen atoms) averaged over 100 ns (pink wire-frame). Density isosurface is shown at 0.05 (on a range of 0-0.15) (B) Atomic water occupancy (oxygen atoms) averaged over 100 ns (green wire-frame). Occupancy isosurface is shown at 50%.

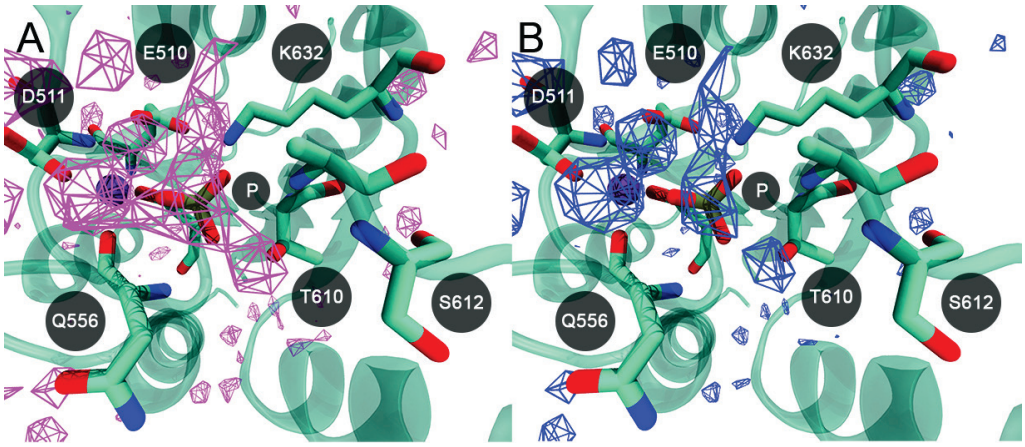

Figure 71. Time-averaged water density and occupancy maps calculated for phosphorylated Ssk1-rec W638A.

(A) Atomic water density (oxygen atoms) averaged over 100 ns (pink wire-frame). Density isosurface is shown at 0.05 (on a range of 0-0.15) (B) Atomic water occupancy (oxygen atoms) averaged over 100 ns (green wire-frame). Occupancy isosurface is shown at 50%.

To test if this protective effect was directly caused by complex formation, Ypd1 was removed and phosphorylated Ssk1-rec W638A was simulated as a monomer in solution (100 ns). **Fig. 71** shows the water density and occupancy maps calculated from this monomeric simulation, indicating consistently high water concentration positioned directly above the phosphoryl group. These results suggest that binding Ypd1 at least partially protects the phosphorylated rec domains by preventing the hydrolytic water from positioning in the proper orientation for an in-line nucleophilic attack. Ypd1's high affinity and preference for Ssk1-rec leads to the formation of an unusually strong transient complex, protecting the phosphoryl group from hydrolysis for an extended period. However, Sln1-rec and Ypd1 dissociate too rapidly to capitalize on any possible protective effect [30]. Recent work demonstrates a

precedent for this mechanism, showing the protection of the response regulator PmrA from phosphatase activity by binding the connector protein PmrD at the rec domain active site [68].

To examine if complex formation alters the chemical environment of the active site, thereby further affecting dephosphorylation rates, the electrostatic potential of each complex was calculated using the PMEpot plugin within VMD [52]. Electrostatic potential was averaged over 100 ns for each simulation. Potential values were mapped (scaled from -15 to +30 *kT/e*) first to a molecular surface representation showing a top-down view of the active site. A volumetric slice projected along the XY plane was also generated, positioned directly above the phosphorylatable aspartate and showing a close-up of the rec domain active site. **Fig. 72** shows this heat map for the phosphorylated and apo Sln1-rec•Ypd1 complexes. The major areas of positive potential are centered around both the metal cation and the conserved active site lysine (Lys1195).

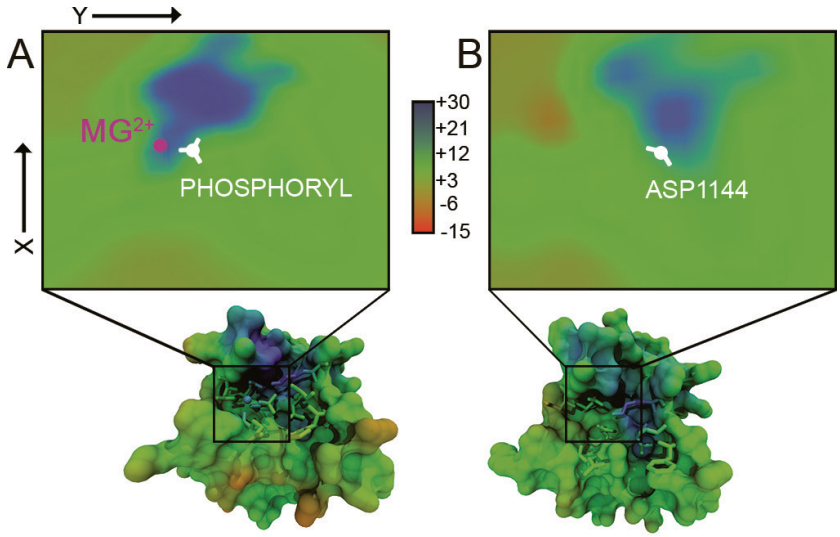

Figure 72. Time-averaged electrostatic potential for phosphorylated and apo Sln1-rec•Ypd1 simulations.

Shown are molecular surface representations of the rec domains, extracted from the complex simulations. Atoms are colored based on average electrostatic potential, scaled from -15 to +30 *kT/e*, computed using the PMEpot plugin. (A) Potential for phosphorylated Sln1-rec•Ypd1. The inset shows a volumetric slice, generated along the XY plane, showing the immediate environment around the phosphoryl group. (B) Potential for apo Sln1-rec•Ypd1. The inset shows a volumetric slice, generated along the XY plane, showing the immediate environment around the phosphorylatable Asp1144.

Fig. 73 shows the same comparison for wild-type Ssk1-rec. Notice the large increase in positive potential surrounding the highly negatively charged phosphoryl group, an effect that is significantly diminished in the Sln1-rec•Ypd1 active site. The effect is even more pronounced in the Ssk1-rec W638A•Ypd1 active site (**Fig. 74**). The positively charged potential surrounding the phosphoaspartate is almost completely absent when phosphorylated Ssk1-rec W638A and wild-type Ssk1-rec were simulated as monomers in solution (**Fig. 75**). Only by binding Ypd1 is the Ssk1-rec active site altered so dramatically. Previous work suggests that autophosphorylation

and phosphotransfer reactions are highly sensitive to changes in salt concentration [47, 65, 69, 70]. Due to the highly charged nature of the rec domain active site, any disruption in the normal electrostatic environment would likely lead to a change in reaction rates and stability. The positively charged potential seen in the Ssk1-rec•Ypd1 active site may be acting to neutralize the negatively charged phosphoryl group, lowering its ground state energy and increasing the activation energy of the dephosphorylation reaction. This would ultimately slow the rate of hydrolysis.

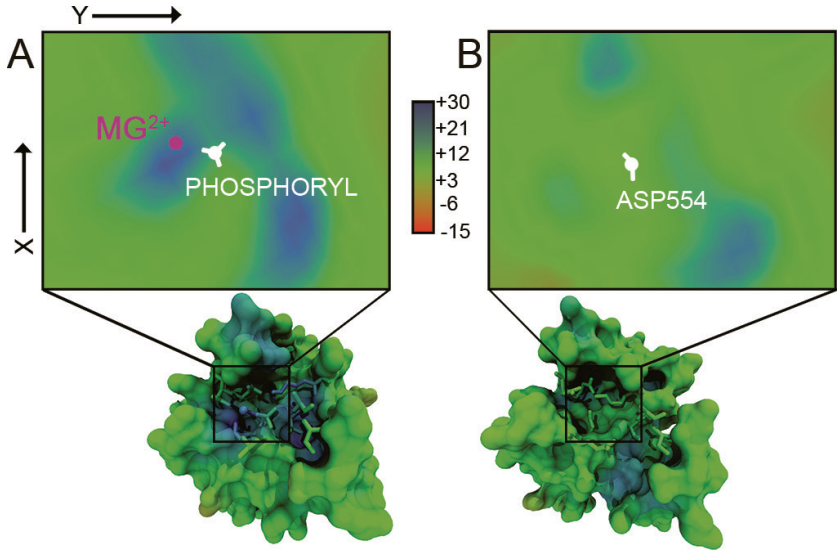

Figure 73. Time-averaged electrostatic potential for phosphorylated and apo wild-type Ssk1-rec•Ypd1.

Shown are molecular surface representations of the rec domains, extracted from the complex simulations. Atoms are colored based on average electrostatic potential, scaled from -15 to +30 *kT/e*, computed using the PMEpot plugin. (A) Potential for phosphorylated Ssk1-rec•Ypd1. The inset shows a volumetric slice, generated along the XY plane, showing the immediate environment around the phosphoryl group. (B) Potential for apo Ssk1-rec•Ypd1. The inset shows a volumetric slice, generated along the XY plane, showing the immediate environment around the phosphorylatable Asp554.

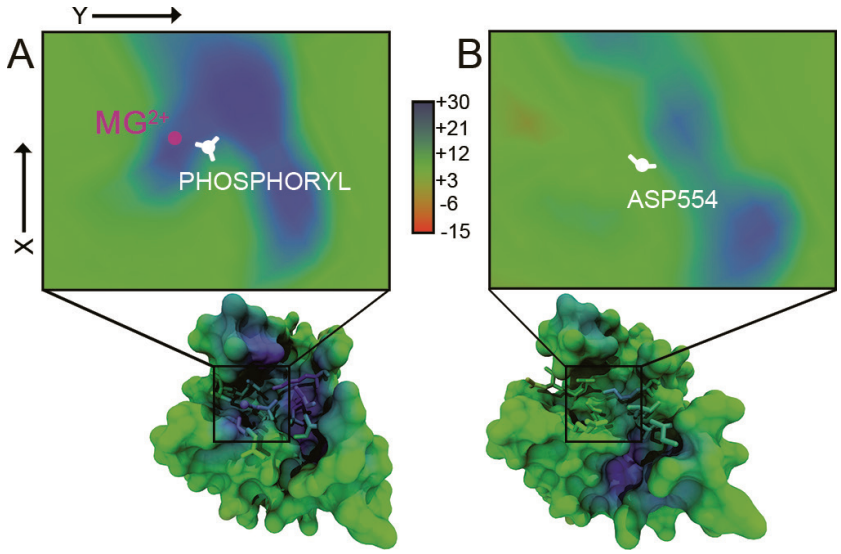

Figure 74. Time-averaged electrostatic potential for phosphorylated and apo Ssk1-rec W638A•Ypd1.

Shown are molecular surface representations of the rec domains, extracted from the complex simulations. Atoms are colored based on average electrostatic potential, scaled from -15 to +30 *kT/e*, computed using the PMEpot plugin. (A) Potential for phosphorylated Ssk1-rec W638A•Ypd1. The inset shows a volumetric slice, generated along the XY plane, showing the immediate environment around the phosphoryl group. (B) Potential for apo Ssk1-rec W638A•Ypd1. The inset shows a volumetric slice, generated along the XY plane, showing the immediate environment around the phosphorylatable Asp554.

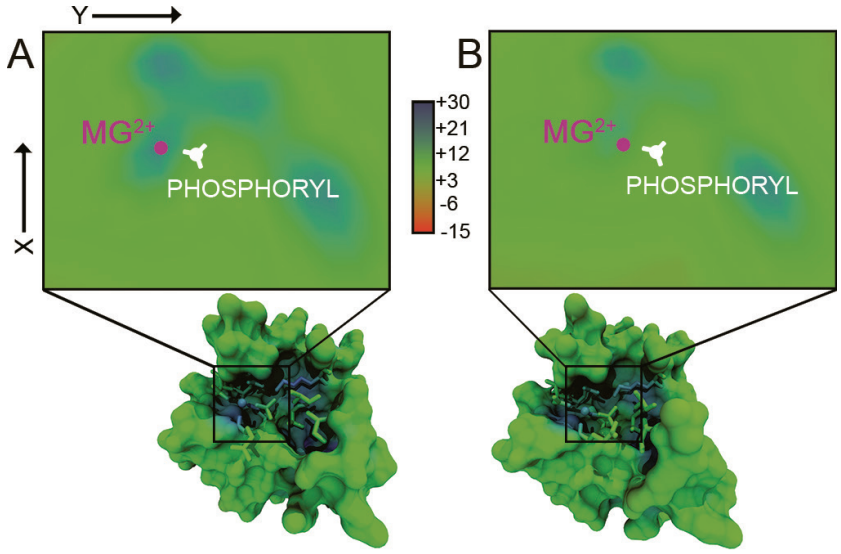

Figure 75. Time-averaged electrostatic potential for phosphorylated wild-type Ssk1-rec and Ssk1-rec W638A (monomers).

Shown are molecular surface representations for both forms of the rec domain, simulated as monomers in solution. Atoms are colored based on average electrostatic potential, scaled from -15 to +30 *kT/e*, computed using the PMEpot plugin. (A) Potential for phosphorylated Ssk1-rec W638A alone. The inset shows a volumetric slice, generated along the XY plane, showing the immediate environment around the phosphoryl group. (B) Potential for phosphorylated wild-type Ssk1-rec. The inset shows a volumetric slice, generated along the XY plane, showing the immediate environment around the phosphoryl group.

3.4 Conclusions

Active site restraints were used to study the activation of rec domains from the fungal osmoregulatory Sln1 pathway in complex with their cognate partner, Ypd1. Observations on interface stability and specific intermolecular contacts suggest that phosphorylation stabilizes the protein-protein interfaces within Sln1-rec•Ypd1 and Ssk1-rec W638A•Ypd1, but destabilizes the interface within wild-type Ssk1-rec•Ypd1. This destabilization is likely related to the poor catalytic and geometric alignment of donor and acceptor atoms (His64-Asp554) within the wild-type Ssk1-rec•Ypd1 active site. These factors may contribute to the perceived absence of reverse phosphotransfer from wild-type Ssk1-rec to Ypd1. The W638A mutation stabilizes both the overall interface and the catalytic alignment of His64-Asp554, likely increasing the propensity for reverse phosphotransfer.

Simulation in explicit solvent provided valuable insight into the accessibility and positioning of active site waters in relation to the phosphoaspartate residue. In both Sln1-rec•Ypd1 and the two Ssk1-rec•Ypd1 complexes, water is prevented from positioning directly above the phosphoryl group in an appropriate orientation for in-line nucleophilic attack. Comparison

with monomeric rec domain simulations indicates that this protective effect is a direct feature of binding Ypd1. As Ssk1-rec forms a relatively stable complex with Ypd1, the active site aspartate residue is partially sequestered from solvent for an extended period of time, dramatically increasing its phosphorylated life-time. The rapid dissociation of Sln1-rec and Ypd1 prevents any such stabilization.

Both Ssk1-rec•Ypd1 active sites were found to be significantly more positively charged than the Sln1-rec•Ypd1 active site. This appears to be an effect of both phosphorylation and complex formation with Ypd1, as it is significantly diminished in the apo and monomeric simulations. Surrounding the highly negatively charged phosphoryl group with positive electrostatic potential may contribute to the stabilization of the phosphoryl group, significantly slowing autodephosphorylation rates by increasing the activation energy of the reaction.

It should be noted that the majority of the complexes used for this study started in initial inactive conformations. Rec domain active site restraints were used to simulate activation post-complex formation to make consistent comparisons. While physiologically this makes sense for Ssk1-rec•Ypd1, Sln1-rec is likely already activated prior to binding its partner. A control was run using a pre-activated Sln1-rec docked to Ypd1. Results suggest no appreciable difference between the original Sln1-rec•Ypd1 and the pre-activated system. The control also demonstrates how an accurate pre-transition state

phosphotransfer complex might be simulated when starting with an imperfectly docked complex model. Application of an extra distance restraint to enforce linear geometry between donor and acceptor atoms was able to reproduce a phosphotransfer-competent active site, despite His64 and the phosphorylatable aspartate residue being out of alignment in the initial model.

These results provide valuable insights into the mechanisms governing complex formation and phosphotransfer within the osmoregulatory Sln1 pathway and demonstrate the value of biased MD simulations for studying events that would otherwise be difficult to observe with traditional methods. However, these studies do not replace the need for experimental validation. While the conclusions presented here may provide intellectually appealing explanations for these various phenomena, they should be used as complementary tools to guide future studies, both experimental and computational.

References

1. Ann M. Stock, Victoria L. Robinson a, Goudreau PN. Two-component signal transduction. Annual Rev Biochemistry. 2000;69(1):183-215.

2. Alex LA, Simon MI. Protein histidine kinases and signal transduction in prokaryotes and eukaryotes. Trends Genet. 1994;10(4):133-8.

3. Fassler JS, West AH. Histidine Phosphotransfer Proteins in Fungal Two-Component Signal Transduction Pathways. Eukaryotic Cell. 2013;12(8):1052-60.

4. West AH, Stock AM. Histidine kinases and response regulator proteins in two-component signaling systems. Trends Biochem Sci. 2001;26(6):369-76.

5. Blair JA, Xu Q, Childers WS, Mathews, II, Kern JW, Eckart M, Deacon AM, Shapiro L. Branched signal wiring of an essential bacterial cell-cycle phosphotransfer protein. Structure. 2013;21(9):1590-601.

6. Hutchison CE, Kieber JJ. Signaling via Histidine-Containing Phosphotransfer Proteins in Arabidopsis. Plant Signaling & Behavior. 2007;2(4):287-9.

7. Porter SW, Xu Q, West AH. Ssk1p response regulator binding surface on histidine-containing phosphotransfer protein Ypd1p. Eukaryot Cell. 2003;2(1):27-33.

8. Brown JL, Bussey H, Stewart RC. Yeast Skn7p functions in a eukaryotic two-component regulatory pathway. EMBO J. 1994;13(21):5186-94.

9. Vetter SM, Schlievert PM. The two-component system Bacillus respiratory response A and B (BrrA-BrrB) is a virulence factor regulator in Bacillus anthracis. Biochemistry. 2007;46(25):7343-52.

10. Dhiman A, Bhatnagar S, Kulshreshtha P, Bhatnagar R. Functional characterization of WalRK: A two-component signal transduction system from Bacillus anthracis(). FEBS Open Bio. 2014;4:65-76.

11. Suárez JM, Edwards AN, McBride SM. The Clostridium difficile cpr Locus Is Regulated by a Noncontiguous Two-Component System in Response to Type A and B Lantibiotics. J Bact. 2013;195(11):2621-31.

12. Martin PK, Li T, Sun D, Biek DP, Schmid MB. Role in Cell Permeability of an Essential Two-Component System in Staphylococcus aureus. J Bact. 1999;181(12):3666-73.

13. Bahn YS, Kojima K, Cox GM, Heitman J. A unique fungal two-component system regulates stress responses, drug sensitivity, sexual development, and virulence of Cryptococcus neoformans. Molecular biology of the cell. 2006;17(7):3122-35.

14. Calera JA, Calderone R. Flocculation of hyphae is associated with a deletion in the putative CaHK1 two-component histidine kinase gene from Candida albicans. Microbiology. 1999;145 (Pt 6):1431-42.

15. Yamada-Okabe T, Mio T, Ono N, Kashima Y, Matsui M, Arisawa M, Yamada-Okabe H. Roles of three histidine kinase genes in hyphal development and virulence of the pathogenic fungus Candida albicans. J Bact. 1999;181(23):7243-7.

16. Desai C, Mavrianos J, Chauhan N. Candida albicans SRR1, a putative two-component response regulator gene, is required for stress adaptation, morphogenesis, and virulence. Eukaryot Cell. 2011;10(10):1370-4.

17. Hinnebusch AG, Johnston M. YeastBook: an encyclopedia of the reference eukaryotic cell. Genetics. 2011;189(3):683-4.

18. Fassler JS, West AH. Genetic and Biochemical Analysis of the SLN1 Pathway in Saccharomyces cerevisiae. Methods Enzymol. 2010;471:291-317.

19. Maeda T, Wurgler-Murphy SM, Saito H. A two-component system that regulates an osmosensing MAP kinase cascade in yeast. Nature. 1994;369(6477):242-5.

20. Posas F, Wurgler-Murphy SM, Maeda T, Witten EA, Thai TC, Saito H. Yeast HOG1 MAP kinase cascade is regulated by a multistep phosphorelay mechanism in the SLN1-YPD1-SSK1 "two-component" osmosensor. Cell. 1996;86(6):865-75.

21. Ota IM, Varshavsky A. A yeast protein similar to bacterial two-component regulators. Science. 1993;262(5133):566-9.

22. Porter SW, West AH. A common docking site for response regulators on the yeast phosphorelay protein YPD1. Biochimica et biophysica acta. 2005;1748(2):138-45.

23. Kato M, Mizuno T, Shimizu T, Hakoshima T. Refined structure of the histidine-containing phosphotransfer (HPt) domain of the anaerobic sensor kinase ArcB from Escherichia coli at 1.57 A resolution. Acta crystallographica Section D, Biological crystallography. 1999;55(Pt 11):1842-9.

24. Mourey L, Da Re S, Pedelacq JD, Tolstykh T, Faurie C, Guillet V, Stock JB, Samama JP. Crystal structure of the CheA histidine phosphotransfer domain that mediates response regulator phosphorylation in bacterial chemotaxis. J Biol Chem. 2001;276(33):31074-82.

25. Zhou H, Dahlquist FW. Phosphotransfer site of the chemotaxis-specific protein kinase CheA as revealed by NMR. Biochemistry. 1997;36(4):699-710.

26. Lukat GS, McCleary WR, Stock AM, Stock JB. Phosphorylation of bacterial response regulator proteins by low molecular weight phospho-donors. Proc Natl Acad Sci. 1992;89(2):718-22.

27. Lu JM, Deschenes RJ, Fassler JS. Saccharomyces cerevisiae histidine phosphotransferase Ypd1p shuttles between the nucleus and cytoplasm

for SLN1-dependent phosphorylation of Ssk1p and Skn7p. Eukaryot Cell. 2003;2(6):1304-14.

28. Saito H, Posas F. Response to Hyperosmotic Stress. Genetics. 2012;192(2):289-318.

29. Xu Q, Porter SW, West AH. The yeast YPD1/SLN1 complex: Insights into molecular recognition in two-component signaling systems. Structure. 2003;11(12):1569-81.

30. Janiak-Spens F, Cook PF, West AH. Kinetic analysis of YPD1-dependent phosphotransfer reactions in the yeast osmoregulatory phosphorelay system. Biochemistry. 2005;44(1):377-86.

31. Perkins JR, Diboun I, Dessailly BH, Lees JG, Orengo C. Transient Protein-Protein Interactions: Structural, Functional, and Network Properties. Structure. 2010;18(10):1233-43.

32. Janiak-Spens F, Sparling JM, Gurfinkel M, West AH. Differential stabilities of phosphorylated response regulator domains reflect functional roles of the yeast osmoregulatory SLN1 and SSK1 proteins. J Bact. 1999;181(2):411-7.

33. Attwood PV, Besant PG, Piggott MJ. Focus on phosphoaspartate and phosphoglutamate. Amino Acids. 2011;40(4):1035-51.

34. Kaserer AO, Andi B, Cook PF, West AH. Effects of osmolytes on the SLN1-YPD1-SSK1 phosphorelay system from *Saccharomyces cerevisiae*. Biochemistry. 2009;48(33):8044-50.

35. Zhao X, Copeland DM, Soares AS, West AH. Crystal structure of a complex between the phosphorelay protein YPD1 and the response regulator domain of SLN1 bound to a phosphoryl analog. J Mol Biol. 2008;375(4):1141-51.

36. The NCBI Reference Sequence collection .http://www.ncbi.nlm.nih.gov/RefSeq.

37. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids. 2000;28(1):235-42.

38. Varughese KI, Tsigelny I, Zhao H. The crystal structure of beryllofluoride Spo0F in complex with the phosphotransferase Spo0B represents a phosphotransfer pretransition state. J Bact. 2006;188(13):4970-7.

39. Zhao R, Collins EJ, Bourret RB, Silversmith RE. Structure and catalytic mechanism of the E. coli chemotaxis phosphatase CheZ. Nat Struct Biol. 2002;9(8):570-5.

40. Carter EA, Ciccotti G, Hynes JT, Kapral R. Constrained reaction coordinate dynamics for the simulation of rare events. Chemical Physics Letters. 1989;156(5):472-7.

41. Grubmüller H. Predicting slow structural transitions in macromolecular systems: Conformational flooding. Phys Rev E. 1995;52(3):2893-906.

42. Huber T, Torda A, van Gunsteren W. Local elevation: A method for improving the searching properties of molecular dynamics simulation. J Computer-Aided Mol Des. 1994;8(6):695-708.

43. Bourret RB. Receiver domain structure and function in response regulator proteins. Current Opinion in Microbiology. 2010;13(2):142-9.

44. Bobay BG, Hoch JA, Cavanagh J. Dynamics and activation in response regulators: The β4-α4 loop. Biomol Concepts. 2012;3(2):175-82.

45. Feher VA, Cavanagh J. Millisecond-timescale motions contribute to the function of the bacterial response regulator protein Spo0F. Nature. 1999;400(6741):289-93.

46. Lee SY, Cho HS, Pelton JG, Yan D, Henderson RK, King DS, Huang L, Kustu S, Berry EA, Wemmer DE. Crystal structure of an activated response regulator bound to its target. Nat Struct Biol. 2001;8(1):52-6.

47. Thomas SA, Brewster JA, Bourret RB. Two variable active site residues modulate response regulator phosphoryl group stability. Mol Micro. 2008;69(2):453-65.

48. Dominguez C, Boelens R, Bonvin AMJJ. HADDOCK: A Protein−Protein Docking Approach Based on Biochemical or Biophysical Information. J Am Chem Soc. 2003;125(7):1731-7.

49. Emsley PL, Bernhard; Scott, W.G.; Cowtan, Kevan. Features and development of Coot. Acta Cryst D. 2010;66:15.

50. Zhang L, Hermans J. Hydrophilicity of cavities in proteins. Proteins. 1996;24(4):433-8.

51. Gumbart J, Trabuco LG, Schreiner E, Villa E, Schulten K. Regulation of the protein-conducting channel by a bound ribosome. Structure. 2009;17(11):1453-64.

52. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. J Mol Graphics. 1996;14(1):33-8.

53. Brooks BR, Brooks CL, 3rd, Mackerell AD, Jr., Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M. CHARMM: The biomolecular simulation program. J Comput Chem. 2009;30(10):1545-614.

54. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B. 1998;102(18):3586-616.

55. Damjanović A, García-Moreno E B, Brooks BR. Self-guided Langevin dynamics study of regulatory interactions in NtrC. Proteins. 2009;76(4):1007-19.

56. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. Scalable molecular dynamics with NAMD. J Comput Chem. 2005;26(16):1781-802.

57. Fiorin G, Klein ML, Hénin J. Using collective variables to drive molecular dynamics simulations. Mol Physics. 2013;111(22-23):3345-62.

58. Team RDC. R: A language and environment for statistical computing. R Foundation. 2013.

59. Grant BJ, Rodrigues APC, ElSawy KM, McCammon JA, Caves LSD. Bio3d: an R package for the comparative analysis of protein structures. Bioinformatics. 2006;22(21):2695-6.

60. Abdel-Azeim S, Chermak E, Vangone A, Oliva R, Cavallo L. MDcons: Intermolecular contact maps as a tool to analyze the interface of protein complexes from molecular dynamics trajectories. BMC Bioinformatics. 2014;15(5):1-11.

61. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh ewald potential. J Chem Phys. 1995;103.

62. Aksimentiev A, Schulten K. Imaging alpha-hemolysin with molecular dynamics: ionic conductance, osmotic permeability, and the electrostatic potential map. Biophys J. 2005;88(6):3745-61.

63. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem. 2004;25(13):1605-12.

64. Bourret RB, Thomas SA, Page SC, Creager-Allen RL, Moore AM, Silversmith RE. Measurement of response regulator

autodephosphorylation rates spanning six orders of magnitude. Methods Enzymol. 2010;471:89-114.

65. Pazy Y, Wollish AC, Thomas SA, Miller PJ, Collins EJ, Bourret RB, Silversmith RE. Matching biochemical reaction kinetics to the timescales of life: Structural determinants that influence the autodephosphorylation rate of response regulator proteins. J Mol Biol. 2009;392(5):1205-20.

66. Stock AM, Martinez-Hackert E, Rasmussen BF, West AH, Stock JB, Ringe D, Petsko GA. Structure of the Mg(2+)-bound form of CheY and mechanism of phosphoryl transfer in bacterial chemotaxis. Biochemistry. 1993;32(49):13375-80.

67. Lukat GS, Stock AM, Stock JB. Divalent metal ion binding to the CheY protein and its significance to phosphotransfer in bacterial chemotaxis. Biochemistry. 1990;29(23):5436-42.

68. Luo S-C, Lou Y-C, Rajasekaran M, Chang Y-W, Hsiao C-D, Chen C. Structural Basis of a Physical Blockage Mechanism for the Interaction of Response Regulator PmrA with Connector Protein PmrD from Klebsiella pneumoniae. J Biol Chem. 2013;288(35):25551-61.

69. Da Re SS, Deville-Bonne D, Tolstykh T, M Vr, Stock JB. Kinetics of CheY phosphorylation by small molecule phosphodonors. FEBS Lett. 1999;457(3):323-6.

70. Mayover TL, Halkides CJ, Stewart RC. Kinetic characterization of CheY phosphorylation reactions: comparison of P-CheA and small-molecule phosphodonors. Biochemistry. 1999;38(8):2259-71.

Chapter 4: Elucidating the role of Gly68 in Ypd1

4.1 Introduction

Histidine phosphotransfer (HPt) proteins play a critical role within expanded TCS pathways. The chemical nature of phosphotransfer requires an alternating pattern of phosphorylatable histidine and aspartate residues for signal transmission (reviewed in [1]). Simple TCS systems achieve this with the upstream HK, autophosphorylating on a histidine residue, and the downstream RR rec domain, catalyzing the transfer of the phosphoryl to its own aspartate residue. The modular nature of these pathways provides tremendous opportunity for variations on the prototypical signaling scheme. The evolution of additional components within the pathway, such as hybrid HKs, containing both histidine and aspartate residues, necessitated the rise of a phosphorylatable histidine-containing intermediate, now referred to as the HPt domain [1-3]. Phylogenetic analysis of multi-step phosphorelays suggests that these expanded systems are likely bacterial in origin, with some evidence indicating a common ancestor with Hsp90, MutL and type II topoisomerases [2, 4, 5]. Though the majority of prokaryotes utilize basic TCS systems, nearly one-third of bacterial genomes contain at least one hybrid HK, indicating the prominence of phosphorelays [2]. Eukaryotic TCS systems are almost exclusively multi-step phosphorelays, incorporating hybrid HKs that are thought to have originated through lateral gene transfer with bacterial systems

[4]. Unlike bacterial systems, however, eukaryotic HPt domains are normally standalone proteins operating downstream from the rec domain of a hybrid HK [1].

One of the most well-characterized of the eukaryotic HPts, Ypd1, is found within the *S. cerevisiae* osmoregulatory Sln1 phosphorelay [6-14]. Ypd1 occupies a branch point within the system, able to receive a phosphoryl from the upstream hybrid HK, Sln1, and transfer to two downstream RRs, the cytoplasmic Ssk1 and the nuclear Skn7 [10, 11, 15-21]. Ssk1 is involved in responding to hyperosmotic stress through regulation of the HOG pathway [22]. Under non-stress conditions, Ypd1 constitutively phosphorylates Ssk1, suppressing its ability to interact with components of the HOG pathway [19, 20, 22, 23]. Hyperosmotic stress prevents Sln1 from autophosphorylating, ultimately contributing to the accumulation of unphosphorylated Ssk1 and the activation of the HOG pathway, which increases glycerol production to restore osmotic balance. Skn7 is involved in the cell wall and/or oxidative stress response [24]. Upon detection, Ypd1 translocates the signal across the nucleus, phosphorylating Skn7 and altering the transcription of stress response-related genes.

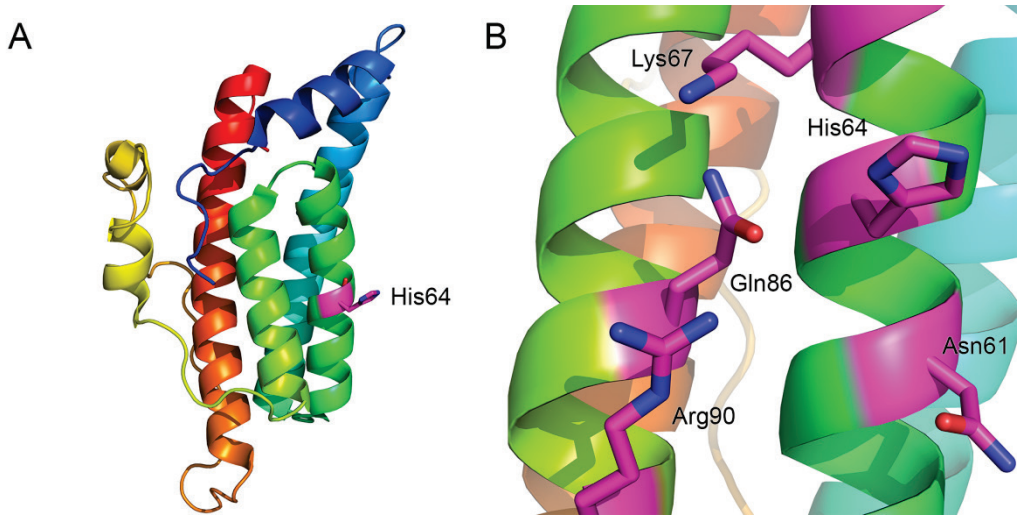

Figure 76. Structure and active site of the *S. cerevisiae* HPt, Ypd1.

(A) Crystal structure of Ypd1 (PDB 1OXB). Shown in stick model and is the phosphorylatable histidine residue (magenta). (B) Active site of Ypd1, with conserved residues shown in stick model (magenta). These residues form part of the common binding surface shared by all HPt domains.

Despite a highly conserved tertiary structure, little sequence homology exists between HPt domains. These proteins typically function non-enzymatically, unable to catalyze their own phosphorylation or the transfer of a phosphoryl group downstream [1]. What little sequence similarity exists is found near the phosphorylatable histidine. HPt domains adopt a common four-helix bundle fold. **Fig. 76A** shows this common structure with the standalone fungal HPt, Ypd1 [12]. In Ypd1, His64 is the conserved site of phosphorylation (**Fig. 76A, stick**). Adjacent residues, such as Asn61, Lys67, Gly68, Gln86 and Arg90, are also conserved and appear to contribute to the overall arrangement and solvent accessibility of the histidine residue (**Fig. 76B**) [14]. Yeast two-hybrid studies and available crystal structures have

revealed the existence of a common hydrophobic docking site around these residues utilized by all three rec domains in the Sln1 pathway [25, 26]. This feature occurs in nearly all HPt proteins [12, 13, 27, 28].

Based on the high-resolution crystal structure of Ypd1 and Sln1, point mutations of several conserved residues at the interface were discovered that disrupt phosphotransfer both to and from Ypd1 [17]. Gly68, located four positions downstream from the phosphorylatable histidine (H+4), was found to have a particularly deleterious effect upon mutation. Yeast two-hybrid studies suggest that its substitution to a glutamine (G68Q) severely impairs binding with all three rec domains within the Sln1 pathway [25, 26]. However, kinetic analysis found that while the same mutation has a negative effect on phosphoryl transfer, the binding itself is relatively unaffected [10]. A bioinformatic survey shows that 88% of all HPt sequences have a glycine at this position, with the remaining sequences containing mainly serine [29, 30]. Taken together, these results emphasize the essential nature of this position, but the somewhat contradictory information makes it challenging to determine the role Gly68 plays in the interactions between Ypd1 and its cognate rec domains.

To address this ambiguity, a two-fold approach was taken. Computational modeling and structural analysis were done for various substitution mutants to investigate Ypd1's interaction with the upstream hybrid HK, Sln1. This was possible due to the relatively large amount of structural

information available for Sln1 and Ypd1 [12, 14]. A more traditional characterization was attempted to study the Ypd1 G68Q mutant's interaction with the downstream RR rec domain, Ssk1-rec, using X-ray crystallography. The first section of this chapter will describe the computational results, and the second section will focus on the co-crystallization of Ypd1 G68Q and Ssk1-rec.

4.2 Computational studies to investigate the role of Gly68 in phosphotransfer between Sln1 and Ypd1

Two possible explanations exist for the effects of the G68Q mutation on the phosphotransfer between Sln1 and Ypd1. They are not mutually exclusive. The first hypothesis is that replacing the small glycine residue with a bulkier glutamine directly disrupts binding between Ypd1 and its partners. Both the position of Gly68 at the common binding interface (**Fig. 76B**) and the yeast two-hybrid data support this hypothesis [25, 26]. However, an *in vitro* fluorescence-based binding assay showed no significant changes in binding affinity between Sln1 and a variety of mutants at position 68 in Ypd1 when compared to wild-type (Skyler Hebdon, unpublished data). The second hypothesis is that substitutions at Gly68 affect the catalysis of phosphotransfer, likely through interactions with residues on the rec domain of Sln1. As explained in earlier sections, RR rec domains are responsible for catalyzing their own phosphotransfer reactions [31]. This is accomplished through a delicate series of conserved inter-atomic interactions at the active site,

involving a bound divalent metal cation and the phosphoryl group itself. Any disruption of these interactions would almost certainly have an effect on phosphotransfer.

A series of Ypd1 point mutants was created and analyzed using an *in vitro*, radiolabeled phosphotransfer assay to determine their phosphotransfer activity (Emily Kennedy, unpublished data). **Table 29** lists the substitutions and their phospho-accepting efficiencies (Emily Kennedy, unpublished data).

Table 29. Relative phospho-accepting abilities of Ypd1 mutants

Protein	Efficiency
Ypd1 Wild-type	100.0
Ypd1 G68S	94.1±3.6
Ypd1 G68A	82.0±15.6
Ypd1 G68V	2.5±3.5
Ypd1 G68L	0.4±0.6
Ypd1 G68E	0.3±0.6
Ypd1 G68Q	40.2±8.3

*Efficiencies are defined as relative to wild-type, which is set to 100%

These results suggest that large and/or hydrophobic side chains, such as glutamine, aspartate, valine and leucine, lead to an inability to receive a phosphoryl group. This partially explains the overwhelmingly conserved presence of a small and/or hydrophilic residue at this position in HPt domains. But this doesn't fully explain why these side chains decrease or eliminate the phosphotransfer ability of Ypd1. To investigate this further, mutations were modeled onto known crystal structures and relaxed using molecular dynamics (MD) simulations to provide estimates of their effects on the local chemical environment.

Materials and Methods

The following Ypd1 G68X mutations were simulated, in addition to a wild-type control: S, V, E, L and Q. Existing crystal structures were obtained for wild-type Ypd1 in complex with the BeF_3^- bound rec domain of Sln1 (PDB 2R25) [13]. These structures were used as templates for each point mutant. Mutations were modeled individually using the PyMOL program (v1.8. 2015, Schrodinger LLC ©2016). For Ypd1 G68Q and G68E, mutant crystal structures (Smita Menon and Krutik Soni, unpublished data) were taken and aligned to wild-type Ypd1 within the complex structure. The mutant Ypd1 and Sln1 structures were then extracted together. For the remaining mutants (G68S, G68L, G68V), side chains were modeled directly onto the BeF_3^- bound co-crystal structure (PDB 2R25) [14].

For each model, the BeF_3^- molecule was replaced with a PO_3^{2-} ligand. Systems were stripped of all other crystallographic waters and ligands, leaving only the Mg^{2+} cation in the active site. Proteins were submerged in an orthorhombic solvent box using the TIP3P water model. Systems were then neutralized and set to a final concentration of 0.15 M NaCl. Proteins were prepared and visualized using the Schrodinger Maestro suite (Maestro, version 10.6, Schrödinger, LLC, New York, NY, 2016).

Simulations were performed with Desmond using the OPLS 2005 force field [32, 33]. The default Desmond relaxation protocol was used, featuring two rounds of energy minimization (restrained and unrestrained), a series of

gradually diminishing restraints over ~270 ps, and a brief 100 ps step of unrestrained NPT simulation. To preserve the pre-phosphotransfer active site and the relative locations of the phosphorylatable histidine and aspartate, a position restraint with a force constant of 100 kcal/mol/Å was applied to the Nε2 atom of His64 and the phosphoryl group during each simulation. Final structures were extracted from trajectories using Maestro and analyzed with PyMOL (v1.8. 2015, Schrodinger LLC ©2016).

Mutations at the Gly68 position lead to changes in the active site residues involved in the catalysis of phosphotransfer

The mechanism of His-to-Asp phosphotransfer in proteins has not been studied directly, due to the difficulty of visualizing the reaction, but non-enzymatic reactions and kinetic information provide important clues as to the nature of the process. The generally accepted view involves a nucleophilic histidine or aspartate attacking the phosphorus atom of the phosphoryl group covalently bound to a donor atom [34]. Transition-state theory describes catalysis as preferential stabilization of a transition state [35-37]. Enzymes like rec domains must be able to facilitate and stabilize the pentavalent transition state that is formed during the phosphotransfer reaction [37]. This is achieved through a complex network of interactions around the phosphoryl group. Rec domains are known to bind a divalent metal cation in their negatively charged active site pocket [14, 38-42]. This metal is likely responsible for neutralizing the negative charge on the pentavalent transition state, stabilizing the high

energy species and ultimately making the reaction more favorable [34]. The same effect can also be accomplished by arranging neutralizing side chains near the phosphoryl group [43]. In rec domains, conserved active site residues, such as the Lys1195, Thr1173 and Ala1174 in Sln1, likely all stabilize the negatively charged transition state formed by transfer between Sln1 and Ypd1 [14].

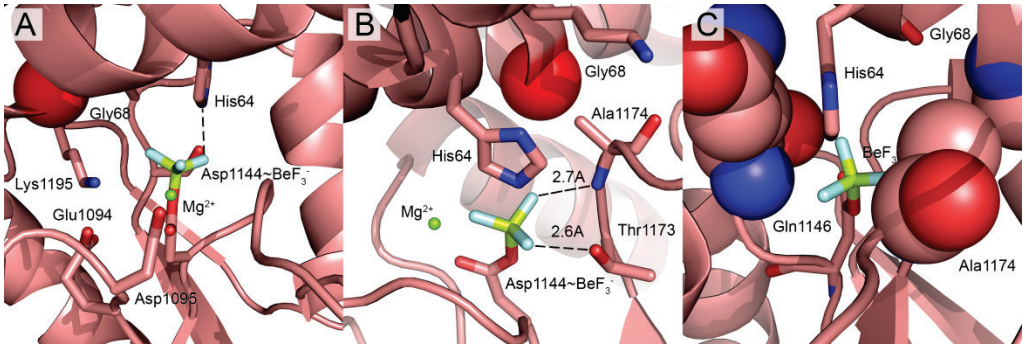

Figure 77. Chemical environment of the Sln1/Ypd1 active site.

(A) Close-up view of the pre-transition state active site from the Sln1/Ypd1~BeF_3^- crystal structure. Shown in stick model are BeF_3^- (yellow) bound to Asp1144, along with the conserved active site residues of Sln1. His64 on Ypd1 is aligned directly above the phosphoryl analog, ready for nucleophilic attack at the "phosphorus" atom. Gly68 of Ypd1 is shown as a sphere (red). Mg^{2+} is shown as a non-bonded sphere (green). (B) Side view of the active site, showing the conserved interactions between Ala1174 (T+1 position) and Thr1173 (switch residue) with the phosphoryl analog. These interactions are critical for stabilizing the negatively charged "phosphoryl" group. (C) Top-down view of the phosphorylatable aspartate (Asp1144). A molecular channel is formed by adjacent residues on both Sln1 and Ypd1. Any disruption to this channel can negatively affect phosphotransfer.

By modeling each mutation onto the "phosphorylated" complex structure of Sln1 and Ypd1, and allowing the structures to relax with MD simulations, the local structural changes caused by each substitution were revealed. Structural analysis was performed on the BeF_3^- bound complex between Sln1 and wild-type Ypd1 to determine the native active site

environment. **Fig. 77A** shows the presumably phosphotransfer-competent active site arrangement of the crystal structure. **Fig. 77B** shows the interactions between the conserved switch residue Thr1174 and the T+1 residue Ala1174 with the BeF_3^- group on Sln1. **Fig. 77C** shows the "channel" through which His64 on Ypd1 arranges to access the phosphoryl group on Sln1 for nucleophilic attack.

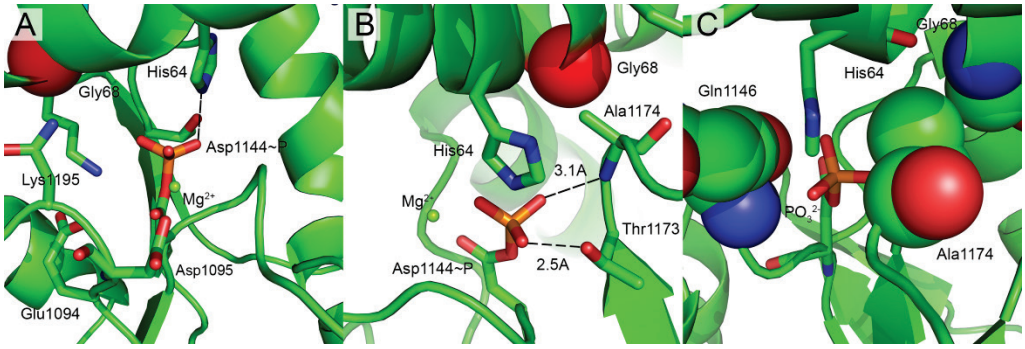

Figure 78. Chemical environment of the simulated Sln1•Ypd1 active site.

(A) Close-up view of the pre-transition state active site from the simulated wild-type control. Shown in stick model are PO_3^{2-} (orange) bound to Asp1144, along with the conserved active site residues of Sln1. His64 on Ypd1 is largely unaffected, still positioned directly above the phosphoryl group, ready for nucleophilic attack at the phosphorus atom. Gly68 of Ypd1 is shown as a sphere (red). Mg^{2+} is shown as a non-bonded sphere (green). (B) Side view of the active site, showing the conserved interactions between Ala1174 (T+1 position) and Thr1173 (switch residue) with the phosphoryl analog. The interactions are maintained in the wild-type control. (C) Top-down view of the phosphorylatable aspartate (Asp1144). The molecular channel is still formed, largely due to the positioning of the β4α4 loop.

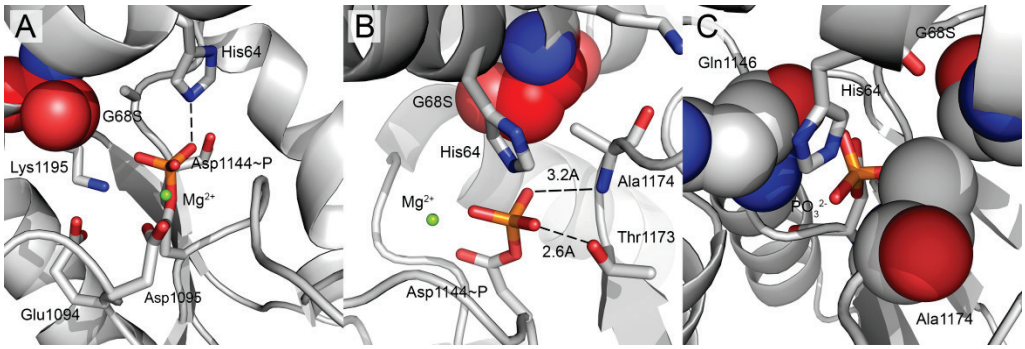

Figure 79. Structural changes caused by the G68S mutation.

(A) Close-up view of the pre-transition state active site from the simulated G68S mutant. Shown in stick model are PO_3^{2-} (orange) bound to Asp1144, along with the conserved active site residues of Sln1. His64 on Ypd1 is largely unaffected (the side chain can freely rotamerize), still positioned above the phosphoryl group. Mutation has caused slight distortion of the phosphoryl geometry. Ser68 of Ypd1 is shown as a sphere (red). Mg^{2+} is shown as a non-bonded sphere (green). (B) Side view of the active site, showing the conserved interactions between Ala1174 (T+1 position) and Thr1173 (switch residue) with the phosphoryl analog. The interactions are maintained, despite the G68S mutation. (C) Top-down view of the phosphorylatable aspartate (Asp1144). The molecular channel is left unaffected.

Wild-type Ypd1 and Ypd1 G68S showed little structural perturbation upon relaxation. In both proteins, the phosphoryl group's linear, upright geometry is largely maintained. Both donor and acceptor atoms remain in the original pre-transfer arrangement (**Figs. 78-79A**). Conserved side chains of Lys1195, Ala1174 and Thr1173 are unchanged as well (**Figs. 78-79B**). Taking these results into consideration, it is logical that Ypd1 G68S possesses a near-wild-type phosphotransfer efficiency.

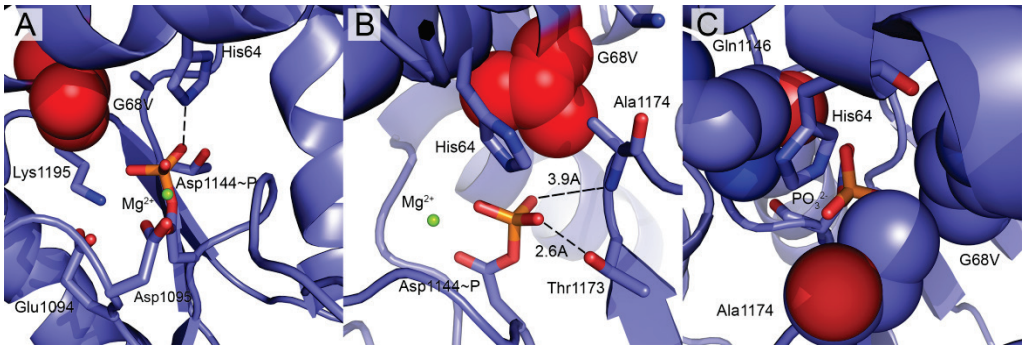

Figure 80. Structural changes caused by the G68V mutation.

(A) Close-up view of the pre-transition state active site from the simulated G68V mutant. Shown in stick model are PO_3^{2-} (orange) bound to Asp1144, along with the conserved active site residues of Sln1. His64 on Ypd1 has shifted out of alignment with the phosphoryl group. Mutation has also distorted of the phosphoryl geometry. Val68 of Ypd1 is shown as a sphere (red). Mg^{2+} is shown as a non-bonded sphere (green). (B) Side view of the active site, showing the conserved interactions between Ala1174 (T+1 position) and Thr1173 (switch residue) with the phosphoryl analog. The amine nitrogen on Ala1174 has been shifted from 2.7 Å in wild-type to 3.9 Å. (C) Top-down view of the phosphorylatable aspartate (Asp1144). The molecular channel has been altered, due to Ala1174 shifting on the β4α4 loop.

Ypd1 G68V caused slight distortion of the phosphoryl geometry (**Fig. 80A**), but led to a significant shift in the β4α4 loop of Sln1, specifically affecting Ala1174 (**Fig. 80B**). Typically, the amine nitrogen atom of Ala1174 forms a hydrogen bond with one oxygen atom of the phosphoryl group, while Thr1173 forms an hydrogen bond with the opposite atom. By replacing the small glycine residue with valine, the nitrogen shifted well outside of hydrogen bonding range with the oxygen. Both this and the phosphoryl group distortion likely contribute to the mutant's low (2.5%) phosphotransfer efficiency.

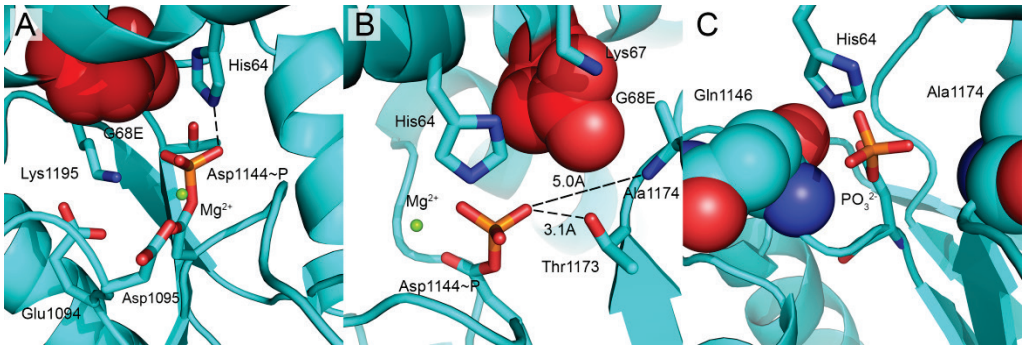

Figure 81. Structural changes caused by the G68E mutation.

(A) Close-up view of the pre-transition state active site from the simulated G68E mutant. Shown in stick model are PO_3^{2-} (orange) bound to Asp1144, along with the conserved active site residues of Sln1. His64 on Ypd1 has shifted out of alignment with the phosphoryl group. Mutation has also distorted the phosphoryl geometry. Glu68 of Ypd1 is shown as a sphere (red). Mg^{2+} is shown as a non-bonded sphere (green). (B) Side view of the active site, showing the conserved interactions between Ala1174 (T+1 position) and Thr1173 (switch residue) with the phosphoryl analog. Both residues have undergone drastic shifts away from the phosphoryl group. In addition, Glu68 can form direct hydrogen bonding interactions with each residue. The acidic side chain also forms a salt-bridge with Lys67. (C) Top-down view of the phosphorylatable aspartate (Asp1144). The molecular channel has been drastically altered, with the β4α4 loop almost completely absent from the region.

Ypd1 G68E generated moderate distortion of the phosphoryl group (**Fig. 81A**), but caused an even larger shift of the β4α4 loop on Sln1 (**Fig. 81B**). Interestingly, the mutant glutamate side chain formed a direct interaction with Sln1 at both Ala1174 and Thr1173 (**Fig. 81B**). These highly conserved residues are responsible for forming critical hydrogen bonds with the phosphoryl group in active rec domains [44]. The acidic side chain was also in close enough proximity to form a salt-bridge with Lys67 on Ypd1 (**Fig 81B**). Previously, mutation of Lys67 to alanine was found to severely impair phosphotransfer between Sln1 and Ypd1 [17]. The authors hypothesized that the positively charged lysine was involved in stabilizing the helical bundle of

Ypd1 through an inter-helical interaction with Glu83 [17]. The mutant salt-bridge formed by G68E appears to interfere with this process.

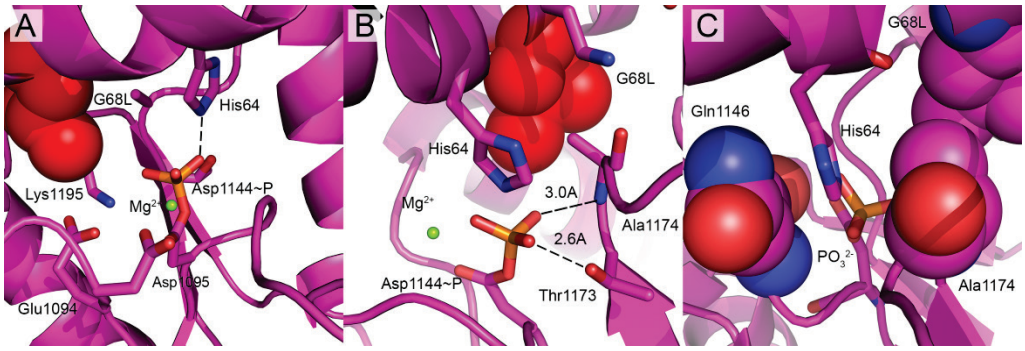

Figure 82. Structural changes caused by the G68L mutation.

(A) Close-up view of the pre-transition state active site from the simulated G68L mutant. Shown in stick model are PO_3^{2-} (orange) bound to Asp1144, along with the conserved active site residues of Sln1. His64 on Ypd1 has shifted out of alignment with the phosphoryl group. Mutation has also severely distorted the phosphoryl geometry. Leu68 of Ypd1 is shown as a sphere (red). Mg^{2+} is shown as a non-bonded sphere (green). (B) Side view of the active site, showing the conserved interactions between Ala1174 (T+1 position) and Thr1173 (switch residue) with the phosphoryl analog. The hydrogen bonds and β4α4 loop are largely unaffected. (C) Top-down view of the phosphorylatable aspartate (Asp1144). The molecular channel remains intact.

Ypd1 G68L showed severe distortion of the phosphoryl group geometry, though the β4α4 loop was seemingly unaffected (**Fig. 82AB**). The distortion is likely caused by repulsive effects between the mutant leucine and the nearby Lys1195 on Sln1, which forms a highly conserved salt-bridge that stabilizes the negatively charged phosphoryl.

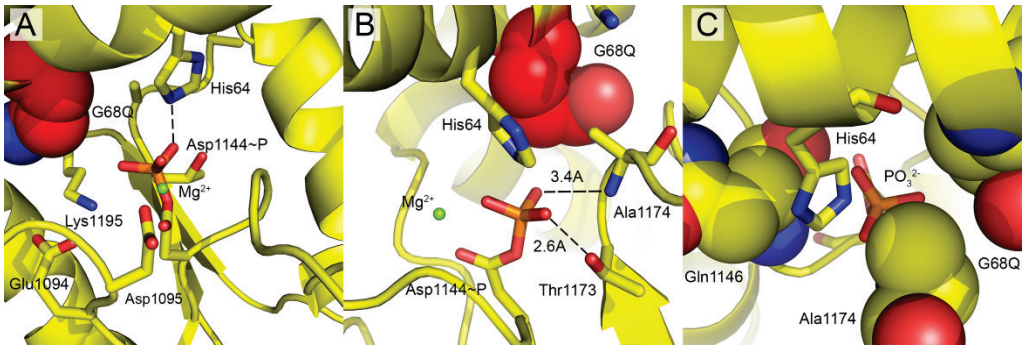

Figure 83. Structural changes caused by the G68Q mutation.

(A) Close-up view of the pre-transition state active site from the simulated G68Q mutant. Shown in stick model are PO_3^{2-} (orange) bound to Asp1144, along with the conserved active site residues of Sln1. His64 on Ypd1 is still in near alignment with the phosphoryl group. Mutation has caused some slight distortion to the phosphoryl geometry. Gln68 of Ypd1 is shown as a sphere (red). Mg^{2+} is shown as a non-bonded sphere (green). (B) Side view of the active site, showing the conserved interactions between Ala1174 (T+1 position) and Thr1173 (switch residue) with the phosphoryl analog. The amine nitrogen on Ala1174 has shifted some, but is still within range of hydrogen bonding with the proper phosphoryl oxygen. (C) Top-down view of the phosphorylatable aspartate (Asp1144). The molecular channel is maintained.

Finally, Ypd1 G68Q caused moderate distortion, but also left the β4α4 loop mostly unaffected (**Fig. 83AB**). Additionally, the mutant glutamine side chain satisfied additional hydrogen bonds with residues on Sln1, which may offset some of its likely negative steric effects on the active site.

Examination of **Figs. 77-83C** shows the active site "channel" between the two proteins through which His64 on Ypd1 and Asp1144 on Sln1 are able to transfer the phosphoryl group. A delicate arrangement of adjacent residues is required for this proper alignment to occur, including the semi-conserved D+2 position Gln1146, the switch residue Thr1173 and the T+1 position Ala1174 on Sln1, as well as Gly68 on Ypd1. Without these residues to form the linear geometry appropriate for the reaction, phosphotransfer is likely

inhibited. Those mutations that abolished phosphotransfer typically disrupted this channel in some way, such as through interference with the β4α4 loop.

Table 30 combines the relevant measurements (phosphor-accepting efficiency, dihedral angle of phosphoryl group, interatomic distances) to compare the effects of the various substitutions. Examination of the structural changes induced by the mutations reveals an approximate trend relating to Ypd1's phospho-accepting ability. As the ideal phosphoryl geometry is distorted, phosphotransfer efficiency is decreased due to the in-line nature of the transfer reaction. This is quantified using the dihedral angle formed between the C_β, C_γ, $O_{\delta 2}$, and P atoms. In addition, disruption of the Thr1173/Ala1174 interactions with the phosphoryl oxygen atoms appears to negatively affect phosphotransfer from Sln1 to Ypd1. These interactions likely serve the same purpose as the stabilizing interaction between the essential Lys1195 and the negatively charged phosphoryl group. Overall, introduction of bulky, hydrophobic side chains within the delicate arrangement of highly charged active site residues interferes with the catalysis of phosphotransfer in numerous ways. The exception to this is the G68E mutation, whose acidic side chain causes similar effects through direct electrostatic interactions with other critical residues on both Ypd1 and Sln1.

Table 30. Comparison of efficiencies and active site properties

Mutation	Phospho-accepting %	Dihedral [1]	Thr1173[2]	Ala1174[3]
Wild-type	100.0	-175.9	2.5	3.1
G68S	94.1±3.6	173.1	2.6	3.2

G68V	2.5±3.5	166.7	2.6	3.9
G68E	0.3±0.5	158.4	3.5	5.0
G68L	0.4±0.6	147.9	2.6	3.0
G68Q	40.2±8.3	151.4	2.6	3.4

[1] Dihedral angle (in degrees) of phosphoryl group, measured for C_β-C_γ-$O_{\delta 2}$-P atoms
[2] Interatomic distance (in Å) between hydroxyl oxygen and phosphoryl oxygen
[3] Interatomic distance (in Å) between amine nitrogen and phosphoryl oxygen

4.3 Co-crystallization of Ssk1 and Ypd1 G68Q

The co-crystal structures of the apo and BeF_3^- bound forms of Sln1/Ypd1 have provided significant insight into the function of eukaryotic phosphorelays. However, the branching nature of the pathway suggests that valuable insights could be obtained from a comparison between this upstream interaction and the interaction of Ypd1 with one of its downstream RR partners. Many years have been spent pursuing a co-crystal structure between the rec domain of Ssk1 (Ssk1-rec) and Ypd1, with limited success. The Ssk1-rec protein has proven largely intransigent to purification and crystallization attempts. Recently, a fortuitous point mutation (W638A) on Ssk1-rec was found to significantly improve solubility and purification yields. The mutant possesses near wild-type properties and was characterized as a pseudo wild-type variant (Katie Branscum, unpublished). These studies led to the successful co-crystallization and structural elucidation of the Ssk1-rec W638A•Ypd1 complex structure (Katie Branscum, Smita Menon, unpublished). However, while invaluable, only the apo form was observed. The active site of the complex is characterized as "loose," in that the phosphorylatable histidine and

aspartate residues are unusually far apart. The lack of a bound metal cation also leaves the rest of the conserved active site side chains disordered. Previous radiolabeled phosphotransfer assays suggest that the Ypd1 G68Q mutant is deficient in both accepting the phosphoryl group from Sln1 and transferring it downstream to Ssk1, though with different efficiencies. In order to visualize why this mutation may be negatively affecting downstream phosphotransfer, co-crystallization attempts were done to obtain the structures of the signaling complex in the apo and BeF_3^- bound forms. This project was carried out in collaboration with Katie Branscum, who originally characterized the fortuitous Ssk1-rec W638A point mutant and succeeded in co-crystallizing the Ssk1-rec W638A•Ypd1 complex.

Mutant constructs and cloning

A construct of Ssk1-rec W638A was generated through QuikChange mutagenesis using an existing wild-type plasmid with a hybrid pET11a/pCYB2 vector design (see **Appendix A**). For wild-type Ypd1, the YPD1 gene from *S. cerevisiae* was inserted into a pME43 vector as detailed in [13]. The Ypd1 G68Q mutant was created with site-directed mutagenesis using this plasmid as a template and primers listed in **Appendix A.**

Expression and purification

Ypd1 G68Q was expressed and purified as previously described for wild-type Ypd1 [13]. Briefly, Ypd1 G68Q was grown in *E. coli* DH5α cells using 1 L of LB with 100 μg/mL of ampicillin. The protein was expressed

constitutively from a modified pUC12 vector, pME43, for 19 hours at 37 °C, 200 rpm. Cells were pelleted and lysed with French Press in a lysis buffer of 100 mM sodium phosphate, pH 7.0, 1 mM EDTA and 1 mM βME. Lysate was clarified through centrifugation for 1 hour at 25,000 x g at 4 °C. Following separation, saturated ammonium sulfate was gradually added to the supernatant up to 55% by volume. The solution was gently stirred for 30 minutes at 4 °C. The resulting mixture was separated again with centrifugation, and the supernatant was discarded. The pellet was resuspended in a dialysis buffer of 20 mM BisTris pH 6.5 and 1 mM βME. This was dialyzed into 2 L of buffer for several hours at 4°C. The buffer was exchanged with fresh dialysis buffer and dialyzed for another 9 hours to remove remaining ammonium sulfate. The protein sample was then subjected to centrifugation to remove precipitated debris. The supernatant was filtered and added to a 5 mL HiTrap Q anion exchange column (GE Healthcare) attached to an AKTA prime FPLC system. The sample was run in 20 mM BisTris pH 6.5 with an increasing concentration of NaCl, 0-1 M). **Fig. 84** shows an SDS-PAGE gel with the Ypd1 G68Q fractions eluted from the column.

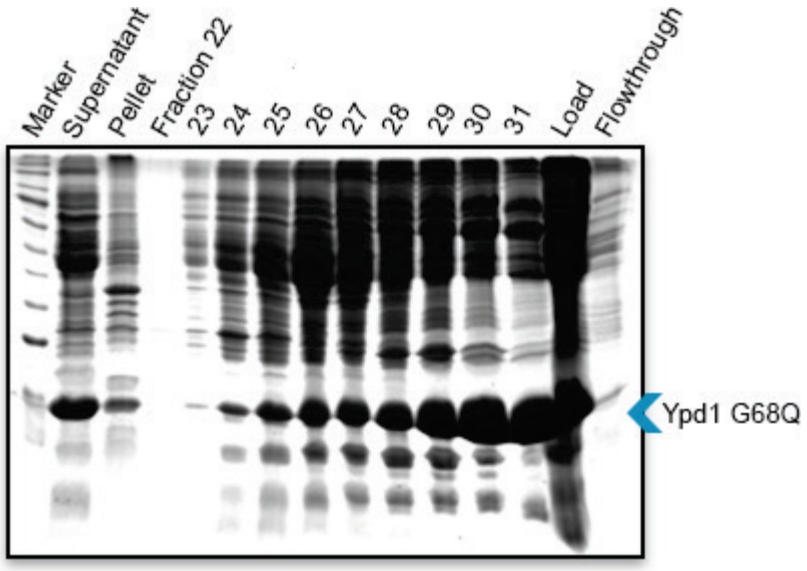

Figure 84. Anion exchange elutions for Ypd1 G68Q.

Fractions collected from a 5 mL HiTrap Q anion exchange column. Ypd1 G68Q (19 kDa) elutes in fractions 24-36. These fractions were collected, concentrated and filtered for gel filtration.

Fractions were pooled and concentrated to a final volume of <0.5 mL or until precipitation was observed. The sample was filtered and added to a Sephadex G-75 column equilibrated in 20 mM Tris, pH 8.0, 100 mM NaCl and 1% glycerol. All solutions were passed through Chelex® resin to remove metal ions beforehand, as this was found to drastically extend the life of Ssk1-rec W638A. Fractions containing Ypd1 were identified and pooled for co-concentration (see **Fig. 85**).

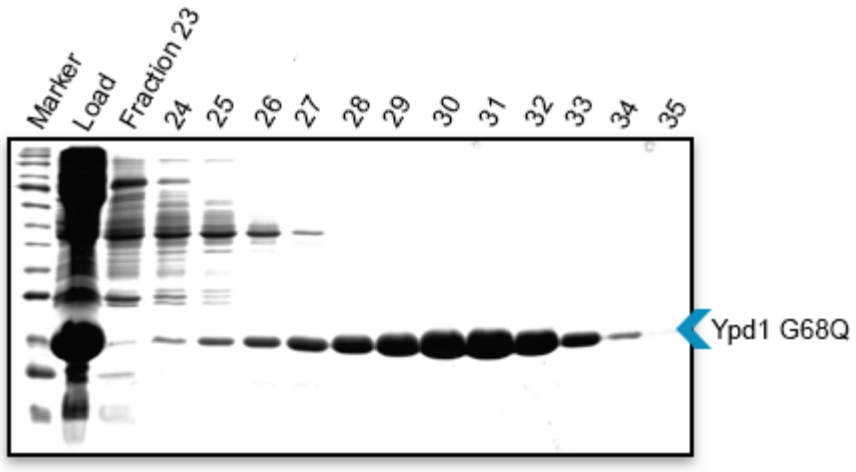

Figure 85. Gel filtration elutions for Ypd1 G68Q.

Fractions collected from a hand-poured Sephadex G-75 column. Ypd1 G68Q (19 kDa) elutes in fractions 28-34. These fractions were collected and pooled for co-concentration.

Ssk1-rec W638A was purified as previously described (Katie Branscum, unpublished data). Briefly, the protein was expressed in *E. coli* BL21 Star cells. Two 1 L flasks of LB with 100 µg/mL of ampicillin were inoculated with 10 mL of saturated cultures and grown at 37 °C to an OD_{600} of ≥0.8. Cultures were then cooled to 16 °C and induced with 1 mM IPTG. These were grown for approximately 22 hours. Cells were pelleted and lysed with French Press, and lysate was clarified by centrifugation at 25,000 x g for 1 hour at 4 °C. Only fresh cells were used, as the freezing process was found to significantly decrease final protein yield. Supernatant was added to a 4 mL chitin column pre-equilibrated in lysis buffer of 20 mM Tris pH 8.0, 500 mM NaCl, 1 mM EDTA, 10% glycerol and 0.1% Triton X-100. Flow through was discarded and samples were washed with 40 mL of lysis buffer and 40 mL of a cleavage buffer with 20 mM Tris pH 8.0, 100 mM NaCl and 10% glycerol.

The column was incubated overnight at 4 ºC with cleavage buffer + 30 mM βME. Cleaved protein was eluted with Chelexed® cleavage buffer (see **Fig. 85**). Samples were then filtered and concentrated to <500 mL. These were then run on a Sephadex G75 column equilibrated in 20 mM Tris pH 8.0, 100 mM NaCl and 1% glycerol. All solutions were Chelexed® beforehand, as this was found to drastically extend the life of Ssk1-rec W638A. Fractions containing Ssk1-rec W638A were identified and pooled for co-concentration (see **Fig. 86**).

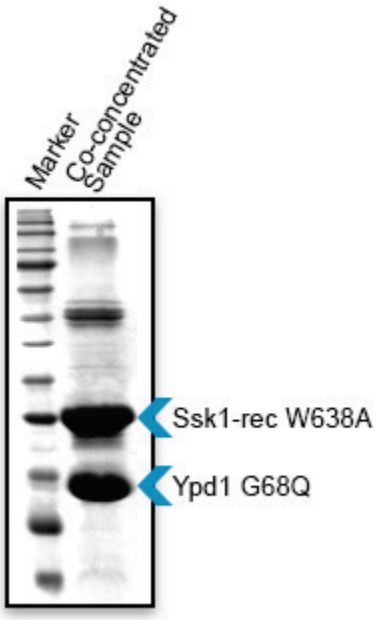

Figure 86. Co-concentrated sample with Ssk1-rec W638A•Ypd1 G68Q.

After gel filtration, equimolar amounts of both proteins were combined and co-concentrated in crystallization buffer. Here, the proteins were concentrated and filtered to a final concentration of 10.63 mg/mL. 24-well hanging drop crystallization trays were then set up with this sample. Note: the upper-band contaminant is characteristic of all Ssk1-rec batches and is likely an endogenous protein from the expression strain.

Co-crystallization attempts

Equimolar amounts of Ssk1-rec W638A and Ypd1 G68Q were combined and co-concentrated to ~9-15 mg/mL using passivated Amicon Ultra Centrifugal filters with a molecular weight cutoff of 10K (EMD Millipore). A final co-concentration buffer of 20 mM Tris pH 8.0, 100 mM NaCl and 1% glycerol was used. **Fig. 86** shows a Coomassie stained SDS-PAGE gel loaded with a typical co-concentrated sample used for crystallization. The initial conditions identified to produce the best results during the co-crystallization of Ssk1-rec W638A and wild-type Ypd1 were 0.1 M CAPS pH 10.5, 1.0-1.2 M NaH_2PO_4, 0.1-0.2 M $LiSO_4$ and 0.8 M K_2HPO_4, based on well A5 of the MCSG-II broad crystallization screen. Hanging-drop, vapor-diffusion 24-well screens for Ssk1-rec and Ypd1 G68Q were set up around similar well conditions:

Table 31. Initial co-crystallization well conditions

	1	2	3	4	5	6
A	0.1 M CAPS pH 10.5 0.8 M NaH_2PO_4 0.1 M $LiSO_4$ 0.8 M K_2HPO_4	0.1 M CAPS pH 10.5 1.0 M NaH_2PO_4 0.1 M $LiSO_4$0.8 M K_2HPO_4	0.1 M CAPS pH 10.5 1.2 M NaH_2PO_4 0.1 M $LiSO_4$0.8 M K_2HPO_4	0.1 M CAPS pH 10.5 1.4 M NaH_2PO_4 0.1 M $LiSO_4$0.8 M K_2HPO_4	0.1 M CAPS pH 10.5 1.6 M NaH_2PO_4 0.1 M $LiSO_4$0.8 M K_2HPO_4	0.1 M CAPS pH 10.5 1.76 M NaH_2PO_4 0.1 M $LiSO_4$0.8 M K_2HPO_4
B	0.1 M CAPS pH 10.5 0.8 M NaH_2PO_4 0.2 M $LiSO_4$0.8 M K_2HPO_4	0.1 M CAPS pH 10.5 1.0 M NaH_2PO_4 0.2 M $LiSO_4$0.8 M K_2HPO_4	0.1 M CAPS pH 10.5 1.2 M NaH_2PO_4 0.2 M $LiSO_4$0.8 M K_2HPO_4	0.1 M CAPS pH 10.5 1.4 M NaH_2PO_4 0.2 M $LiSO_4$0.8 M K_2HPO_4	0.1 M CAPS pH 10.5 1.6 M NaH_2PO_4 0.2 M $LiSO_4$0.8 M K_2HPO_4	0.1 M CAPS pH 10.5 1.76 M NaH_2PO_4 0.2 M $LiSO_4$0.8 M K_2HPO_4

C	0.1 M CAPS pH 10.5 1.2 M NaH_2PO_4 0.1 M $LiSO_4$0.4 M K_2HPO_4	0.1 M CAPS pH 10.5 1.2 M NaH_2PO_4 0.1 M $LiSO_4$0.6 M K_2HPO_4	0.1 M CAPS pH 10.5 1.2 M NaH_2PO_4 0.1 M $LiSO_4$0.8 M K_2HPO_4	0.1 M CAPS pH 10.5 1.2 M NaH_2PO_4 0.1 M $LiSO_4$1.0 M K_2HPO_4	0.1 M CAPS pH 10.5 1.2 M NaH_2PO_4 0.1 M $LiSO_4$1.2 M K_2HPO_4	0.1 M CAPS pH 10.5 1.2 M NaH_2PO_4 0.1 M $LiSO_4$1.4 M K_2HPO_4
D	0.1 M CAPS pH 10.5 1.2 M NaH_2PO_4 0.2 M $LiSO_4$0.4 M K_2HPO_4	0.1 M CAPS pH 10.5 1.2 M NaH_2PO_4 0.2 M $LiSO_4$0.6 M K_2HPO_4	0.1 M CAPS pH 10.5 1.2 M NaH_2PO_4 0.2 M $LiSO_4$0.8 M K_2HPO_4	0.1 M CAPS pH 10.5 1.2 M NaH_2PO_4 0.2 M $LiSO_4$1.0 M K_2HPO_4	0.1 M CAPS pH 10.5 1.2 M NaH_2PO_4 0.2 M $LiSO_4$1.2 M K_2HPO_4	0.1 M CAPS pH 10.5 1.2 M NaH_2PO_4 0.2 M $LiSO_4$1.4 M K_2HPO_4

Additional fine-screening was performed to identify optimal crystallization conditions for the mutant co-complex. The following conditions were found to produce the largest, best-diffracting crystals based on 24-well fine screens:

Table 32. Further optimized co-crystallization well conditions.

	1	2	3	4	5	6
A	0.1 M CAPS pH 10.5 1.15 M NaH_2PO_4 0.2 M $LiSO_4$0.65 M K_2HPO_4	0.1 M CAPS pH 10.5 1.16 M NaH_2PO_4 0.2 M $LiSO_4$0.65 M K_2HPO_4	0.1 M CAPS pH 10.5 1.17 M NaH_2PO_4 0.2 M $LiSO_4$0.65 M K_2HPO_4	0.1 M CAPS pH 10.5 1.18 M NaH_2PO_4 0.2 M $LiSO_4$0.65 M K_2HPO_4	0.1 M CAPS pH 10.5 1.19 M NaH_2PO_4 0.2 M $LiSO_4$0.65 M K_2HPO_4	0.1 M CAPS pH 10.5 1.20 M NaH_2PO_4 0.2 M $LiSO_4$0.65 M K_2HPO_4
B	0.1 M CAPS pH 10.5 1.15 M NaH_2PO_4 0.2 M $LiSO_4$0.73 M K_2HPO_4	0.1 M CAPS pH 10.5 1.16 M NaH_2PO_4 0.2 M $LiSO_4$0.73 M K_2HPO_4	0.1 M CAPS pH 10.5 1.17 M NaH_2PO_4 0.2 M $LiSO_4$0.73 M K_2HPO_4	0.1 M CAPS pH 10.5 1.18 M NaH_2PO_4 0.2 M $LiSO_4$0.73 M K_2HPO_4	0.1 M CAPS pH 10.5 1.19 M NaH_2PO_4 0.2 M $LiSO_4$0.73 M K_2HPO_4	0.1 M CAPS pH 10.5 1.20 M NaH_2PO_4 0.2 M $LiSO_4$0.73 M K_2HPO_4
C	0.1 M CAPS pH 10.5 1.15 M NaH_2PO_4 0.2 M $LiSO_4$0.78 M K_2HPO_4	0.1 M CAPS pH 10.5 1.16 M NaH_2PO_4 0.1 M $LiSO_4$0.6 M K_2HPO_4	0.1 M CAPS pH 10.5 1.17 M NaH_2PO_4 0.1 M $LiSO_4$0.8 M K_2HPO_4	0.1 M CAPS pH 10.5 1.18 M NaH_2PO_4 0.1 M $LiSO_4$1.0 M K_2HPO_4	0.1 M CAPS pH 10.5 1.19 M NaH_2PO_4 0.1 M $LiSO_4$1.2 M K_2HPO_4	0.1 M CAPS pH 10.5 1.20 M NaH_2PO_4 0.1 M $LiSO_4$1.4 M K_2HPO_4

D	0.1 M CAPS pH 10.5 1.15 M NaH_2PO_4 0.2 M $LiSO_4$0.85 M K_2HPO_4	0.1 M CAPS pH 10.5 1.6 M NaH_2PO_4 0.2 M $LiSO_4$0.6 M K_2HPO_4	0.1 M CAPS pH 10.5 1.17 M NaH_2PO_4 0.2 M $LiSO_4$0.8 M K_2HPO_4	0.1 M CAPS pH 10.5 1.18 M NaH_2PO_4 0.2 M $LiSO_4$1.0 M K_2HPO_4	0.1 M CAPS pH 10.5 1.2 M NaH_2PO_4 0.2 M $LiSO_4$1.2 M K_2HPO_4	0.1 M CAPS pH 10.5 1.2 M NaH_2PO_4 0.2 M $LiSO_4$1.4 M K_2HPO_4

The following sample conditions/additives were also attempted in order to optimize crystal quality and/or produce a co-crystal complex with a pre-transition state active site arrangement:

Table 33. Sample buffer conditions and additives for co-crystallization.

Sample conditions	Additives
20 mM Tris pH 8.0, 75 mM NaCl, 1% glycerol	N/A
20 mM Tris pH 8.0, 100 mM NaCl, 1% glycerol	N/A
20 mM Tris pH 8.0, 125 mM NaCl, 1% glycerol	N/A
20 mM Tris pH 8.0, 100 mM NaCl, 1% glycerol	15 mM $MgCl_2$
20 mM Tris pH 8.0, 75 mM NaCl, 1% glycerol	15 mM $MgCl_2$
20 mM Tris pH 8.0, 100 mM NaCl, 1% glycerol	15 mM $MgCl_2$, 5 mM $BeCl_2$, 35 mM NaF
20 mM Tris pH 8.0, 75 mM NaCl, 1% glycerol	15 mM $MgCl_2$, 5 mM $BeCl_2$, 35 mM NaF
20 mM Tris pH 8.0, 150 mM NaCl, 1% glycerol	N/A

Note: Buffer was exchanged during final co-concentration step with Amicon filter. Additives were introduced to samples post-concentration.

Results of co-crystallization screening

1:1 μL hanging-drop screens were found to rapidly produce crystals ~50-150 μm in diameter, typically within a week. Crystal formation was surprisingly sensitive to NaCl concentration, requiring 100 mM NaCl to produce crystals of any significant size; 75 mM NaCl produced numerous small crystals unsuitable for screening; >100 mM NaCl failed to produce crystals under any attempted conditions. Most large crystals were grown under

conditions with no additives at room temperature. A single condition was found to produce adequate crystals with 15 mM $MgCl_2$, 5 mM $BeCl_2$, 35 mM NaF included in the sample: 0.1 M CAPS pH 10.5, 1.15 M NaH_2PO_4, 0.2 M $LiSO_4$, 0.65 M K_2HPO_4 (well A1). Addition of 15 mM $MgCl_2$ alone appeared to inhibit crystal growth.

After approximately 7 days, crystallization drops formed a resilient surface film, likely composed of denatured protein. Attempts to prevent this by varying salt concentration and protein purity were unsuccessful. Once exposed to air, the film rapidly thickened. This "skin" made manipulation challenging. In addition, most crystals were fragile and firmly adhered to the cover slip, despite using a hanging drop approach.

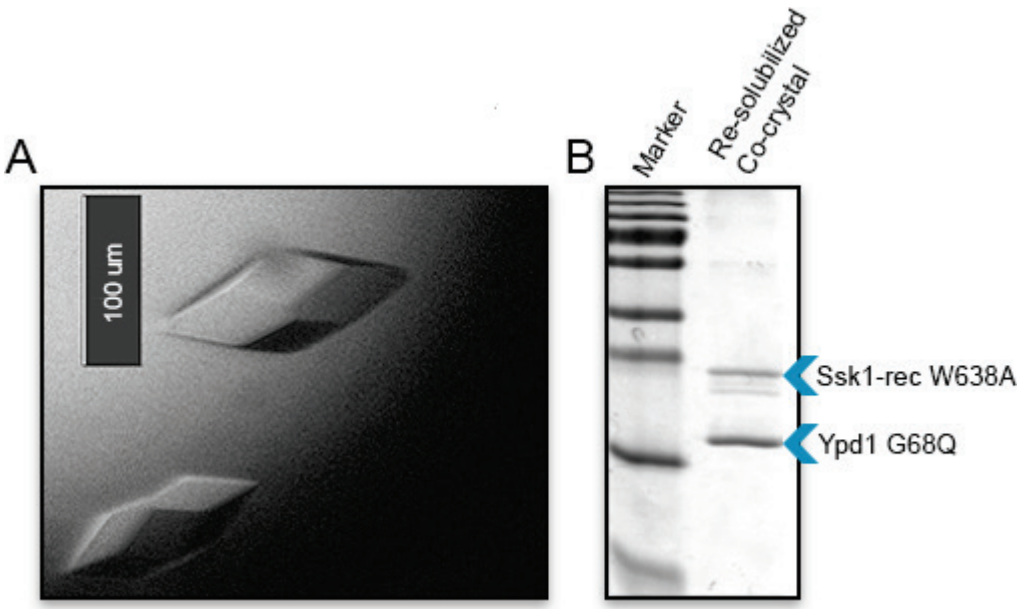

Figure 87. Two-week crystal growth with $MgCl_2$, $BeCl_2$ and NaF.

(A) Co-crystals grown at room temperature with 15 mM $MgCl_2$, 5 mM $BeCl_2$ and 35 mM NaF in the sample. (B) Crystals were harvested and rinsed in 1X Laemmli buffer, then run on an SDS-PAGE gel to confirm the presence of both Ssk1-rec W638A and Ypd1 G68Q.

Fig. 87A shows a representative co-crystal from sample containing 15 mM $MgCl_2$, 5 mM $BeCl_2$, 35 mM NaF, 20 mM Tris pH 8.0, 100 mM NaCl and 1% glycerol. Crystals were produced at room temperature under the following conditions: 0.1 M CAPS pH 10.5, 1.15 M NaH_2PO_4, 0.2 M $LiSO_4$, 0.65 M K_2HPO_4. The same crystals were harvested, rinsed in 1X Laemmli buffer and run on a SDS-PAGE gel to verify the presence of both proteins. **Fig. 87B** shows a Coomassie stain of this sample, confirming that both Ssk1-rec W638A and Ypd1 G68Q exist in roughly equimolar concentrations. **Fig. 88** shows typical crystals grown at room temperature from a sample containing 20 mM Tris pH 8.0, 100 mM NaCl and 1% glycerol with no additives under the following conditions: 0.1 M CAPS pH 10.5, 1.15 M NaH_2PO4, 0.2 M $LiSO_4$ and 0.65 M K_2HPO_4.

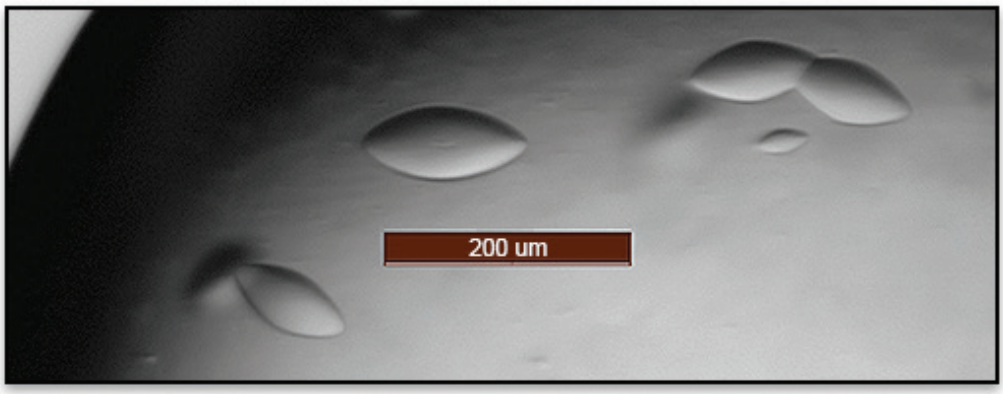

Figure 88. Typical two-week crystal growth with no additives.

Co-crystals grown at room temperature from a sample with 20 mM Tris pH 8.0, 100 mM NaCl and 1% glycerol, and no additives. The rounded appearance was frequently observed under these conditions.

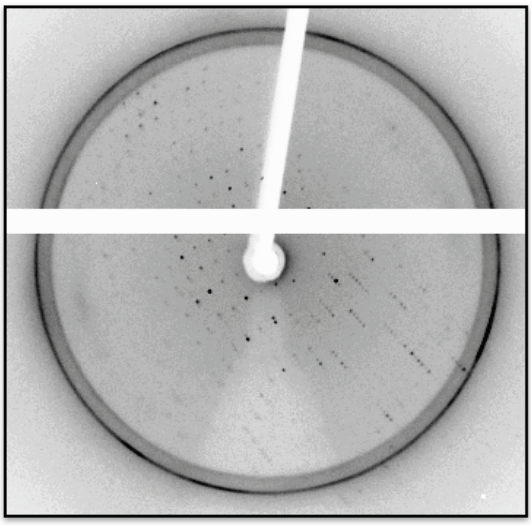

Figure 89. Co-crystal room temperature diffraction (3.5-4 Å).

A crystal from the well seen above (**Fig. 88**) was harvested and placed into a MiTeGen Room Temperature capillary mount. The proteins produced diffraction out to approximately 3.5 Å at room temperature.

Cryoprotectant optimization

The optimal cryoprotectant conditions for the original Ssk1-rec W638A•Ypd1 complex were found to be 1.2 M NaH_2PO_4, 0.8 M K_2HPO_4, 0.2 M $LiSO_4$, 0.1 M CAPS pH 10.5 and 9% glycerol (Katie Branscum, Smita Menon, unpublished data). These conditions were tested for the Ssk1-rec W638A•Ypd1 G68Q crystals with poor results. Most crystals cracked when exposed to freezing temperatures. In addition, attempts to directly mount crystals under a cold stream rapidly produced ice. Mounting crystals at room temperature and collecting for several frames produced diffraction to 3.5-4 Å (see **Fig. 89; crystal from well A2, Table 32**) using the home source X-ray generator (1.541 Å Cu Kα rotating anode) with a Pilatus 200K detector

(Rigaku). The following conditions were then attempted to optimize cryoprotection:

Table 34. Cryoprotection optimization trials.

Condition	Result
15% ethylene glycol in mother liquor	5-6 Å diffraction (poor)
Stepwise: 5 -> 10 -> 15% ethylene glycol in mother liquor	5-6 Å diffraction (poor)
100% sodium malonate	5-6 Å diffraction (poor)
14% triethylene glycol in mother liquor	Rapid ice formation (poor)
30% trehalose in mother liquor	5-6 Å diffraction (poor)
30% sucrose in mother liquor	5-6 Å diffraction, distinct reflections (moderate)
2.5 M $LiSO_4$	Rapid ice formation
100% 1-2-propanediol	5-6 Å diffraction (poor)
Stepwise: 15 -> 30% sucrose, wait 1 hour for equilibration each step	5-6 Å diffraction, distinct reflections (moderate)
CryoProtX (Molecular Dimensions)	Rapid ice formation (poor)
PEG 3350	Crystals lose integrity (poor)
*Saturated sucrose in mother liquor (1 hour equilibration time)	4.5-5 Å diffraction, distinct reflections (moderate)
*Saturated sucrose in mother liquor (5 hours equilibration time)	4.5-5 Å diffraction, distinct reflections (moderate)
**Krytox Fluorinated oil	3-4 Å diffraction (good)

*2 μL drop of mother liquor was placed next to crystallization drop. A solid sucrose crystal was added directly to this. The two drops were bridged, and the well was resealed to allow for equilibration.
**Drops were completely submerged in Krytox oil.

Both direct mounting of crystals onto the goniometer under a cryostream and prior submersion in liquid nitrogen were attempted. The methods produced similar results. The most promising initial condition was a 30% sucrose/mother liquor solution. This showed improvement in crystal diffraction by producing distinct, round reflections. However, the resolution was too poor for characterization of a single point mutant side chain.

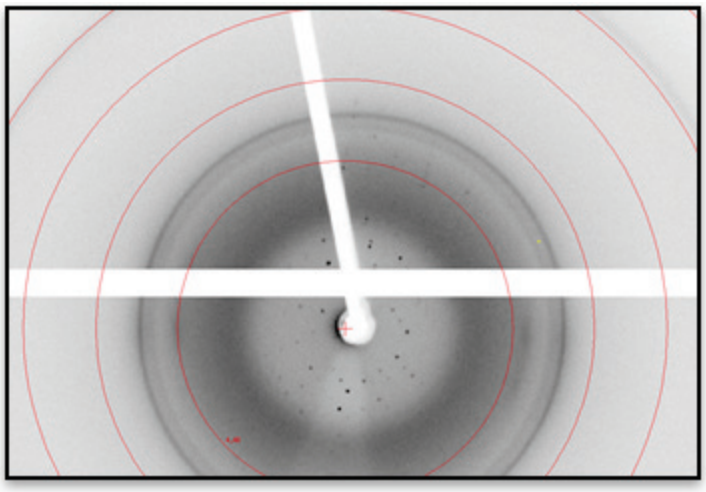

Figure 90. Diffraction to 3.5-4 Å by co-crystal containing additives while using Krytox oil as cryoprotectant.

Pattern produced by crystals seen in **Fig. 87,** from sample containing 15 mM $MgCl_2$, 5 mM $BeCl_2$, 35 mM NaF, 20 mM Tris pH 8.0, 100 mM NaCl and 1% glycerol.

The next significant improvement was observed using Krytox Fluorinated oil. As previously described, crystallization drops invariably formed a surface "skin" after approximately one week. This skin made manipulating and harvesting crystals extremely challenging. Most crystallization conditions produced fragile crystals, likely due to a high solvent content (suggested by the round appearance and lack of facets on most crystals). Combined, these likely contributed to the poor diffraction that was observed during cryotrials. Skin formation was rapidly increased when drops were exposed to air. In an attempt to prevent this, wells were unsealed and drops were submerged completely in Krytox oil. Crystals were then harvested from this submersion and directly mounted under a cryostream. The oil not only prevented skin formation by minimizing air exposure, but also provided

an effective method for buffer exchange and cryoprotection. **Fig. 90** shows a diffraction pattern obtained using Krytox oil for a drop containing 15 mM $MgCl_2$, 5 mM $BeCl_2$, 35 mM NaF (see **Fig. 87A** for an image of the crystals). The oil improved resolution to approximately 3.5-4 Å for 2 frames with a 120 second exposure time. Additional crystals were attempted without the additives and produced slightly better results (see **Fig. 91**). Unfortunately, the data for these crystals would not index properly, making them unsuitable for a full dataset collection.

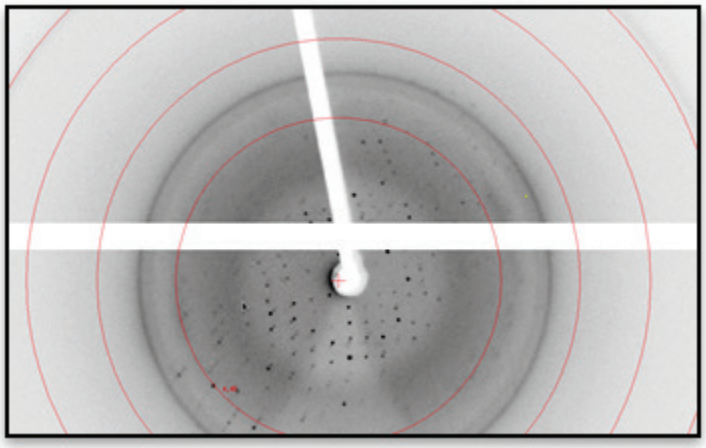

Figure 91. Diffraction to 3.5-4 Å by co-crystal containing no additives while using Krytox oil as cryoprotectant.

Pattern produced by similar crystals as seen in **Fig. 88,** from sample containing 20 mM Tris pH 8.0, 100 mM NaCl and 1% glycerol.

Future directions for characterizing the co-crystal structure

Submersion in Krytox Fluorinated oil appears to be an effective cryoprotectant for the Ssk1-rec W638A•Ypd1 G68Q crystals. Screening produced diffraction data between 3.5-4 Å for some of the larger harvested

crystals. Best results were from crystals grown in the following conditions: 0.1 M CAPS pH 10.5, 1.16 M NaH_2PO_4, 0.2 M $LiSO_4$, 0.65 M K_2HPO_4 for samples with no additives (in 20 mM Tris pH 8.0, 100mM NaCl, 1% glycerol); 0.1 M CAPS pH 10.5, 1.15 M NaH_2PO_4, 0.2 M $LiSO_4$, 0.65 M K_2HPO_4 for samples containing 15 mM $MgCl_2$, 5 mM $BeCl_2$, 35 mM NaF (in 20 mM Tris pH 8.0, 100 mM NaCl, 1% glycerol). Poor results were obtained for crystals containing 15 mM $MgCl_2$ alone. The success of using Krytox oil to prevent skin formation prompted an attempt to grow crystals in drops covered with Al's oil, though this was unsuccessful. The crystals used to test Krytox as a cryoprotectant were several months old and despite the oil, had already formed a partial skin. Harvesting these likely damaged the crystals and incorporated skin onto the loop before mounting, as was repeatedly seen when testing the other cryoprotectants. Both could affect the quality of the diffraction data. The next step for this project is to grow additional crystals for diffraction screening. More attempts are needed to find an intact crystal suitable for the collection of a full dataset.

4.4 Conclusions

In conclusion, MD relaxation simulations of Ypd1 G68X mutations were used to reveal clues about the role that this residue plays in phosphotransfer between the upstream Sln1-rec and the intermediate HPt. Bulky or hydrophobic amino acids were found to disrupt critical side chain geometries involved in catalysis, but not complex formation. These results

explain the evolutionary conservation of this position in nearly all HPt domains. In addition, attempts to co-crystallize Ypd1 G68Q with the downstream Ssk1-rec W638A were successful. Optimal growth and cryoprotection conditions were identified. The cryo-conditions not only protect the crystals from ice formation, but also inhibit their characteristic skin growth upon exposure to air, making manipulation of the fragile constructs more successful.

References

1. Fassler JS, West AH. Histidine Phosphotransfer Proteins in Fungal Two-Component Signal Transduction Pathways. Eukaryotic Cell. 2013;12(8):1052-60.

2. Zhang W, Shi L. Distribution and evolution of multiple-step phosphorelay in prokaryotes: lateral domain recruitment involved in the formation of hybrid-type histidine kinases. Microbiology. 2005;151(Pt 7):2159-73.

3. Wuichet K, Cantwell BJ, Zhulin IB. Evolution and phyletic distribution of two-component signal transduction systems. Current Opinion in Microbiology. 2010;13(2):219-25.

4. Koretke KK, Lupas AN, Warren PV, Rosenberg M, Brown JR. Evolution of two-component signal transduction. Mol Biol Evol. 2000;17.

5. Dutta R, Qin L, Inouye M. Histidine kinases: diversity of domain organization. Mol Microbiol. 1999;34(4):633-40.

6. Saito H, Posas F. Response to Hyperosmotic Stress. Genetics. 2012;192(2):289-318.

7. Levin DE. Cell Wall Integrity Signaling in Saccharomyces cerevisiae. Microbiology and Molecular Biology Reviews. 2005;69(2):262-91.

8. Fassler JS, West AH. Genetic and Biochemical Analysis of the SLN1 Pathway in Saccharomyces cerevisiae. Methods Enzymol. 2010;471:291-317.

9. Hohmann S. Osmotic stress signaling and osmoadaptation in yeasts. Microbiol Mol Biol Rev. 2002;66(2):300-72.

10. Janiak-Spens F, Cook PF, West AH. Kinetic analysis of YPD1-dependent phosphotransfer reactions in the yeast osmoregulatory phosphorelay system. Biochemistry. 2005;44(1):377-86.

11. Lu JM, Deschenes RJ, Fassler JS. Saccharomyces cerevisiae histidine phosphotransferase Ypd1p shuttles between the nucleus and cytoplasm for SLN1-dependent phosphorylation of Ssk1p and Skn7p. Eukaryot Cell. 2003;2(6):1304-14.

12. Xu Q, Porter SW, West AH. The yeast YPD1/SLN1 complex: Insights into molecular recognition in two-component signaling systems. Structure. 2003;11(12):1569-81.

13. Xu Q, West AH. Conservation of structure and function among histidine-containing phosphotransfer (HPt) domains as revealed by the crystal structure of YPD1. J Mol Biol. 1999;292(5):1039-50.

14. Zhao X, Copeland DM, Soares AS, West AH. Crystal structure of a complex between the phosphorelay protein YPD1 and the response regulator domain of SLN1 bound to a phosphoryl analog. J Mol Biol. 2008;375(4):1141-51.

15. Horie T, Tatebayashi K, Yamada R, Saito H. Phosphorylated Ssk1 prevents unphosphorylated Ssk1 from activating the Ssk2 mitogen-activated protein kinase kinase kinase in the yeast high-osmolarity glycerol osmoregulatory pathway. Molecular and cellular biology. 2008;28(17):5172-83.

16. Janiak-Spens F, Sparling JM, Gurfinkel M, West AH. Differential stabilities of phosphorylated response regulator domains reflect functional roles of the yeast osmoregulatory SLN1 and SSK1 proteins. J Bact. 1999;181(2):411-7.

17. Janiak-Spens F, West AH. Functional roles of conserved amino acid residues surrounding the phosphorylatable histidine of the yeast phosphorelay protein YPD1. Mol Micro. 2000;37(1):136-44.

18. Kaserer AO, Andi B, Cook PF, West AH. Effects of osmolytes on the SLN1-YPD1-SSK1 phosphorelay system from *Saccharomyces cerevisiae*. Biochemistry. 2009;48(33):8044-50.

19. Posas F, Saito H. Osmotic activation of the HOG MAPK pathway via Ste11p MAPKKK: scaffold role of Pbs2p MAPKK. Science. 1997;276(5319):1702-5.

20. Posas F, Wurgler-Murphy SM, Maeda T, Witten EA, Thai TC, Saito H. Yeast HOG1 MAP kinase cascade is regulated by a multistep phosphorelay mechanism in the SLN1-YPD1-SSK1 "two-component" osmosensor. Cell. 1996;86(6):865-75.

21. Song HK, Lee JY, Lee MG, Moon J, Min K, Yang JK, Suh SW. Insights into eukaryotic multistep phosphorelay signal transduction revealed by the crystal structure of Ypd1p from Saccharomyces cerevisiae. J Mol Biol. 1999;293(4):753-61.

22. Maeda T, Wurgler-Murphy SM, Saito H. A two-component system that regulates an osmosensing MAP kinase cascade in yeast. Nature. 1994;369(6477):242-5.

23. Reiser V, Raitt DC, Saito H. Yeast osmosensor Sln1 and plant cytokinin receptor Cre1 respond to changes in turgor pressure. The Journal of cell biology. 2003;161(6):1035-40.

24. Shankarnarayan S, Malone CL, Deschenes RJ, Fassler JS. Modulation of yeast Sln1 kinase activity by the CCW12 cell wall protein. J Biol Chem. 2008;283(4):1962-73.

25. Porter SW, West AH. A common docking site for response regulators on the yeast phosphorelay protein YPD1. Biochimica et biophysica acta. 2005;1748(2):138-45.

26. Porter SW, Xu Q, West AH. Ssk1p response regulator binding surface on histidine-containing phosphotransfer protein Ypd1p. Eukaryot Cell. 2003;2(1):27-33.

27. Kato M, Mizuno T, Shimizu T, Hakoshima T. Refined structure of the histidine-containing phosphotransfer (HPt) domain of the anaerobic sensor kinase ArcB from Escherichia coli at 1.57 A resolution. Acta crystallographica Section D, Biological crystallography. 1999;55(Pt 11):1842-9.

28. Bauer J, Reiss K, Veerabagu M, Heunemann M, Harter K, Stehle T. Structure-function analysis of Arabidopsis thaliana histidine kinase AHK5 bound to its cognate phosphotransfer protein AHP1. Molecular plant. 2013;6(3):959-70.

29. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL, Studholme DJ, Yeats C, Eddy SR. The Pfam protein families database. Nucleic Acids Res. 2004;32.

30. Tatusova T, Ciufo S, Fedorov B, O'Neill K, Tolstoy I. RefSeq microbial genomes database: New representation and annotation strategy. Nucleic Acids Res. 2014;42(Database issue):D553-9.

31. Lukat GS, McCleary WR, Stock AM, Stock JB. Phosphorylation of bacterial response regulator proteins by low molecular weight phospho-donors. Proc Natl Acad Sci. 1992;89(2):718-22.

32. Banks JL, Beard HS, Cao Y, Cho AE, Damm W, Farid R, Felts AK, Halgren TA, Mainz DT, Maple JR, Murphy R, Philipp DM, Repasky MP, Zhang LY, Berne BJ, Friesner RA, Gallicchio E, Levy RM. Integrated Modeling Program, Applied Chemical Theory (IMPACT). J Comput Chem. 2005;26(16):1752-80.

33. Bowers KJ, Chow DE, Xu H, Dror RO, Eastwood MP, Gregersen BA, Klepeis JL, Kolossvary I, Moraes MA, Sacerdoti FD, Salmon JK, Shan Y, Shaw DE, editors. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. SC 2006 Conference, Proceedings of the ACM/IEEE; 2006 11-17 Nov. 2006.

34. Silversmith RE, Appleby JL, Bourret RB. Catalytic mechanism of phosphorylation and dephosphorylation of CheY: kinetic

characterization of imidazole phosphates as phosphodonors and the role of acid catalysis. Biochemistry. 1997;36(48):14965-74.

35. Wolfenden R. Analog approaches to the structure of the transition state in enzyme reactions. Accounts of Chemical Research. 1972;5(1):10-8.

36. Gutsche CD. Catalysis in Chemistry and Enzymology. William P. Jencks. McGraw-Hill, New York, 1969. xvi + 656 pp. McGraw-Hill Series in Advanced Chemistry. Science. 1970;168(3935):1080-1.

37. Lassila JK, Zalatan JG, Herschlag D. Biological phosphoryl-transfer reactions: understanding mechanism and catalysis. Annu Rev Biochem. 2011;80:669-702.

38. Cho H, Wang W, Kim R, Yokota H, Damo S, Kim SH, Wemmer D, Kustu S, Yan D. BeF_3^- acts as a phosphate analog in proteins phosphorylated on aspartate: structure of a BeF_3^- complex with phosphoserine phosphatase. Proc Natl Acad Sci. 2001;98(15):8525-30.

39. Kojetin DJ, Thompson RJ, Benson LM, Naylor S, Waterman J, Davies KG, Opperman CH, Stephenson K, Hoch JA, Cavanagh J. Structural analysis of divalent metals binding to the *Bacillus subtilis* response regulator Spo0F: The possibility for in vitro metalloregulation in the initiation of sporulation. Biometals. 2005;18(5):449-66.

40. Lukat GS, Stock AM, Stock JB. Divalent metal ion binding to the CheY protein and its significance to phosphotransfer in bacterial chemotaxis. Biochemistry. 1990;29(23):5436-42.

41. Stock AM, Martinez-Hackert E, Rasmussen BF, West AH, Stock JB, Ringe D, Petsko GA. Structure of the Mg(2+)-bound form of CheY and mechanism of phosphoryl transfer in bacterial chemotaxis. Biochemistry. 1993;32(49):13375-80.

42. Varughese KI, Tsigelny I, Zhao H. The crystal structure of beryllofluoride Spo0F in complex with the phosphotransferase Spo0B represents a phosphotransfer pretransition state. J Bact. 2006;188(13):4970-7.

43. Lukat GS, Lee BH, Mottonen JM, Stock AM, Stock JB. Roles of the highly conserved aspartate and lysine residues in the response regulator of bacterial chemotaxis. J Biol Chem. 1991;266(13):8348-54.

44. Stock AM, Guhaniyogi J. A new perspective on response regulator activation. J Bact. 2006;188(21):7328-30.

Chapter 5: Appendix A

5.1 List of Abbreviations

General Abbreviations

AST	Dianionic phosphoaspartate
ATP	Adenosine triphosphate
βME	β-mercaptoethanol
CA	Catalytic and ATP-binding (kinase domain)
cAMP	Cyclic adenosine monophosphate
cGMP	Cyclic guanine monophosphate
CLARA	Clustering for large applications
DCCM	Dynamical cross-correlation map
DHp	Dimerization and histidine phosphotransfer (kinase domain)
EDTA	Ethylenediaminetetraacetic acid
HK	Histidine kinase
HPt	Histidine phosphotransfer (domain)
IC	Intermolecular contact
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB	Luria broth
MAP3K	Mitogen activated protein kinase kinase kinase
MD	Molecular dynamics
NAMD	Nanoscale Molecule Dynamics program
NCBI	National Center for Biotechnology Information
PCA	Principal component analysis
PDB	Protein Data Bank
RCSB	Research Collaboratory for Structural Bioinformatics
Rec	Receiver (domain)
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
RR	Response regulator
SASA	Solvent accessible surface area

SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
TCS	Two-component signaling

Protein Designations

CheY	Chemotaxis protein Y; *E. coli*
FixJ	Nitrogen fixation protein J; *R. meliloti*
HOG1	High osmolarity glycerol response; *S. cerevisiae*
PhoP	Transcriptional regulatory protein PhoP; *E. coli*
Skn7	Suppressor of Kre Null; *S. cerevisiae*
Sln1	Synthetic lethal of the N-end rule 1; *S. cerevisiae*
Spo0F	Sporulation initial phosphotransferase F; *B. subtilis*
Ssk1	Suppressor of the sensor kinase 1; *S. cerevisiae*
Ssk2/22	Suppressor of sensor kinase; *S. cerevisiae*
Ypd1	Tyrosine phosphatase dependent 1; *S. cerevisiae*

5.2 Constructs

Organism	Protein Product	Oligo No.	Restriction Sites	Plasmid	Storage Strain No.	Expression Strain No.
S.c.	Ssk1-rec W638A	AW776/ AW777	NdeI/SmaI	pl430	OU668 (DH5α)	OU738 - BL21 Star
S.c.	Ssk1-rec W638A	AW776/ AW777	NdeI/SmaI	pl430	OU668 (DH5α)	OU669 - BL21 Gold
S.c.	Ypd1	N/A	NdeI/PstI	pl15	OU830 (DH5α)	OU831 - DH5α
S.c.	Ypd1 G68Q	AW118/ AW155	NdeI/PstI	pl51	OU832 (DH5α)	OU833 - DH5α

*S.c. represents *Saccharomyces cerevisiae*

5.3 Primers

Oligo No.	5' or 3'	Protein	Sequence (5' to 3')
AW776	5'	Ssk1-rec W638A	CTGACTAAACCAGTGAATTTACACGCGCTTAGTAAGAAAATTACAGAGTG
AW777	3'	Ssk1-rec W638A	CACTCTGTAATTTTCTTACTAAGCGCGTGTAAATTCACTGGTTTAGTCAG
AW118	5'	Ypd1 G68Q	AATCTGGGCCATTTTTTAAAGCAATCTTCTGCT
AW155	3'	Ypd1 G68Q	GCCTAATGCAGCAGAAGATTGCTTTAAAAA