Elucidation of a red light sensing two component system from non-photosynthetic bacteria

By

Anna W. Baker

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Microbiology)

At the

UNIVERSITY OF WISCONSIN-MADISON

2016

Date of final oral examination: 7/12/2016

This dissertation is approved by the following members of the Final Oral Committee:

Katrina T. Forest, Professor, Bacteriology

Timothy J. Donohue, Professor, Bacteriology

Katherine D. McMahon, Professor, Bacteriology, Civil and Environmental Engineering

Garret Suen, Assistant Professor, Bacteriology

Yongna Xing, Associate Professor, Oncology

Table of Contents ……i

Abstract

Bacteriophytochrome (BphP) histidine kinases (HKs) and their cognate response regulators (BRRs) form two-component signaling systems (TCSs) that sense and respond to visible light in bacteria. Immediate signal transduction is accomplished through a two-step phosphorelay in which the light-regulated HK first autophosphorylates and then transfers a phosphate group to the RR protein. In most BphP systems, the RR has no output domain and the consequent signal transduction mechanism is unknown. Furthermore, for non-photosynthetic bacteria, no common phenotype has been discovered for BphP-BRR TCSs. This work characterizes the BphP-BRR TCS from the chemotrophic desert bacterium *Ramlibacter tataouinensis* and analyzes BphP systems from other non-photosynthetic bacteria to address these knowledge gaps. RtBphP1 was spectrally and biochemically characterized as a red light sensitive histidine kinase with suppressed autokinase activity in the illuminated state. The structures of its cognate RR protein, RtBRR, and a homolog from *Agrobacterium tumefaciens* were solved and determined to dimerize in a unique arm-in-arm quaternary fold, an arrangement that promotes phosphotransfer from RtBphP1 to RtBRR. Based on a conserved amino acid sequence motif that forms the interaction, arm-in-arm dimers are predicted to be common to many BRRs from non-photosynthetic bacteria. During development of a non-radioactive BphP kinase assay, several unexpected post-translational modifications (PTMs) were discovered in the full length BphPs RtBphP1 and Agp1. Lysine acetylation and serine phosphorylation were present at multiple sites, including regions of the BphP involved in ligand binding, regulatory conformational change, and autokinase activities. A combination of bioinformatics analyses allowed prediction of likely interaction partners for BRRs and inference of phenotypic consequences to red light signal transduction. BRR interaction targets included hybrid histidine kinases, a conserved uncharacterized protein, oxidative stress response, DNA

repair, and energy storage enzymes. Taken together, these findings structurally elucidate BphP-BRR signal transduction, suggest possible impacts on enzyme activity by PTM beyond reversible histidine/aspartate phosphorylation, and define targets for BRR protein interaction and phenotypic studies.

Acknowledgements

My completion of this work was supported by excellent funding from the NSF Graduate Research Fellowship Program and the NIH-sponsored Biotechnology Training Program. Having a fellowship every year in graduate school allowed me to pursue topics of great interest to me personally, and I felt liberated to study a basic science question with the NSF's support. I thank the BTP for the opportunity to intern at a biotechnology company, which solidified my future interest in working in the research and development industry.

I thank my advisor, Katrina Forest, for attracting me to the lab and for being open and enthusiastic to me beginning a new line of research, for helping me to grow as a researcher, and for her personal support when needed. I am grateful to my thesis committee members: Tim Donohue, Trina McMahon, Garret Suen, and Yongna Xing for their helpful ideas, support, and constancy through all five years of my degree. I believe they represented exactly what a thesis committee should be to a student.

The members of the Forest Lab past and present have been a great talent pool to draw on for brainstorming, troubleshooting, and friendship. I especially thank Ken Satyshur and Neydis Moreno whose assistance on the X-ray beam and in the lab was invaluable. Special thanks also to lab members Michele Auldridge, Lorraine McLaughlin, Carolina Sepulveda, Shy Bhattacharya, Pun Wangkanont, Doug Bosley, Heli Lehtivuori, Katiria Gonzalez-Rivera, Peter Newhouse, and Jeff Dwulit-Smith. Thank you all for contributing to a great work environment.

Many thanks to our excellent collaborators abroad in the bacteriophytochrome and *Ramlibacter* fields. The work presented here was based off of essential groundwork isolating and characterizing *R. tataouinensis* as a species completed by Thierry Heulin, Gilles DeLuca, and Andre Vermeglio among others. I am very grateful to Patrick Scheerer and Bilal Qureshi for

our collaboration on BphP-BRR complexes, and for the purified Agp1 protein provided by Patrick's lab.

I have been fortunate to receive logistical support from many sources within the Department of Bacteriology and the Microbiology Doctoral Training Program at UW-Madison. It has been a pleasure to be a part of this extremely well-run department that puts its students first. Thanks especially to Rick Gourse, Cathy Davis-Gray, Gary Pine, Dean Maly, and Janet Newlands. Many thanks to Michael Thomas and his lab and Jade Wang and her lab for access to equipment essential to completion of this work. Special thanks to Sarah Stevens of the McMahon Lab for her contribution to Chapter IV of this work.

Two very special mentors from my past deserve thanks for leading me to this point: Emina Stojkovic and Carrie Harwood. Thank you for believing in me and giving me the opportunities I needed to make it to the PhD. I honestly would not be here without your support.

Last and most important, thank you to my family for their endless support. To my mom, Gretchen Winkler Baker, thank you for supporting me for my entire life, not just through this degree. To my dad, Joe Baker, I miss you and know that you would be proud of this achievement. Thanks to my sister Brigitte Baker, who has always been there with love and straight talk and thanks to my soon to be brother in law John Seyferth. To my husband Geoff Baker-McAlister, I am in awe of how much love and thought you give to me every single day and I can't thank you enough, so I will just show you. To my kids, Ephraim McAlister, Ellasyn McAlister, and our baby Baker-McAlister, thank you for your love and for giving me something greater than myself to work for. I also thank Deb McAlister, Ian McAlister, Graeme McAlister, and Christy Simmons for welcoming me into their awesome family.

List of Figures and Tables

List of Abbreviations and Acronyms

BphP- bacteriophytochrome protein

Cph- BphP from cyanobacteria

HK- histidine kinase

RR- response regulator

SDRR- single domain response regulator

BRR- bacteriophytochrome response regulator (BRRs are also typically SDRRs)

RtBRR- BRR from *R. tataouinensis*

AtBRR- BRR from *A. tumefaciens*

TCS- two component system

PTM- post translational modification

RtBphP1- BphP #1 from *R. tataouinensis*

RtBphP2- BphP #2 from *R. tataouinensis*

Agp1- BphP #1 from *A. tumefaciens*

DrBphP- BphP from *D. radiodurans*

RtBphP1- PAS-GAF- RtBphP1 lacking the PHY and HK domains

DrBphP- PAS-GAF- DrBphP lacking the PHY and HK domains

BV- biliverdin Ixα, the chromophore for BphPs

HO- heme oxygenase, enzyme that forms open chain BV from heme

CheY- a SDRR regulating chemotaxis in *E. coli* and other species

FliM- flagellar motor protein bound by CheY

eSTK- eukaryotic-like serine/threonine kinase

YihE- non eSTK type serine/threonine kinase found in E. coli and other species

Pat- acetyltransferase found in *S. enterica* and other species

SDS-PAGE- sodium dodecyl sulfate polyacrylamide gel electrophoresis

PAS- PER/ARNT/SIM a large protein domain family named for representative proteins period clock protein, aryl hydrocarbon receptor nuclear translocator, and single-minded protein

GAF- protein domain family named for representative proteins cyclic GMP phosphodiesterase, adenyl cyclases, and FhlA

PHY- phytochrome specific domain

HisKA- coiled coil domain bearing a phosphoacceptor histidine

HATPase C- catalytic domain which binds and hydrolyzes ATP and transfers a phosphate group to the phosphoacceptor histidine of HisKA

PISA- proteins, interfaces, structures, and assemblies web server

MS/MS- tandem mass spectrometry in which digested peptides are first ionized and analyzed for mass/charge, then dissociated by collision to further fragment and determine mass/charge of fragments

MASCOT- MS/MS peptide identification software and server

HHK- hybrid histidine kinase protein which bears both histidine kinase (HisKA, HATPase C) and receiver (RR) domains

DUF- domain of unknown function

ATP- adenosine triphosphate

P2CS- prokaryotic two component systems database

Pr- protein absorbing in the red spectrum or the dark adapted state of a canonical BphP

Pfr- protein absorbing in the far red spectrum, or the illuminated state of a canonical BphP

N-terminus- the amino terminus of a polypeptide chain

C-terminus- the carboxy- terminus of a polypeptide chain

BLAST- basic local alignment search tool, an algorithm for finding DNA or protein sequence homology

SAD- single wavelength anomalous dispersion, method for finding phases of X-ray intensities

PDB- protein data bank, dataset storage for protein structures. Example: PDB:5IC5 corresponds to the entry for the RtBRR structure.

k_{cat}- rate of turnover of an enzyme, in units of 1/sec

K_m- the concentration of substrate that leads to a reaction velocity of ½ V_{max}

Å- Ångstrom, or 0.1 nm

R_{work}- In crystallography, a residual factor based on the difference between the observed data and the proposed structural model

R_{free}- a residual factor similar to Rwork but based on preselected, unrefined data

kDa- kilodalton, one thousand atomic mass units

^{32}P- radioactive isotope of phosphorus, used in labeling of ATP and protein molecules

NCBI- national center for biotechnology information

IMG- integrated microbial genomes database

BeF_3- beryllium fluoride, a phosphate group mimic

DRaCALA- differential radial capillary action of ligand assay

CD- circular dichroism

Chapter I

Introduction to bacteriophytochrome two component systems

Two component systems: signal transduction systems consisting of a sensory kinase and response regulator

Bacteria rely on signal sensing and signal transduction to adjust to their environments. Bacteria are varied in their abilities to cope with stresses and environments are continually shifting. How can bacterial cells best use their varied molecular skills to survive in rapidly changing surroundings? A major signal transduction mechanism in bacteria is the two component signal transduction system (TCS). These are molecular phosphorelays comprised in the simplest form as a sensor histidine kinase (HK) protein whose autophosphorylation activity is set by the amount of a signal input, and a response regulator (RR) protein that accepts phosphate from the sensor and carries the signal to other systems in the cell that interpret the signal input into useful survival output (Stock, 2000).

A classic example of a TCS from the model organism *Escherichia coli* illustrates the mechanism of signaling. *E coli* is sensitive to changes in osmolarity in its environment by use of the EnvZ-OmpR two component system (Forst, 1989; Mizuno, 1990). When the concentration of solutes outside the cell differs from that inside, the concentration on either side must be equilibrated by osmosis, which can kill cells by rupture in the worst case. The sensor HK EnvZ spans the inner cell membrane and senses changes in membrane surface tension caused by high osmolarity, changes its conformation and becomes active for autophosphorylation in its cytoplasmic histidine kinase domain (Tokishita, 1992; Tomomori,1999; Khorchid, 2005). HK autophosphorylation is accomplished by ATP hydrolysis in the catalytic core of the protein and transfer of a phosphate group to a histidine residue on the same protein (Roberts, 1994). The RR protein OmpR accepts phosphate from the phosphorylated histidine of EnvZ on an aspartate residue, and forms phosphorylation-dependent dimers which bind to promoters upstream of the porin genes *ompC* and *ompF* and tune their expression to match the needs of the cell (Aiba, 1990; Huang, 1997). In high osmotic stress, the EnvZ-OmpR TCS prompts the

cell to express more porins which allow the passage of solutes across the cell membranes and preserves the cell's barrier. High expression of porins can also be turned off by EnvZ-OmpR when osmolarity returns to equilibrium. When surface tension no longer prompts the EnvZ sensory domain to adopt the active conformation, autophosphorylation is halted and phosphate is no longer transferred to OmpR. Now, the kinase domain of EnvZ performs the reverse reaction, actually dephosphorylating OmpR dimers, allowing them to return to the monomeric status and turning off porin expression (Tokishita, 1990; Russo, 1991). Through this system, two proteins (HK and RR) are able to sense environmental change, rapidly adjust the cellular contents to avoid death, and reset the system when the danger has passed. Through reversible post translational modification, in this case histidine and aspartate phosphorylation, proteins can form effective signaling relays.

Beyond sensing osmolarity, TCS are responsible for sensing a variety of inputs and controlling myriad physiological responses across the Bacteria. Additionally, two component systems are present in plants (Hwang, 2002) and fungi (Ota, 1993), although TCS are not the predominant signaling pathways in eukaryotes as in Bacteria. TCS are more prevalent in bacteria from fluctuating environments than in bacteria from constant environments (Galperin, 2005), underlining their importance to environmental response and survival. In bacteria, TCS regulate essential processes such as chemotaxis (Hess, 1988; Bourret, 1990), sporulation (Perego, 1989), and cell differentiation (Quon, 1996). TCS are also crucial to bacterial stress responses (Aguilar, 2001 Danese, 1995). Additionally, certain bacteria utilize TCS to regulate virulence and antibiotic resistance (Gooderham, 2009; Beier, 2006). TCS are extremely common in Bacteria. In fact, of the 6386 bacterial genomes currently profiled (as of May, 2016) in the P2CS server (Barakat, 2011), 97.3% encode at least one HK-RR pair. *Stigmatella aurantiaca* encodes the most TCS proteins, with 192 putative HKs and 143 RRs, although the

median number of TCS genes across all catalogued species is more modest, at 14 HKs and 16 RRs per genome (Figure 1).

The linear, nearly one-to-one relationship between the number of HKs and RRs in bacterial genomes (Figure 1) alludes to another hallmark of TCS, which is specificity between the HK and RR components which are part of the same signaling system (Laub, 2007). Although phosphorelay between HK and RR pairs proceeds by the same histidine-to-aspartate phosphotransfer mechanism in most TCSs, and most bacterial species encode for tens to even hundreds of HK-RR pairs, each pair is able to specifically interact and exchange phosphate groups with great fidelity. How is this accomplished without significant cross-talk occurring between discrete signaling pairs in the cell? One way to avoid unwanted signal crossing is that HKs that can phosphorylate a non-cognate RR also act as phosphatase enzymes on that non cognate RR, effectively reversing the unwanted phosphotransfer and avoiding activation of a pathway by the wrong input signal (Alves, 2003). Another factor that minimizes cross-talk is the low stoichiometry of HKs present in the cell in relation to RRs. In the case of the EnvZ-OmpR TCS there are approximately 35 molecules of RR for each one HK molecule (Cai, 2002), which limits the number of non-specific phosphotransfer events. This is compounded by the observation that cognate RRs out-compete non-cognate RRs for phosphotransfer from HKs, both *in vitro* (Biondi 2006) and *in vivo* (Verhamme, 2002). Overall, the picture of TCS specificity within the cell is that occasional nonspecific phosphotransfer may occur, but the effects of specific phosphotransfer far outweigh any unwanted crosstalk. Distinct from the concept of unwanted cross talk, there are examples of cross regulation between TCS in some bacterial systems (Matubara, 2000; Stewart, 2003). These can be branched signaling systems, in which one HK can phosphorylate more than one RR specifically, or convergent signaling systems, in which more than one HK can specifically phosphorylate a common RR (Laub, 2007). An example of a convergent TCS in a bacteriophytochrome system (Wojnowska, 2013) will be

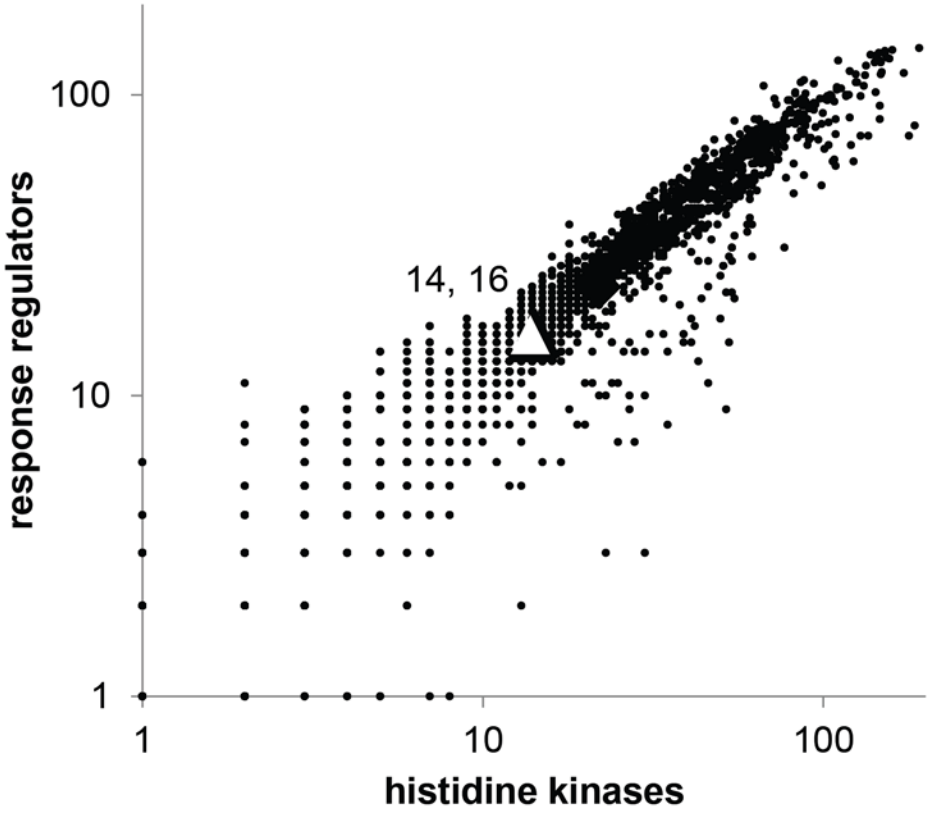


Figure 1. Prevalence of histidine kinase and response regulator genes in bacterial genomes. The number of putative HK and RR genes from 6286 bacterial genomes were taken from the P2CS database (Barakat, 2010) and plotted on a logarithmic scale to illustrate the relationship between TCS signaling partners. The median number of HKs and RRs in bacteria (14 and 16, respectively) is represented as a white triangle.

discussed later in this section, and the concept of branched signaling becomes relevant to this work in chapter 4.

An understanding of the protein domains and their structures underpins an understanding of how these TCS proteins efficiently allow bacteria to adjust to their surroundings (Figure 2). Beginning from the amino terminus, a typical sensor histidine kinase is comprised of a transmembrane sensory domain, which can sense changes in the periplasm or membrane (Mascher, 2006; Cheung, 2010); a cytoplasmic HAMP domain, which has dual functions as a dimerization and signal transducing domain (Hulko, 2006); a cytoplasmic HisKA domain, which is a coiled coil bearing the critical histidine residue that can be reversibly phosphorylated; and a cytoplasmic HATPase c domain, which binds to and catalyzes the hydrolysis of ATP to transfer one phosphate group to the critical histidine (Klumpp, 2002). The positioning of the sensory domain in the periplasm and membrane allows the signal transducing domains to respond to the presence of molecules or other signals from the exterior without direct exposure. In order for this intramolecular signal transduction to be accomplished, there must be conformational changes transmitted along the long axis of the sensor kinase protein, ultimately activating or inactivating catalytic activity of the HisKA domain in response to the specific signal.

The typically membrane-tethered sensor kinase is accompanied by a specific response regulator protein, which is completely cytoplasmic and free to carry signals further into the cell's interior. A typical response regulator is comprised of an N-terminal receiver domain and a C-terminal effector domain, often a helix-turn-helix fold which binds to specific regions of DNA to act as a transcription factor (Galperin, 2010). The receiver domain is comprised of an internal parallel five-stranded β-sheet with hydrophobic character surrounded by five α-helices. A conserved aspartate residue protrudes from the C-terminus of β3 and serves as the phosphoacceptor site (Bourret, 2010).

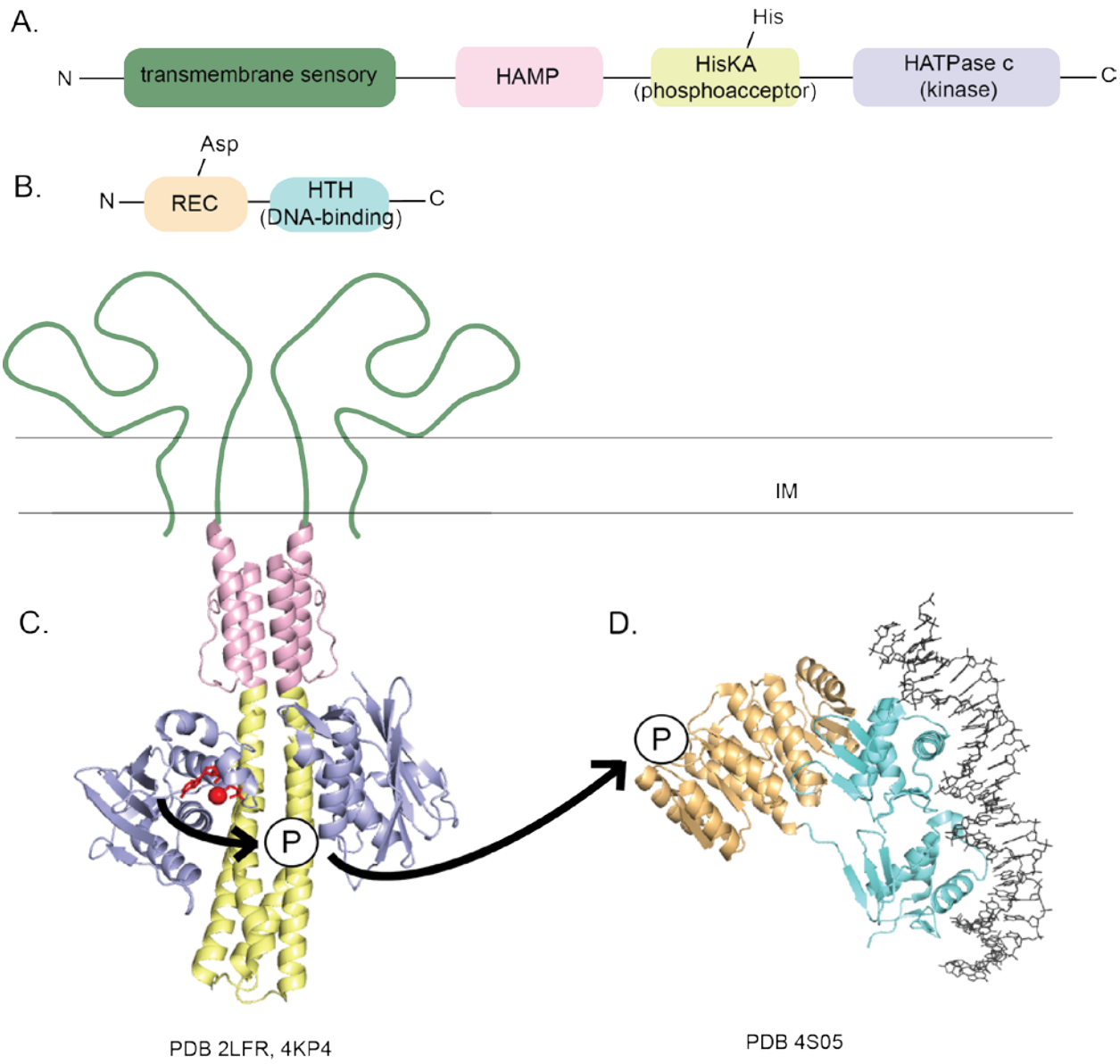


Figure 2. Structural domains of typical TCS proteins. A &C. Domain organization and structural model of a transmembrane histidine kinase. In green: transmembrane sensory domain represents diverse folds that span the inner bacterial cell membrane and sense various signals; pink: HAMP coiled coil domain that mediates dimerization and intramolecular signal transduction; yellow: HisKA coiled coil domain that bears the phosphoaccepting histidine residue; blue: HATPase c catalytic domain that binds and hydrolyzes ATP to release phosphate for transfer to histidine in the HisKA domain. B &D. Domain organization and structural model of a typical response regulator. In orange: receiver domain that bears the phosphoaccepting aspartate residue and often can dimerize in response to phosphorylation; cyan: helix-turn-helix fold that binds to a specific sequence of DNA when dimers are formed. Other terms: N-: amino terminus; –C: carboxy terminus; IM: inner membrane; Circle P: phosphate group.

The TCS signaling partners described in this work differ from the classic models described above (Figure 3). The sensor histidine kinases known as bacteriophytochromes are not inserted in the inner cell membrane but are instead cytoplasmic (Davis, 1999). Due to the visible light input that they sense, which can penetrate through the cell, these HKs do not require access to the periplasm in order to set their autophosphorylation state. Further, most BphP RR proteins are in a class known as single-domain response regulators (SDRRs), which have only a phosphate receiver domain and lack a dedicated output domain (Galperin, 2006). The archetype for SDRRs is the *E. coli* CheY protein, in which the phosphorylation state of the RR regulates a protein-protein interaction that controls chemotaxis motility (Barak, 1992). Unlike the classic RR described above, these SDRRs lack a DNA-binding domain. Outside of CheY and a few other well characterized systems (Hecht, 1995; Tzeng, 1997; Metzger, 2013), the molecular mechanisms of signaling for most SDRRs are unknown.

Bacteriophytochromes: red light sensing histidine kinases

Visible light is an important environmental signal that bacteria have the capacity to sense and adapt to through two component signaling. Members of the phytochrome family of photoreceptor proteins are sensor HKs that absorb and respond to red light to relay intracellular signals. Phytochromes have been extensively studied in terms of their signaling outcomes in plants, in which they regulate developmental processes (Quail, 2002, Hughes, 2013). Within the last 20 years, phytochrome homologs known as bacteriophytochrome proteins (BphPs) were discovered in cyanobacteria (Yeh, 1997) and in the non-photosynthetic bacterium *Deinococcus radiodurans* (Davis, 1999) and since then BphPs have been found through sequence homology in 307 bacterial species or strains (Mandalari, 2013). The retention of BphP coding sequences in the genomes of bacteria from diverse environments suggests: 1) that red light sensing is

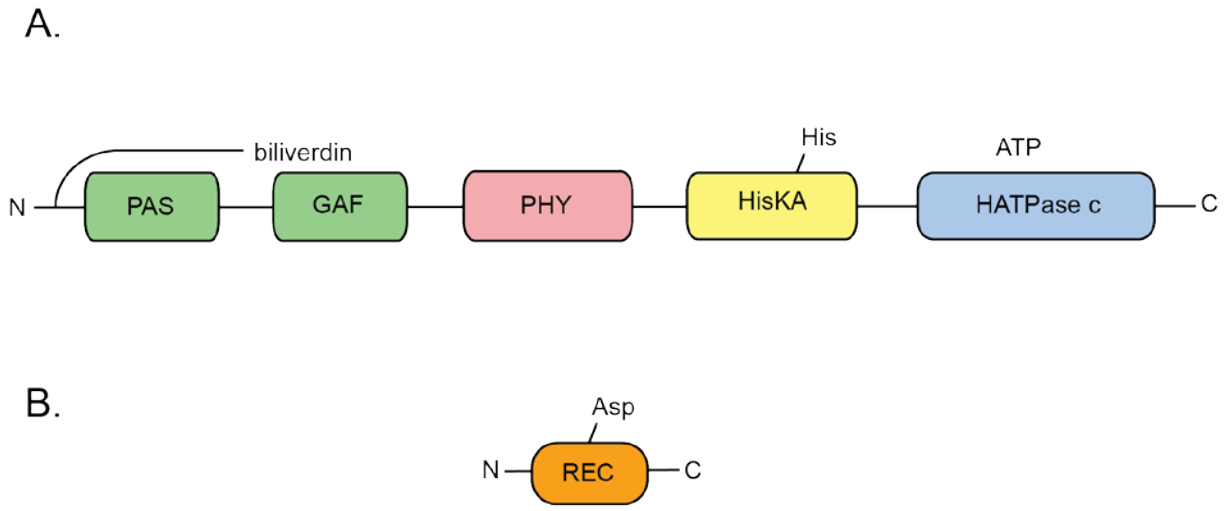


Figure 3. Domains of bacteriophytochrome histidine kinase and response regulator proteins. A. The light sensing domains (in green) include the PAS domain, the site of covalent chromophore binding, and the GAF domain, which surrounds the chromophore. Together these two domains form a light sensing figure of eight knot. The PHY domain (in pink) includes the tongue region that refolds in response to chromophore rearrangement during photoconversion and transduces a structural signal to the HK domains. The HisKA domain (in yellow) is a long central coiled-coil which bears the conserved histidine residue that becomes phosphorylated by the HATPase C domain (in blue) which is the site of ATP binding and catalysis. B. Bacteriophytochrome response regulators are single domain proteins consisting solely of a receiver domain. The BRR bears a conserved aspartate residue, the site of phosphorylation during phosphotransfer.

coupled to a variety of physiological outputs and 2) red light (or other signal) sensing by BphP confers a fitness advantage to those bacterial species which have retained functional BphPs.

BphPs are widespread in both photosynthetic and chemotrophic bacteria, although the studies on the mechanism of signaling and the resultant phenotypes have historically been more complete for photosynthetic organisms (Table 1). One of the best understood system for BphP sensing and signaling is the purple photosynthetic bacterium *Rhodopseudomonas palustris*. In *R. palustris* two BphPs phosphorylate a common response regulator to control expression of an antenna light harvesting complex (Giraud, 2005). This is an example of the convergent cross-regulation mentioned earlier which is possible, even beneficial for regulation, in some TCSs. A similar regulation of light harvesting genes by a BphP was also observed in *Bradyrhizobium* sp. BTAI1 (Jaubert 2008). In the cyanobacterium *Synechocystis* sp. PCC 6803, the BphP known as Cph1 upregulates a number of genes in response to red light, including those controlling chlorophyll biosynthesis and nitrogen metabolism (Hubschmann 2005). In these systems, the control of light harvesting machinery by a light sensing TCS makes a great deal of intuitive sense, but these phenotypes cannot be universally applied to the large number of non-photosynthetic bacteria that encode BphPs.

The phenotypes attributed to BphP TCSs in non-photosynthetic organisms are more varied in type, and the exact molecular details of signaling pathways have been more difficult to pin down. In chemotrophic *D. radiodurans*, the BphP has been shown to regulate production of a species-specific carotenoid, deinoxanthin, which is proposed to protect *D. radiodurans* from visible light damage (Davis, 1999). In *Azospirillum brasilense*, a BphP was shown to regulate a general stress response comprised of a variety of genes but no carotenoid genes (Kumar 2012). In the plant pathogens *Agrobacterium tumefaciens* and *Pseudomonas syringae*, the BphPs have been linked to control of swarming motility that can enhance colonization of host plants (Oberpichler, 2008; Wu, 2013). In the human opportunistic pathogen *Pseudomonas aeruginosa*,

a BphP was shown to regulate elements of the LasR quorum sensing system (Barkovitz, 2011). It seems clear that in non-photosynthetic BphP systems, the phenotypic outputs are more varied than in photosynthetic bacteria. Study of BphP TCSs from chemotrophs has also been complicated by the abundance of SDRRs in these systems, which cannot be easily be traced to regulation of a specific genetic operon as is sometimes the case in the phototrophs. Thus, the complete BphP-mediated signal transduction pathway is not defined in any non-photosynthetic bacterial system. Because the photosynthetic BphP study systems have yielded the most complete understanding of how light sensing is coupled to molecular signaling to result in specific phenotypic responses, this work will focus on the lesser known non-photosynthetic BphP-RR systems.

Table 1. Bacterial systems for BphP studies from the literature. Summary of the energy harvesting properties, number of encoded BphPs, summary of biochemical and phenotypic experimental data

species	**energy source**	**number of BphPs (canonical)**	**biochemically characterized TCS?**	**phenotype**	**reference**
Agrobacterium tumefaciens	chemotroph	2	y	motility/plant pathogenicity	Oberpichler, 2008;
Azospirillum brasilense	chemotroph	2	n	general stress response	Kumar, 2012
Bradyrhizobium sp. BTAI1	facultative phototroph	1	n	light harvesting complex	Jaubert, 2008
Deinococcus radiodurans	chemotroph	1	n	carotenoid dependent stress response	Davis, 1999
Pseudomonas aeruginosa	chemotroph	1	n	quorum sensing	Barkovitz, 2011
Pseudomonas syringae	chemotroph	1-2*	y	motility/plant pathogenicity	Wu, 2013
Rhodopseudomonas palustris	facultative phototroph	2-3*	y	light harvesting complex	Giraud, 2005
Synechocystis sp. PCC 6803	phototroph	1	y	chlorophyll synthesis, stress, nitrogen metabolism	Hubschmann, 2005

Much is known about the structure and photochemistry of BphPs while little is known about their signal transduction pathways and resultant phenotypes in bacteria (Auldridge, 2011). Canonical BphPs have an N-terminal photosensory module comprised of PAS, GAF, and PHY domains and a C-terminal histidine kinase (HK) output module (Figure 3) (Montgomery, 2002). BphPs follow the modular domain architecture of other HK proteins, and thus other output domains besides HK domains are found in BphP proteins, but this work will focus on the canonical BphPs with PAS-GAF-PHY-HK architecture. The first structure of a dimeric BphP from *D. radiodurans* (DrBphP) revealed the position of a covalently bound chromophore, biliverdin (Figure 4), attached to a conserved cysteine in the PAS domain and interacting with multiple amino acids in the GAF domain (Wagner, 2005). BphPs are characterized as having two stable photochromic states as revealed by UV-Vis spectroscopy: absorbing red light (called the Pr state) and absorbing far red light (called Pfr) (Bhoo, 2001). BphPs are prompted to transition between these two states by absorption of photons at the holoprotein's λ_{max} in the Q-band of the electromagnetic spectrum (typically ~700 nm). Transition between Pr and Pfr states is associated with conformational change which regulates HK autophosphorylation, and controls subsequent phosphorylation of a response regulator protein by the histidine kinase domain of the BphP (Giraud, 2005). Structures of BphPs from cyanobacteria revealed that the secondary structure of a region of the PHY domain termed the tongue refolds from β-sheet to α-helix during photoconversion from the dark to illuminated states (Anders, 2013; Anders, 2014). Structures of the PAS-GAF-PHY domains of DrBphP in both the dark and illuminated states determined that tongue refolding prompts flexibility in a hinge region of the PHY domain that allows the PHY and likely the HK domains to reposition relative to one another in order to control autophosphorylation in response to light (Figure 5) (Takala, 2014; Takala, 2015). The chromophore binding domains (PAS, GAF) and mechanism of photoisomerization of the biliverdin chromophore have been extensively studied in part because BphPs are more static and thus crystallize more easily when the dynamic PHY and HK domains are removed. There

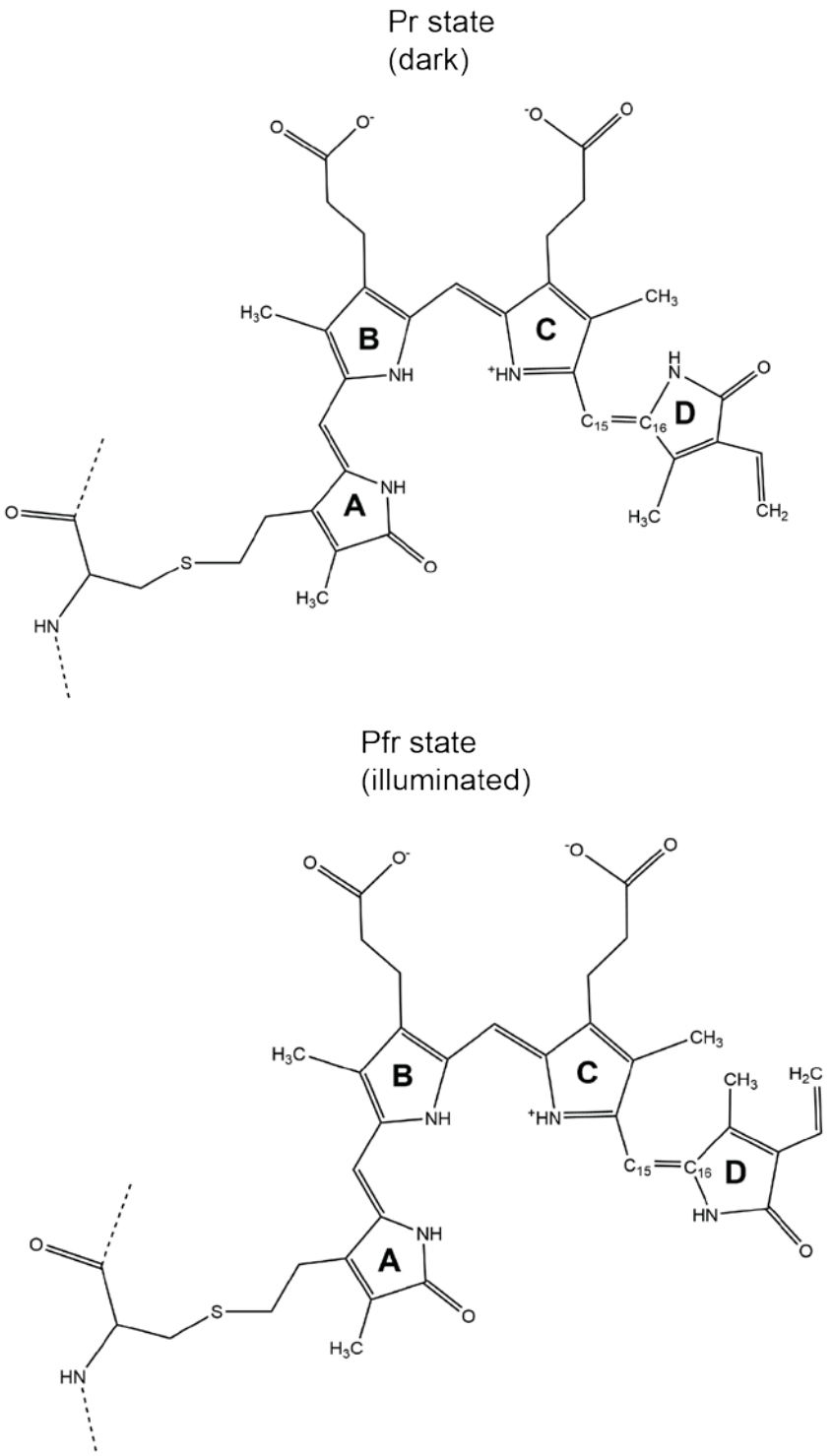


Figure 4. Biliverdin IXα, the light absorbing chromophore bound by BphP proteins. The A-ring of biliverdin is covalently bound to a cysteine residue in the PAS domain of the BphP. The C15=C16 double bond near the D-ring is photoisomerized on absorption of a red photon. Changes in the chromophore conformation are detected by the surrounding GAF domain of the BphP and transmitted throughout the protein to regulate kinase activity.

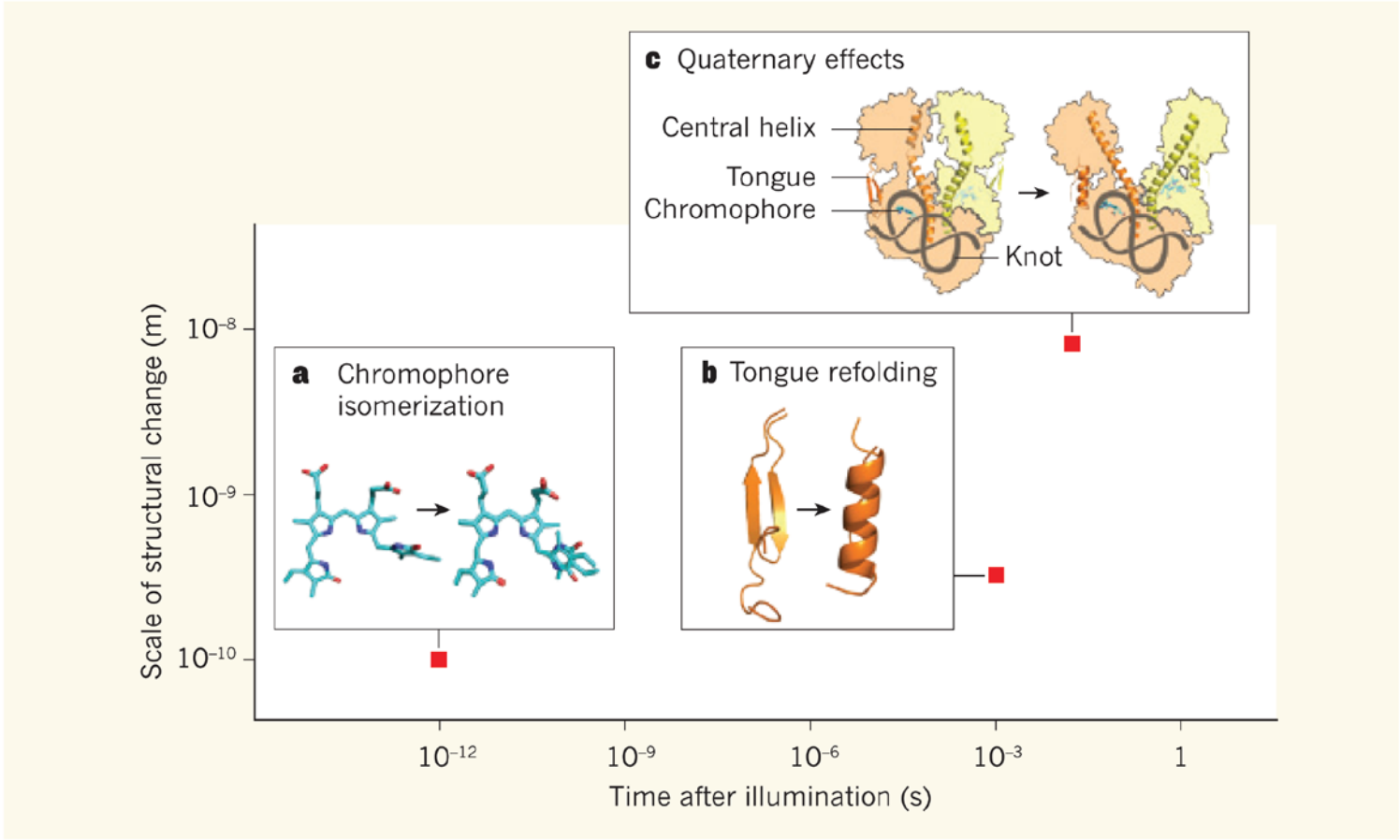


Figure 5. Signal magnification in phytochrome light sensors. A. Light-induced isomerization of the chromophore in phytochromes causes atomic-scale structural changes within picoseconds of illumination. B. These changes lead to rearrangement of a crucial secondary-structure element in the 'tongue' region of the phytochrome: a β-hairpin unravels and an α-helix forms. C. Within 30 milliseconds, the closed quaternary structure adopted in the dark state opens to a Y-shaped conformation in the illuminated state, with effector domains (structure unknown) at the distal tips of the Y. The phytochrome's characteristic figure-of-eight knot may rigidify the chromophore-binding domain, thereby restricting light-driven motion to the hinge in the central helix of each dimer subunit. The two subunits of the phytochrome are shown in orange and yellow; red squares indicate approximate times and length scales of the changes. Taken from Baker, 2014.

has been little research focus on intracellular signaling by BphPs and although some physiological responses controlled by these proteins are known (Yeh, 1997; Davis 1999; Bhoo, 2001; Karniol, 2003; Giraud 2005; Hubschman, 2005; Jaubert, 2008; Barkovits, 2011; Kumar, 2012), the responses are varied (Table 1). Since BphPs were discovered, the number of bacterial genomes available in public databases has increased to thousands (Benson, 2012). This allowed the bioinformatic discovery of BphP-like coding sequences in hundreds of bacteria (Van der Horst, 2007, Mandalari, 2013), many of them non-photosynthetic, and has reignited questions like what are these red light sensors doing in these organisms, and how does sensing of photons ultimately result in environmental adaptation?

Bacteriophytochrome response regulators (BRRs): SDRRs from BphP systems

BRRs are SDRRs, which consist of a receiver domain with no obvious C-terminal regulatory domain such as a DNA-binding motif. Previous to this work, four BRRs were characterized by X-ray crystallography: Rcp1 from *Synechocystis* sp. PCC 6803 (Im, 2002) (PDB: 1I3C); RcpA, and RcpB both from *Calothrix* sp. PCC 7601 (Benda, 2004) (PDB: 1K66, 1K68); and Rpa3017 from *Rhodopseudomonas palustris* (Yang, 2015) (PDB: 4ZYL). All of these structures came from species capable of photosynthesis, and no BRR structures were known from non-photosynthetic bacteria. BRR proteins share a common overall topology with other SDRRs. They consist of an internal parallel five-stranded β-sheet with hydrophobic character surrounded by five α-helices. A conserved aspartate residue protrudes from the C-terminus of β3 and serves as the phosphoacceptor site. Prior to this work, four dimerization modes were known for receiver domains, named for the α-helix and β-strand numbers involved in packing (Figure 6) (Gao, 2010). The Rcp-RRs and Rpa3017 all crystallized previously unknown inverted 4-5-5 homodimers (Figure 6D), a quaternary arrangement that has only been seen in BRRs. Several bacterial response regulators are known to have stoichiometries regulated by the

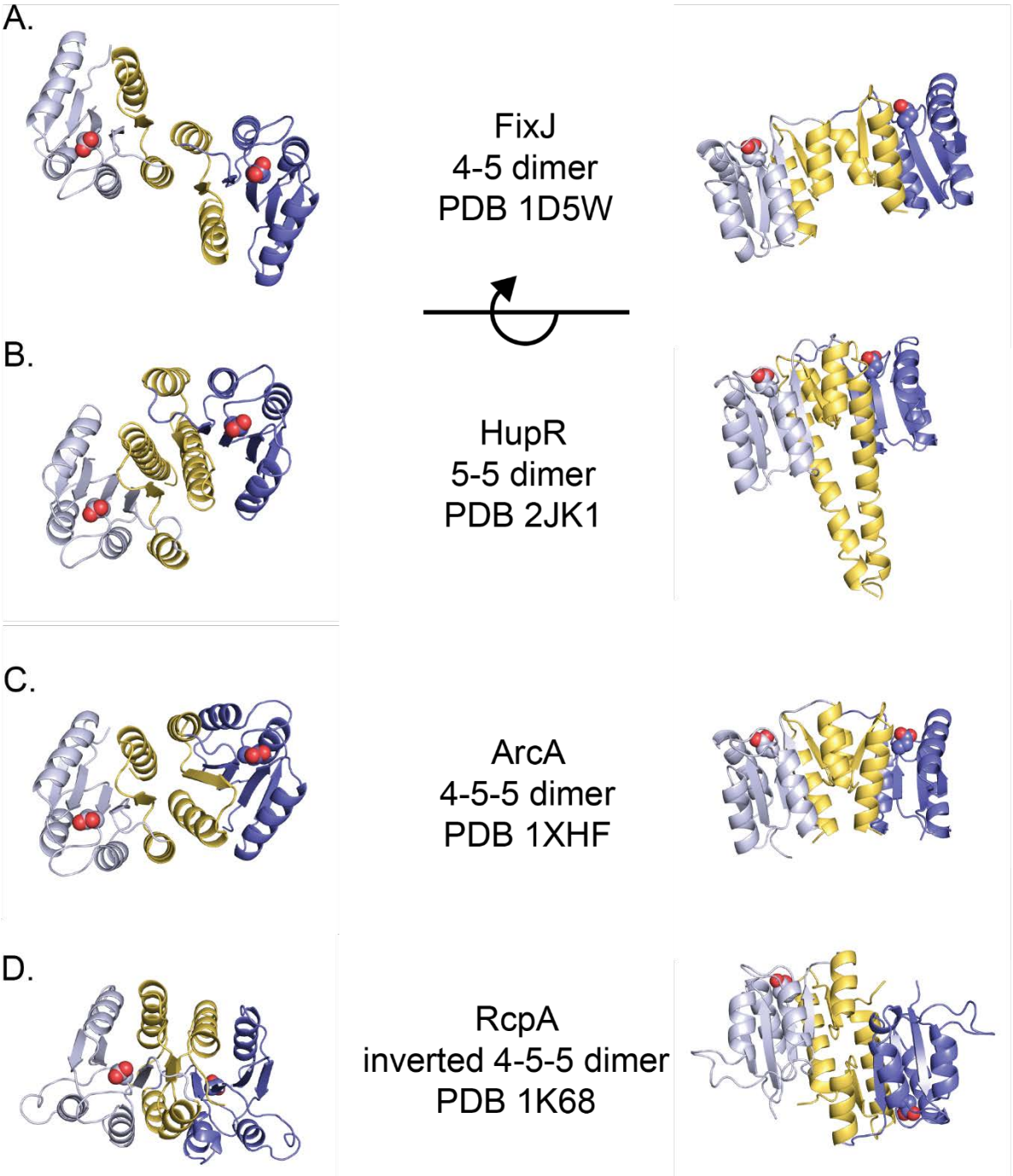
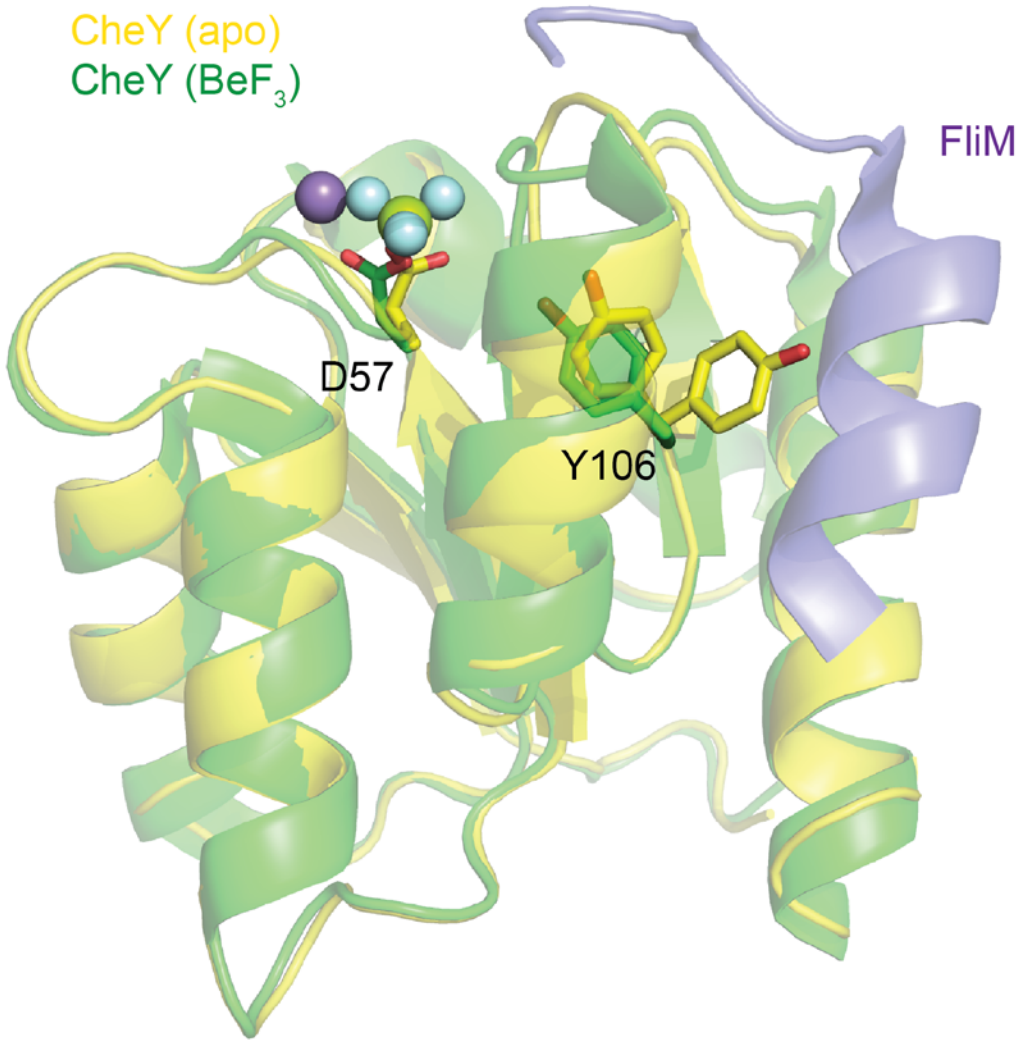


Figure 6. Previously known dimerization modes of response regulator proteins. All types are named for the numbers of the terminal β-strands and α-helices that form the dimer interaction (colored in yellow). The phosphoacceptor aspartate residues are depicted as red and blue spheres to mark the active sites. The 4-5 dimer (A), 5-5 dimer (B) and 4-5-5 dimer (C) structures are all from non-BphP TCSs and have output domains (not shown). These three dimerize head-to-head, with the phosphoacceptor sites facing the same direction. The inverted 4-5-5 dimer structure (D), was observed for four BphP SDRR proteins. This unusual arrangement dimerizes head-to-tail, with the two phosphoacccep-tor sites facing opposite directions. Adapted from Gao, 2010.

phosphorylation state (Fielder, 1995; Maris, 2002; Lewis, 2002), however dimers were observed for RcpA and B irrespective of the phosphorylation state (Benda, 2004). The same work noted conservation of consecutive Phe and Trp residues in a C-terminal extension in phytochrome RRs not found in other SDRRs. These residues fold as an α-helix to form an exposed hydrophobic patch which interacts with the adjacent protomer via an aromatic cluster which is presumably essential to formation of the inverted 4-5-5 dimer.

The best understood SDRR is part of the *E. coli* chemotaxis TCS, comprised of the HK CheA and the RR CheY. Because of the extensive and detailed structural and mechanistic literature on CheY, most SDRRs in bacterial genomes including BRRs are annotated as CheY or CheY-like proteins because they share the SDRR fold (Ashby, 2004). CheY in its unphosphorylated state can adopt two conformations for the residue Tyr 106 on β5, one of which points out toward the solvent and occludes binding to the flagellar switch protein FliM (Figure 7). When CheY is phosphorylated by CheA in response to chemosensing, Tyr 106 adopts the inward conformation and allows stable binding of CheY to FliM (Zhu, 1996; Lee, 2001). This protein-protein interaction changes the direction of the flagellum from clockwise to counter-clockwise and thus alters the direction of cellular movement in response to chemosensing (Barak, 1992). Importantly, most BRR proteins include the conserved “switch” tyrosine residue on β5, or have a phenylalanine in its place. Although the same mechanism of signaling has not yet been shown for a BRR protein in response to the phosphorylation state, the presence of the tyrosine suggests a similar mechanism may occur in BRRs. However, no protein interaction partners besides the BphP HKs are yet known for BRRs from any system.



PDB: 1FQW, 2B1J, 3CHY

Figure 7. Mechanism of structural signaling changes in CheY in response to phosphorylation. Superposition of two CheY crystal structures: the apo, or non-phosphorylated form (in yellow) and the phosphor-mimic bound structure (in green). Phosphorylation of D57 in CheY prompts conformational change that results in a single conformation for Y106, which allows the N-terminus of FliM (in purple) to bind CheY. This protein:protein interaction changes the direction of flagellar motility in *E. coli* cells in response to chemosensing.

Ramlibacter tataouinensis: a non-photosynthetic bacterium with a BphP TCS

To address the knowledge gaps of how red light signals are received and relayed in bacteria by BphPs and what the signaling outputs may be, this work focuses on bacteriophytochromes from non-photosynthetic bacteria, particularly the desert isolate *Ramlibacter tataouinensis*, which encodes a striking abundance of putative photoreceptor proteins (DeLuca, 2011, Van der Horst, 2007, Mandalari, 2013). The chemotrophic β-Proteobacterium *R. tataouinensis* was first isolated from sandy soil in the Sahara Desert and is remarkable for tolerance to desiccation and dual cell morphologies (Heulin, 2003, Gommeaux, 2005). Pure cultures of *R. tataouinensis* contain both rod-shaped cells which are presumed to be motile via Type IV pili (as *R. tataouinensis* lacks flagellar genes) and small (approximately 1 μm in diameter) coccoid cells which are desiccation tolerant but lack motility. The genome sequence and annotation revealed that *R. tataouinensis* encodes four putative blue light photoreceptors and two putative BphP red light photoreceptors (DeLuca, 2011). Given the scarcity of water in desert soil, as well as extremes of temperature and radiation in desert environments a hypothesis is proposed that these sensors allow *R. tataouinensis* to inhabit its harsh environment by cueing physiological adaptation. Some physiological adaptations which are influenced by blue or red photoreceptors in bacteria include phototaxis, general stress response, carotenoid production, DNA repair, and transition between planktonic and a host adapted or multicellular (biofilm) lifestyle (Van der Horst, 2007, Purcell, 2008, Gomelsky, 2011). These, and the interesting desiccation tolerance and cell morphology transitions observed in *R. tataouinensis*, are candidates for physiological responses to be controlled by the *R. tataouinensis* BphPs.

R. tataouinensis encodes two putative bacteriophytochromes (DeLuca, 2011), designated RtBphP1 and RtBphP2 following the established BphP nomenclature. The duplication of BphP genes in bacterial genomes appears to be a common occurrence and

several of the well-studied species with BphPs encode two or more (Van der Horst, 2007, Mandalari, 2013). The RtBphP1 gene is flanked by genes encoding a putative heme oxygenase and a putative single-domain response regulator (Figure 8). Both of these types of proteins would be required for RtBphP1 to function as a light-responsive kinase. Heme oxygenase carries out the cleavage of heme to produce biliverdin which binds autocatalytically via a thioether linkage to a conserved cysteine near the N-terminus of the BphP and allows it to sense light (Bhoo, 2001). In certain bacterial species, there is evidence that heme oxygenase and BphP must form a complex in order to load the chromophore (Shah, 2012). A response regulator is required to accept the phosphate from the HK domain of the BphP and carry the signal to the next node in the signal transduction pathway (Davis, 1999). In bacterial two-component systems, transient contact between a kinase and response regulator is required for specific partner recognition and for phosphotransfer (West, 2001). Interestingly, these three genes in the RtBphP1 gene cassette are in different reading frames and overlap by several base pairs which suggests they are co-regulated and co-transcribed. The gene for RtBphP2 is not in the same operon as RtBphP1 and it is not surrounded by heme oxygenase or response regulator genes. Analysis of gene transcription by reverse-transcriptase PCR indicates that the three overlapping genes are well expressed in both dark and red light culture conditions relative to a RecA control (Figure 8). Compared to the three overlapping genes, lower expression was detected for a hybrid histidine kinase gene following the BRR and for RtBphP2. Bioinformatic analysis of the amino acid sequence of RtBphP1 and RtBphP2 indicates that both have the features typical of BphPs which are known to be required for photosensing and kinase activity. Both have canonical PAS-GAF-PHY-HK domain architecture, both encode a cysteine residue near the N-terminus which is the presumed site of biliverdin ligation, and the primary sequences contain motifs known to be important for chromophore interactions (Karniol, 2005). Based on gene proximity and amino acid sequence homology to BphP and BRR proteins, the gene

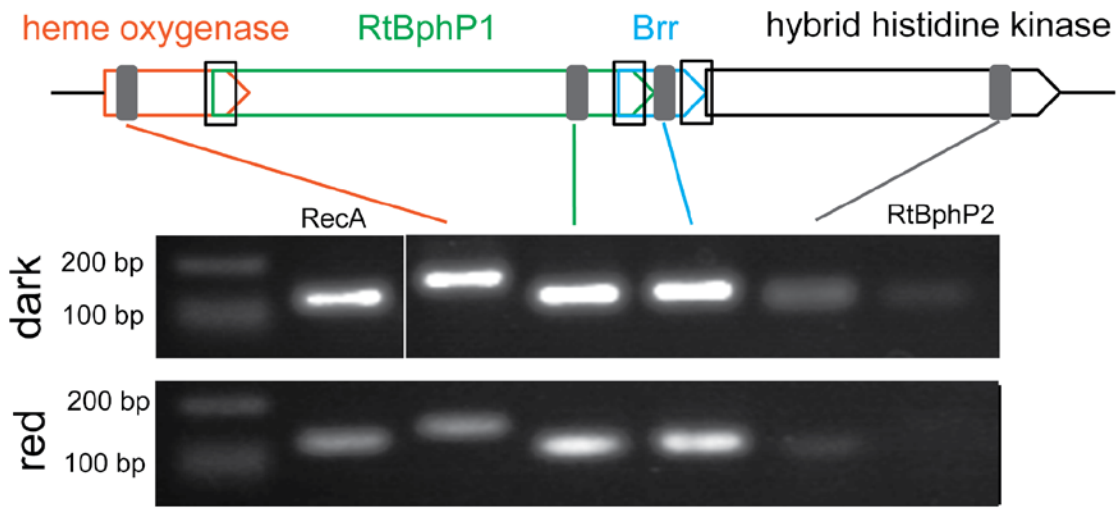


Figure 8. The *Ramlibacter tataouinensis* BphP gene cassette and transcription of genes shown by RT-PCR. The first three genes in the cassette (Rta_25460, heme oxygenase; Rta_25470, RtBphP1; and Rta_25480, RtBRR) are encoded overlapping in three different reading frames. Rta_25490, a putative hybrid histidine kinase follows the BRR gene. Two conserved hypothetical proteins complete the cassette. Rta_28950, RtBphP2, is encoded in a distant genomic region (DeLuca, 2011). *R. tataouinensis* cultures were grown to mid-log phase and then further cultured in darkness or 700 nm light before extraction of total RNA and cDNA synthesis. 100-200 bp amplicons from specific genes were generated from the cDNA templates to detect gene transcription.

products from the RtBphP1 cassette were proposed to form an active red light responsive TCS (Figure 9), a hypothesis which is borne out in this work.

Posttranslational modifications in Bacteria: focus on phosphorylation and acetylation

Reversible post-translational modification (PTM) of proteins (in the form of histidine and aspartate phosphorylation) is at the heart of TCS signaling (Figure 10 A,B). This sensor-regulated addition and removal of phosphate groups from specific amino acids is such an effective mode of signaling that TCS is the predominant form of signal transduction in bacteria. Up until the last couple of decades, histidine and aspartate phosphorylation was considered the extent of PTM in bacterial systems (Bakal, 2000). Although other PTMs such as serine/threonine/tyrosine phosphorylation and acetylation were known (Figure 10 C,D), they were considered strictly relevant in eukaryotic systems. Due to the relatively late arrival of bacterial systems to the wider world of PTMs, researchers have been able to rapidly discover the extent of novel phosphorylation, acetylation, glycosylation, lipidation, nitrosylation and more by taking advantage of detection methodologies pioneered by eukaryotic PTM researchers.

Recent work has revealed that a number of PTMs, including serine/threonine/tyrosine phosphorylation and acetylation, are common in bacterial systems and appear to regulate a variety of cellular processes. Studies of the phosphoproteome of *E. coli* revealed 79 proteins with Ser/Thr/Tyr phosphorylation with a distribution of 68%/23%/9% respectively (Macek, 2008). Similar Ser/Thr/Tyr phosphoproteome results have been reported for *R. palustris*, which revealed 54 phosphorylated proteins including five HK proteins (Hu, 2012). Another study detected 57 phosphorylated proteins in *P. aeruginosa* and 56 phosphorylated proteins in *P. putida*, with a consensus of just nine commonly phosphorylated proteins between the two species (Ravichandran, 2009). Detection of Ser/Thr/Tyr phosphopeptides relies on molecular

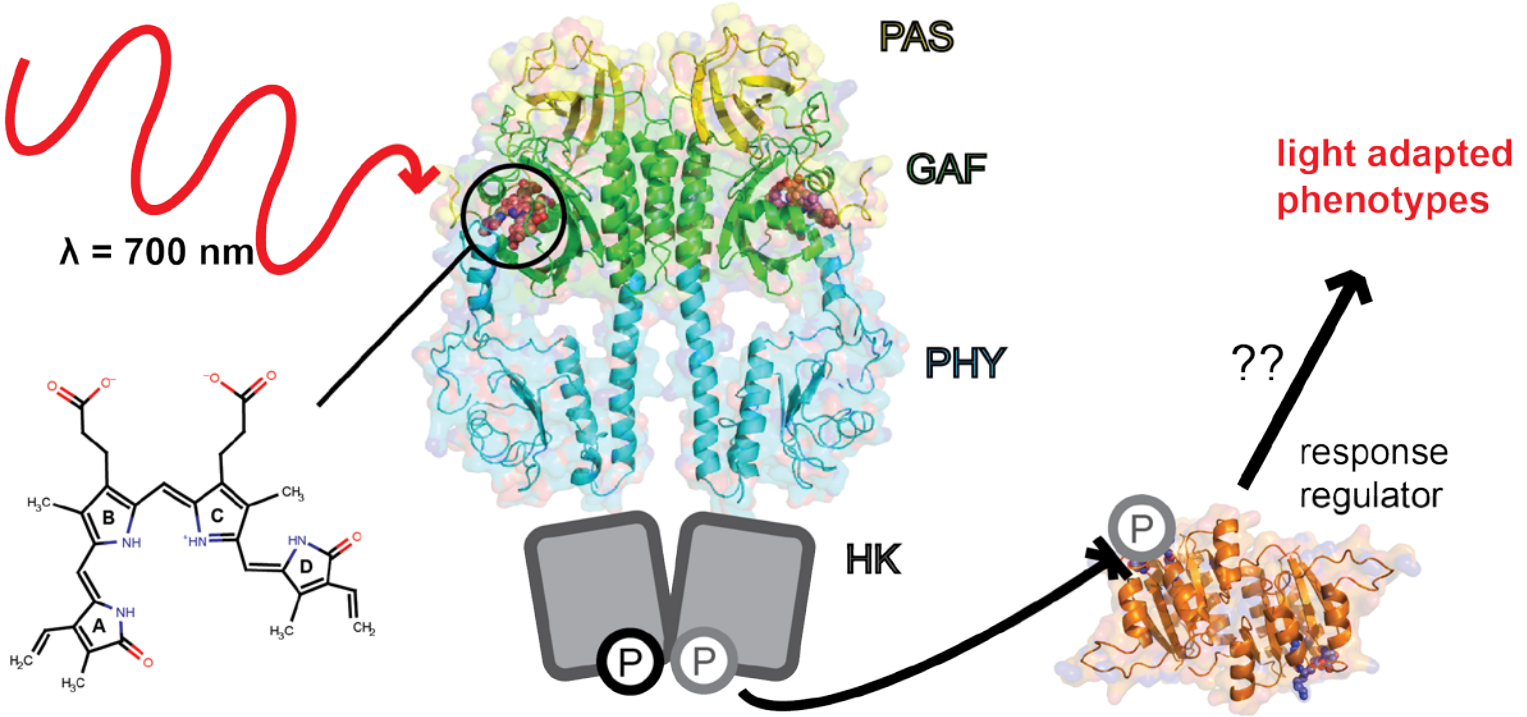
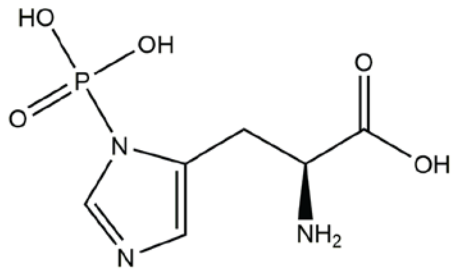
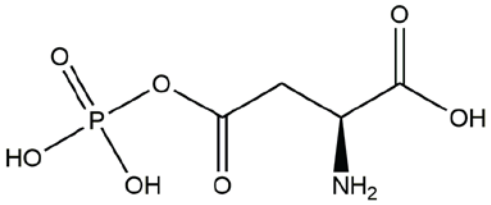


Figure 9. Working model for *R. tataouinensis* red light signal transduction. *R. tataouinensis* senses red light via a biliverdin chromophore covalently bound to the bacteriophytochrome RtBphP1 (homology model based on PaBphP from *Pseudomonas aeruginosa*). RtBphP1 undergoes light-induced conformational change to regulate kinase activity. The histidine kinase (HK) domain of RtBphP1 phosphorylates a response regulator protein (homology model based on Rcp1 from *Synechocystis* PCC6803) which transduces the signal by an unknown pathway. We are currently investigating phenotypes resulting from red light stimuli.

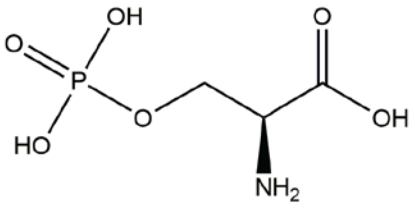
A. phosphohistidine



B. phosphoaspartate



C. phosphoserine



D. acetyllysine

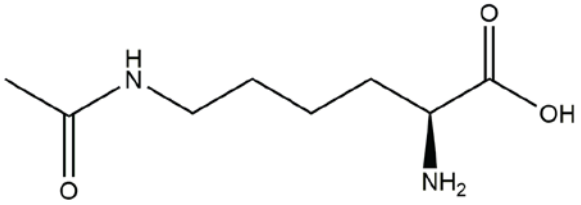


Figure 10. Post translational modifications on amino acids discussed in this work. Formation of phosphohistidine (A) on the HisKA domain during autophosphorylation and transfer of the phosphate group to form phosphoaspartate (B) on the RR are central reactions in two component signaling. Other common post translational modifications on bacterial proteins include phosphorylation of serine (C) and acetylation of lysine (D).

enrichment techniques and sufficient signal to be detected by mass spectrometry, therefore these experiments may yet underreport the abundance of these PTMs in bacterial cells.

Lysine acetylation is also prevalent in bacterial proteomes. In *E. coli* DH5, 91 proteins with lysine acetylation were detected, primarily on metabolic enzymes (Zhang, 2009). An even greater abundance of lysine acetylation, on 195 proteins, was reported for the proteome of *S. enterica* (Wang, 2010). Studies of the *R. palustris* acetylproteome revealed a more modest 24 acetylated proteins, however this more stringent study verified acetylation of target proteins through biochemical assays (Crosby, 2012). Although lysine acetylation appears on many metabolic enzymes, there is evidence that this PTM can impact signal transduction as well. In the *E. coli* chemotaxis system, the SDRR CheY is acetylated on two lysine residues which enhances binding to the FliM flagellar protein and subsequent flagellar motility (Barak, 2001). Thus, CheY undergoes additional PTM to the reversible phosphorylation by the HK protein CheA. Taken together with the HK phosphorylation data from *R. palustris*, here is precedence in the literature for both Ser/Thr/Tyr phosphorylation and lysine acetylation occurring in TCS proteins in bacteria. Later in this work, novel PTMs detected on BphP HK proteins will be discussed, specifically serine phosphorylation and lysine acetylation.

All of the BphP and BRR proteins purified in this study were expressed in *E. coli* BL21 strains, and indeed most if not all characterized BphPs from the literature are also from E. *coli* recombinant systems (Yeh, 1997; Karniol, 2003; Giraud, 2005; Jaubert, 2008; Kumar, 2012). In order to consider novel phosphorylation and acetylation events, it is important to catalog the serine/threonine/tyrosine kinases and lysine acetyltransferase enzymes present in *E. coli* and also consider non-enzymatic modes of protein modification. In bacteria, serine or threonine phosphorylation can be accomplished either by eukaryotic-like serine/threonine kinases (eSTKs), such as the enzymes PknA and PknB from *Mycobacterium tuberculosis*, or by non-eSTKs unique to bacteria such as YihE from *E. coli, and* SpoIIAB, RsbT and RsbW from

Bacillus species (Pereira, 2011). BLAST searches of the BL21(DE3) genome indicate that of these ser/thr kinases, only a homolog of YihE is encoded in the protein expression strain used in this work (Figure 11). YihE is a kinase upregulated in response to activation of the Cpx TCS, which responds to extracytoplasmic stresses on the cell membrane and in the periplasm (Duguay, 2004). Lysine acetylation is known to proceed in bacteria by one well-characterized enzyme, denoted Pat/YfiQ/AcuA (Starai, 2004), though other putative uncharacterized acetyltransferases have been identified (Vetting, 2008). *E. coli* BL21 encodes a Pat lysine acetyltransferase homolog (Figure 12) in addition to an YihE serine/threonine kinase, and therefore can be considered competent to carry out both modifications if these genes are expressed. In addition to enzymatic PTM activity, the ubiquitous small molecule acetyl phosphate is also known to phosphorylate (McCleary, 1994) and acetylate (Weinert, 2013) bacterial proteins *in vivo* and *in vitro*.

Score	Expect	Method	Identities	Positives	Gaps
616 bits(1589)	0.0	Compositional matrix adjust.	296/300(99%)	299/300(99%)	0/300(0%)

```
K12 YihE    MNNSAFTFQTLHPDTIMDALFEQGIRVDSGLTPLNSYENRVYQFQDEDRRRFVVKFYRPE  60
            MNNSAFTFQTLHPDTIMDALFE GIRVDSGLTPLNSYENRVYQFQDEDRRRFVVKFYRPE
BL21 YihE   MNNSAFTFQTLHPDTIMDALFEHGIRVDSGLTPLNSYENRVYQFQDEDRRRFVVKFYRPE  60

K12 YihE    RWTADQILEEHQFALQLVNDEVPVAAPVAFNGQTLLNHQGFYFAVFPSVGGRQFEADNID  120
            RWTADQILEEHQFALQLVNDEVPVAAPVAFNGQTLLNHQGFYFAVFPSVGGRQFEADNID
BL21 YihE   RWTADQILEEHQFALQLVNDEVPVAAPVAFNGQTLLNHQGFYFAVFPSVGGRQFEADNID  120

K12 YihE    QMEAVGRYLGRMHQTGRKQLFIHRPTIGLNEYLIEPRKLFEDATLMPSGLKAAFLKATDE  180
            QMEAVGRYLGRMHQTGRKQLFIHRPTIGLNEYLIEPRKLFEDATL+PSGLKAAFLKATDE
BL21 YihE   QMEAVGRYLGRMHQTGRKQLFIHRPTIGLNEYLIEPRKLFEDATLIPSGLKAAFLKATDE  180

K12 YihE    LIAAVTAHWREDFTVLRLHGDCHAGNILWRDGPMFVDLDDARNGPAIQDLWMLLNGDKAE  240
            LIAAVTAHWREDFTVLRLHGDCHAGNILWRDGPMFVDLDDARNGPA+QDLWMLLNGDKAE
BL21 YihE   LIAAVTAHWREDFTVLRLHGDCHAGNILWRDGPMFVDLDDARNGPAVQDLWMLLNGDKAE  240

K12 YihE    QRMQLETIIEAYEEFSEFDTAEIGLIEPLRAMRLVYYLAWLMRRWADPAFPKNFPWLTGK  300
            QRMQLETIIEAYEEFSEFDTAEIGLIEPLRAMRLVYYLAWLMRRWADPAFPKNFPWLTG+
BL21 YihE   QRMQLETIIEAYEEFSEFDTAEIGLIEPLRAMRLVYYLAWLMRRWADPAFPKNFPWLTGE  300
```

Figure 11. Sequence alignment of the enzymatically characterized ser/thr kinase YihE from *E. coli* K12 with its homolog in BL21(DE3).

```
 Score = 1628 bits (4217),  Expect = 0.0, Method: Compositional matrix adjust.
 Identities = 815/886 (92%), Positives = 851/886 (96%), Gaps = 0/886 (0%)

Query  1    MSQQGLEALLRPKSIAVIGASMKPHRAGYLMMRNLLAGGFNGPVLPVTPAWKAVLGVMAWPDIASLPFTPDLAILCTNASRNLALLDALGAKGCKTCIILSAPTSQHEELLACARHYKMR  120
            MSQ+GLEALLRPKSIAVIGASMKP+RAGYLMMRNLLAGGFNGPVLPVTPAWKAVLGV+AWPDIASLPFTPDLA+LCTNASRNLAL++ LG KGCKTCIILSAP SQHE+L ACA  + MR
Sbjct  1    MSQRGLEALLRPKSIAVIGASMKPNRAGYLMMRNLLAGGFNGPVLPVTPAWKAVLGVLAWPDIASLPFTPDLAVLCTNASRNLALMEELGEKGCKTCIILSAPASQHEDLRACALRHNMR  120

Query  121  LLGPNSLGLLAPWQGLNASFSPVPIKQGKLAFISQSAAVSNTILDWAQQREMGFSYFIALGDSLDIDVDELLDYLARDSKTSAILLYLEQLsdarrfvsaarsasrNKPILVIKSGRSPA  240
            LLGPNSLGLLAPWQGLNASFSPVPIK+GKLAFISQSAAVSNTILDWAQQREMGFSYFIALGDSLDIDVDELLDYLARDSKTSAILLYLEQLSDARRFVSAARSASRNKPILVIKSGRSPA
Sbjct  121  LLGPNSLGLLAPWQGLNASFSPVPIKRGKLAFISQSAAVSNTILDWAQQREMGFSYFIALGDSLDIDVDELLDYLARDSKTSAILLYLEQLSDARRFVSAARSASRNKPILVIKSGRSPA  240

Query  241  AQRLLNTSAGMDPAWDAAIQRAGLLRVQDTHELFSAVETLSHMRPLRGDRLMIISNGAAPAALALDELWSRNGKLATLSEETCLQLRQALPAHIDIANPLDLCDDASSEHYVKTLDILLA  360
            AQRLLNT+AGMDPAWDAAIQRAGLLRVQDTHELFSAVETLSHMRPLRGDRLMIISNGAAPAALALD LWSRNGKLAT SEETC +LR ALP H+ ++NPLDL DDASSEHYVKTLDILL
Sbjct  241  AQRLLNTTAGMDPAWDAAIQRAGLLRVQDTHELFSAVETLSHMRPLRGDRLMIISNGAAPAALALDALWSRNGKLATPSEETCQKLRDALPEHVAVSNPLDLRDDASSEHYVKTLDILLH  360

Query  361  SQDFDALMVIHSPSAAAPGTESAHALIETIKRHPRGKFVTLLTNWCGEFSSQEARRLFSEAGLPTYRTPEGTITAFMHMVEYRRNQKQLRETPALPSNLTSNTAEAHNLLQRAIAEGAAS  480
            SQDFDALMVIHSPSAAAP TESA  LIE +K HPR K+V+LLTNWCGE SSQEARRLFSEAGLPTYRTPEGTITAFMHMVEYRRNQKQLRETPALPSNLTSNTAEAH LLQ+AIAEGA S
Sbjct  361  SQDFDALMVIHSPSAAAPATESAQVLIEAVKHHPRSKYVSLLTNWCGEHSSQEARRLFSEAGLPTYRTPEGTITAFMHMVEYRRNQKQLRETPALPSNLTSNTAEAHLLLQQAIAEGATS  480

Query  481  LDTHEVQPILHAYGLHTLPTWIASDSAEAVHIAEQIGYPVALKLRSPDIPHKSEVQGVMLYLRTASEVQQAANAIFDRVKMAWPQARIHGLLVQSMANRAGAQELRVVVEHDPVFGPLIM  600
            LDTHEVQPIL AYG++TLPTWIASDS EAVHIAEQIGYPVALKLRSPDIPHKSEVQGVMLYLRTA+EVQQAANAIFDRVKM WPQAR+HGLLVQSMANRAGAQELRVVVEHDPVFGPLIM
Sbjct  481  LDTHEVQPILQAYGMNTLPTWIASDSTEAVHIAEQIGYPVALKLRSPDIPHKSEVQGVMLYLRTANEVQQAANAIFDRVKMTWPQARVHGLLVQSMANRAGAQELRVVVEHDPVFGPLIM  600

Query  601  LGEGGVEWRPEEQAVVALPPLNMNLARYLVIQGIKQRKIRARSALRPLDIVGLSQLLVQVSNLIVDCPEIQRLDIHPLLASASEFTALDVTLDIAPFDGDNESRLAVRPYPHQLEEWVEM  720
            LGEGGVEWRPE+QAVVALPPLNMNLARYLVIQGIK +KIRARSALRPLD+ GLSQLLVQVSNLIVDCPEIQRLDIHPLLAS SEFTALDVTLDIAPF+GDNESRLAVRPYPHQLEEWVE+
Sbjct  601  LGEGGVEWRPEDQAVVALPPLNMNLARYLVIQGIKSKKIRARSALRPLDVAGLSQLLVQVSNLIVDCPEIQRLDIHPLLASGSEFTALDVTLDIAPFEGDNESRLAVRPYPHQLEEWVEL  720

Query  721  KNGDRCLFRPILPEDEPQLRQFIAQVTKEDLYYRYFSEINEFTHEDLANMTQIDYDREMAFVAVRRMDNAEEILGVTRAISDPDNVDAEFAVLVRSDLKGLGLGRRLMEKLIAYTRDHGL  840
            KNG+RCLFRPILPEDEPQL+QFI++VTKEDLYYRYFSEINEFTHEDLANMTQIDYDREMAFVAVRR+D  EEILGVTRAISDPDN+DAEFAVLVRSDLKGLGLGRRLMEKLI YTRDHGL
Sbjct  721  KNGERCLFRPILPEDEPQLQQFISRVTKEDLYYRYFSEINEFTHEDLANMTQIDYDREMAFVAVRRIDQTEEILGVTRAISDPDNIDAEFAVLVRSDLKGLGLGRRLMEKLITYTRDHGL  840

Query  841  KRLNGITMPNNRGMVALARKLGFQVDIQLEEGIVGLTLNLAKCDES  886
            +RLNGITMPNNRGMVALARKLGF VDIQLEEGIVGLTLNLA+ +ES
Sbjct  841  QRLNGITMPNNRGMVALARKLGFNVDIQLEEGIVGLTLNLAQREES  886
```

Figure 12. Sequence alignment of the enzymatically characterized lysine acetyltransferase Pat from *S. enterica* and its homolog in BL21(DE3).

Biochemical, structural, and bioinformatic elucidation of the BphP-BRR TCS from non-photosynthetic bacteria

The aims throughout this project have been to better understand the *molecular mechanism* of BphP-BRR signal transduction and the *outcomes of the signaling* in non-photosynthetic bacteria. This work focuses on the BphP TCS from non-photosynthetic bacteria, particularly *R. tataouinensis*, and establishes groundwork to make this a model system for BphP-BRR study. In chapter two, a BphP from *R. tataouinensis* is biochemically characterized and the structure of its cognate BRR is described. A novel mode of dimerization for BRRs from *R. tataouinensis* and *A. tumefaciens* is described and shown to have consequences to signal transduction between the BphP and the BRR proteins. In chapter three, unexpected post-translational modifications on BphPs from several species are discovered and a model for PTM effect on enzyme activities is proposed. In chapter four, bioinformatic methods are combined to predict potential protein-protein interaction partners for the novel BRR dimers described in chapter two. These *in silico* predictions will guide future *in vivo* and *in vitro* work.

References

1. **Aguilar PS**, **Hernandex-Arriaga AM**, **Cybulski LE**, **Erazo AC**, **de Mendoza D**. 2001. Molecular basis of thermosensing: a two-component signal trasduction thermometer in Bacillus subtilis. Embo J **20**:1681–1691.

2. **Alves R**, **Savageau MA**. 2003. Comparative analysis of prototype two-component systems with either bifunctional or monofunctional sensors: Differences in molecular structure and physiological function. Mol Microbiol **48**:25–51.

3. **Anders K**, **Daminelli-Widany G**, **Mroginski MA**, **Von Stetten D**, **Essen LO**. 2013. Structure of the cyanobacterial phytochrome 2 photosensor implies a tryptophan switch for phytochrome signaling. J Biol Chem **288**:35714–35725.

4. **Anders K**, **Gutt A**, **Gartner W**, **Essen LO**. 2014. Phototransformation of the red light sensor cyanobacterial phytochrome 2 from synechocystis species depends on its tongue motifs. J Biol Chem **289**:25590–25600.

5. **Ashby MK**. 2004. Survey of the number of two-component response regulator genes in the complete and annotated genome sequences of prokaryotes. FEMS Microbiol Lett **231**:277–281.

6. **Auldridge ME**, **Forest KT**. 2011. Bacterial phytochromes: more than meets the light. Crit Rev Biochem Mol Biol **46**:67–88.

7. **Bakal CJ**, **Davies JE**. 2000. No longer an exclusive club: Eukaryotic signalling domains in bacteria. Trends Cell Biol **10**:32–38.

8. **Baker AW**, **Forest KT**. 2014. Action at a distance in a light receptor. Nature **509**:174–175.

9. **Barak R**, **Eisenbach M**. 2001. Acetylation of the response regulator, CheY, is involved in bacterial chemotaxis. Mol Microbiol **40**:731–743.

10. **Barak R**, **Eisenbach M**. 1992. Correlation between Phosphorylation of the Chemotaxis Protein CheY and Its Activity at the Flagellar Motor ? Biochemistry **31**:1821–1826.

11. **Barakat M**, **Ortet P**, **Whitworth DE**. 2011. P2CS: A database of prokaryotic two-component systems. Nucleic Acids Res **39**:771–776.

12. **Barkovits K**, **Schubert B**, **Heine S**, **Scheer M**, **Frankenberg-Dinkel N**. 2011. Function of the bacteriophytochrome BphP in the RpoS/Las quorum-sensing network of Pseudomonas aeruginosa. Microbiology **157**:1651–1664.

13. **Beier D**, **Gross R**. 2006. Regulation of bacterial virulence by two-component systems. Curr Opin Microbiol **9**:143–152.

14. **Benda C**, **Scheufler C**, **Tandeau de Marsac N**, **Gärtner W**. 2004. Crystal structures of two cyanobacterial response regulators in apo- and phosphorylated form reveal a novel dimerization motif of phytochrome-associated response regulators. Biophys J **87**:476–487.

15. **Benson DA**, **Cavanaugh M**, **Clark K**, **Karsch-Mizrachi I**, **Lipman DJ**, **Ostell J**, **Sayers EW**. 2013. GenBank. Nucleic Acids Res **41**:1–7.

16. **Bhoo SH**, **Davis SJ**, **Walker J**, **Karniol B**, **Vierstra RD**. 2001. Bacteriophytochromes are photochromic histidine kinases using a biliverdin chromophore. Nature **414**:776–779.

17. **Biondi EG**, **Reisinger SJ**, **Skerker JM**, **Arif M**, **Perchuk BS**, **Ryan KR**, **Laub MT**. 2006. Regulation of the bacterial cell cycle by an integrated genetic circuit. Nature **444**:899–904.

18. **Bourret RB**, **Hess JF**, **Simon MI**. 1990. Conserved aspartate residues and phosphorylation in signal transduction by the chemotaxis protein CheY. Proc Natl Acad Sci U S A **87**:41–45.

19. **Bourret RB**. 2010. Receiver domain structure and function in response regulator proteins. Curr Opin Microbiol **13**:142–149.

20. **Cheung J**, **Hendrickson WA**. 2010. Sensor domains of two-component regulatory systems. Curr Opin Microbiol **13**:116–123.

21. **Crosby HA**, **Pelletier DA**, **Hurst GB**, **Escalante-Semerena JC**. 2012. System-wide studies of N-lysine acetylation in Rhodopseudomonas palustris reveal substrate specificity of protein acetyltransferases. J Biol Chem **287**:15590–15601.

22. **Danese PN**, **Snyder WB**, **Cosma CL**, **Davis LJB**, **Silhavy TJ**. 1995. The Cpx two-component signal transduction pathway of Escherichia coli regulates degP transcription. Genes Dev **9**:387–398.

23. **Davis SJ**. 1999. Bacteriophytochromes: Phytochrome-Like Photoreceptors from Nonphotosynthetic Eubacteria. Science (80-) **286**:2517–2520.

24. **de Luca G**, **Barakat M**, **Ortet P**, **Fochesato S**, **Jourlin-Castelli C**, **Ansaldi M**, **Py B**, **Fichant G**, **Coutinho PM**, **Voulhoux R**, **Bastien O**, **Maréchal E**, **Henrissat B**, **Quentin Y**, **Noirot P**, **Filloux A**, **Méjean V**, **DuBow MS**, **Barras F**, **Barbe V**, **Weissenbach J**, **Mihalcescu I**, **Verméglio A**, **Achouak W**, **Heulin T**. 2011. The cyst-dividing bacterium ramlibacter tataouinensis TTB310 genome reveals a well-stocked toolbox for adaptation to a desert environment. PLoS One **6**:e23784.

25. **Duguay AR**, **Silhavy TJ**. 2004. Quality control in the bacterial periplasm. Biochim Biophys Acta - Mol Cell Res **1694**:121–134.

26. **Fiedler U**, **Weiss V**. 1995. A common switch in activation of the response regulators NtrC and PhoB: phosphorylation induces dimerization of the receiver modules. EMBO J **14**:3696–3705.

27. **Forst S**, **Delgado J**, **Rampersaud A**, **Inouye M**. 1990. In vivo phosphorylation of OmpR, the transcription activator of the ompF and ompC genes in Escherichia coli. J Bacteriol **172**:3473–3477.

28. **Forst S**, **Delgado J**, **Inouye M**. 1989. Phosphorylation of OmpR by the osmosensor EnvZ modulates expression of the ompF and ompC genes in *Escherichia coli*. Proc Natl Acad Sci U S A **86**:6052–6056.

29. **Galperin MY**. 2005. A census of membrane-bound and intracellular signal transduction proteins in bacteria: bacterial IQ, extroverts and introverts. BMC Microbiol **5**:35.

30. **Galperin MY**. 2010. Diversity of structure and function of response regulator output domains. Curr Opin Microbiol **13**:150–159.

31. **Galperin MY**. 2006. Structural classification of bacterial response regulators: Diversity of output domains and domain combinations. J Bacteriol **188**:4169–4182.

32. **Gao R**, **Stock AM**. 2010. Molecular strategies for phosphorylation-mediated regulation of response regulator activity. Curr Opin Microbiol **13**:160–167.

33. **Giraud E**, **Zappa S**, **Vuillet L**, **Adriano JM**, **Hannibal L**, **Fardoux J**, **Berthomieu C**, **Bouyer P**, **Pignol D**, **Vermeglio A**. 2005. A new type of bacteriophytochrome acts in tandem with a classical bacteriophytochrome to control the antennae synthesis in Rhodopseudomonas palustris. J Biol Chem **280**:32389–32397.

34. **Gomelsky M**, **Hoff WD**. 2011. Light helps bacteria make important lifestyle decisions. Trends Microbiol **19**:441–448.

35. **Gommeaux M**, **Barakat M**, **Lesourd M**, **Thiéry J**, **Heulin T**. 2005. A morphological transition in the pleomorphic bacterium Ramlibacter tataouinensis TTB310. Res Microbiol **156**:1026–1030.

36. **Gooderham WJ**, **Hancock REW**. 2009. Regulation of virulence and antibiotic resistance by two-component regulatory systems in Pseudomonas aeruginosa. FEMS Microbiol Rev **33**:279–294.

37. **Hecht GB**, **Lane T**, **Ohta N**, **Sommer JM**, **Newton A**. 1995. An essential single domain response regulator required for normal cell division and differentiation in Caulobacter crescentus. EMBO J **14**:3915–3924.

38. **Hess JF**, **Bourret RB**, **Simon MI**. 1988. Histidine phosphorylation and phosphoryl group transfer in bacterial chemotaxis. Nature **336**:139–143.

39. **Heulin T**, **Barakat M**, **Christen R**, **Lesourd M**, **Sutra L**, **De Luca G**, **Achouak W**. 2003. Ramlibacter tataouinensis gen. nov., sp. nov., and Ramlibacter henchirensis sp. nov., cyst-producing bacteria isolated from subdesert soil in Tunisia. Int J Syst Evol Microbiol **53**:589–594.

40. **Hu CW**, **Lin MH**, **Huang HC**, **Ku WC**, **Yi TH**, **Tsai CF**, **Chen YJ**, **Sugiyama N**, **Ishihama Y**, **Juan HF**, **Wu SH**. 2012. Phosphoproteomic analysis of rhodopseudomonas palustris reveals the role of pyruvate phosphate dikinase phosphorylation in lipid production. J Proteome Res **11**:5362–5375.

41. **Huang KJ**, **Lan CY**, **Igo MM**. 1997. Phosphorylation stimulates the cooperative DNA-binding properties of the transcription factor OmpR. Proc Natl Acad Sci U S A **94**:2828–2832.

42. **Hubschmann T**, **Jorissen HJMM**, **Borner T**, **Gartner W**, **De Marsac NT**. 2001. Phosphorylation of proteins in the light-dependent signalling pathway of a filamentous cyanobacterium. Eur J Biochem **268**:3383–3389.

43. **Hubschmann T**, **Yamamoto H**, **Gieler T**, **Murata N**, **Borner T**. 2005. Red and far-red light alter the transcript profile in the cyanobacterium Synechocystis sp. PCC 6803: Impact of cyanobacterial phytochromes. FEBS Lett **579**:1613–1618.

44. **Hughes J**. 2013. Phytochrome cytoplasmic signaling. Annu Rev Plant Biol **64**:377–402.

45. **Hulko M**, **Berndt F**, **Gruber M**, **Linder JU**, **Truffault V**, **Schultz A**, **Martin J**, **Schultz JE**, **Lupas AN**, **Coles M**. 2006. The HAMP Domain Structure Implies Helix Rotation in Transmembrane Signaling. Cell **126**:929–940.

46. **Hwang I**, **Chen HH**, **Sheen J**. 2002. Two-component signal transduction pathways in Arabidopsis. Plant Physiol **129**:500–515.

47. **Im YJ**, **Rho SH**, **Park CM**, **Yang SS**, **Kang JG**, **Lee JY**, **Song PS**, **Eom SH**. 2002. Crystal structure of a cyanobacterial phytochrome response regulator. Protein Sci **11**:614–624.

48. **Jaubert M**, **Lavergne J**, **Fardoux J**, **Hannibal L**, **Vuillet L**, **Adriano JM**, **Bouyer P**, **Pignol D**, **Giraud E**, **Vermeglio A**. 2007. A singular bacteriophytochrome acquired by lateral gene transfer. J Biol Chem **282**:7320–7328.

49. **Jaubert M**, **Vuillet L**, **Hannibal L**, **Adriano JM**, **Fardoux J**, **Bouyer P**, **Bonaldi K**, **Fleischman D**, **Giraud E**, **Vermeglio A**. 2008. Control of peripheral light-harvesting complex synthesis by a bacteriophytochrome in the aerobic photosynthetic bacterium Bradyrhizobium strain BTAi1. J Bacteriol **190**:5824–5831.

50. **Karniol B**, **Vierstra RD**. 2003. The pair of bacteriophytochromes from *Agrobacterium tumefaciens* are histidine kinases with opposing photobiological properties. Proc Natl Acad Sci USA **100**:2807–2812.

51. **Khorchid A**, **Inouye M**, **Ikura M**. 2005. Structural characterization of Escherichia coli sensor histidine kinase EnvZ: the periplasmic C-terminal core domain is critical for homodimerization. Biochem J **385**:255–264.

52. **Klumpp S**, **Krieglstein J**. 2002. Phosphorylation and dephosphorylation of histidine residues in proteins. Eur J Biochem **269**:1067–1071.

53. **Kojima A**. 1992. Transmembrane Signal Transduction and Osmoregulation in Escherichia coli : Functional Importance of the Transmembrane Regions of Membrane-Located Protein protein kinase which modulates expression of the ompF and signal transduction in Escherichia that thi. J Biochem **111**:707–713.

54. **Kumar S**, **Kateriya S**, **Singh VS**, **Tanwar M**, **Agarwal S**, **Singh H**, **Khurana JP**, **Amla DV**, **Tripathi AK**. 2012. Bacteriophytochrome controls carotenoid-independent response to photodynamic stress in a non-photosynthetic rhizobacterium, Azospirillum brasilense Sp7. Sci Rep **2**:872.

55. **Laub MT**, **Goulian M**. 2007. Specificity in two-component signal transduction pathways. Annu Rev Genet **41**:121–145.

56. **Laub MT**, **Biondi EG**, **Skerker JM**. 2007. Phosphotransfer Profiling: Systematic Mapping of Two-Component Signal Transduction Pathways and Phosphorelays. Methods Enzymol **423**:531–548.

57. **Lee SY**, **Cho HS**, **Pelton JG**, **Yan D**, **Berry EA**, **Wemmer DE**. 2001. Crystal structure of activated CheY: Comparison with other activated receiver domains. J Biol Chem **276**:16425–16431.

58. **Lewis RJ**, **Scott DJ**, **Brannigan J a**, **Ladds JC**, **Cervin M a**, **Spiegelman GB**, **Hoggett JG**, **Barák I**, **Wilkinson AJ**. 2002. Dimer formation and transcription activation in the sporulation response regulator Spo0A. J Mol Biol **316**:235–245.

59. **Macek B**, **Gnad F**, **Soufi B**, **Kumar C**, **Olsen J V**, **Mijakovic I**, **Mann M**. 2008. Phosphoproteome Analysis of E . coli Reveals Evolutionary Conservation of Bacterial Ser / Thr / Tyr Phosphorylation. Mol Cell Proteomics **7**:299–307.

60. **Macek B**, **Mijakovic I**, **Olsen J V**, **Gnad F**, **Kumar C**, **Jensen PR**, **Mann M**. 2007. The serine/threonine/tyrosine phosphoproteome of the model bacterium Bacillus subtilis. Mol Cell Proteomics **6**:697–707.

61. **Mandalari C**, **Losi A**, **Gärtner W**. 2013. Distance-tree analysis, distribution and co-presence of bilin- and flavin-binding prokaryotic photoreceptors for visible light. Photochem Photobiol Sci **12**:1144–57.

62. **Maris AE**, **Sawaya MR**, **Kaczor-Grzeskowiak M**, **Jarvis MR**, **Bearson SMD**, **Kopka ML**, **Schröder I**, **Gunsalus RP**, **Dickerson RE**. 2002. Dimerization allows DNA target site recognition by the NarL response regulator. Nat Struct Biol **9**:771–778.

63. **Mascher T**, **Helmann JD**, **Unden G**. 2006. Stimulus perception in bacterial signal-transducing histidine kinases. Microbiol Mol Biol Rev **70**:910–938.

64. **Matsubara M**, **Kitaoka SI**, **Takeda SI**, **Mizuno T**. 2000. Tuning of the porin expression under anaerobic growth conditions by his-to-Asp cross-phosphorelay through both the EnvZ-osmosensor and ArcB-anaerosensor in Escherichia coli. Genes Cells **5**:555–569.

65. **McCleary WR**, **Stock JB**. 1994. Acetyl phosphate and the activation of two-component response regulators. J Biol Chem **269**:31567–31572.

66. **Metzger LC**, **Francez-Charlot A**, **Vorholt JA**. 2013. Single-domain response regulator involved in the general stress response of Methylobacterium extorquens. Microbiol (United Kingdom) **159**:1067–1076.

67. **Mizuno T**, **Mizushima S**. 1990. Signal transduction and gene regulation through the phosphorylation of two regulatory components: the molecular basis for the osmotic regulation of the porin genes. Mol Microbiol **4**:1077–1082.

68. **Montgomery BL**, **Lagarias JC**. 2002. Phytochrome ancestry: Sensors of bilins and light. Trends Plant Sci **7**:357–366.

69. **Oberpichler I**, **Rosen R**, **Rasouly A**, **Vugman M**, **Ron EZ**, **Lamparter T**. 2008. Light affects motility and infectivity of Agrobacterium tumefaciens. Environ Microbiol **10**:2020–2029.

70. **Ota IM**, **Varshavsky A**. 1993. A yeast protein similar to bacterial two-component regulators. Science (80-) **262**:566–9.

71. **Perego M**, **Cole SP**, **Burbulys D**, **Trach K**, **Hoch JA**. 1989. Characterization of the gene for a protein kinase which phosphorylates the sporulation-regulatory proteins Spo0A and Spo0F of Bacillus subtilis. J Bacteriol **171**:6187–6196.

72. **Pereira SFF**, **Goss L**, **Dworkin J**. 2011. Eukaryote-like serine/threonine kinases and phosphatases in bacteria. Microbiol Mol Biol Rev **75**:192–212.

73. **Purcell EB**, **Crosson S**. 2008. Photoregulation in prokaryotes. Curr Opin Microbiol **11**:168–178.

74. **Quail PH**. 2002. Phytochrome Photosensory Signalling Networks. Nat Rev Mol Cell Biol **3**:85–93.

75. **Quon KC**, **Marczynski GT**, **Shapiro L**. 1996. Cell cycle control by an essential bacterial two-component signal transduction protein. Cell **84**:83–93.

76. **Ravichandran A**, **Sugiyama N**, **Tomita M**, **Swarup S**, **Ishihama Y**. 2009. Ser/Thr/Tyr phosphoproteome analysis of pathogenic and non-pathogenic Pseudomonas species. Proteomics **9**:2764–2775.

77. **Roberts DL**, **Bennett DW**, **Forst SA**. 1994. Identification of the site of phosphorylation on the osmosensor, EnvZ, of Escherichia coli. J Biol Chem **269**:8728–8733.

78. **Russo FD**, **Silhavy TJ**. 1991. EnvZ controls the concentration of phosphorylated OmpR to mediate osmoregulation of the porin genes. J Mol Biol **222**:567–580.

79. **Shah R**, **Schwach J**, **Frankenberg-Dinkel N**, **Gärtner W**. 2012. Complex formation between heme oxygenase and phytochrome during biosynthesis in Pseudomonas syringae pv. tomato. Photochem Photobiol Sci **11**:1026–31.

80. **Starai VJ**, **Escalante-Semerena JC**. 2004. Identification of the protein acetyltransferase (Pat) enzyme that acetylates acetyl-CoA synthetase in Salmonella enterica. J Mol Biol **340**:1005–1012.

81. **Stewart V**, **Stewart V**. 2003. sensors NarX and NarQ of proteobacteria. Society **31**:1–10.

82. **Stock AM**, **Robinson VL**, **Goudreau PN**. 2000. Two-Component Signal Transduction. Reactions **69**:183–215.

83. **Takala H**, **Björling A**, **Berntsson O**, **Lehtivuori H**, **Niebling S**, **Hoernke M**, **Kosheleva I**, **Henning R**, **Menzel A**, **Ihalainen J a**, **Westenhoff S**. 2014. Signal amplification and transduction in phytochrome photosensors. Nature **509**:245–8.

84. **Takala H**, **Bjorling A**, **Linna M**, **Westenhoff S**, **Ihalainen JA**. 2015. Light-induced changes in the dimerization interface of bacteriophytochromes. J Biol Chem **290**:16383–16392.

85. **Tokishita S-I**, **Yamada H**, **Aiba H**, **Mizuno2 T**. 1990. Transmembrane Signal Transduction and Osmoregulation in Escherichia coli: II. The Osmotic Sensor, EnvZ, Located in the Isolated Cytoplasmic Membrane Displays Its Phosphorylation and Dephosphorylation Abilities as to the Activator Protein, OmpR1. J Biochem **108**:488–493.

86. **Tomomori C**, **Tanaka T**, **Dutta R**, **Park H**, **Saha SK**, **Zhu Y**, **Ishima R**, **Liu D**, **Tong KI**, **Kurokawa H**, **Qian H**, **Inouye M**, **Ikura M**. 1999. Solution structure of the homodimeric core domain of Escherichia coli histidine kinase EnvZ. Nat Struct Biol **6**:729–34.

87. **Tzeng YL**, **Hoch J a**. 1997. Molecular recognition in signal transduction: the interaction surfaces of the Spo0F response regulator with its cognate phosphorelay proteins revealed by alanine scanning mutagenesis. J Mol Biol **272**:200–212.

88. **van der Horst MA**, **Key J**, **Hellingwerf KJ**. 2007. Photosensing in chemotrophic, non-phototrophic bacteria: let there be light sensing too. Trends Microbiol **15**:554–562.

89. **Verhamme DT**, **Arents JC**, **Postma PW**, **Crielaard W**, **Hellingwerf KJ**. 2002. Investigation of *in vivo* cross-talk between key two-component systems of *Escherichia coli*. Microbiology **148**:69–78.

90. **Vetting MW**, **Errey JC**, **Blanchard JS**. 2008. Rv0802c from Mycobacterium tuberculosis: The first structure of a succinyltransferase with the GNAT fold. Acta Crystallogr Sect F Struct Biol Cryst Commun **64**:978–985.

91. **Wagner JR**, **Brunzelle JS**, **Forest KT**, **Vierstra RD**. 2005. A light-sensing knot revealed by the structure of the chromophore-binding domain of phytochrome. Nature **438**:325–331.

92. **Wang Q**, **Zhang Y**, **Yang C**, **Xiong H**, **Lin Y**, **Yao J**, **Li H**, **Xie L**, **Zhao W**, **Yao Y**, **Ning Z-B**, **Zeng R**, **Xiong Y**, **Guan K-L**, **Zhao S**, **Zhao G-P**. 2010. Acetylation of Metabolic Enzymes Coordinates Carbon Source Utilization and Metabolic Flux. Science (80-) **327**:1004–1007.

93. **Weinert B**, **Iesmantavicius V**, **Wagner S**, **Schölz C**, **Gummesson B**, **Beli P**, **Nyström T**, **Choudhary C**. 2013. Acetyl-Phosphate is a critical determinant of Lysine Acetylation in E.coli. Mol Cell **51**:265–272.

94. **West AH**, **Stock AM**. 2001. Histidine kinases and response regulator proteins in two-component signaling systems. Trends Biochem Sci **26**:369–376.

95. **Wu L**, **McGrane RS**, **Beattie GA**. 2013. Light regulation of swarming motility in Pseudomonas syringae integrates signaling pathways mediated by a bacteriophytochrome and a LOV protein. MBio **4**:e00334–13.

96. **Yang X**, **Zeng X**, **Moffat K**, **Yang X**. 2015. Structure of the response regulator RPA3017 involved in red-light signaling in Rhodopseudomonas palustris. Acta Crystallogr Sect Struct Biol Commun **71**:1215–1222.

97. **Yeh K-C**, **Wu S-H**, **Murphy JT**, **Lagarias JC**. 1997. A Cyanobacterial Phytochrome Two-Component Light Sensory System. Science (80-) **277**:1505–1508.

98. **Zhang J**, **Sprung R**, **Pei J**, **Tan X**, **Kim S**, **Zhu H**, **Liu C-F**, **Grishin N V.**, **Zhao Y**. 2008. Lysine Acetylation Is a Highly Abundant and Evolutionarily Conserved Modification in Escherichia Coli. Mol Cell Proteomics **8**:215–225.

99. **Zhang J**, **Sprung R**, **Pei J**, **Tan X**, **Kim S**, **Zhu H**, **Liu C-F**, **Grishin N V.**, **Zhao Y**. 2008. Lysine Acetylation Is a Highly Abundant and Evolutionarily Conserved Modification in Escherichia Coli. Mol Cell Proteomics **8**:215–225.

100. **Zhu X**, **Amsler CD**, **Volz K**, **Matsumura P**. 1996. Tyrosine 106 of CheY plays an important role in chemotaxis signal transduction in Escherichia coli. J Bacteriol **178**:4208–4215.

Chapter II

Biochemical and structural characterization of bacteriophytochromes and their response regulators from non-photosynthetic bacteria

originally published as

Arm-in-arm response regulator dimers promote intermolecular signal transduction

J. Bacteriol. April 2016 vol. 198 no. 8 1218-1229

Co-authors: Kenneth A. Satyshur, Neydis Moreno Morales, Katrina T. Forest

Abstract

Bacteriophytochrome photosensors (BphPs) and their cognate response regulators make up two-component signal transduction systems which direct bacteria to mount phenotypic responses to changes in environmental light quality. Most of these systems utilize single-domain response regulators to transduce signals through unknown pathways and mechanisms. Here we describe the photocycle and autophosphorylation kinetics of RtBphP1, a red light regulated histidine kinase from the desert bacterium *Ramlibacter tataouinensis*. RtBphP1 undergoes red to far-red photoconversion with rapid thermal reversion to the dark state. RtBphP1 is autophosphorylated in the dark; this activity is inhibited under red light. The RtBphP1 cognate response regulator, RtBRR, and a homolog, AtBRR from *Agrobacterium tumefaciens,* crystallize unexpectedly as arm-in-arm dimers, reliant on a conserved hydrophobic motif hFWAhL. RtBRR and AtBRR dimerize distinctly from four structurally characterized phytochrome response regulators found in photosynthetic organisms, and from all other receiver domain homodimers in the Protein Data Bank. A unique cacodylate-zinc-histidine tag metal organic framework yielded SAD phases and may be of general interest. Examination of the effect of BRR stoichiometry on signal transduction found that phosphorylated RtBRR is accumulated more efficiently than the engineered monomeric $RtBRR_{mon}$ in phosphotransfer reactions. Thus, we conclude that arm-in-arm dimers are a relevant signaling intermediate in this class of two component regulatory systems.

Importance

BphP histidine kinases and their cognate response regulators comprise widespread red light sensing two-component systems. Much work on BphPs has focused on structural understanding of light sensing, and on enhancing the natural infrared fluorescence of these proteins, rather than on signal transduction or resultant phenotypes. To begin to address this

knowledge gap we solved the crystal structures of two single domain response regulators encoded immediately downstream of BphPs. We observed a previously unknown “arm-in-arm” dimer linkage. Monomerization via deletion of the C-terminal dimerization motif had an inhibitory effect on net response regulator phosphorylation, underlining the importance of these unusual dimers for signal transduction.

Introduction

Bacteria utilize two-component sensory and signaling systems (TCS) to monitor and respond to diverse signals in the environment, including the spectrum and intensity of visible light. Light of a specific wavelength may control directed responses, as in the case of phototaxis (1, 2), or can control generalized stress responses (3). Bacteriophytochromes (BphPs) are soluble cytoplasmic red light sensing modules, often histidine kinases (HKs) encoded adjacent to a single domain response regulator (SDRR) to which phosphate is transferred in a light-regulated fashion (4–6). While the mechanism of red light reception by the sensory domains of BphPs is the topic of much active research, and recent work has added to our understanding of the mode of intramolecular signal transduction by the sensory domains of phytochrome (7–9), the effects of light on phosphate flux through entire pathways are less well-studied. Few phenotypic responses have been conclusively attributed to BphP TCSs, and those that are known suggest the physiological responses controlled by these photoreceptors are as diverse as the environments bacteria inhabit (3, 10–14). Certainly critical to connecting red light sensing to appropriate cellular responses are the bacteriophytochrome RR proteins (BRRs). Thus, we characterized two BRRs structurally and placed them in context of the biochemical activity of a cognate BphP from the *Ramlibacter tataouinensis* red light sensing TCS. This TCS was revealed by annotation of the genome sequence of this chemotrophic desert microbe, which encodes a high number of putative light sensing proteins (15).

Typical BphPs and cyanobacterial phytochromes (Cphs) are dimeric proteins capable of maximal absorbance of red light in the dark state (Pr) which convert to a far-red light absorbing state (Pfr) after exposure to red light. Canonical domain architecture is PAS-GAF-PHY-HK (16). The PAS domain is the site of covalent chromophore attachment in BphPs, which utilize biliverdin IXα [BV], whereas in Cphs phycocyanobilin is covalently linked to the GAF domain (16). In all cases, the GAF domain amino acids surround the chromophore. Signal transduction

is initiated by a light-driven isomerization of a double bond in this tetrapyrrole (17), and the PHY domain transduces conformational changes to the HK domain. The HK catalyzes ATP hydrolysis and transfers the Ɣ-phosphate to a conserved histidine residue; the transfer is assumed based on sequence comparisons to other trans-acting HKs to be to the sister protomer in the dimer (18, 19). The kinetics of autophosphorylation of a Cph and subsequent phosphotransfer to its cognate RR have been measured for one example, but have not been reported in detail for non-photosynthetic systems. Psakis, et al. (2011) reported K_m and k_{cat} values for the Cph1/Rcp1 TCS from *Synechocystis* sp. PCC 6803 and demonstrated that efficiency of both Cph1 autophosphorylation and phosphotransfer to the RR were four-fold higher in the Pr state (20). Other groups' results support the finding that kinase activity is higher in the Pr state than in the Pfr state for BphPs as well. Giraud, *et al*. (2005) reported that autophosphorylation by *Rhodopseudomonas palustris* RpBphP2 was 83% reduced in the Pfr state (6). Phosphotransfer to the cognate RR Rpa3017 was also more efficient in the Pr state. Karniol, *et al*. (2003) reported that Agp1 from *Agrobacterium tumefaciens* had two-fold greater autophosphorylation activity in the Pr state, and 10-fold greater phosphotransfer to the response regulator AtRR1 (here called AtBRR) in the Pr state (5). The lack of kinetic parameters for BphPs prompted us to characterize the autophosphorylation activities of RtBphP1 as the earliest light regulated step in this TCS signaling cascade.

Another hallmark of BV-binding BphPs is their reverse photoconversion from the Pfr state to the Pr state upon exposure to far red light (16). Many are stable in both states, and can be switched to the alternative ground state with the appropriate light stimulus. BphPs can also thermally revert to the Pr state in the absence of light, with a reversion half-life that is characteristic of the particular phytochrome and can be temperature dependent. An example of a poorly stable Pfr form is Agp1, which is relatively unstable at room temperature (21), and

displays accelerated thermal reversion at 30° C (22). Here we describe reversion kinetics for RtBphP1 and show that RtBphP1 and Agp1 share rapid thermal reversion.

Most putative BRRs are annotated as CheY-like SDRRs, which consist of a receiver domain with no obvious regulatory domain such as a DNA-binding motif. For cyanobacterial Cph systems, three such RRs have been characterized by X-ray crystallography: Rcp1 from *Synechocystis* sp. PCC 6803 (23) (PDB: 1I3C), RcpA, and RcpB both from *Calothrix* sp. PCC 7601 (24) (PDB: 1K66, 1K68). This cyanobacterial set has recently been augmented by the structure of Rpa3017 (25). These proteins share a common overall topology with other SDRRs. They consist of an internal parallel five-stranded β-sheet with hydrophobic character surrounded by five α-helices. A conserved aspartate residue protrudes from the terminus of β3 and serves as the phosphoacceptor site. Receiver domains can dimerize by one of four known structural arrangements, named for the α-helix and β-strand numbers involved in packing (26). The Rcp-RRs and Rpa3017 all crystallized as unusual inverted 4-5-5 homodimers, and Benda, *et al* observed dimers for RcpA and B irrespective of the phosphorylation state (24). The same work noted conservation of consecutive Phe and Trp residues in a C-terminal extension in phytochrome RRs not found in other SDRRs. These residues fold as an α-helix to form a solvent-exposed hydrophobic patch which interacts with the sister protomer via an aromatic cluster. Here we report the crystal structure of dimeric BRRs from two chemotrophic bacterial species which share the conserved FW key residues, yet form homodimers through a distinct crossover linkage which we propose represents a novel category of RRs. These and arm-in-arm RR dimers may be common to a set of BphP TCSs from non-photosynthetic bacteria.

We investigate the effects of these arm-in-arm dimers on signal transduction, and underline their importance to intermolecular signal transduction in BphP-BRR TCS.

Results

Photoproperties of RtBphP1

In order to begin characterization of the red light-sensing TCS from *R. tataouinensis*, we cloned, expressed, and purified full-length RtBphP1. Size-exclusion chromatography separated RtBphP1 molecules into four distinct size populations corresponding to 771, 300, 189, and 108 kDa (Fig. 1A). The predominant 189 kDa dimer fraction had a molar proportion of BV to BphP of 0.78 as determined by the relative absorbance and extinction coefficients of RtBphP1 (ε = 98,937 at 708 nm) and free BV (ε = 39,900 at 391 nm). This fraction was used to carry out all dark reversion, autophosphorylation and phosphotransfer experiments.

Zinc-binding fluorescence assays (data not shown) demonstrated that RtBphP1 covalently binds the added BV chromophore. UV-Vis spectroscopy verified that RtBphP1 is a functional red light-sensing BphP with a Pr dark ground state (λ_{max} = 708 nm) (Fig. 1B). Exposure of fully dark-adapted RtBphP1 to 700 nm red light for 1 minute induces a relatively unstable Pfr state (λ_{max} ~750 nm), which thermally reverts to the Pr state (Fig. 1B, 1C) with biphasic behavior and half-lives of 0.7 and 20.4 minutes (Fig. 1D).

RtBphP1 is a light-regulated autokinase

RtBphP1 acts as a light-regulated autokinase as evidenced by radiolabeled phosphorylation reactions and their associated Michaelis-Menten kinetic constants (Fig. 2). The protein is autophosphorylated in the dark, and this activity is modestly suppressed by 700 nm light. The initial rate of autophosphorylation is slow, and is retarded by two-fold under red light (dark $k_{cat} = 2.0 \pm 0.1 \times 10^{-5}$ s^{-1} versus red $k_{cat} = 0.9 \pm 0.3 \times 10^{-5}$ s^{-1}). Red light acts as a

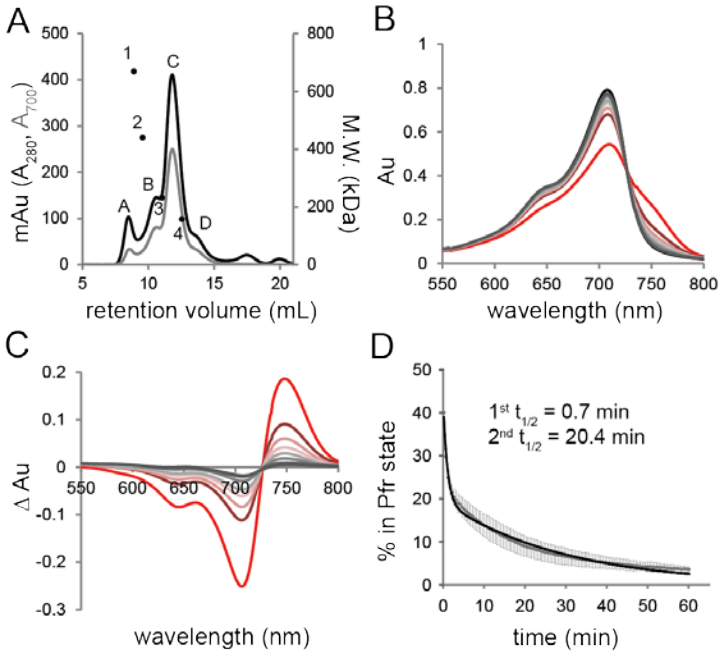


Figure 1. Photoproperties of RtBphP1. A. Gel filtration of RtBphP1. Peaks with both 280 nm and 700 nm absorbance correspond to RtBphP1 oligomers. Peak C is the dimer fraction used in all experiments. Numbered circles represent sizing standards: (1) thyroglobulin (669 kDa); (2) ferritin (440 kDa); (3) catalase (232 kDa), and (4) aldolase (158 kDa). B. Visible spectra of RtBphP1 in the fully dark-adapted, Pr state (black line), after 1 minute illumination with 700 nm light (red line) and 5-60 minute interval recovery products (gradient of red to dark grey). C. Difference spectra (illuminated minus dark state absorbance) of the data from A. D. Time course and biexponential modeling of Pfr thermal decay at 24° C. The mean data and derived half-lives from three independent experiments are shown.

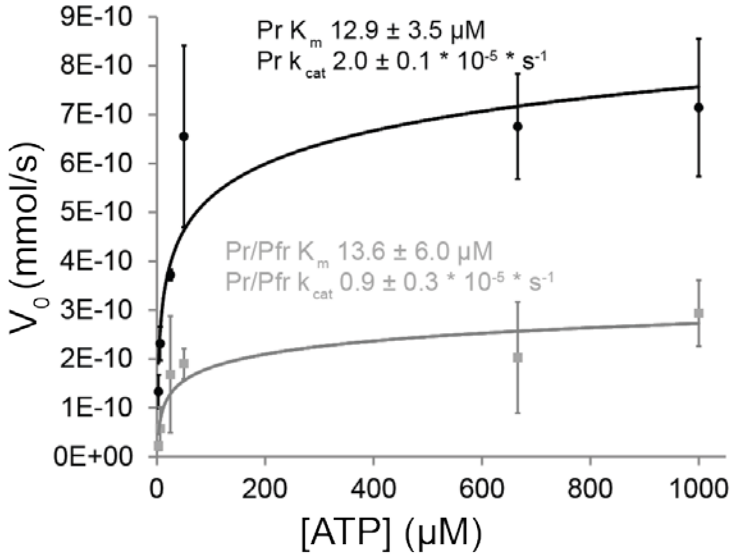


Figure 2. Red-light regulated autophosphorylation of RtBphP1. Michaelis-Menten plot of five independent experiments. Initial rates of formation (mean ± S.D.) of ^{32}P-labeled RtBphP1 vs. ATP concentration are compared for activity in the dark state (black) or in the red illuminated state (grey).

noncompetitive inhibitor of autokinase activity as indicated by unchanged K_m values for ATP (for Pr and Pr/Pfr K_m = 12.9 ± 3.5 μM and 13.6 ± 6.0 μM, respectively) despite modulated k_{cat}.

AtBRR and RtBRR$_{HIS}$ are novel arm-in-arm dimers

In order to probe intermolecular signal transduction by BphPs, we solved the crystal structure of RtBRR, the putative cognate RR of RtBphP1, in the non-phosphorylated state. Crystal growth and phasing were made possible by the inclusion of a C-terminal 14-residue tag ending in hexahistidine, which is well-ordered and coordinated by zinc and cacodylate ($(CH_3)_2AsO_2H$) (Fig. 3A). In fact, the zinc and As atoms provided strong anomalous signal, and we were able to phase the RtBRR$_{HIS}$ structure with a Zn^{2+}/As single-wavelength anomalous dispersion data set. This technique should be of general use to solve protein structures that have a histidine tag by crystallizing the cacodylate/zinc/His complex, since the number of structures in PDB with His tags is rapidly increasing (42). This strategy is similar to a new general method that has been proposed for coating the surface of a protein with zinc atoms to solve the phase problem (43). AtBRR phases were obtained by molecular replacement with RtBRR$_{HIS}$ as the search model (Fig. 3B). RtBRR$_{HIS}$ was refined to a resolution of 1.9 Å with R_{work}/R_{free} values of 18% and 23%. AtBRR was refined to a resolution of 1.9 Å with R_{work}/R_{free} values of 20% and 24% (Table 1).

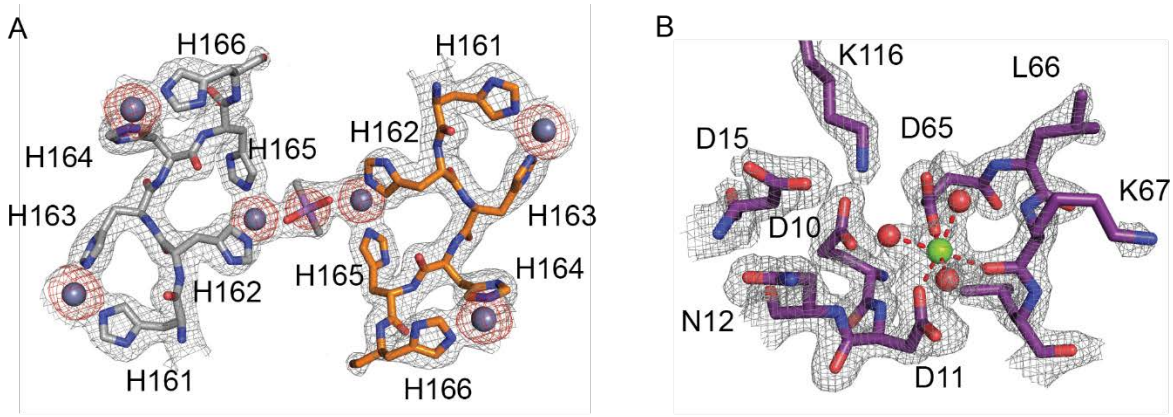


Figure 3. Electron density maps and models for $RtBRR_{HIS}$ and AtBRR. A. Ordered hexahistidine tags from two $RtBRR_{HIS}$ molecules (one with C grey, one with C orange, both with O red, N blue) across the crystallographic twofold axis are coordinated by cacodylate (As purple) and zinc ions (slate spheres). Grey mesh: Fo electron density map phased with anomalous signal, contoured at 1σ. Red mesh: anomalous difference map contoured at 3 σ. B. Detail of the AtBRR phospho-acceptor site including Mg ion (green sphere) coordinated by side chain and main chain atoms and three ordered water molecules (red spheres). The electron density shown is a 2Fo-Fc simulated annealing composite omit map contoured at 1σ.

Table 1. X-ray data collection and structure determination statistics

Data Collection	**RtBRR$_{HIS}$ (5BRI)**	**AtBRR (5BRJ)**
Wavelength, Å	0.9791	0.9787
Resolution*, Å	38.3-1.83 (1.85-1.83)	35.0-1.92 (1.95-1.92)
Space Group	$P4_122$	$P4_122$
Unit Cell (a, b, c (Å))	47.7, 47.7, 193.5	41.0, 41.0, 187.7
Completeness, %	99.8 (98.1)	98.7 (97.0)
#Unique/#measured Reflections	20790/20756	13163/12959
#Anomalous Reflections	37570	-
Redundancy	1.8 (1.8)	27.2 (26.4)
<I/σI>	22.9 (1.1)	53.2 (23.0)
Wilson B value, Å^2	33.0	23.7
R_{sym}†, %	2.6 (105.)	7.2 (15.5)
Refinement		
Resolution, Å	38.36-1.90 (1.93-1.90)	30.9-1.92 (1.99-1.92)
Reflections/# anomalous	18619/33646	12959
R_{work}/R_{free}, ‡ %	17.5/21.3 (31.5/40.4)	19.6/24.0 (21.7/27.4)
Rms deviations		
Bonds, Å	0.02	0.07
Angles, °	1.64	1.03
Ramachandran statistics, %		
Allowed	98.8	96.3
Generously allowed	1.2	3.7
# atoms		
protein	1375	1107
ligand	9	1
water	123	135
<B factor>, Å^2		
protein	38.4	26.8
ligand	37.3	20.2
water	52.8	36.6

* The highest resolution bin is indicated in parentheses.

† $R_{sym} = \Sigma\Sigma j|Ij - \langle I \rangle|\Sigma Ij$, where Ij is the intensity measurement for reflection j and $\langle I \rangle$ is the mean intensity for multiply recorded reflections.

‡ $R_{work}/R_{free} = \Sigma||F_{obs}| - |F_{calc}||/|F_{obs}|$, where the working and free R factors are calculated by using the working and free reflection sets, respectively. For the R_{free}, 5-10% of the total reflections were held aside throughout refinement.

As expected, the basic topology of $RtBRR_{HIS}$ is similar to the archetypal SDRR, *E. coli* CheY (44) (PDBID: 3CHY, RMSD = 1.3 Å over 70 C_α atoms within secondary structure elements). The common structural features are a hydrophobic five-stranded parallel β sheet core (2-1-3-4-5) surrounded by five α helices (Fig 4A). The sequence-conserved phosphoaccepting aspartate (D64 in RtBRR, D65 in AtBRR) is situated at the C-terminal end of β3 and faces into the solvent, accessible for interaction with the cognate BphP HK (Fig. 4A). Both BRR structures had clear electron density for magnesium ions in the phosphorylation site (Fig. 3B), which are required for the phosphoryltransfer reaction (45).

The crystal structure of $RtBRR_{HIS}$ revealed an unusual crossover dimer interface geometry that links sister monomers (Fig 4A). The dimer relies on a C-terminal extension with marked structural deviation from known SDRR structures; a bulky hydrophobic β6 (IFWAVL) that extends from α5, threading through the adjacent monomer before turning back toward the originating protomer via a proline/glycine rich sequence (Fig. 4B). Two antiparallel strands, one from each monomer, form an intermolecular β sheet, evidenced by the presence of six main chain hydrogen bonds. This small sheet is widened by main chain interactions with the loop between α1 and β2 and stapled in place by the side chain of N29, itself invariant among a group of approximately 150 RRs that share the DLGhFWAhLNEPPPG sequence (where h is a hydrophobic M/V/L/I residue). Conserved P4 positions the N-terminus for additional interactions at the interface (Fig. 4B). The $RtBRR_{HIS}$ polypeptide ends in a sixth α helix (α6), which packs against α1 and β2 and is not found in most SDRR structures. Surface electrostatic potential models of RtBRR reveal a sizable cleft between protomers and an extensive positively charged stripe across the dimer surface, either or both of which might serve as an interaction site for signaling partners (Fig. 4C).

To ensure that the observed $RtBRR_{HIS}$ dimer interface was not a consequence of stabilizing effects of helix promoting residues or histidine-metal interactions in the cloning tag

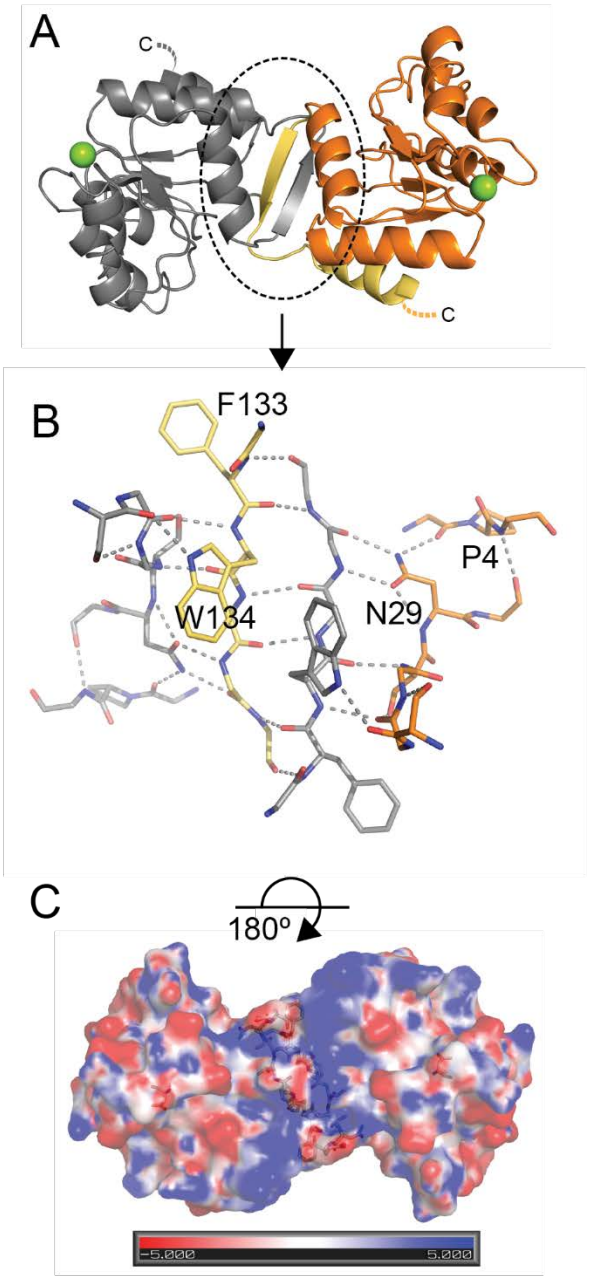


Figure 4. Structure of arm-in-arm BRR dimers. A. Overview of the $RtBRR_{HIS}$ dimer in which one protomer (grey) links with another (orange, with β6 and α6 highlighted in yellow) to form the arm-in-arm dimer (cloning tag and histidine residues intimated with dashed line). The green sphere represents the Mg ion that marks the phospho-acceptor surface. B: Detail of the $RtBRR_{HIS}$ arm-in-arm dimer interface. β6 from each monomer participates in an intermolecular antiparallel β-sheet; the extensive hydrogen-bonded network also involves N-terminal residues and the loop between α1 and β2. Conserved N29 bridges the quaternary arrangement by hydrogen bonding to main chain atoms of residues 3 and 136. Conserved P4 positions the N-terminus for these interactions. C. Surface interaction potential for the RtBRR dimer. The hFWAhL dimerization motif and D64 are shown as sticks.

and to explore BRR stoichiometry in other BphP systems, we also analyzed the interface in the crystal structure of the native sequence AtBRR from *A. tumefaciens* strain F2. This protein contains all three of the noted N-terminal Pro, α1-β2 loop LxN, and C-terminal DLGhFWAhLNEPPG motifs (Fig. 5). AtBRR dimers were linked by the same bulky hydrophobic β6 (VFWALL) although the following proline-rich turn and terminal helix residues could not be modeled due to weak electron density. Given the stabilizing effect of zinc in the crystal structure of $RtBRR_{HIS}$, size exclusion chromatography runs for $RtBRR_{HIS}$ and AtBRR were performed in metal-free buffer and revealed peaks corresponding to dimers in solution for both species (Fig. 6A, 6B).

Analysis of packing interactions of $RtBRR_{HIS}$ and AtBRR using the PISA server (37) confirmed the dominant nature of this hydrophobic crossover strand in stabilizing the dimer. The interface has solvent-inaccessible areas of 1310 $Å^2$ (ΔG = -20.7 kcal/mol) and 1400 $Å^2$ (ΔG = -17.9 kcal/mol) per monomer for RtBRR (regardless of inclusion of tag residues) and AtBRR, respectively, with both hydrophobic packing and hydrogen bonding playing important roles in the interaction (Table 2).

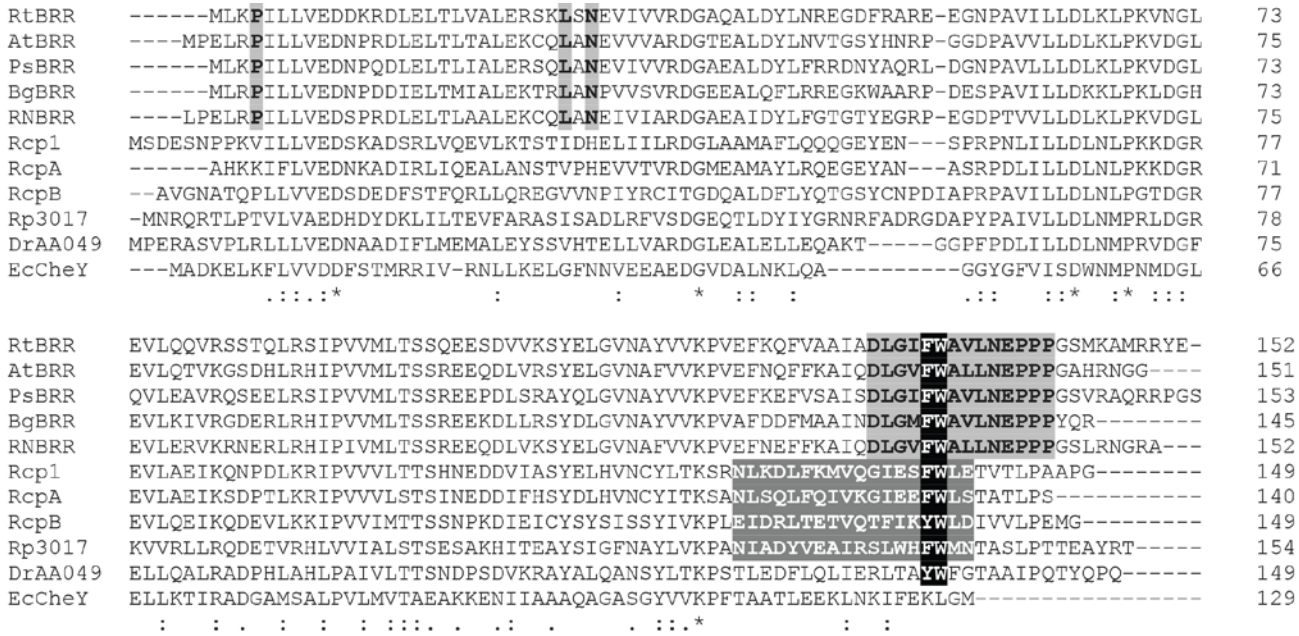


Figure 5. Multiple sequence alignment of SDRRs encoded near bacteriophytochromes. The F/Y and W conserved in bacteriophytochrome RRs are highlighted (black background, white text). Residues crucial to the arm-in-arm dimer interface observed in RtBRR and AtBRR are also conserved in *Pseudomonas syringae* (PsBRR), *Burkholderia glumae* (BgBRR), and *Rhizobium* NT-26 (RNBRR) RRs (grey background, black text). SDRRs adopting the inverted 4-5-5 dimer interface (Rcp1, RcpA, RcpB, and Rpa3017; α5 residues are shown with grey background, white text) lack most or all of these residues, as does the RR from *D. radiodurans* (DrAA049).

Table 2. Statistics for phytochrome cognate RR dimer interfaces (calculated using ProtParam and PISA).

Protein (PDB ID)	Crossover motif	Interface area per monomer (Å^2)	ΔG (kcal/mol)	Hydropathicity score[1] for motif	# residues in interface (per monomer)	# h-bonds in interface
RtBRR (5IC5)	IFWAVL	1309.7	-20.7	2.70	30	16
AtBRR (5BRJ)	VFWALL	1399.9	-17.9	2.58	30	24
Rcp1 (1I3C)	SFWLET	1172.5	-13.3	0.12	29	19
RcpA (1K68)	EFWLSY	1104.9	-15.3	0.02	27	11
RcpB (1K66)	KYWLDI	1248.4	-13.0	-0.22	33	18
Rpa3017 (4ZYL)	HFWMNT	1164.9	-7.5	-0.60	33	16

[1] Grand average of hydropathicity (38).

BRR dimers promote signal transduction

In order to investigate the relevance of the unique arm-in-arm BRR dimers formed by the crossover strand in phosphorelay from BphPs to BRRs, we engineered a monomeric BRR missing the C-terminal 21 amino acids of the native RtBRR sequence. These amino acids correspond approximately to an extension in the sequence compared to canonical SDRRs (Fig.5). Size exclusion chromatography verified that $RtBRR_{mon}$ has a MW in solution of 16 kDa, versus 36 kDa for the $RtBRR_{HIS}$ dimer (Fig. 6A, 6C).

Dark-adapted RtBphP1 dimers were allowed to autophosphorylate at room temperature before the addition of $RtBRR_{HIS}$, $RtBRR_{mon}$, or AtBRR, and the transfer of radiolabeled phosphate was visualized by autoradiography. Phosphotransfer was evident to both $RtBRR_{HIS}$ dimer and $RtBRR_{mon}$ monomer (Fig. 7A), indicating that specificity between the two components was not negatively impacted by the monomer deletion. Specificity for the cognate RR was evident, as RtBphP1 did not transfer phosphate to AtBRR at appreciable levels (Fig. 7A).

To assess the relative efficiency of phosphotransfer from RtBphP1 to each BRR, we independently measured both loss of phosphorylated HK (HK*) (Fig. 7A, B) and accumulation of phosphorylated BRR (BRR*) (Fig. 7C, D). In the presence of $RtBRR_{HIS}$ dimer, RtBphP1 was more efficiently dephosphorylated than the HK^* only phosphorolysis control, with 65% of HK^* remaining after 30 seconds (Fig 7A, B). However, in the presence of $RtBRR_{mon}$ the amount of remaining HK^* was actually higher than the basal level (130%), suggesting monomeric RtBRR inhibits phosphotransfer from HK* and/or phosphorolysis in the movement of phosphate through the TCS (Fig 8). RtBphP1 phosphorylation in the presence of AtBRR was equivalent to the buffer only control, demonstrating phosphotransfer specificity between the RtBphP1 and its cognate BRR.

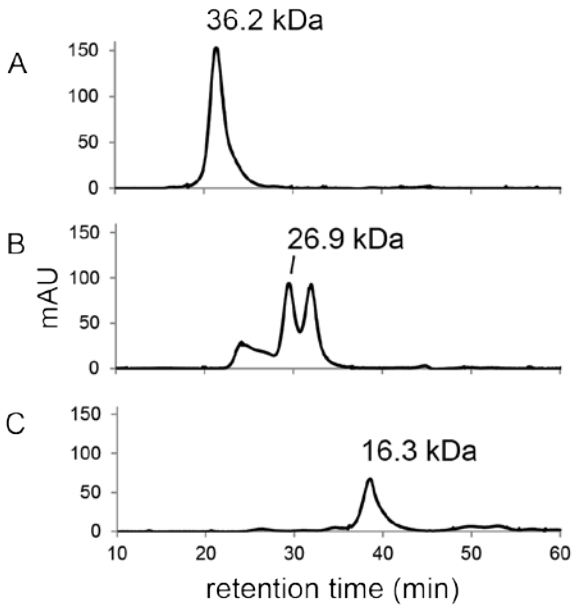


Figure 6. Oligomeric status of BRR proteins determined by gel filtration chromatography. Molecular weights are reported for the highest magnitude peak in each panel based on a standard curve. (Non-labeled peaks in panel B are 33.1 kDa and 23.9 kDa).The labeled peaks were isolated and used for downstream experiments. A: $RtBRR_{HIS}$, B: AtBRR, C:$RtBRR_{mon}$.

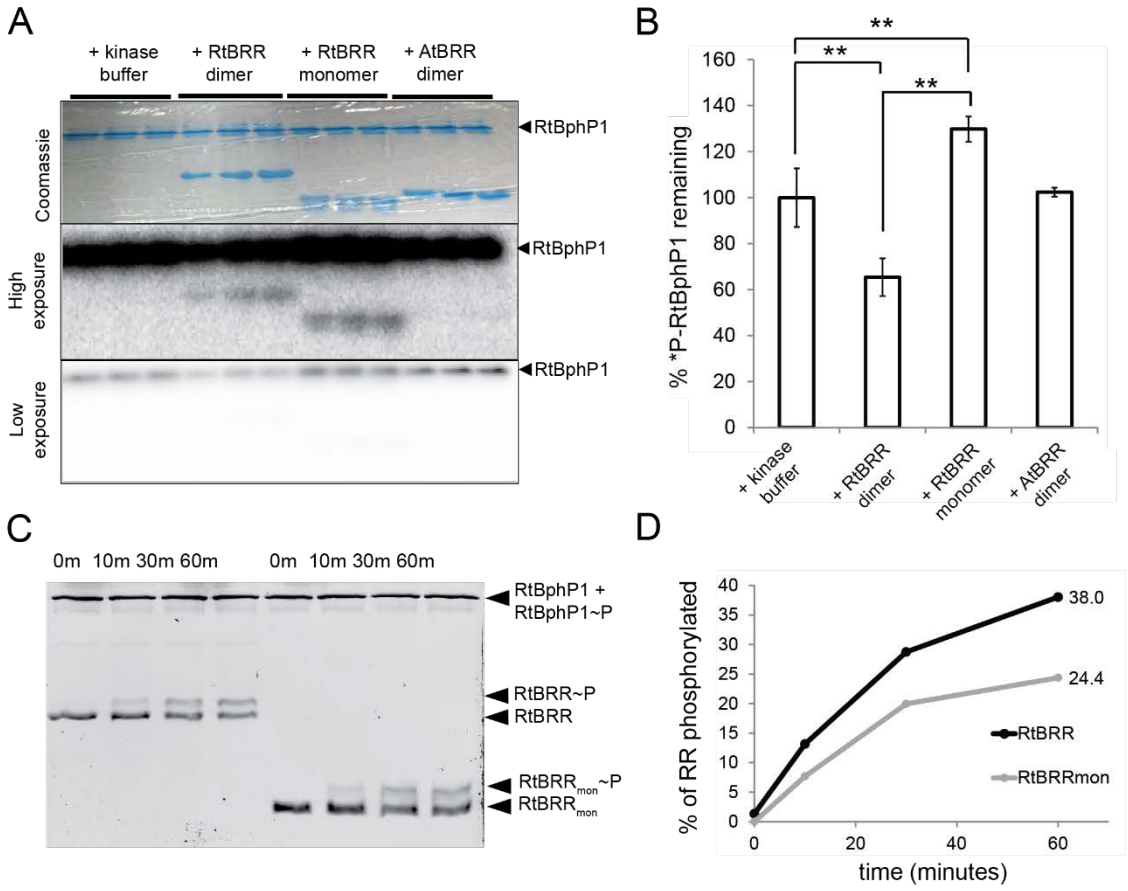


Figure 7. Phosphorylation state of TCS partners during phosphotransfer reactions. A. After preincubation with ATP-^{32}P, RtBphP1 was incubated with listed RR as described in Materials and Methods and visualized by Coomassie stain and phosphoimaging of an SDS-PAGE gel (see Materials and Methods). B. Phosphorylation of RtBphP1 was quantified using the low exposure image in A, normalized to the intensity of Coomassie-stained BphP bands, and plotted as % of basal phosphate remaining. Three independent experiments are plotted as mean ± S.D.; two asterisks denote $P < 0.01$ in a Tukey's HSD test. C. Phosphorylated RtBphP1 was incubated with $RtBRR_{HIS}$ dimers (first four lanes) or engineered monomers (lanes 5-8) and ATP and imaged on a Phos-tag acrylamide gel (see Materials and Methods). D. SYPRO Ruby stained bands from C were quantified with ImageJ and plotted as % of RR phosphorylated.

In order to examine signal transduction via accumulation of BRR*, RtBphP1 dimers were incubated with $RtBRR_{HIS}$ or $RtBRR_{mon}$ and an excess of ATP before analyzing the level of RR phosphorylation using Phos-tag acrylamide gels (46–48). Phosphotransfer was evident to both $RtBRR_{HIS}$ dimer and $RtBRR_{mon}$ monomer (Fig. 7C), providing confirmation that specificity between the two components was not impacted by the monomerization. Analysis of relative phosphorylation levels of BRRs indicated that $RtBRR_{HIS}$ more rapidly accumulated greater levels of phosphate than $RtBRR_{mon}$ for all time points tested (Fig. 7D). At the final 60 minute time point, $RtBRR_{HIS}$ was 38% phosphorylated compared to 24% $RtBRR_{mon}$ phosphorylated. Thus, monomerization of RtBRR inhibits phosphotransfer from HK* and/or enhances intrinsic BRR* dephosphorylation (Fig. 8). One or more of five steps may be impacted by BRR stoichiometry: HK*-RR association, phosphotransfer, HK-RR* dissociation, or autodephosphorylation of either signaling partner (Fig 8).

Discussion

Our characterization the *R. tataouinensis* bacteriophytochrome TCS provides autophosphorylation kinetics of a red light repressed HK with rapid thermal dark reversion, reveals structural details of a previously unknown arm-in-arm dimer association for the RR, and demonstrates the importance of the arm-in-arm dimer for efficient signal transduction through the TCS.

The thermal reset of RtBphP1 to the dark, kinetically accelerated signaling state proceeds rapidly (Fig 1D). This behavior, manifested as weak or unstable absorbance at 750 nm in the Pfr state, is also observed in Agp1 from *A. tumefaciens* (21), the cognate BphP for AtBRR. This is contrasted with most other characterized BphPs and Cphs including *D. radiodurans* DrBphP and *Synechocystis* sp. PCC 6803 Cph1, which achieve stable Pfr absorbance states and more slowly decay back to the Pr state, yet can be photoswitched to the Pr state with far red light. The former class may act as single state light sensing switches, which in the absence of light are rapidly activated. The advantage of the latter class is that such BphPs can act as two-state light sensors contained in a single enzyme, advantageous in environments where sensing the ratio of two light wavelength ranges confers a survival advantage. Potentially, BphPs with unstable Pfr absorbance require a second signal or binding partner to stabilize the far red state, thereby integrating multiple input signals.

RtBphP1 autophosphorylation measurements support the model that red light acts as a non-competitive inhibitor of the BphP transphosphorylation activity (Fig 2). BphP HK dimers with covalently attached BV in the sensory domain can bind ATP in the kinase domain with similar affinity in the Pr and Pfr states. Autophosphorylation proceeds more rapidly in the dark, thus catalytic steps of ATP hydrolysis and/or transfer of the γ-phosphate to histidine are potentially light-regulated. Conformational changes originating in the sensory domains of BphP must be

transduced to regulate trans-phosphorylation. The PHY domain tongue-refolding mechanism (7) demonstrated for DrBphP likely contributes to regulation of HK domain activity by repositioning the HK domains on each protomer relative to one other. Perhaps it is surprising that autokinase activity is reduced only two-fold by red light (*in vitro*) if *R. tataouinensis* utilizes this sensor kinase to regulate a process vital to environmental survival. However, ATP hydrolysis and transfer of phosphate to histidine form but one measurable kinetic step that contributes to overall phosphate flux in the TCS (Fig. 8).

The arm-in-arm dimer interface observed for $RtBRR_{HIS}$ and AtBRR (Fig. 9A, 9B) differs from the dimer interface observed in three Cph cognate RR structures and the *R. palustris* Rpa3107 (Fig 9C, 9D) (PDB: 1I3C, 1K66, 1K68, 4ZYL) (23–25) and differs from three other known RR homodimerization modes (26). Although all six structural examples of BphP-associated RRs form dimers mediated by conserved Phe/Tyr and Trp amino acids, conservation of these two aromatic residues alone does not result in equivalent quaternary structures. The hydropathicity of surrounding amino acids and their influence on local secondary structure determine the quaternary arrangement for these SDRRs as arm-in-arm (Fig 9 A,B) or inverted 4-5-5 dimers (Fig. 9 C,D).

In the cyanobacterial RRs, charged and polar residues surround FW and result in high hydrophilicity, quantified by “grand average of hydropathicity” scores of -0.2-0.1 (38) (Table 2). The dimerization motif residues form α-helix 5 near the C-terminus, where Phe and Trp jut out to pack against the sister protomer (Fig 9 C,D). In our newly solved BRR structures, bulky hydrophobic residues surround FW and result in high hydrophobicity (hydropathicity scores of 2.6-2.7) (Table 2); α-helices do not form and instead hydrophobic residues are buried within the folded core, leaving main chain atoms to form a hydrogen-bond dominated β-sheet interface (Fig. 9 A,B). More than 150 bacterial SDRRs in currently searchable databases carry the extended C-terminal hydrophobic crossover motif hGhFWAhLNEPP. All of these also carry a

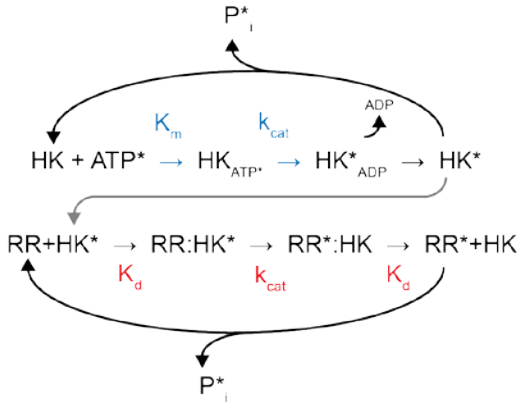


Figure 8. Simplified schematic for phosphate flux through the RtBphP1 (HK) - RtBRR (RR) TCS, showing some of the rate constants and affinities that contribute to autophosphorylation and phosphotransfer steps during signal transduction. K_m and k_{cat} for RtBphP1 (blue) were directly measured in this study. One or more of the phosphotransfer steps (red) or phosphorolysis steps is likely to be impacted by BRR stoichiometry, thus accounting for greater accumulation of phosphate on $RtBRR_{HIS}$ vs. $RtBRR_{mon}$. (Putative additional members of this TCS are not taken into account here.)

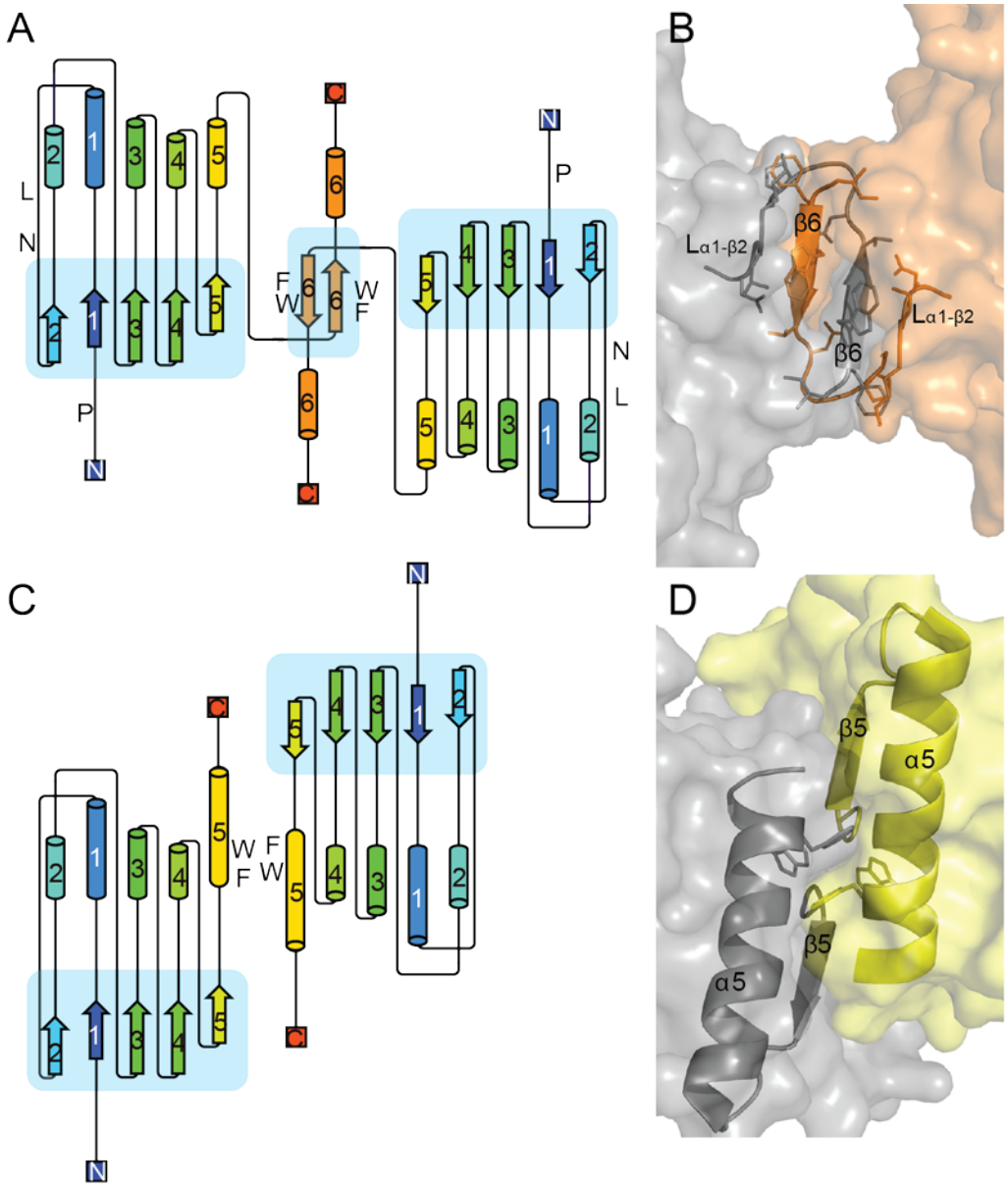


Figure 9. Uniqueness of arm-in-arm response regulator dimers and the FW role. A. Topology of the arm-in-arm dimer seen in RtBRR and AtBRR, characterized by the hGhFWAhLNEPP motif which interacts with N-terminal segment (marked by P4) and α1-β2 connecting loop (marked by L27, N29). B. Structural view of these motifs in the context of RtBRR. C. Topology of the inverted 4-5-5 dimer seen in Rcp1, RcpA, RcpB, and Rpa3017, characterized by hydrophilic residues surrounding the FW key. D. Structural view of the β5-α5 packing interaction in the context of Rcp1 (PDBID: 1I3C) (23). Topology diagrams were generated using Pro-origami (54).

Pro near the N-terminus and the LxN motif between α1 and β2. Notably, all of these predicted arm-in-arm dimer SDRRs are found in non-photosynthetic bacteria, most of which are plant pathogens or commensals such as *Pseudomonas syringae* and *Burkholderia glumae* (Fig. 5). Future work to elucidate the biological roles of these signaling pathways is needed.

The BRR dimers possess greater solvent-inaccessible surface area and greater negative ΔG of dimerization than the six previous examples (Table 2), which suggests the arm-in-arm dimers require a substantial input of energy to dissociate and may maintain the arm-in-arm arrangement regardless of signaling state. Regulation of monomer-dimer equilibrium in response to phosphorylation has been suggested as a possible mechanism of signal transduction for the Cph1-Rcp1 TCS (23). Such a transition cannot be ruled out for RtBphP1 and RtBRR, however the predicted thermodynamics for the dimer interface disfavor such a mechanism. Alternatively, one can consider a model in which phosphorylation promotes conversion between inverted 4-5-5 and arm-in-arm dimers, impacting quaternary structure arrangement. How phosphorylation affects other, potentially signal-transducing, regions of the BRR structure remains an open question. The BRR proteins studied here retain the switch tyrosine essential for chemotaxis signaling in the *E. coli* CheY system (49), and the 6th helix in RtBRR does not occlude the position occupied by the FliM α-helix, the binding partner for CheY (50). Further structural studies of variants and/or phosphate-analog bound BRRs are warranted to address these questions.

The marked biochemical consequences of disrupting the arm-in-arm dimer interface imply that this oligomeric arrangement is the relevant signal receiver in the TCS *in vivo* (51). Although we cannot define the precise molecular block, we observe an inhibitory effect on both dephosphorylation of RtBphP1 and accumulation of phosphorylated RtBRR in the presence of $RtBRR_{mon}$ compared to dimeric RtBRR. What new functionalities could be conferred by the novel arm-in-arm dimer compared to a canonical monomeric SDRR? In both experiments

conducted, phosphotransfer proceeded to monomeric as well as dimeric forms of RtBRR, thus HK recognition cannot be carried by the C-terminal dimerization motif. *In vivo*, additional BRR interactions that couple the TCS to an appropriate cellular response may require the novel surface generated by the dimer interface or may take advantage of the larger molecular size of a dimer compared to a monomeric SDRR. The BphP TCS from *Rhizobium* NT-26 has been shown to branch, with the BphP1 serving as a phosphodonor to both the cognate arm-in-arm BRR (Fig. 5) and an adjacent hybrid HK containing a RR domain (48); one might postulate that such complex networks of protein interactions could capitalize on the multiple binding sites presented by a dimeric BRR. Alternatively, arm-in-arm BRR dimers may provide a cooperative mechanism for signal transduction by doubling the local concentration of phosphoacceptor sites. This mechanism might increase efficiency of a RR that acts as a phosphate sink in a multicomponent TCS to fine tune signaling, as has been proposed for the LovR RR in the blue-light regulated LOVHK pathway (52, 53). Generating a molecular picture of how BRRs couple light sensing by BphPs to phenotypic responses is the next major knowledge gap to be addressed. Identifying signal transduction steps beyond phosphorylation of the SDRR remains one of the most elusive areas of BphP research, which will be aided by biochemical and structural knowledge of the pathway proteins.

Materials and Methods

Cloning, expression, and purification of BphPs and BRRs

R. tataouinensis strain TTB310 was grown as described (27) and genomic DNA was extracted using a Wizard genomic DNA kit (Promega, Madison, WI). The genes for RtBphP1 and RtBRR were amplified by PCR from *R. tataouinensis* genomic DNA with primers encoding BamHI and HindIII restriction enzyme sites (RtBphP1 F: 5'-CGAAGGATCCATGAACCTTCCGCCGCCTGACCTGG-3', RtBphP1 R: 5'-CGAAAAGCTTTTAAGCATGGTTCCTGTCCTCTTTCCTCTTGGGCGGC-3', RtBRR F: 5'-GGTTGGATCCATGCTTAAACCCATCTTGCTTGTCGAGGACGACAAGC-3', RtBRR R: 5'-CCTTAAGCTTTGCTTCGTAGCGGCGCATGGCCTTCATGGACC-3'). These sites were used to clone the genes into plasmid pJ414 encoding an N-terminal hexahistidine tag and tobacco etch virus (TEV) protease site (DNA 2.0, Menlo Park, CA) or pET23a encoding a C-terminal hexahistidine tag (EMD Millipore, Billerica, MD), creating pJ414RtBphP1 and pET23aRtBRR$_{HIS}$, respectively. The gene for RtBRR$_{mon}$ was amplified from pJ414RtBRR using the original forward primer and a reverse primer (R: 5'-GGTTAAAAGCTTTTAGCCCAGGTCGGCGATGGCGGCG-3') and then religated into empty pJ414, which resulted in the 5' truncation of 21 codons. The AtBRR gene was codon optimized for expression in *E. coli*, synthesized and cloned into pJ414, resulting in pJ414AtBRR (DNA 2.0, Menlo Park, CA).

Vectors with sequence-confirmed inserts were transformed into BL21(DE3) (AtBRR) or BL21-Codon Plus(DE3)-RP (*R. tataouinensis* genes) (Agilent Technologies, Santa Clara, CA). Overexpression was induced with 1 mM IPTG at OD_{600} of 0.5-0.8 and carried out for 16-18 hours at 18° C. Cell pellets from 2 L cultures were resuspended in 30 mL lysis buffer (30 mM Tris pH 8.0, 300 mM NaCl, 10 mM imidazole) and were lysed in a French press. RtBphP1 lysate

clarified by centrifugation (30 minutes at 39,190 rcf) was incubated in the dark for 1 hour on ice with 200 µL of 20 mM BV HCl in DMSO (Frontier Scientific, Logan, UT).

All chromatography steps were run at 4° C on an Akta Explorer (GE Healthcare, Pittsburgh, PA). RtBphP1, $RtBRR_{mon}$, and AtBRR were enriched first on a Ni-NTA column (Qiagen, Hilden, Germany) followed by buffer exchange on a 50,000 MWCO or 10,000 MWCO filter (Amicon, EMD Millipore, Billerica, MD) before digestion with TEV protease purified in-house as previously described (28) at 8° C for 16-18 hours. Cleaved species were collected from the flow-through fraction of a second Ni-NTA column and the buffer was changed to 30 mM Tris, pH 8.0 (RtBphP1) or 30 mM tris pH 8.0, 100 mM NaCl, 5% glycerol (BRRs). Proteins were further purified and analyzed for stoichiometry by size exclusion columns Superdex 200 (RtBphP1) or Superdex 75 (BRRs) (GE Healthcare, Pittsburgh, PA). The 189 kDa dimer fraction of RtBphP1 was isolated for use in all experiments. Of the proteins in this study, only dimeric $RtBRR_{HIS}$ retained the C-terminal hexahistidine tag, and thus underwent two-step purification with Ni-NTA affinity purification followed by Superdex 75 size exclusion.

BphP spectrophotometry and Pfr half-life determination

UV-Vis spectra and single-wavelength measurements of RtBphP1 were collected on a Beckman Coulter DU-640B spectrophotometer (Pasadena, CA). All dark state experiments were carried out under dim green light. Spectra were collected on dark-adapted protein which was protected from light for ≥3 hours (dark/Pr state), or illuminated with red light for 1 minute (mixed Pr/Pfr state). Proteins were illuminated with light from a Fostec ACE source filtered with a 700 ± 5 nM bandpass filter (Andover Corp., Salem, NH) which delivered irradiance of 140 µmol/m^2/s as previously described (29). Illuminated state measurements were limited to spectra recorded no less than 10 seconds after light stimulus removal, by which time thermal reversion was underway. Dark-reversion spectra were recorded on an illuminated RtBphP1 sample for a time

course of 10 seconds to 60 minutes after light stimulus removal. Difference spectra were calculated as illuminated absorbance minus Pr absorbance and plotted for each time point.

The extinction coefficient (ε_{708}) of RtBphP1 in the Pr state was calculated by solving $c_{280} = A_{280}/\varepsilon_{280}$ using the theoretical ε_{280}. Assuming $c_{280} = c_{708}$, we then measured A_{708} and solved $\varepsilon_{708}=A_{708}/c_{708}$. This yielded ε_{708} of 98,937 M^{-1} cm^{-1} for Pr RtBphP1. Notably this simplified method does not take into account contamination by apo-phytochrome, which was estimated at 22% using the extinction coefficient for BV (39,900 M^{-1} cm^{-1}) at 388 nm (30). Pfr half-lives were determined by measuring A_{708} for the Pr state on a dark-adapted RtBphP1 sample, followed by 1 minute illumination with 700 nm light, then subsequent A_{708} measurement every 30 seconds for 1 hour at 24° C. Three independent experiments were conducted, and a biexponential curve $y = Aexp(bt) + Cexp(dt)$ was fit to the mean data using SigmaPlot dynamic curve fitting (Systat Software, San Jose, CA). The mean and standard deviation of ln(2)/b and ln(2)/d yielded the reported first and second half-lives of the Pfr state, respectively.

$RtBRR_{HIS}$ and AtBRR X-ray crystal structure determination and protein interface analysis

Purified $RtBRR_{HIS}$ and AtBRR were concentrated to 6.3 mg/mL and 30.0 mg/mL, respectively, in 3000 MWCO centrifugation filters (EMD Millipore, Darmstadt, Germany), and the buffer was changed to 30 mM tris pH 8.0, 100 mM NaCl, 5% glycerol. Hanging drops (31) were set up with 1 µL BRR and 1 µL reservoir solution. $RtBRR_{HIS}$ crystallized with a reservoir solution of 200 mM zinc chloride, 100 mM sodium cacodylate pH 6.5, 10 mM magnesium chloride, and 10% isopropanol. AtBRR crystallized with a reservoir solution of 100 mM tris pH 8.5, 200 mM magnesium chloride, 20% PEG 8000. Three-dimensional, birefringent crystals grew at 19° C within one week (multiple diamond-shaped crystals for $RtBRR_{HIS}$, single large rods for AtBRR). For data collection, crystals were cryoprotected in mother liquor with 15% glycerol for one minute before vitrification in liquid nitrogen.

Diffraction data were collected at the Advanced Photon Source (Argonne National Laboratory, Argonne, IL) on LS-CAT 21-ID-D on a Rayonix MX300 detector ($RtBRR_{HIS}$) and on LS-CAT 21-ID-F on a Rayonix MX-225 detector (AtBRR). The data were integrated and scaled with HKL2000 (32). For $RtBRR_{HIS}$, Phaser-MR (33) was used for initial phasing by molecular replacement with search model PDB 1I3C. The structure revealed an interesting dimethylarsenate/zinc interaction with an ordered C-terminal His tag. Even though we collected data at the peak wavelength for Se (0.97910 Å, LS-CAT 21-ID-D) and not As (1.0417 Å) or Zn (1.2837 Å), we were able to use the anomalous signal and the protein sequence in AutoSolve (34) to independently phase the reported structure. The measurability of the anomalous signal was 0.0619 at 2.58 Å resolution (35) with peaks found for one As and three Zn^{2+} ions. The experimentally-phased electron density map revealed all but the N-terminal 6 of the 166 amino acids in the protein. The cacodylate sits on a two-fold rotation axis and was refined at half occupancy. For AtBRR, Phaser-MR was used for phasing by molecular replacement using $RtBRR_{HIS}$ as the search model. Model fitting was done in Coot (36) and refinement was done using Phenix.refine (35). Coordinate and structure factor files were deposited in the Protein Data Bank under codes 5BRI and 5BRJ, for $RtBRR_{HIS}$ and AtBRR, respectively.

Coordinate files for AtBRR and $RtBRR_{HIS}$ were submitted to the Proteins, Interfaces, Surfaces, and Assemblies (PISA) web server (37) to predict biologically relevant dimer interfaces from the crystal structures; the same analysis was performed on BRR structures from cyanobacteria: Rcp1 (23), RcpA and RcpB (24). Grand average of hydropathicity (GRAVY) scores were computed for the hFWAhL motif from AtBRR, RtBRR, Rcp1, RcpA, RcpB, *Deinococcus radiodurans* A0049, and *R. palustris* Rpa3017 using ExPasy ProtParam (38). Surface electrostatics for RtBRR (without cloning tag residues) were generated using the APBS server (39, 40) and visualized in PyMol.

Autophosphorylation kinetics and phosphotransfer profiling

Autophosphorylation reactions were carried out at room temperature under green safe lights or over a 700 nm LED panel (230 μmol/m^2/s). Reactions contained 5 μM RtBphP1 in kinase buffer (150 mM MES pH 7.5, 150 mM KCl, 5% glycerol, 5 mM $MnCl_2$, 5 mM $MgCl_2$, 2.5 mM DTT) were started with the addition of ATP (6.25-1000 μM cold ATP with 0.03-0.15% γ-^{32}P ATP [6000 Ci/mmol, 150 mCi/mL]), and were stopped after 10 seconds with an equal volume of 2X SDS loading buffer (100 mM Tris-HCl pH 6.8, 4% w/v sodium dodecyl sulfate, 0.2% bromophenol blue, 20% v/v glycerol, 200 mM DTT). Samples were not heated or vortexed prior to 15% SDS-PAGE to minimize loss of phosphorylated species. After Coomassie staining, RtBphP1 band slices were cut and added to 4 mL of Bio-safe II scintillation fluid (Research Products International, Mt. Prospect, IL) before recording counts per minute with a Packard Tri-Carb 2100TR Liquid Scintillation Analyzer (Perkin-Elmer, Waltham, MA). Initial rates (mmol ^{32}P~RtBphP1 /sec) were plotted *vs.* [ATP] to derive V_{max}, K_m, and k_{cat} values (20). The mean and standard deviation from five independent experiments are reported. Previous experiments carried out at constant [ATP] and varying [RtBphP1] were used to ensure this assay was done in the linear range of activity.

Radioactive phosphotransfer reactions were carried out at room temperature under green safe lights. Reactions contained 5 μM RtBphP1 plus 15 μM of RR in kinase buffer. First, the autophosphorylation reaction was run for 30 seconds by addition of ATP (1000 μM, 0.05% γ-^{32}P ATP) to RtBphP1, then phosphotransfer was initiated by adding the RR for an additional 30 seconds. Reactions were stopped with an equal volume of 2X SDS loading buffer. Samples were run on 20% SDS-PAGE gels, stained prior to exposure to phosphor screen (2-16 hours exposure time), and imaged on a GE Typhoon FLA-9000. Phosphor band intensities were quantified with ImageJ software (41), including background estimation and normalization for BphP protein amount based on Coomassie staining. Values reported are the mean and

standard deviation of three independent experiments. A one-way ANOVA test determined there was a statistically significant difference between groups. A Tukey's honestly significant difference test determined the statistical difference between individual groups.

Phosphotransfer reactions for Phos-tag acrylamide gels contained 5 μM RtBphP1 plus 15 μM of RR in kinase buffer with or without 1 mM ATP and were incubated for 0, 10, 30, and 60 minutes before the addition of 3X SDS loading buffer to stop the reaction. Potentially phosphorylated proteins were placed on ice after the reactions were stopped and separated on 12% SDS-PAGE gels with 100 μM Phos-tag acrylamide (Wako Pure Chemicals, Osaka, Japan) within one hour. Gels were run at 4° C until the dye front ran off, and then stained overnight with SYPRO Ruby before imaging on a GE Typhoon FLA-9000. Band intensities (I) minus background intensity were quantified using ImageJ software. Percent of RR phosphorylated was calculated as $I_{upper}*100 / I_{upper} + I_{lower}$ for each lane.

References

1. **Wilde A**, **Fiedler B**, **Borner T**. 2002. The cyanobacterial phytochrome Cph2 inhibits phototaxis towards blue light. Mol Microbiol **44**:981–988.

2. **Ng W-O**, **Grossman AR**, **Bhaya D**. 2003. Multiple light inputs control phototaxis in Synechocystis sp. strain PCC6803. J Bacteriol **185**:1599–1607.

3. **Kumar S**, **Kateriya S**, **Singh VS**, **Tanwar M**, **Agarwal S**, **Singh H**, **Khurana JP**, **Amla DV**, **Tripathi AK**. 2012. Bacteriophytochrome controls carotenoid-independent response to photodynamic stress in a non-photosynthetic rhizobacterium, Azospirillum brasilense Sp7. Sci Rep **2**:872.

4. **Bhoo SH**, **Davis SJ**, **Walker J**, **Karniol B**, **Vierstra RD**. 2001. Bacteriophytochromes are photochromic histidine kinases using a biliverdin chromophore. Nature **414**:776–779.

5. **Karniol B**, **Vierstra RD**. 2003. The pair of bacteriophytochromes from Agrobacterium tumefaciens are histidine kinases with opposing photobiological properties. Proc Natl Acad Sci USA **100**:2807–2812.

6. **Giraud E**, **Zappa S**, **Vuillet L**, **Adriano J-M**, **Hannibal L**, **Fardoux J**, **Berthomieu C**, **Bouyer P**, **Pignol D**, **Vermeglio A**. 2005. A new type of bacteriophytochrome acts in tandem with a classical bacteriophytochrome to control the antennae synthesis in Rhodopseudomonas palustris. J Biol Chem **280**:32389–32397.

7. **Takala H**, **Björling A**, **Berntsson O**, **Lehtivuori H**, **Niebling S**, **Hoernke M**, **Kosheleva I**, **Henning R**, **Menzel A**, **Ihalainen JA**, **Westenhoff S**. 2014. Signal amplification and transduction in phytochrome photosensors. Nature **509**:245–248.

8. **Burgie ES**, **Wang T**, **Bussell AN**, **Walker JM**, **Li H**, **Vierstra RD**. 2014. Crystallographic and electron microscopic analyses of a bacterial phytochrome reveal local and global rearrangements during photoconversion. J Biol Chem **289**:24573–24587.

9. **Yang X**, **Ren Z**, **Kuk J**, **Moffat K**. 2011. Temperature-scan cryocrystallography reveals reaction intermediates in bacteriophytochrome. Nature **479**:428–432.

10. **Evans K**, **Fordham-Skelton AP**, **Mistry H**, **Reynolds CD**, **Lawless AM**, **Papiz MZ**.

2005. A bacteriophytochrome regulates the synthesis of LH4 complexes in Rhodopseudomonas palustris. Photosynth Res **85**:169–180.

11. **Barkovits K**, **Schubert B**, **Heine S**, **Scheer M**, **Frankenberg-Dinkel N**. 2011. Function of the bacteriophytochrome BphP in the RpoS/Las quorum-sensing network of Pseudomonas aeruginosa. Microbiology **157**:1651–1664.

12. **Davis SJ**, **Vener A V**, **Vierstra RD**. 1999. Bacteriophytochromes: phytochrome-like photoreceptors from nonphotosynthetic eubacteria. Science **286**:2517–2520.

13. **Jaubert M**, **Vuillet L**, **Hannibal L**, **Adriano J-M**, **Fardoux J**, **Bouyer P**, **Bonaldi K**, **Fleischman D**, **Giraud E**, **Vermeglio A**. 2008. Control of Peripheral Light-Harvesting Complex Synthesis by a Bacteriophytochrome in the Aerobic Photosynthetic Bacterium Bradyrhizobium Strain BTAi1. J Bacteriol **190**:5824–5831.

14. **Wu L**, **McGrane RS**, **Beattie GA**. 2013. Light regulation of swarming motility in Pseudomonas syringae integrates signaling pathways mediated by a bacteriophytochrome and a LOV protein. MBio **4**:e00334–13.

15. **De Luca G**, **Barakat M**, **Ortet P**, **Fochesato S**, **Jourlin-Castelli C**, **Ansaldi M**, **Py B**, **Fichant G**, **Coutinho PM**, **Voulhoux R**, **Bastien O**, **Maréchal E**, **Henrissat B**, **Quentin Y**, **Noirot P**, **Filloux A**, **Méjean V**, **DuBow MS**, **Barras F**, **Barbe V**, **Weissenbach J**, **Mihalcescu I**, **Verméglio A**, **Achouak W**, **Heulin T**. 2011. The Cyst-Dividing Bacterium Ramlibacter tataouinensis TTB310 Genome Reveals a Well-Stocked Toolbox for Adaptation to a Desert Environment. PLoS One **6**:e23784.

16. **Auldridge ME**, **Forest KT**. 2011. Bacterial phytochromes: more than meets the light. Crit Rev Biochem Mol Biol **46**:67–88.

17. **Wagner JR**, **Brunzelle JS**, **Forest KT**, **Vierstra RD**. 2005. A light-sensing knot revealed by the structure of the chromophore-binding domain of phytochrome. Nature **438**:325–331.

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

19. **Casino P**, **Miguel-Romero L**, **Marina A**. 2014. Visualizing autophosphorylation in histidine kinases. Nat Commun **5**:1–10.

20. **Psakis G**, **Mailliet J**, **Lang C**, **Teufel L**, **Essen L-O**, **Hughes J**. 2011. Signaling Kinetics of Cyanobacterial Phytochrome Cph1, a Light Regulated Histidine Kinase. Biochemistry **50**:6178–6188.

21. **Lamparter T**, **Michael N**, **Mittmann F**, **Esteban B**. 2002. Phytochrome from Agrobacterium tumefaciens has unusual spectral properties and reveals an N-terminal chromophore attachment site. Proc Natl Acad Sci U S A **99**:11628–11633.

22. **Njimona I**, **Lamparter T**. 2011. Temperature effects on Agrobacterium phytochrome Agp1. PLoS One **6**:e25977.

23. **Im YJ**, **Rho S-H**, **Park C-M**, **Yang S-S**, **Kang J-G**, **Lee JY**, **Song P-S**, **Eom SH**. 2002. Crystal structure of a cyanobacterial phytochrome response regulator. Protein Sci **11**:614–624.

24. **Benda C**, **Scheufler C**, **De Marsac NT**, **Gartner W**. 2004. Crystal structures of two cyanobacterial response regulators in apo- and phosphorylated form reveal a novel dimerization motif of phytochrome-associated response regulators. Biophys J **87**:476–487.

25. **Yang X**, **Zeng X**, **Moffat K**, **Yang X**. 2015. Structure of the response regulator RPA3017 involved in red-light signaling in Rhodopseudomonas palustris. Acta Crystallogr Sect F, Struct Biol Commun **71**:1215–1222.

26. **Gao R**, **Stock AM**. 2010. Molecular strategies for phosphorylation-mediated regulation of response regulator activity. Curr Opin Microbiol **13**:160–167.

27. **Heulin T**, **Barakat M**, **Christen R**, **Lesourd M**, **Sutra L**, **De Luca G**, **Achouak W**. 2003. Ramlibacter tataouinensis gen. nov., sp. nov., and Ramlibacter henchirensis sp. nov., cyst-producing bacteria isolated from subdesert soil in Tunisia. Int J Syst Evol Microbiol **53**:589–594.

28. **Tropea JE**, **Cherry S**, **Waugh DS**. 2009. Expression and Purification of Soluble His6-Tagged TEV Protease, p. 297–307. *In* Methods in Molecular Biology: High Throughput Protein Expression and Purification. Springer.

29. **Auldridge ME**, **Satyshur KA**, **Anstrom DM**, **Forest KT**. 2012. Structure-guided

engineering enhances a phytochrome-based infrared fluorescent protein. J Biol Chem **287**:7000–7009.

30. **Shu X**, **Royant A**, **Lin MZ**, **Aguilera TA**, **Lev-Ram V**, **Steinbach PA**, **Tsien RY**. 2009. Mammalian expression of infrared fluorescent proteins engineered from a bacterial phytochrome. Science **324**:804–7.

31. **McPherson A**. 2009. Introduction to Macromolecular Crystallography, 2nd Ed. John Wiley & Sons, Hoboken.

32. **Otwinowski Z**, **Minor W**. 1997. Processing of X-ray Diffraction Data Collected in Oscillation Mode. Methods Enzym **276**:307–326.

33. **McCoy AJ**, **Grosse-Kunstleve RW**, **Adams PD**, **Winn MD**, **Storoni LC**, **Read RJ**. 2007. Phaser crystallographic software. J Appl Crystallogr **40**:658–674.

34. **Terwilliger TC**, **Adams PD**, **Read RJ**, **McCoy AJ**, **Moriarty NW**, **Grosse-Kunstleve RW**, **Afonine P V**, **Zwart PH**, **Hung L-W**. 2009. Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. Acta Crystallogr Sect D Biol Crystallogr **65**:582–601.

35. **Adams PD**, **Afonine P V**, **Bunkóczi G**, **Chen VB**, **Davis IW**, **Echols N**, **Headd JJ**, **Hung L-W**, **Kapral GJ**, **Grosse-Kunstleve RW**, **McCoy AJ**, **Moriarty NW**, **Oeffner R**, **Read RJ**, **Richardson DC**, **Richardson JS**, **Terwilliger TC**, **Zwart PH**. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr **66**:213–221.

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

37. **Krissinel E**, **Henrick K**. 2007. Inference of macromolecular assemblies from crystalline state. J Mol Biol **372**:774–797.

38. **Gasteiger E**, **Hoogland C**, **Gattiker A**, **Duvaud S**, **Wilkins MR**, **Appel RD**, **Bairoch A**. 2005. Protein Identification and Analysis Tools on the ExPASy Server. Proteomics Protoc Handb 571–607.

39. **Dolinsky TJ**, **Czodrowski P**, **Li H**, **Nielsen JE**, **Jensen JH**, **Klebe G**, **Baker NA**. 2007. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. Nucleic Acids Res **35**:W522–525.

40. **Baker NA**, **Sept D**, **Joseph S**, **Holst MJ**, **McCammon JA**. 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci U S A **98**:10037–10041.

41. **Rasband WS**. ImageJ [Software, Version 1.49]. Natl Institutes Heal Bethesda, MD, USA.

42. **Carson M**, **Johnson DH**, **McDonald H**, **Brouillette C**, **DeLucas LJ**. 2007. His-tag impact on structure. Acta Crystallogr Sect D Biol Crystallogr **63**:295–301.

43. **Cha S-S**, **An YJ**, **Jeong C-S**, **Kim M-K**, **Lee S-G**, **Lee K-H**, **Oh B-H**. 2012. Experimental phasing using zinc anomalous scattering. Acta Crystallogr Sect D Biol Crystallogr **68**:1253–1258.

44. **Volz K**, **Matsumura P**. 1991. Crystal structure of Escherichia coli CheY refined at 1.7-Å resolution. J Biol Chem **266**:15511–15519.

45. **Stock AM**, **Martinez-Hackert E**, **Rasmussen BF**, **West AH**, **Stock JB**, **Ringe D**, **Petsko GA**. 1993. Structure of the magnesium-bound form of CheY and mechanism of phosphoryl transfer in bacterial chemotaxis. Biochemistry **32**:13375–13380.

46. **Barbieri CM**, **Stock AM**. 2008. Universally applicable methods for monitoring response regulator aspartate phosphorylation both in vitro and in vivo using Phos-tag-based reagents. Anal Biochem **376**:73–82.

47. **Kinoshita E**, **Kinoshita-Kikuta E**. 2011. Improved Phos-tag SDS-PAGE under neutral pH conditions for advanced protein phosphorylation profiling. Proteomics **11**:319–323.

48. **Wojnowska M**, **Yan J**, **Sivalingam GN**, **Cryar A**, **Gor J**, **Thalassinos K**, **Djordjevic S**. 2013. Autophosphorylation activity of a soluble hexameric histidine kinase correlates with the shift in protein conformational equilibrium. Chem Biol **20**:1411–1420.

49. **Zhu X**, **Amsler CD**, **Volz K**, **Matsumura P**. 1996. Tyrosine 106 of CheY plays an important role in chemotaxis signal transduction in Escherichia coli. J Bacteriol **178**:4208–

4215.

50. **Dyer CM**, **Dahlquist FW**. 2006. Switched or not?: the structure of unphosphorylated CheY bound to the N terminus of FliM. J Bacteriol **188**:7354–7363.

51. **Laub MT**, **Biondi EG**, **Skerker JM**. 2007. Phosphotransfer profiling: systematic mapping of two-component signal transduction pathways and phosphorelays. Methods Enzymol **423**:531–548.

52. **Foreman R**, **Fiebig A**, **Crosson S**. 2012. The LovK-LovR two-component system is a regulator of the general stress pathway in Caulobacter crescentus. J Bacteriol **194**:3038–3049.

53. **Sycz G**, **Carrica MC**, **Tseng T-S**, **Bogomolni RA**, **Briggs WR**, **Goldbaum FA**, **Paris G**. 2015. LOV Histidine Kinase Modulates the General Stress Response System and Affects the virB Operon Expression in Brucella abortus. PLoS One **10**:e0124058.

54. **Stivala A**, **Wybrow M**, **Wirth A**, **Whisstock JC**, **Stuckey PJ**. 2011. Automatic generation of protein structure cartoons with Pro-origami. Bioinformatics **27**:3315–3316.

Chapter III

Novel post translational modifications in recombinantly expressed bacteriophytochromes

Tandem mass spectrometry was completed by the Mass Spectrometry and Proteomics Facility at the University of Wisconsin-Madison Biotechnology Center

Abstract

Bacteriophytochromes are TCS proteins are known to undergo reversible post translational modification, specifically phosphorylation on a conserved histidine, as part of red light signal transduction. Most spectral, fluorescence, and enzymatic characterization of BphPs to date has taken place on protein which was expressed recombinantly in *E. coli* cell lines. Based on the previous success of tracking BRR phosphorylation on aspartate with Phos-tag acrylamide gels, we intended to optimize the Phos-tag system to track BphP autophosphorylation but instead discovered multiple protein isoforms for full length BphPs which were purified from *E. coli* BL21(DE3) cells. Two hypotheses about the identity of these isoforms were investigated: that the isoforms could be due to the chromophore binding state of the BphP, and that novel post-translational modifications other than the expected histidine site were present on BphPs. It was determined that Phos-tag does not have affinity for biliverdin bound BphP or apo-BphP by comparing the Phos-tag profiles of apo-DrBphP and biliverdin bound DrBphP. Next, Western blot with anti-phospho-serine and anti-acetyl lysine probes suggested that the recombinant full-length BphPs RtBphP1 and Agp1 are acetylated on one or more lysines and that RtBphP1 and Agp1 are phosphorylated on one or more serine residues. A tandem mass spectrometry experiment confirmed the presence of acetylated lysines on RtBphP1 and Agp1 and confirmed the presence of phosphorylated serine on RtBphP1. Acetyl groups were detected on 14 lysine sites and phosphate groups were detected on two serine sites. One of the acetylated lysines was observed in both proteins in all replicates and is near the catalytic site of the histidine kinase domain. Given the widespread use of *E. coli* expression in the BphP field, the PTMs reported here are likely relevant to the majority of previously characterized BphPs and the possible effects of these PTMs on enzyme activity is a concern. A model for post-translationally modified full length RtBphP1 is presented and the implications of these novel PTMs on enzyme activity will form the basis for future investigations.

Introduction

In Chapter 2, the autokinase activity of RtBphP1 was characterized by tracking the incorporation of radiolabeled ^{32}P, and this method has been used for multiple histidine kinase proteins and some BphPs (Georgellis, 1997; Giraud, 2005; Baker, 2016). Based on the success of tracking the appearance of phosphorylated forms of BRR, adapting the Phos-tag gel system to tracking appearance of histidine-phosphorylated BphPs was an attractive possibility. The Phos-tag reagent is composed of a pendant tag that binds two metal ions (Mn^{2+} or Mg^{2+}) which then coordinate phosphate groups on post translationally modified amino acids (Figure 1A) (Kinoshita, 2006; Kinoshita, 2009). When sample proteins are run through an acrylamide gel matrix containing Phos-tag in an electric field, phosphorylated forms of a single protein are retained longer in the matrix due to the affinity of phosphate for Phos-tag (Figure 1B). The expectation for purified, unreacted BphP samples was that they would run as a single, unphosphorylated band on Phos-tag gels. With the addition of ATP and in appropriate reaction conditions identified in Chapter 2, a second, histidine phosphorylated band was expected to appear above the original band (Yamada, 2007). Unexpectedly, purified BphPs which were not reacted with ATP did not run as a single band but as multiple isoforms. This chapter is composed of experiments that determine the cause of these isoforms.

Because Phos-tag has affinity for phosphorylated proteins, a hypothesis that the isoforms represented single or multiple phosphorylations on BphP was considered. The reagent has affinity for phosphate groups on serine, threonine, tyrosine, histidine, and aspartate amino acids and therefore phosphorylated BphPs could be modified on any of these amino acids (Phos-tag PAGE Guidebook). Histidine and aspartate phosphorylation are the hallmark of chemical signaling in TCS and are expected to occur in the BphP-BRR system (Figure 1C) (Swanson, 1994). Serine and threonine phosphorylation, and to a lesser degree tyrosine phosphorylation, has been reported as common in several bacterial proteomes and serine

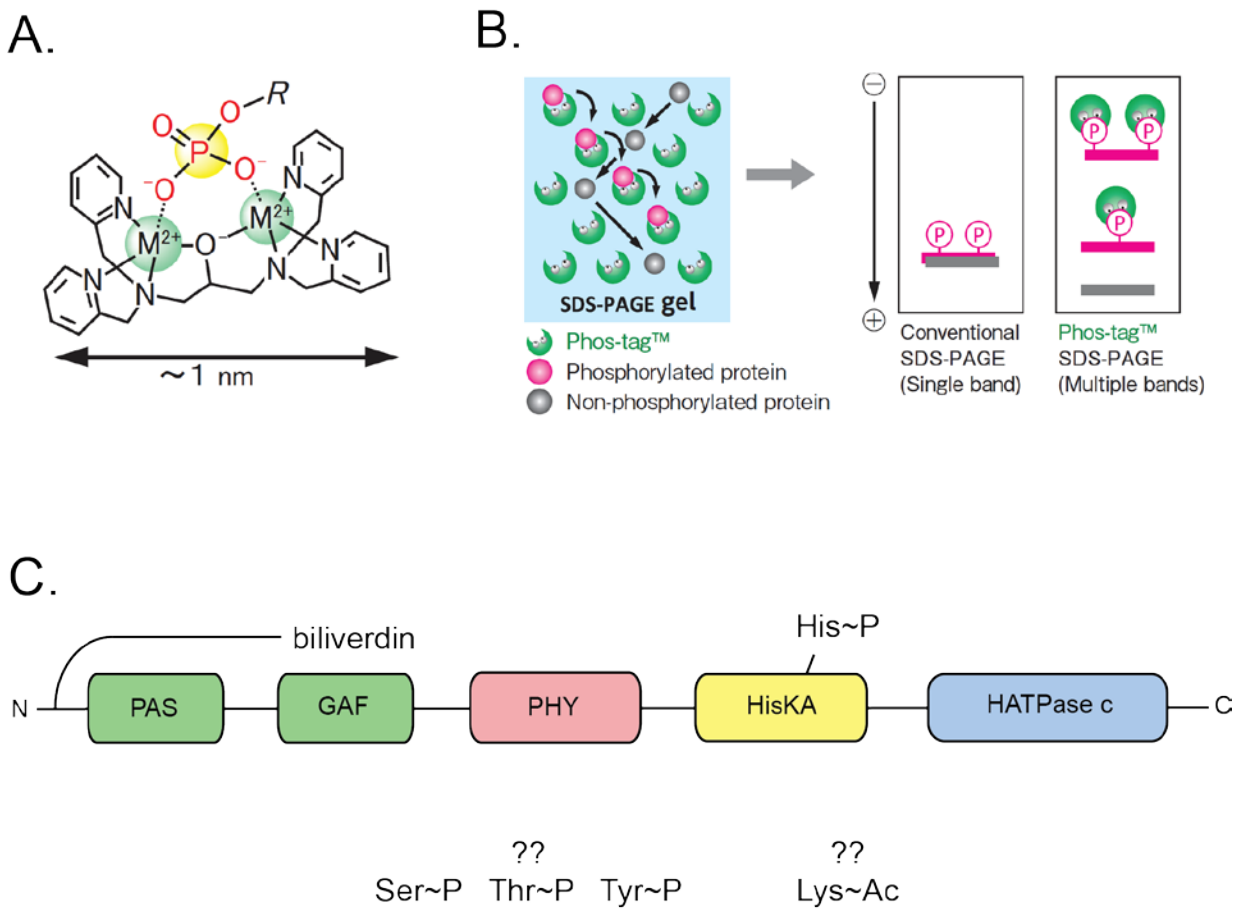


Figure 1. Phos-tag gel methodology and expected post translational modifications on bacteriophytochromes. A. Chemical structure of the Phos-tag reagent bound to two metal ions (M2+) and a phosphate group. B. Method of phosphate-affinity separation with Phos-tag acrylamide gels. Phosphorylated species migrate slower through the Phos-tag acrylamide matrix and appear as upshifted bands when stained for protein (A and B were taken from the Phos-tag PAGE Guidebook). C. Domain architecture of a BphP and the sites of expected modifications including covalently bound biliverdin (BV) and reversible phosphorylation of His 510 in the HisKA domain. Additional modifications that Phos-tag may detect include phsopho-serine (Ser~P), phospho-threonine (Thr~P), phosphor-tyrosine (Tyr~P) and acetyl-lysine (Lys~Ac).

phosphorylation has been detected in TCS proteins (Macek, 2007; Macek, 2008; Ravichandran, 2009; Hu, 2012). Another post-translational modification, acetylation, was considered due to the recently described prevalence of acetyl modifications in bacterial proteomes (Zhang, 2009; Wang, 2010; Crosby, 2012; Kim, 2013; Okanishi, 2013; Mo, 2015). In bacterial cells, acetylation is known to commonly occur on lysine residues (Thao, 2011). Lysine acetylation has been reported as common in bacterial proteomes and is also known to co-occur with phosphorylation on the same protein in some cases (Barak, 2001; Zhang, 2009; Crosby, 2012; Hu, 2013). Ser/Thr phosphorylation and acetylation have both been detected on proteins from species which encode BphPs (Zhang, 2009; Ravichandran, 2009; Crosby, 2012). Therefore, alternative PTM (other than His/Asp phosphorylation) was considered as a possible cause of the BphP isoforms detected on Phos-tag gels and the following work describes preliminary evidence for serine phosphorylation and lysine acetylation on BphPs.

The known chromophore-bound heterogeneity of BphP proteins described in Chapter 2 was also considered as a possible cause of Phos-tag isoform bands. It was determined that purified RtBphP1 has biliverdin bound to 78% of protein molecules, with the remaining 22% being apoprotein (Baker, 2016). Although Phos-tag is not designed to separate proteins based on this quality, the divalent cations are key to its affinity for phosphate groups, and biliverdin is also known to chelate the Zn^{2+} cation (Berkelman, 1986). Furthermore, apo- and holo- forms of BphP have been separated by various chromatography methods in the literature (Wagner, 2007; Bornschlogl, 2009). Therefore, a hypothesis that the biliverdin-binding state could cause multiple isoforms of BphP on Phos-tag gels was considered although ultimately not supported.

The post translationally modified BphPs studied in this chapter were purified from the *E. coli* protein expression strain BL21(DE3) and work by others characterized BphPs purified from this and other *E. coli* strains (Yeh, 1997; Karniol, 2003; Giraud, 2005; Jaubert, 2008; Kumar, 2012). We observed alternative PTMs on BphPs from different species, either co-expressed

with heme oxygenase or with exogenous biliverdin added, and purified by the Forest lab and by other researchers (Scheerer, 2005). The persistence of PTM on the BphPs studied here, and the presence of serine kinase and acetyltransferase enzymes in *E. coli* suggests that many recombinantly expressed BphPs from the literature also bear these PTMs. PTMs are known to regulate enzyme activities (Strahl, 2000; Westermann, 2003; Gallego, 2007), and this raises the question of whether reported activities for BphPs by us and others reflect the true properties of these proteins? The PTM data collected here are summarized and applied to a full scale BphP structural model and hypotheses about specific site PTMs and their potential effects on enzyme activities are postulated.

Results

Phos-tag gels reveal BphP isoforms

The original goal of this line of research was to develop Phos-tag acrylamide gels as a non-radioactive method to determine autophosphorylation kinetics for full-length BphPs to be used in place of the ^{32}P-labeled ATP method employed in Chapter two (Psakis, 2011; Baker, 2016). The expectation was that purified BphPs would run as one band before treatment with the ATP phospho-donor, and would run as two bands after ATP incubation and that the top band would represent BphP phosphorylated on the conserved His 510. However, we found that unlike the BRR proteins, which ran as a single band before treatment with a phospho-donor (see Chapter two, Figure 7), the BphPs ran as multiple bands on Phos-tag gels before any treatment (Figure 2). Multiple bands were detected for three different full length BphPs: RtBphP1 and RtBphP2 from *R. tataouinensis*, and Agp1 from *A. tumefaciens*. Consequently, Phos-tag cannot be used to track the appearance of histidine phosphorylated BphPs at this time. However, the multiple bands raised several questions about potential isoforms of full length BphPs which this chapter will address.

In the cases of RtBphP1 and Agp1, three discernable bands were detected (Figure 2A). One interpretation is that the gels separate BphPs based on affinity for covalently bound phosphate groups, and that there are three different phosphorylation states. Alternatively, the isoforms could be un-phosphorylated (bottom band), single site phosphorylated (middle band), and doubly phosphorylated at two sites (top band). Or, the middle and top bands could represent two different single phosphorylation sites that run on the gels with varying affinity. Possibly, Phos-tag may separate BphP isoforms based on another property, like the biliverdin binding state of the protein, or by another post-translational modification such as acetylation.

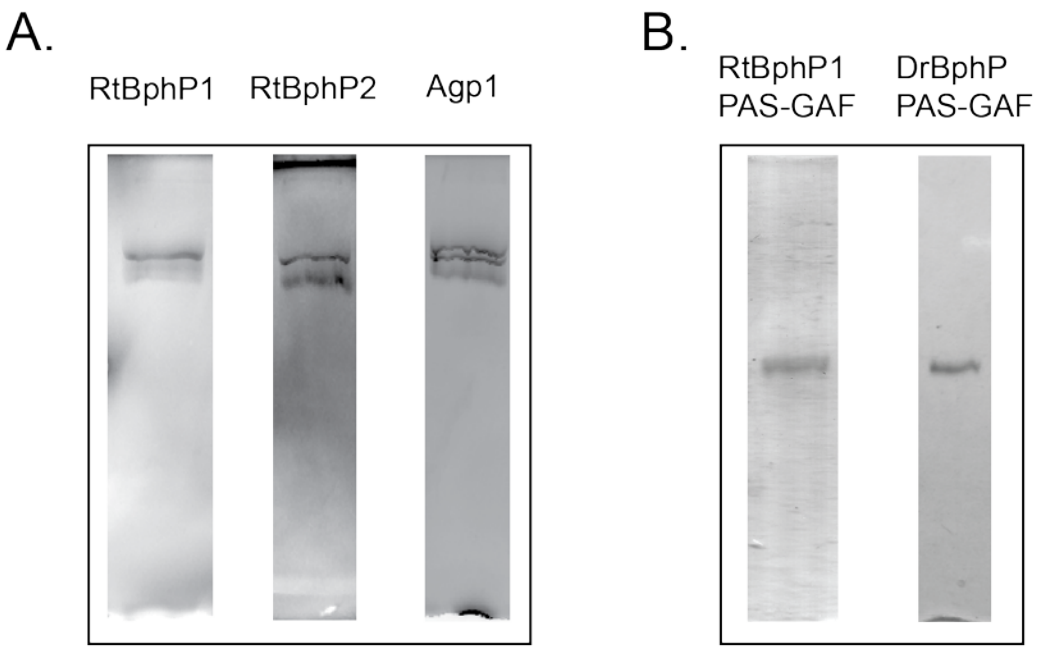


Figure 2. Phos-tag gel separation of full length and truncated BphPs. A. Purified, full length BphPs RtBphP1, RtBphP2, and Agp1 were separated on 6% acrylamide Phos-tag gels and stained with SYPRO Ruby. Multiple isoforms were detected for each protein. B. Purified, truncated BphPs RtBphP1 PAS-GAF and DrBphP PAS-GAF were separated on 10% acrylamide Phos-tag gels and stained with SYPRO Ruby. Multiple distinct isoforms were not detected in truncated BphPs.

The Phos-tag effect on truncated forms of RtBphP1 and DrBphP, which lack the PHY and HK domains was also examined. The truncated proteins, RtBphP1 PAS-GAF and DrBphP PAS-GAF, ran as single, more diffuse bands than the full-length BphPs (Figure 2B). This suggests that the property that caused differential migration of isoforms for the full-length BphP (phosphorylation, acetylation, biliverdin-binding, or another unknown feature) was lacking in this construct.

Because RtBRR was previously demonstrated to accept phosphate from RtBphP1, the Phos-tag banding pattern of RtBphP1 prior to and after phosphotransfer reactions was examined (Figure 3). Consistent with previous results, a second phosphorylated band for RtBRR appeared after 10 and 30 minute reaction times. However, RtBphP1 remained as multiple isoforms after phosphotransfer, indicating that the cause of the isoforms was not solely due to histidine phosphorylation of the BphP.

Phos-tag gels do not have affinity for BV-bound BphPs

In the case of RtBphP2, two bands were detected instead of three (Figure 2A). The preparation of this protein differed in that BphP overexpression was accompanied by the expression of heme oxygenase, which produces the biliverdin chromophore in the *E. coli* BL21(DE3) protein expression cell line. In the case of RtBphP1 and Agp1, the biliverdin cofactor was assembled with the protein after protein overexpression. Two questions arose: Does Phos-tag acrylamide separate apo-BphP from biliverdin bound BphP, and this accounts for the different isoform bands? And, does the assembly of BphP with the chromophore in *E. coli* prevent any post-translational modification (phosphorylation or acetylation) that causes the multiple bands? The Phos-tag profiles of three forms of DrBphP from *D. radiodurans* were examined to address this (Figure 4). In the absence of biliverdin, the DrBphP apoprotein did not run as a single band on Phos-tag gels, suggesting that the same PTM was present as in the

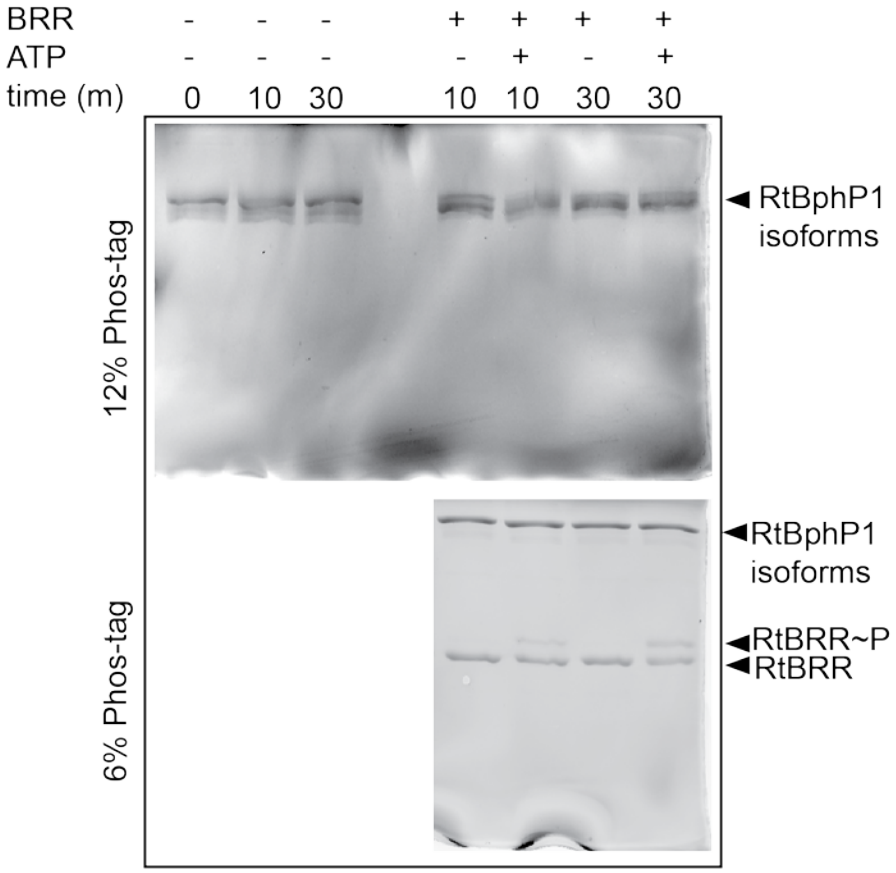


Figure 3. Phos-tag gel analysis of RtBphP1 to RtBRR phosphotransfer reactions. Full length RtBphP1 was incubated at room temperature in the dark for a time course of 30 minutes (lanes 1-3) and separated on 12% Phos-tag gels to reveal multiple isoforms when stained with SYPRO Ruby. Similarly, RtBphP1 was incubated with RtBRR in the presence and absence of 100 μM ATP for 10 and 30 minutes (lanes 4-7). The reacted proteins were separated on 12% and 6% Phos-tag gels to reveal isoforms. RtBphP1 bands did not decrease to a single, non-phosphorylated band as RtBRR gained a second, phosphorylated band.

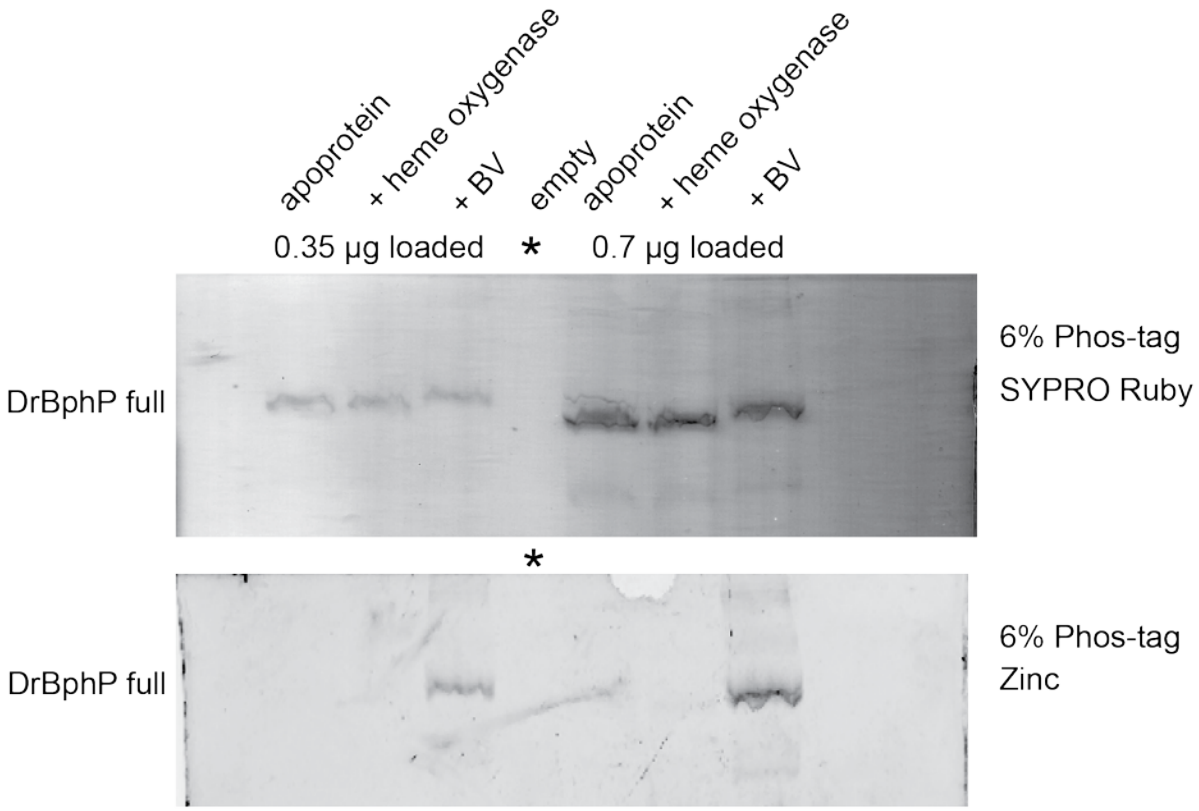


Figure 4. Phos-tag gel separation of apo-BphPs and biliverdin bound BphPs. Full length DrBphP was purified as apo-protein, coexpressed with heme oxygenase, or bound to biliverdin during purification and separated on a 6% acrylamide Phos-tag gel. The gel was stained with SYPRO Ruby (top panel) or visualized for zinc fluorescence at 575 nm (bottom panel). Multiple isoforms were detected for all three DrBphP samples, and only the exogenous BV sample displayed significant fluorescence due to BV binding. The empty lane (asterisk) and adjoining lanes fluoresced due to a crease in the gel. A slight upshift in all isoform bands was detected in the exogenous BV sample.

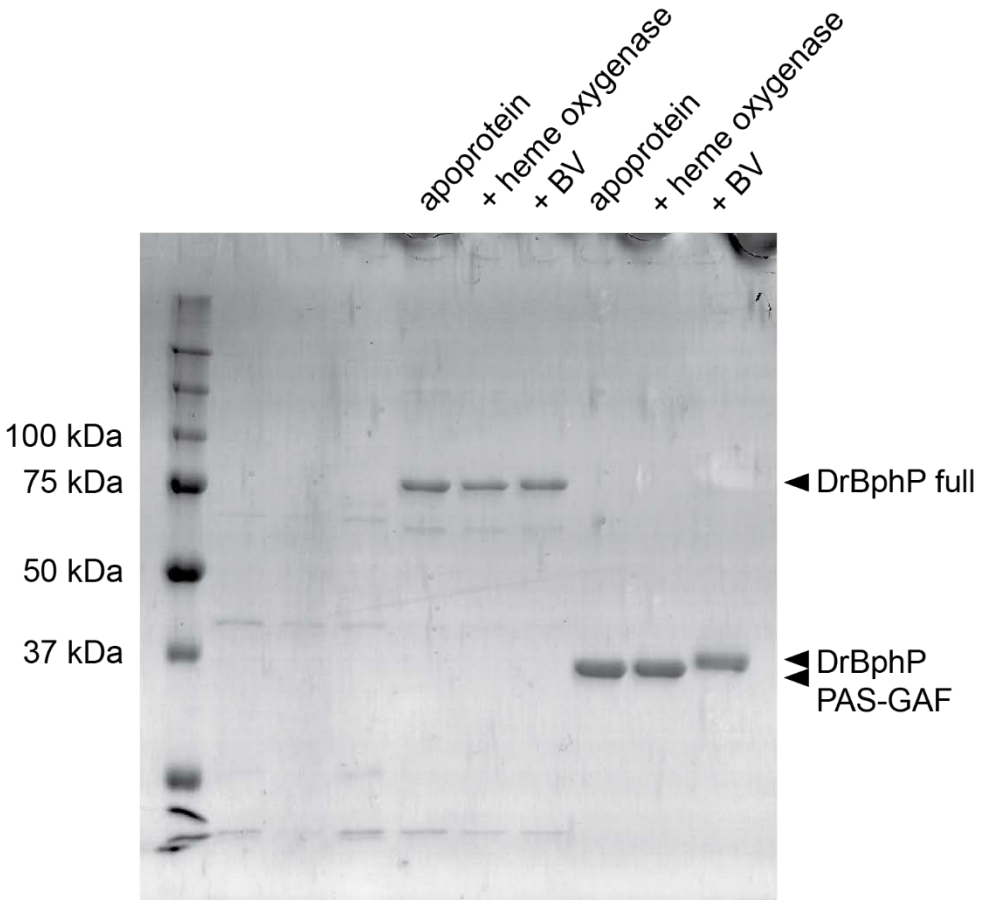


Figure 5. Standard SDS-PAGE gel separation of apo-BphPs and biliverdin bound BphPs. Full length DrBphP and DrBphP PAS-GAF were expressed as apoprotein, coexpressed with heme oxygenase, or bound to biliverdin during purification and separated on a standard 12% acrylamide SDS PAGE gel before staining with Coomassie blue. A slight upshift in the purified protein band was detected in the exogenous BV samples.

RtBphP1 and Agp1 samples. When DrBphP was co-expressed with heme oxygenase, or had excess biliverdin added *in vitro* the bands took on a more collapsed appearance and the distinct top band seen in the apoprotein was not present, yet still did not run as a single band. A zinc fluorescence image of the same gel indicates that biliverdin bound to the protein to a high degree only in the exogenous biliverdin sample. There was an observable effect of bands running slightly higher for BphP samples where biliverdin was present, but this effect was present also in SDS-PAGE gels without Phos-tag (Figure 5).

Full-length BphPs are serine phosphorylated and lysine acetylated

To determine if the isoforms of full-length BphPs seen on Phos-tag gels were due to post-translational modifications, and to determine what amino acid residues were modified, multiple BphPs were probed by Western blotting with anti-acetyl-lysine or anti-phospho-serine antibodies (Figure 6). A positive signal for acetyl-lysine epitopes was detected for RtBphP1 and Agp1, and no signal was detected for RtBphP2. A positive signal for phospho-serine was detected for RtBphP1 and Agp1, and no signal was detected for RtBphP2. Because the detection method for these Western blots relied on fluorescence at a wavelength that might theoretically overlap with BphP fluorescence, we considered these data suggestive and followed up the observation with another experimental technique.

To confirm that post translational modifications were present on full length BphPs and map them to specific amino acids of the protein, purified RtBphP1 and Agp1 were subjected to trypsin digestion and tandem mass spectrometry (MS/MS). The digested peptides were fragmented and PTMs were assigned to specific residues based on their mass to charge ratio using the MASCOT server (Table 1, Figure 7) (Perkins, 1999). The same purification batch of RtBphP1 underwent two separate MS/MS runs and MASCOT analyses, and Agp1 underwent one run and MASCOT analysis. The MS/MS data sets for RtBphP1 and Agp1 resulted in 96%

(both replicates) and 86% sequence coverage, respectively. Serine phosphorylation was detected on two sites of RtBphP1, one unique site found per replicate, and lysine acetylation was detected on eight sites, with only the acetylation at K628 in common for both replicates. Serine phosphorylation was not detected on Agp1 peptides, and lysine acetylation was detected on six sites, including the homologous site to K628. The amino acid sequences of the full-length BphPs used in this study were aligned in order to determine if there were commonly modified sites between RtBphP1 and Agp1 and if any of the modified residues are conserved. Only one lysine residue, K628 in RtBphP1, was found to be acetylated in both replicates of RtBphP1 and Agp1 and is conserved at that site in both species. The other sites of lysine acetylation are spread across the protein sequence and do not repeat between RtBphP1 and Agp1, though most sites occur in the PHY and HK domains. PTMs on the following residues in the RtBphP1 numbering system were detected in this study: K82, K116, S232, S263, K308, K325, K432, K451, K483, K499, K579, K628, K642, K663, K667, and K692.

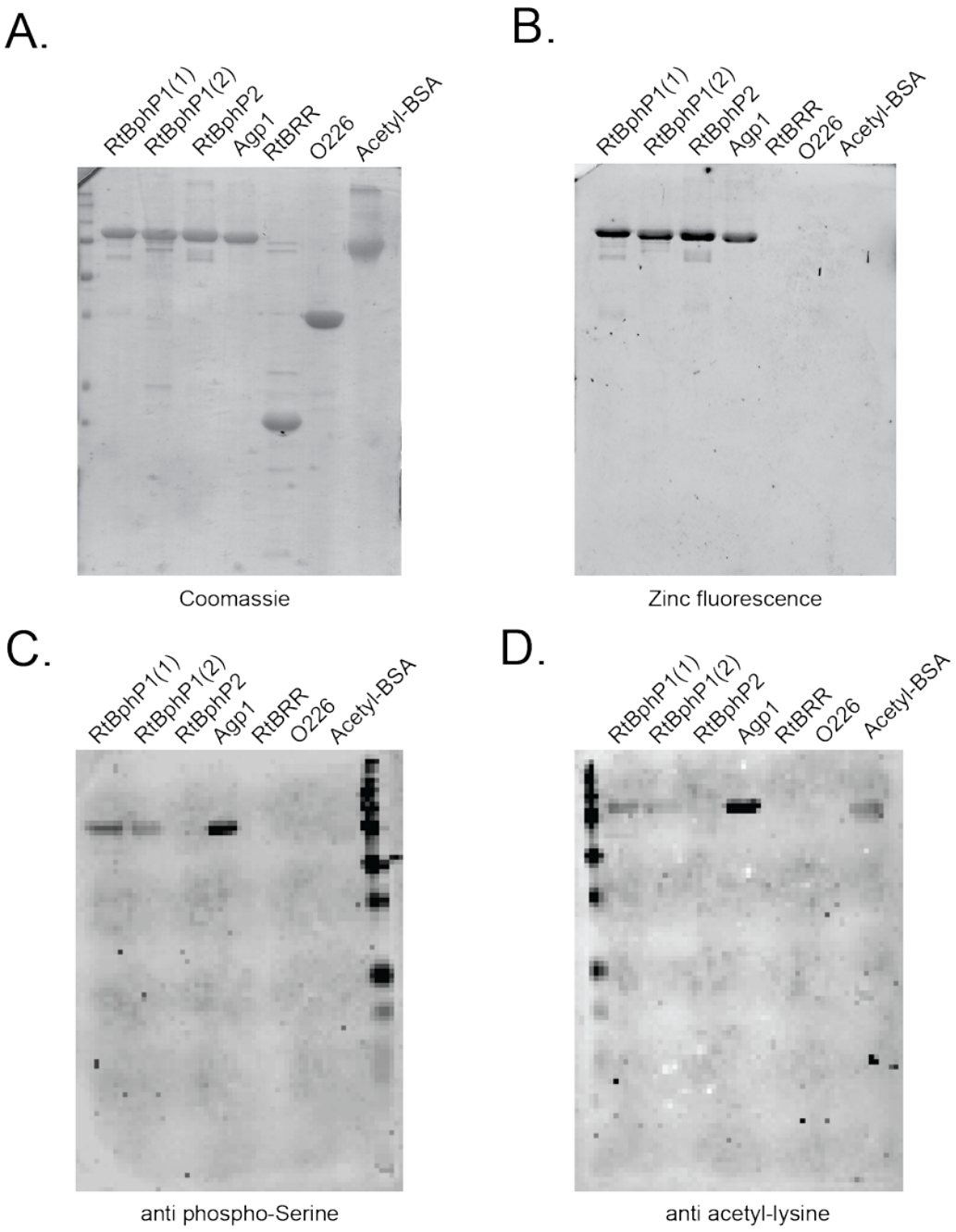


Figure 6. Western blot of full length BphPs with phospho-serine and acetyl lysine probes. A. Purified full length BphPs (RtBphP1 [from two different purification batches], RtBphP2, and Agp1) and controls (RtBRR, Actinobacterium geranyl-geranyl diphosphate synthase [O226], and acetylated bovine serum albumin) were separated on a standard 12% SDS PAGE gel before staining with Coomassie blue. C. Proteins were visualized for zinc fluorescence indicative of covalently bound biliverdin. C and D. Replicate gels were transferred to a nitrocellulose membrane and probed with anti-phosphoserine (C) and anti-acetyllysine (D) antibodies before probing with Alexafluor 647 secondary antibodies and visualization at 700 nm.

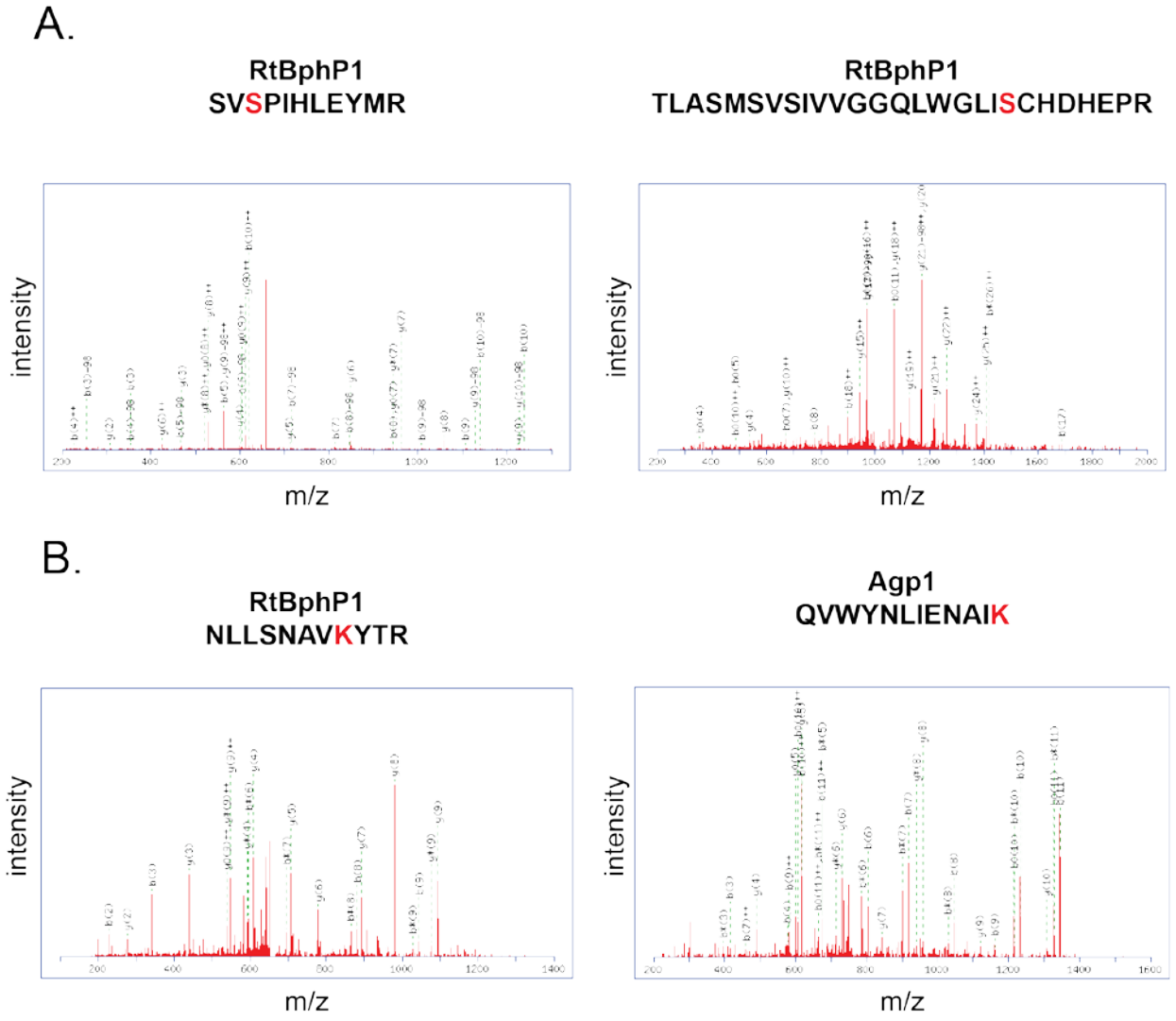


Figure 7. Representative fragmentation spectra for BphP peptides with post-translational modifications. Fragment ions are labeled with the amino acid position where peptide bond cleavage occured (number in parentheses), on which side of the cleavage charge was retained (b if N-terminus reatined charge, y if C-terminus retained charge), and the charge state (no notation for singly charged, ++ for doubly charged) A. Fragmentations of peptides from RtBphP1 with phosphoserine modifications. B. Fragmentations of peptides from RtBphP1 and Agp1 with acetyllysine modifications. Modified residues are highlighted in red.

Table 1. MASCOT analysis of trypsin digested and fragmented peptides from RtBphP1 and Agp1 and assignment of modified amino acids. RtBphP1 (two replicates) and Agp1 were digested with trypsin and subjected to tandem mass spectrometry. Peptides were assigned to sequence fragments with MASCOT and hits with ions score > 25 are reported. The ions score for an MS/MS match is based on the calculated probability, P, that the observed match between the experimental data and the database sequence is a random event. The reported score is -10Log(P). Expect value is the number of times one could expect to get this score or better by chance. Modified residues are highlighted in red.

# detected	amino acids	peptide sequence	missed cleavages	expected mass	experimental mass	modification	site	ions score	expect value
RtBphP1 technical replicate 1									
1	1-30	-.SGSMNLPPPDLAQCADEPIR-VPGAIQPHGR.L	1	3321.528	3319.5002	Acetyl (Protein N-term)	N-term	33	0.1
1	233 - 243	R.SV**S**PIHLEYMR.N	0	1410.6345	1410.6367	Phospho (ST)	S3	50	0.00055
1	311 - 328	R.**K**LTLEIVSHLANSDATLK.M	1	1996.097	1994.1048	Acetyl (K)	K1	43	0.0016
11	435 - 443	R.**K**QVQASDGR.I	1	1029.52-030.5144	1029.5203-1030.5043	Acetyl (K)	K1	31-69	0.0000037-0.025
1	500 - 516	R.VN**K**ELEAFSYTVSHDLR.A	1	2049.0104	2049.0167	Acetyl (K)	K3	52	0.0007
2	624 - 634	R.NLLSNAV**K**YTR.Q	1	1319.7171-1320.7099	1319.7197-1320.7037	Acetyl (K)	K8	25-53	0.00018-0.12
1	641 - 647	R.IQV**K**AVR.R	1	854.5311	854.5338	Acetyl (K)	K4	36	0.0013
2	658 - 670	R.DNGVGFQM**K**YVGK.L	1	1484.7099-1485.7072	1483.713-1484.697	Acetyl (K)	K9	32-33	0.029-0.035
1	667 - 677	K.YVG**K**LFGVFQR.L	1	1354.737	1354.7398	Acetyl (K)	K4	43	0.0014
2	678 - 695	R.LHQAEDFEGTGIGLASV**K**.R	0	1914.9453-1915.9454	1912.953-1913.9371	Acetyl (K)	K18	29-34	0.035-0.1
2	678 - 696	R.LHQAEDFEGTGIGLASV**K**R.I	1	2070.0391-2070.0458	2069.0542	Acetyl (K)	K18	32-59	0.00011-0.056
RtBphP1 technical replicate 2									
1	247 - 273	R.TLASMSVSIVVGGQLWGLI**S**C-HDHEPR.G	0	3029.4558	3028.43	Phospho (ST)	S20	29	0.098

6	311 - 328	R.KLTLEIVSHLANSDATLK.M	1	1994.1082-1994.1095	1994.1048	Acetyl (K)	K1	27-35	0.0033-0.019
1	312 - 328	K.LTLEIVSHLANSDATLK.M	0	1867.0047	1866.0098	Acetyl (K)	K17	28	0.026
1	435 - 443	R.KQVQASDGR.I	1	1030.518	1030.5043	Acetyl (K)	K1	28	0.044
2	500 - 516	R.VNKELEAFSYTVSHDLR.A	1	2049.015-2049.0174	2049.0167	Acetyl (K)	K3	48-52	0.00086-0.0004
1	624 - 631	R.NLLSNAVK.Y	0	900.503	899.5076	Acetyl (K)	K8	26	0.079
1	624 - 634	R.NLLSNAVKYTR.Q	1	1319.7201	1319.7197	Acetyl (K)	K8	42	0.0014
1	641 - 647	R.IQVKAVR.R	1	854.5342	854.5338	Acetyl (K)	K4	32	0.0031
1	658 - 670	R.DNGVGFQMKYVGK.L	1	1483.7124	1483.713	Acetyl (K)	K9	58	6.70E-05
4	678 - 695	R.LHQAEDFEGTGIGLASVK.R	0	1913.9483-1913.9494	1912.953	Acetyl (K)	K18	28-47	0.074-0.00087
1	678 - 696	R.LHQAEDFEGTGIGLASVKR.I	1	2071.0582	2070.0382	Acetyl (K)	K18	31	0.039
AgP1 technical replicate 1									
2	89 - 96	K.KLDVSAHR.S	1	967.5187-967.5194	966.5247	Acetyl (K)	K1	29-35	0.0051-0.022
1	117 - 126	K.LMGELTSLAK.Y	0	1104.5849	1103.5896	Acetyl (K)	K10	25	0.098
2	455 - 464	K.SFEIWKEQLR.N	1	1376.7072-1376.7075	1376.7088	Acetyl (K)	K6	36-40	0.0033-0.01
3	492 - 501	R.KTEEMADLTR.E	1	1235.5796-1235.5802	1234.5863	Acetyl (K)	K1	35-53	0.00017-0.01
2	576 - 587	R.TQLTLKPVDMQK.V	0	1443.7732-1443.7751	1443.7643	Acetyl (K)	K12	25-28	0.046-0.087
1	624 - 635	R.QVWYNLIENAIK.Y	0	1532.7972	1532.7874	Acetyl (K)	K12	38	0.0052

PTMs are localized to functional regions of GAF, PHY and HK domains

In order to understand the potential impact PTM may have on the function of BphPs, the cumulative MS/MS data from Agp1 and RtBphP1 (Figure 8) were mapped onto a homology model of full-length RtBphP1 (Figure 9). The two sites of serine phosphorylation, S232 and S263, are located in the GAF domain in the vicinity of the biliverdin binding site (Figure 10). S232 is located on an α-helix that hangs over the chromophore, and the affected serine is adjacent to the biliverdin A-ring and the site of the covalent linkage between the protein and chromophore. S263 is located on a β-sheet that forms the back wall of the chromophore binding pocket, and is near the D-ring of biliverdin that undergoes a flip in response to red light sensing. Two acetylated lysine sites, K432 and K451, localize near the region of the PHY domain called the tongue, which makes contact with the biliverdin binding pocket and refolds in response to red light sensing (Figure 10). The HATPase_C domain, the catalytic core of the histidine kinase region is acetylated on six sites (Figure 11). Three acetylations are on lysines (K628, K663 and K667) that are on the surface that would come into contact with the long helix of the adjacent protomer in the process of transferring hydrolyzed phosphate to H510 during the autophosphorylation reaction. The aforementioned K628, which was detected as acetylated in both RtBphP1 and Agp1, is positioned just below the ATP binding site of the catalytic core.

Figure 8. Amino acid sequence alignment of BphPs used in this study and sites of post translational modification. Alignment was performed on RtBphP1, RtBphP2, Agp1, and DrBphP with ClustalOmega. Residues identified as phosphorylated or acetylated are highlighted in red.

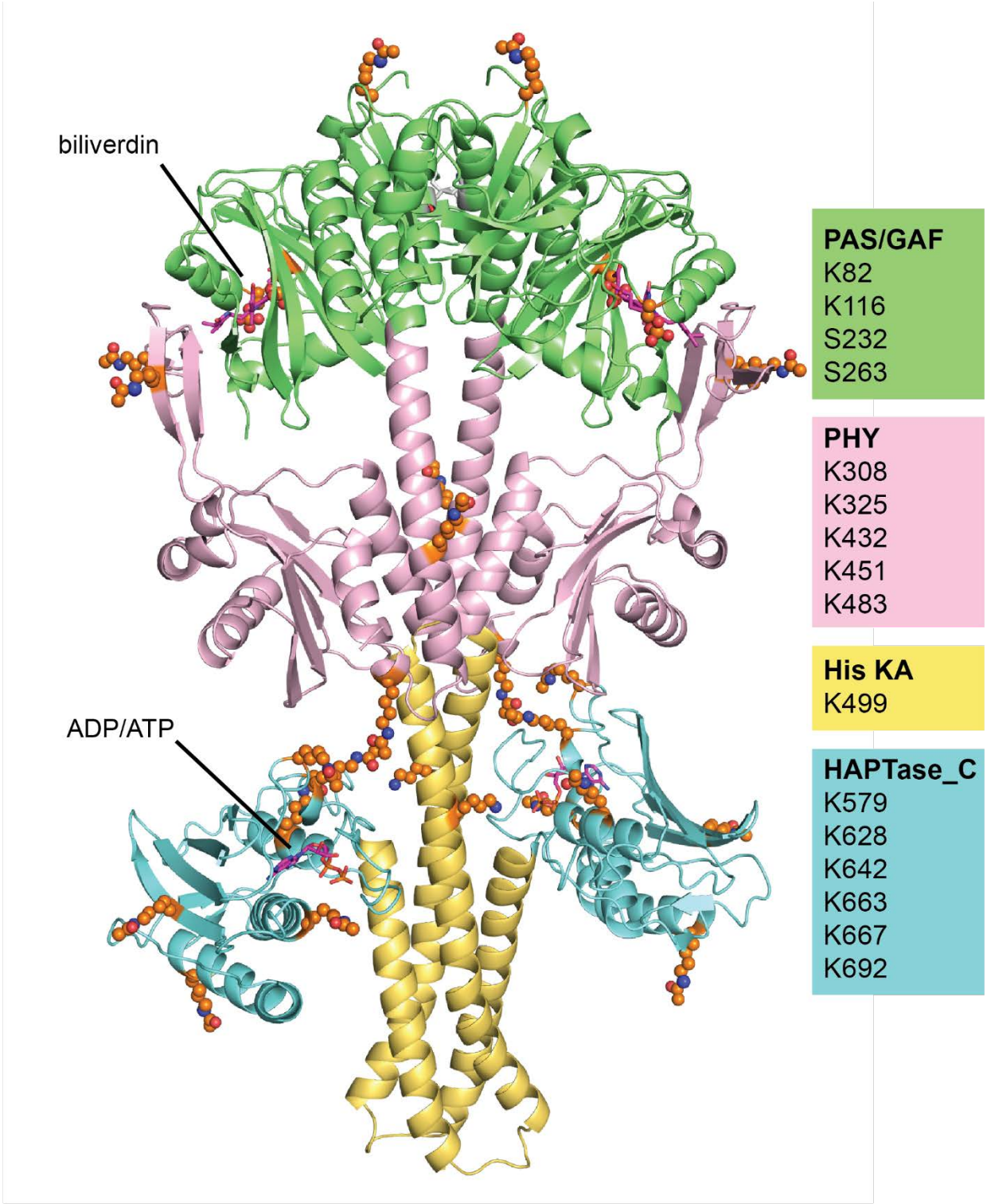


Figure 9. Homology model of full length RtBphP1 with post translational modifications. The RtBphP1 model was based on PDB structures 4O0P (PAS, GAF and PHY domains) and 2C2A (HisKA and HATPase_C domains) and assembled in pymol. Sites of PTM are depicted as ball and stick side chains with phospho- and acetyl- groups attached. Each PTM site is listed in the RtBphP1 numbering scheme, although some PTMs were detected on Agp1. Biliverdin and ADP ligands are depicted as magenta stick models.

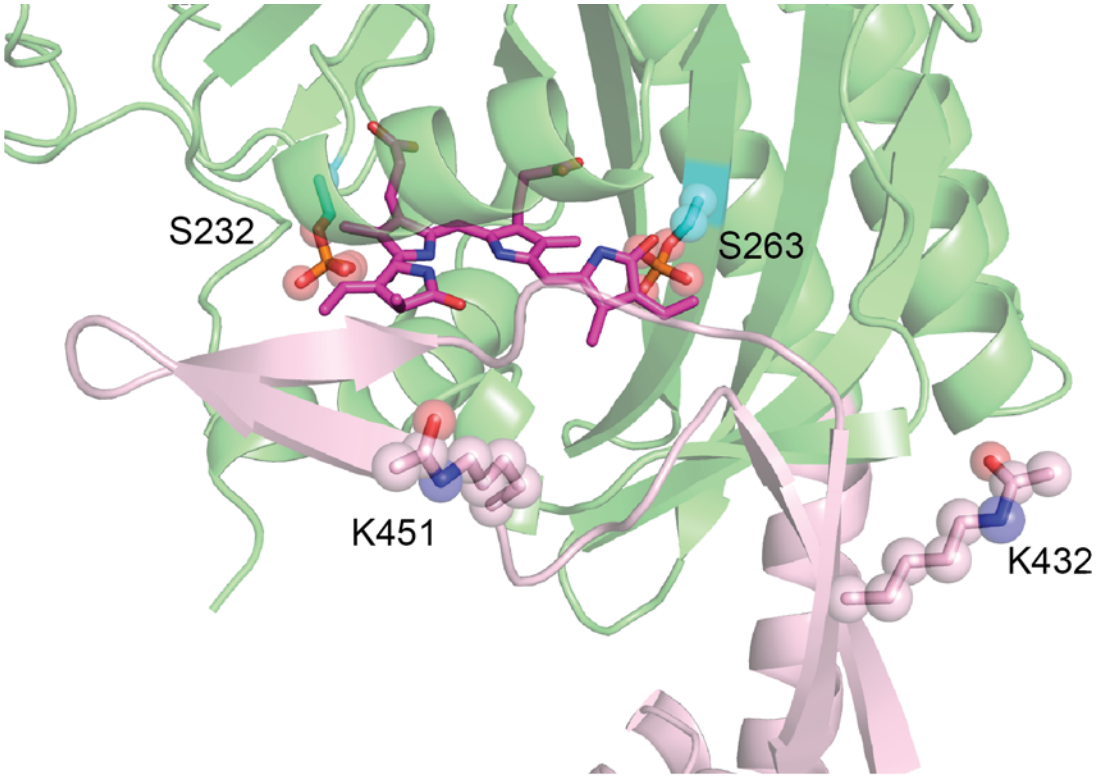


Figure 10. Homology model of RtBphP1 GAF and PHY domain post translational modifications. PTMs detected in this study with possible consequences on red light sensing and signal transduction include S232~P, S263~P, K432~Ac, and K451~Ac. The phosphorylated serines flank biliverdin in its binding pocket and may effect chromophore ligation and photoconversion. K451~Ac and K432~Ac in the PHY domain may influence signal transduction between the light sensing region and catalytic region of the protein.

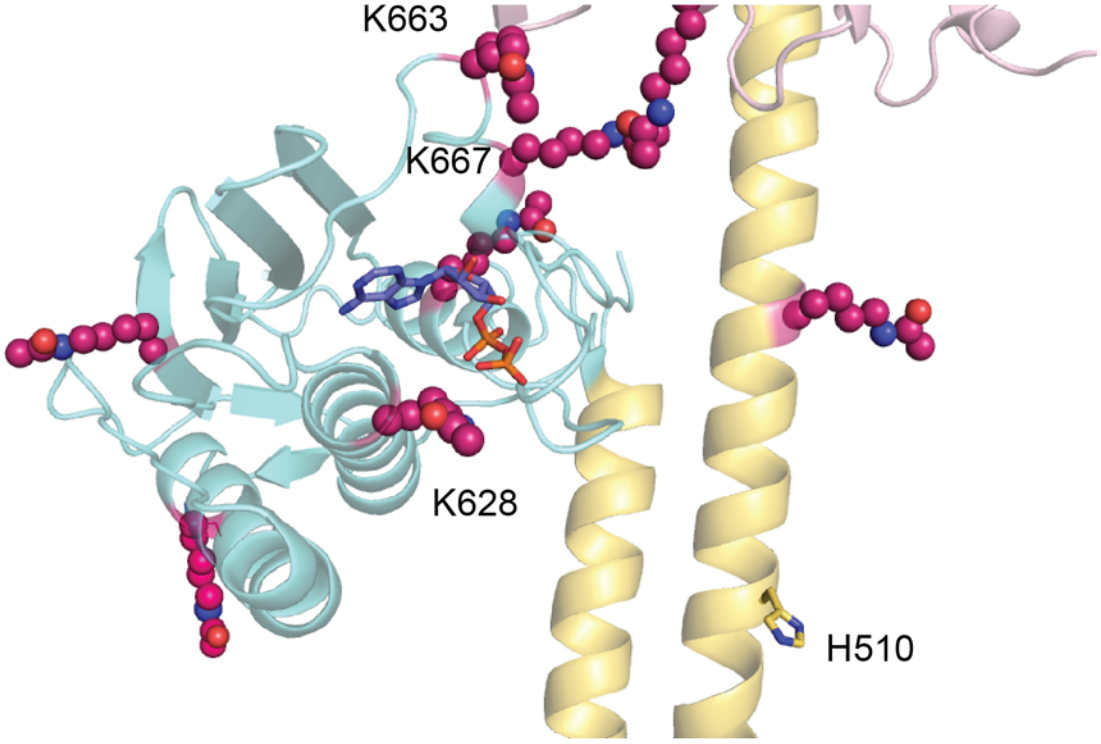


Figure 11. Homology model of RtBphP1 HK domains and post translational modifications. PTMs detected in this study which may impact ATP binding, catalysis, and autophosphorylation include K628~Ac, K663~Ac, and K667~Ac. K628, which was detected as acetylated in both RtBphP1 and Agp1 has a known role in stabilizing ATP in its binding pocket in histidine kinase proteins, and its modification may influence the affinity of BphPs for ATP. K628, K663, and K667 are all situated on the surface of the catalytic HATPase_C domain that must move to contact H510 of the adjacent protomer during autophosphorylation. Modification of these lysines may impact the rate of BphP autophosphorylation, and consequently, phosphotransfer to BRR.

Discussion

This work began with an observation that full-length BphPs run as multiple bands on Phos-tag affinity gels. The Phos-tag system was used to characterize the phosphotransfer activity between RtBphP1 and RtBRR successfully (see Chapter 3), and adapting this assay to evaluate autokinase activity of BphPs without a radioactive label was an attractive possibility. However, given that BphPs which have not yet been exposed to the ATP phosphodonor already run as multiple bands, this excludes the possibility of tracking the appearance of H510 phosphorylation without extensive optimization. However, the investigation into the probable causes of BphP isoforms on Phos-tag gels yielded the novel finding that full-length BphPs recombinantly expressed in *E. coli* contain phosphorylated serine and acetylated lysine residues, presumably in addition to the expected histidine phosphorylation that is required for signal transduction in the TCS. Phos-tag gels of full-length BphP isoforms also suffered in general from poor resolution, due to the relatively high molecular weight of BphPs, which could only be partially compensated for by using a low concentration of acrylamide in the gels (Kinoshita, 2009). As such, reproducibility of the Phos-tag data was limited, and other methods such as Western blots and tandem mass spectrometry were employed later in the study.

Protein isoforms were detected on Phos-tag gels for four different full-length BphPs: RtBphP1 and RtBphP2 from *R. tataouinensis*, Agp1 from *A. tumefaciens*, and DrBphP from *D. radiodurans*. RtBphP1 and Agp1 consistently ran as three distinct bands, whereas RtBphP2 and DrBphP were not reproducibly well resolved enough to determine how many isoforms were present. In contrast, when truncated BphPs which contained only the PAS and GAF domains were run on Phos-tag gels, a single diffuse band was detected both in the case of RtBphP1 PAS-GAF and DrBphP PAS-GAF. The observation that BphP isoforms are lacking in the truncated BphPs suggests that when those limited domains are expressed in the *E. coli* system, they lack the post translational modifications present in the full-length proteins. This could be

because the sites of PTM causing the isoform bands are lacking in these truncated proteins, and are only present in the PHY and HK domains. Another interpretation is that the lack of PHY and or HK domains does not allow modification of the PAS or GAF domains to proceed as in the case of the full length protein. The apparent lack of PTM on the PAS-GAF domains may explain why no studies on BphPs to date have reported phosphoserine or acetyl lysine observations, as most structural and fluorescence studies on these proteins have focused on PAS-GAF or PAS-GAF-PHY constructs (Wagner, 2005; Yang, 2007; Yang, 2008; Essen, 2008; Takala, 2014). Since multiple crystal structures of truncated BphPs exist in the protein data bank, the electron density maps of several of these (PDB: 2O9C, 3IBR, 4IJG, 4O0P, 4Q0J, and 4R6L) were examined for the presence of a representative PTM detected in the full-length BphPs. All of the truncated BphP structures examined lacked density for a phosphate group in the vicinity of Ser 232 which was identified as phosphorylated in full length RtBphP1, indicating that a significant portion of the BphP molecules in those crystal lattices were not phosphorylated at that site. This is in agreement with the Phos-tag gel profile of the PAS-GAF constructs of RtBphP1 and DrBphP (Figure 2B), and further supports that truncated BphPs lack unexpected PTMs. No crystal structures for full length PAS-GAF-PHY-HK BphPs yet exist in the literature, therefore structural evidence for PTMs on the sites identified in this study are currently absent.

Since RtBphP1 was known to bind biliverdin to only 78% of molecules (see Chapter 2) and because previous work has shown that apo and holo forms of BphP can be separated by methods such as hydrophobic interaction chromatography, it was important to investigate the possibility that the Phos-tag reagent had differential affinity for biliverdin bound molecules of BphP and thus was separating apo-protein and holo-protein as isoforms. The full length DrBphP was expressed in *E. coli* BL21(DE3) as both apoprotein and co-expressed with heme oxygenase, which supplies the biliverdin chromophore. Some apoprotein was incubated after cell lysis with an excess of BV. Of the three conditions, no DrBphP sample ran as a single band

on Phos-tag gels, indicating that whether BV is present or absent, isoforms of BphP persist. When the isoform bands were visualized for fluorescence of zinc, which is chelated by biliverdin, all of the isoform bands were visible in the exogenous BV sample, indicating that all of the isoforms separated by Phos-tag contain some bound BV. A zinc signal for the heme oxygenase coexpressed DrBphP was unexpectedly absent, but could be due to poor expression of the heme oxygenase protein, a low abundance of the heme precursor in the expression cells, or poor incorporation of the chromophore *in vivo*. The general phenomenon of slower migration of BV-bound BphPs seen on standard SDS-PAGE gels, and consequently on Phos-tag gels, is unsurprising, given that the slightly greater mass and charge difference added to the protein by BV. Due to the presence of isoforms both in the presence and absence of BV, it was concluded that the isoforms of BphP seen on Phos-tag gels were not due to affinity of the reagent for biliverdin-bound BphPs, and the possibility of alternative PTMs was further explored with methods other than Phos-tag.

Lysine acetylation was detected by Western blot on two different BphPs: RtBphP1 and Agp1. MS/MS confirmed the presence of lysine acetylation on RtBphP1 and Agp1 and determined the specific lysine residues that were acetylated. Serine phosphorylation was detected by Western blot on two of three BphPs studied: RtBphP1 and Agp1. A possible pitfall of the Western Blot imaging technique employed here involves the detection of AlexaFluor 647 secondary antibodies with 685 nm excitation/700 nm emission wavelengths on a LiCor scanner. These wavelengths overlap with the excitation/emission fluorescence window of BphPs, therefore positive signal in the Western Blot images may be interpreted as antibody-conjugate fluorescence or as BphP inherent fluorescence. Preliminary dot-blot analysis of BphPs indicates that under the relatively gentle treatment of proteins for dot blot, BphPs retain their fluorescence relative to the amount of biliverdin bound (data not shown). It is unknown whether BphPs denatured during electrophoresis and transfer to nitrocellulose membranes also retain their

inherent fluorescence. In light of this, the development of alternative imaging methods for these Western Blots, such as HRP chemiluminescence, is desired for future studies.

MS/MS confirmed the presence of serine phosphorylation only in RtBphP1 and determined the specific sites of serine phosphorylation. Given that anti-phosphoserine antibodies bind to Agp1 purified protein, why were phosphorylated serines not detected by MS/MS? Phosphopeptide detection by mass spectrometry suffers from low detection rates due to the low abundance of phosphopeptides in biological samples and because peptides with negatively charged phosphate groups do not ionize as easily as non-phosphorylated peptides (Wysocki, 2005). Strategies to overcome this detection problem include ensuring full or nearly complete coverage of the amino acid sequence, enrichment of the sample for phosphopeptides (McLachin, 2001), and chemical modification of phosphorylated residues to a more easily ionized and detected chemical group (Arrigoni, 2006). The MS/MS data sets for RtBphP1 and Agp1 resulted in 96% and 86% sequence coverage, respectively, with the site corresponding to S263 in Agp1 not being covered. The poorer coverage of the Agp1 dataset may explain why phosphoserine sites were not detected. Phosphopeptide enrichment or phosphate modification steps were not employed in this study, which may aid in better detection of phosphoserines on BphPs in future work. Conversely, the possibility exists that some of the PTMs assigned in this study represent false positive results. The MASCOT MS/MS ions search algorithm employs a decoy peptide search of random amino acid strings and scores those masses alongside the experimental masses from each data set (Elias, 2005). The false positive detection rate of decoy peptides factors into the ions and expect value scores reported in Table 1, and high false positive scores contribute to low ions/expect scores. Since the results in Table 1 were filtered to remove low scoring hits, the main threat of false positives in this study would come from real peptides derived from contaminating proteins in the BphP preparations, which happen to have masses matching BphP + PTM peptide masses.

Phosphorylated serines and acetylated lysines were mapped to functional regions of BphPs using a RtBphP1 homology model (Figure 9). Given the novel discovery of these PTMs, no studies of enzyme activity on modified vs. non-modified BphPs have yet been conducted. However, hypotheses about the impact of PTM at specific regions of BphPs that have known structure and function can be generated to direct future experiments.

S232 and S263, both phosphorylated in RtBphP1, are on the interior of the biliverdin binding pocket (Figure 10). S232 is located near the A ring of the chromophore, which bears the vinyl group that forms a thioether bond with C11, a reaction essential to covalent binding of biliverdin to the BphP. In order for the thioether to form, the proper distance and charge environment of the reactive atoms must be maintained. Examination of the high resolution structure of DrBphP PAS-GAF (PDB 2O9C; Wagner, 2007) demonstrates that S232 normally comes within hydrogen bonding distance of Glu14 (2.7 A) and this interaction assists in bringing the N-terminal region of the PAS domain, and C11 into proximity of the biliverdin A-ring vinyl group. The modification of S232 with a phosphate would preclude the hydrogen bond between E14 and S232, and possibly sterically inhibit the N-terminus and C11 from reaching the A-ring of BV, which would have negative consequences to biliverdin binding to BphP. Therefore, a hypothesis that BphPs with unmodified S232 will bind biliverdin more efficiently than BphPs with S232~P is proposed. S263 is on the interior of the biliverdin binding pocket, positioned behind the D-ring of the chromophore. In its nonmodified state, this serine would not block the rotating D-ring as it undergoes photoconversion from the Pr to Pfr state, or as the D ring returns to the Pr state from Pfr. Once phosphorylated however, the added bulk and negative charge of the phosphate group could negatively impact the freedom of D-ring rotation. Therefore, a hypothesis that BphPs with S263~P will not undergo photoconversion or thermal reversion normally as unmodified BphPs is proposed.

Lysines 432 and 451 were detected as acetylated in RtBphP1 and Agp1, respectively, and these two residues map to critical regions of the PHY domain (Figure 10). K432 is positioned on a dynamic loop of the PHY domain that makes close contact with the chromophore binding region of the GAF domain. The loop forms a lid over the open portion of the BV binding pocket and is thought to be important in sensing the initial conformational changes that take place in the chromophore and GAF domain during photoconversion and transmit the light signal through the rest of the protein (Yang, 2008; Essen, 2008; Takala, 2014). The corresponding residue in the two-state crystal structure is a leucine, and unfortunately this portion of the loop could only be modeled for the dark state (Takala, 2014). Therefore, the position of the leucine, or a substituted lysine, cannot be known in the illuminated state. However, acetylation of the lysine could have potential effects on conformational change that results in regulation of autophosphorylation in response to light. K451 in the BphP dark state model is on the β hairpin that is known to refold into an α-helix after the transition to the illuminated state (Stojkovic, 2014). The implications of acetylation on this lysine are difficult to predict, because the corresponding residue in DrBphP, which was crystallized in both states, is not lysine but leucine. However, this residue is part of both the dark state beta sheet and light state alpha helix, so PTM at this site could have an effect on refolding and resulting signal transduction to the HK domains of BphP. Acetylation at either K432 or K451 could manifest as a lessening in red light control of the autophosphorylation rate of the BphP, despite chromophore binding and normal spectral photoconversion behavior.

Six lysine residues on the HATPase_C domain, the site of ATP binding and catalysis, were detected as acetylated in RtBphP1 and Agp1 (Figure 11). The only acetylated lysine common to both the RtBphP1 and Agp1 datasets is K628, which is positioned below the ATP binding site. Mutations at this conserved lysine in other HK proteins has been linked to decreased ATP binding, due to the stabilizing effect of lysine on the adjacent tyrosine, which

participates in a stacking interaction with ATP and because lysine participates in a direct interaction with the γ phosphate of ATP (Marina, 2001; Marina 2005). Therefore, lower affinity for ATP in BphPs with acetyl-K628 compared to nonmodified BphP is proposed as a potential effect of acetylation. The Compendium of Protein Lysine Modifications database (Liu, 2013) contains records for six acetylated histidine kinase proteins from bacteria, including *E. coli*, *R. palustris*, and *Desulfovibrio vulgaris*. None of these acetylations are on a lysine homologous to RtBphP1 K628, therefore this is the first report of histidine kinase acetyl modification at this ATP-binding proximal site. Other potentially impactful acetylations are on K663 and K667 which are positioned on the surface of the HATPase_C domain that would contact the His KA domain in order to transfer the hydrolyzed phosphate group to H510. Acetylation of these lysines could negatively impact the rate of autophosphorylation, even in the dark, more active state.

Most of the BphPs characterized so far in the literature have been purified from *E. coli* recombinant systems Yeh, 1997; Karniol, 2003; Giraud, 2005; Jaubert, 2008; Kumar, 2012). In this study, four different full length BphPs exhibited characteristics of unexpected posttranslational modification beyond the expected histidine phosphorylation, irrespective of the chromophore binding state. Therefore, it is reasonable to suspect that previous studies on BphP enzymatic activities have been carried out on proteins with similar unexpected PTMs. This warrants future experimenters to carefully characterize the following activities on both modified and non-modified BphPs: biliverdin binding, photoconversion, ATP binding, autophosphorylation, and phosphotransfer. Several strategies for producing non-modified BphPs exist, including enzymatic dephosphorylation and deacetylation, expression of BphPs in serine kinase/acetyltransferase deficient strains of *E. coli*, or *in vitro* translation of proteins. Another question that remains to be explored is whether the PTMs discovered in these recombinantly expressed BphPs also occur in the organisms that utilize BphPs for environmental sensing and signaling, and if these PTMs are regulatory in nature. Strategies to

explore this possibility include purification of BphPs from their native organisms, or Phos-tag separation of cell lysates from these native organisms, followed by BphP-specific antibody probing. The experiments presented here will form the foundation toward answering these important, new questions in the BphP field.

Methods

Recombinant expression and purification of BphPs

Full-length RtBphP1 and RtBphP1 PAS-GAF were purified from *E. coli* BL21(DE3) as previously described (Baker, 2016) with biliverdin added after cell lysis. Agp1 was purified as described with biliverdin added after cell lysis (Scheerer, 2003); this protein was purified in the laboratory of Patrick Scheerer at Charité-University Medicine Berlin. RtBphP2 was purified from *E. coli* BL21(DE3) and expressed from a pBADHisB vector (constructed by Gilles DeLuca, Andre Vermeglio, and Thierry Heulin, Cadarache, France) carrying genes for both *R. tataouinensis* RtBphP2 and *hmuO* from *Bradyrhizobium* sp. BTAI1. 1L of culture with 25 µg/mL ampicillin was incubated at 37° C shaking at 150 rpm until it reached an OD_{600} of 0.5, then protein expression was induced with addition of 0.01% arabinose. Induced cultures grew overnight at 30° C before resuspension in lysis buffer (10 mM HEPES pH 7.0, 500 mM NaCl, 20 mM imidazole) with addition of DNAseI before lysis in a French press. Crude lysate was clarified by centrifugation for 30 minutes at 7,000 g at 4° C. Nickel column purification was carried out by equilibrating a 5 mL column with lysis buffer, binding the soluble cell lysate to the resin, and eluting with elution buffer (10 mM HEPES pH 7.0, 500 mM NaCl, 500 mM imidazole). Purified RtBphP2 was transferred to a storage buffer of 25 mM Tris HCl pH 7.0, 500 mM NaCl using a 50,000 MWCO centrifugation filter. Full length DrBphP and DrBphP PAS-GAF were expressed in *E. coli* BL21(DE3) with and without an additional pRSF-DUET plasmid with the *hmuO* gene in 5 mL cultures of LB with 25 µg/mL ampicillin (and 50 µg/mL kanamycin if PRSF-DUET-*hmuO* was present) and purified using the Protein miniprep protocol (The Qiaexpressionist Handbook, Fifth Ed. P. 85).

Phos-tag acrylamide gel separation of BphP isoforms

Purified full length BphPs and truncated BphPs were thawed on ice from -80° storage before the addition of 3X SDS loading buffer (1.5 M tris pH 6.8, 50% glycerol, 15% SDS, 0.75%

bromophenol blue, 0.75% β-mercaptoethanol) before immediately loading onto 6% (full length BphPs) and 10% (PAS-GAF BphPs) SDS-PAGE gels with 100 μM Phos-tag acrylamide (Wako Pure Chemicals, Osaka, Japan). Gels were run at 4° C until the dye front ran off, and then stained overnight with SYPRO Ruby before imaging on a GE Typhoon FLA-9000 (GE Healthcare Life Sciences, Pittsburgh, PA) with 632 nm laser excitation and LPR(red) filter.

Phosphotransfer reactions for Phos-tag acrylamide gels contained 5 μM RtBphP1 plus 15 μM of RR in kinase buffer with or without 1 mM ATP and were incubated for 10 or 30 minutes before the addition of 3X SDS loading buffer to stop the reaction. Potentially phosphorylated proteins were placed on ice after the reactions were stopped and separated on 6% and 12% SDS-PAGE gels with 100 μM Phos-tag acrylamide within one hour. Gels were run at 4° C until the dye front ran off, and then stained overnight with SYPRO Ruby before imaging on a GE Typhoon FLA-9000 with 632 nm laser excitation and LPR(red) filter.

Zinc fluorescence of BphPs

Standard SDS-PAGE or Phos-tag acrylamide gels were cast with the addition of 1 mM zinc acetate and electrophoresis of protein samples was carried out used a 1X SDS running buffer with the addition of 1 mM zinc acetate. Gels were visualized on a GE Typhoon FLA-900 with 532 nm laser excitation and Cy3 (orange) filter. After zinc fluorescence imaging gels could be stained for protein as usual.

Western blot of full length BphPs

Purified full length RtBphP1, RtBphP2, and Agp1 were thawed on ice from storage at -80° C and mixed with an equal volume of 2X SDS loading dye (1M tris pH 6.8, 50% glycerol, 10% SDS, 0.5% bromophenol blue, 0.5% β-mercaptoethanol) before loading 1 μg of protein on a 15% SDS-PAGE gel. Gels were run until the dye front ran off and proteins were transferred to a nitrocellulose membrane in transfer buffer (25 mM tris pH 8.3, 190 mM glycine, 20%

methanol). Membranes were blocked with 5% milk in 1X PBS and probed with either rabbit anti-phosphoserine polyclonal antibodies (1:5000 dilution) or rabbit anti acetyl lysine antibodies (1:5000 dilution) (Sigma Aldrich, St. Louis, MO) followed by probing with goat anti-rabbit antibodies conjugated to AlexaFluor 647 (Thermo Fisher Scientific, Waltham, MA). Membranes were visualized on a LiCor Odyssey CLX system (LiCor, Lincoln, NE) on the 700 nm channel. Duplicate SDS-PAGE gels were not transferred but instead stained with Coomassie blue and scanned.

Tandem mass spectrometry and analysis

Full length RtBphP1 and Agp1 were submitted to the University of Wisconsin Biotechnology Center Mass Spectrometry/Proteomics facility for trypsin digestion and subsequent processing on a Thermo Scientific LTQ Orbitrap XL (Thermo Scientific, Waltham, MA). Trypsin-cut peptides were ionized by electrospray in positive mode and peptides were fragmented by collision induced dissociation. Output files of m/z and intensities were analyzed by MASCOT MS/MS Ions search (Matrix Science, Boston, MA) for fixed modifications (carbamidomethylation) and variable modification (deamidation, oxidation of methionines, acetylation of lysines, acetylation of protein N-terminus, and phosphorylation of serine/threonine). Results for peptide assignments were filtered for table 1 by limiting the minimum Ions score to 25.

Sequence alignments and homology modeling

The amino acid sequences of RtBphP1, RtBphP1, Agp1, and DrBphP were aligned with ClustalOmega (Sievers, 2011).

Homology models were generated with iTasser (Roy, 2010) and visualized with pyMOL (Schroedinger LLC, Cambridge, MA). The PAS, GAF, and PHY portions of the RtBphP1 amino acid sequence was modeled based on the dark state structure of DrBphP (PDB 40OP) (Takala,

2014). The HisKA and HATPase C domain sequences of RtBphP1 were modeled based on TM0853 from *Thermotoga maritima* (PDB 2C2A) (Marina, 2005). Post translational modifications were added to the model with the plugin PyTMs (Warnecke, 2014).

References

1. **Arrigoni G**, **Resjö S**, **Levander F**, **Nilsson R**, **Degerman E**, **Quadroni M**, **Pinna LA**, **James P**. 2006. Chemical derivatization of phosphoserine and phosphothreonine containing peptides to increase sensitivity for MALDI-based analysis and for selectivity of MS/MS analysis. Proteomics **6**:757–766.

2. **Baker AW**, **Satyshur KA**, **Moreno Morales N**, **Forest KT**. 2016. Arm-in-arm response regulator dimers promote intermolecular signal transduction. J Bacteriol **198**:1218–1229.

3. **Barak R**, **Eisenbach M**. 2001. Acetylation of the response regulator, CheY, is involved in bacterial chemotaxis. Mol Microbiol **40**:731–743.

4. **Bornschlögl T**, **Anstrom DM**, **Mey E**, **Dzubiella J**, **Rief M**, **Forest KT**. 2009. Tightening the knot in phytochrome by single-molecule atomic force microscopy. Biophys J **96**:1508–1514.

5. **Crosby HA**, **Pelletier DA**, **Hurst GB**, **Escalante-Semerena JC**. 2012. System-wide studies of N-lysine acetylation in Rhodopseudomonas palustris reveal substrate specificity of protein acetyltransferases. J Biol Chem **287**:15590–15601.

6. **Elias JE**, **Haas W**, **Faherty BK**, **Gygi SP**. 2005. Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. Nat Methods **2**:667–675.

7. **Essen L-O**, **Mailliet J**, **Hughes J**. 2008. The structure of a complete phytochrome sensory module in the Pr ground state. Proc Natl Acad Sci U S A **105**:14709–14714.

8. **Gallego M**, **Virshup DM**. 2007. Post-translational modifications regulate the ticking of the circadian clock. Nat Rev Mol Cell Biol **8**:139–148.

9. **Georgellis D**, **Lynch AS**, **Lin ECC**, **Lin ECC**. 1997. In vitro phosphorylation study of the arc two-component signal transduction system of Escherichia coli **179**:5429–5435.

10. **Giraud E**, **Zappa S**, **Vuillet L**, **Adriano JM**, **Hannibal L**, **Fardoux J**, **Berthomieu C**, **Bouyer P**, **Pignol D**, **Vermiglio A**. 2005. A new type of bacteriophytochrome acts in tandem with a classical bacteriophytochrome to control the antennae synthesis in Rhodopseudomonas palustris. J Biol Chem **280**:32389–32397.

11. **Hu CW**, **Lin MH**, **Huang HC**, **Ku WC**, **Yi TH**, **Tsai CF**, **Chen YJ**, **Sugiyama N**, **Ishihama Y**, **Juan HF**, **Wu SH**. 2012. Phosphoproteomic analysis of rhodopseudomonas palustris reveals the role of pyruvate phosphate dikinase phosphorylation in lipid production. J Proteome Res **11**:5362–5375.

12. **Jaubert M**, **Lavergne J**, **Fardoux J**, **Hannibal L**, **Vuillet L**, **Adriano JM**, **Bouyer P**, **Pignol D**, **Giraud E**, **Vermeglio A**. 2007. A singular bacteriophytochrome acquired by lateral gene transfer. J Biol Chem **282**:7320–7328.

13. **Karniol B**, **Vierstra RD**. 2003. The pair of bacteriophytochromes from *Agrobacterium tumefaciens* are histidine kinases with opposing photobiological properties. Proc Natl Acad Sci USA **100**:2807–2812.

14. **Kim D**, **Yu BJ**, **Kim JA**, **Lee YJ**, **Choi SG**, **Kang S**, **Pan JG**. 2013. The acetylproteome of Gram-positive model bacterium Bacillus subtilis. Proteomics **13**:1726–1736.

15. **Kinoshita E**. 2005. Phosphate-binding Tag, a New Tool to Visualize Phosphorylated Proteins. Mol Cell Proteomics **5**:749–757.

16. **Kinoshita E**, **Kinoshita-Kikuta E**, **Koike T**. 2009. Separation and detection of large phosphoproteins using Phos-tag SDS-PAGE. Nat Protoc **4**:1513–21.

17. **Kumar S**, **Kateriya S**, **Singh VS**, **Tanwar M**, **Agarwal S**, **Singh H**, **Khurana JP**, **Amla DV**, **Tripathi AK**. 2012. Bacteriophytochrome controls carotenoid-independent response to photodynamic stress in a non-photosynthetic rhizobacterium, Azospirillum brasilense Sp7. Sci Rep **2**:872.

18. **Liu Z**, **Wang Y**, **Gao T**, **Pan Z**, **Cheng H**, **Yang Q**, **Cheng Z**, **Guo A**, **Ren J**, **Xue Y**. 2014. CPLM: A database of protein lysine modifications. Nucleic Acids Res **42**:531–536.

19. **Macek B**, **Gnad F**, **Soufi B**, **Kumar C**, **Olsen J V**, **Mijakovic I**, **Mann M**. 2008. Phosphoproteome Analysis of E . coli Reveals Evolutionary Conservation of Bacterial Ser / Thr / Tyr Phosphorylation. Mol Cell Proteomics **7**:299–307.

20. **Macek B**, **Mijakovic I**, **Olsen J V**, **Gnad F**, **Kumar C**, **Jensen PR**, **Mann M**. 2007. The serine/threonine/tyrosine phosphoproteome of the model bacterium Bacillus subtilis. Mol Cell Proteomics **6**:697–707.

21. **Marina A**, **Mott C**, **Auyzenberg A**, **Hendrickson WA**, **Waldburger CD**. 2001. Structural and mutational analysis of the PhoQ histidine kinase catalytic domain. Insight into the reaction mechanism. J Biol Chem **276**:41182–41190.

22. **Marina A**, **Waldburger CD**, **Hendrickson WA**. 2005. Structure of the entire cytoplasmic portion of a sensor histidine-kinase protein. EMBO J **24**:4247–4259.

23. **McLachlin DT**, **Chait BT**. 2001. Analysis of phosphorylated proteins and peptides by mass spectrometry. Curr Opin Chem Biol **5**:591–602.

24. **Mo R**, **Yang M**, **Chen Z**, **Cheng Z**, **Yi X**, **Li C**, **He C**, **Xiong Q**, **Chen H**, **Wang Q**, **Ge F**. 2015. Acetylome analysis reveals the involvement of lysine acetylation in photosynthesis and carbon metabolism in the model cyanobacterium synechocystis sp. PCC 6803. J Proteome Res **14**:1275–1286.

25. **Okanishi H**, **Kim K**, **Masui R**, **Kuramitsu S**. 2013. Acetylome with Structural Mapping Reveals the Signi fi cance of Lysine Acetylation in Thermus thermophilus. J Proeteome Res **12**:3952–3968.

26. **Pappin DJC**, **Creasy DM**, **Cottrell JS**, **Perkins DN**. 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis **20**:3551–67.

27. **Psakis G**, **Mailliet J**, **Lang C**, **Teufel L**, **Essen LO**, **Hughes J**. 2011. Signaling kinetics of cyanobacterial phytochrome Cph1, a light regulated histidine kinase. Biochemistry **50**:6178–6188.

28. **Qiagen**. 1991. The QIA expressionist Handbook, 5th Ed.Diagen GmbH, Diisseldorf, Germany.

29. **Ravichandran A**, **Sugiyama N**, **Tomita M**, **Swarup S**, **Ishihama Y**. 2009. Ser/Thr/Tyr phosphoproteome analysis of pathogenic and non-pathogenic Pseudomonas species. Proteomics **9**:2764–2775.

30. **Roy A**, **Kucukural A**, **Zhang Y**. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc **5**:725–738.

31. **Scheerer P**, **Michael N**, **Park JH**, **Noack S**, **Förster C**, **Hammam MAS**, **Inomata K**, **Choe HW**, **Lamparter T**, **Krauß N**. 2006. Crystallization and preliminary X-ray crystallographic analysis of the N-terminal photosensory module of phytochrome Agp1, a biliverdin-binding photoreceptor from Agrobacterium tumefaciens. J Struct Biol **153**:97–102.

32. **Schroedinger**. The PyMOL Molecular Graphics System Version 1.8.

33. **Sievers F**, **Wilm A**, **Dineen D**, **Gibson TJ**, **Karplus K**, **Li W**, **Lopez R**, **McWilliam H**, **Remmert M**, **Söding J**, **Thompson JD**, **Higgins DG**. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol **7**:539.

34. **Stojkovic EA**, **Toh KC**, **Alexandre MTA**, **Baclayon M**, **Moffat K**, **Kennis JTM**. 2014. FTIR spectroscopy revealing light-dependent refolding of the conserved tongue region of bacteriophytochrome. J Phys Chem Lett **5**:2512–2515.

35. **Strahl BD**, **Allis CD**. 2000. The language of covalent histone modifications. Nature **403**:41–45.

36. **Swanson R V.**, **Alex LA**, **Simon MI**. 1994. Histidine and aspartate phosphorylation: two-component systems and the limits of homology. Trends Biochem Sci **19**:485–490.

37. **Takala H**, **Björling A**, **Berntsson O**, **Lehtivuori H**, **Niebling S**, **Hoernke M**, **Kosheleva I**, **Henning R**, **Menzel A**, **Ihalainen J a**, **Westenhoff S**. 2014. Signal amplification and transduction in phytochrome photosensors. Nature **509**:245–8.

38. **Thao S**, **Escalante-Semerena JC**. 2011. Control of protein function by reversible Nε-lysine acetylation in bacteria. Curr Opin Microbiol **14**:200–4.

39. **Wagner JR**, **Brunzelle JS**, **Forest KT**, **Vierstra RD**. 2005. A light-sensing knot revealed by the structure of the chromophore-binding domain of phytochrome. Nature **438**:325–331.

40. **Wagner JR**, **Zhang J**, **Brunzelle JS**, **Vierstra RD**, **Forest KT**. 2007. High resolution structure of Deinococcus bacteriophytochrome yields new insights into phytochrome architecture and evolution. J Biol Chem **282**:12298–12309.

41. **Wako Pure Chemical Industries**. 2013. Phos-tag^{TM} PAGE GUIDEBOOK.

42. **Warnecke A**, **Sandalova T**, **Achour A**, **Harris RA**. 2014. PyTMs: a useful PyMOL plugin for modeling common post-translational modifications. BMC Bioinformatics **15**:370.

43. **Westermann S**, **Weber K**. 2003. Post-translational modifications regulate microtubule function. Nat Rev Mol Cell Biol **4**:938–947.

44. **Wysocki VH**, **Resing KA**, **Zhang Q**, **Cheng G**. 2005. Mass spectrometry of peptides and proteins. Methods **35**:211–222.

45. **Yamada S**, **Nakamura H**, **Kinoshita E**, **Kinoshita-Kikuta E**, **Koike T**, **Shiro Y**. 2007. Separation of a phosphorylated histidine protein using phosphate affinity polyacrylamide gel electrophoresis. Anal Biochem **360**:160–162.

46. **Yang X**, **Kuk J**, **Moffat K**. 2008. Crystal structure of Pseudomonas aeruginosa bacteriophytochrome: photoconversion and signal transduction. Proc Natl Acad Sci U S A **105**:14715–14720.

47. **Yang X**, **Stojkovic E a**, **Kuk J**, **Moffat K**. 2007. Crystal structure of the chromophore binding domain of an unusual bacteriophytochrome, RpBphP3, reveals residues that modulate photoconversion. Proc Natl Acad Sci U S A **104**:12571–12576.

48. **Yeh K-C**, **Wu S-H**, **Murphy JT**, **Lagarias JC**. 1997. A Cyanobacterial Phytochrome Two-Component Light Sensory System. Science (80-) **277**:1505–1508.

49. **Zhang J**, **Sprung R**, **Pei J**, **Tan X**, **Kim S**, **Zhu H**, **Liu C-F**, **Grishin N V.**, **Zhao Y**. 2008. Lysine Acetylation Is a Highly Abundant and Evolutionarily Conserved Modification in Escherichia Coli. Mol Cell Proteomics **8**:215–225.

Chapter IV

Bioinformatic analyses of red light signal transduction partners

Gene Clustering Analysis Computational work was performed by Sarah Stevens, Katherine McMahon Laboratory, Department of Bacteriology, University of Wisconsin-Madison

This research was performed using the compute resources and assistance of the UW-Madison Center for High Throughput Computing (CHTC) in the Department of Computer Sciences. The CHTC is supported by UW-Madison, the Advanced Computing Initiative, the Wisconsin Alumni Research Foundation, the Wisconsin Institutes for Discovery, and the National Science Foundation, and is an active member of the Open Science Grid, which is supported by the National Science Foundation and the U.S. Department of Energy's Office of Science.

Abstract:

Putative BphP red light sensors are widespread in phylogenetically diverse bacterial genomes. Despite intensive structural and biochemical characterization of BphP signal transduction systems, relatively little is known about bacterial responses to red light. Additionally, no signal transduction partners downstream of BRRs have been identified for canonical BphP systems. This chapter employs several bioinformatics approaches to generate testable hypotheses about gaps in signal transduction knowledge. Amino acid sequence analyses predict the presence of arm-in-arm BRRs in 102 bacteria with sequenced genomes, and a common soil-dwelling lifestyle is assigned to the majority of these. A targeted survey of ORFs encoded in BphP gene neighborhoods reveals that additional histidine kinase proteins may play a critical role in signal integration along with BphPs and BRRs. A multiple genome clustering search unveils potential signal transduction partners which are common to arm-in-arm species, but lacking in outgroups, including: two uncharacterized proteins; oxidative stress response proteins; DNA repair proteins; and an energy storage polymer synthase. We conclude with a new interaction map for BphP-BRR signaling that will inform future *in vivo* and *in vitro* experiments to elucidate red light signal transduction.

Introduction

The work presented in the previous chapters focused on biochemical investigations of BphP-BRR signal transduction and those *in vitro* methods have yielded important results. However, work done here and by others has not fully addressed important downstream questions about red light signal transduction, such as what partners does BRR interact with after phosphorylation by the BphP and what is the ultimate effect on the cell from red light sensing? In this chapter several bioinformatics approaches are used to predict potential signal transduction partners and phenotypic effects by mining the wealth of bacterial genomic information which is publicly available.

Proteins with common amino acid sequences often adopt common structures, and perform common functions in the cell based on those structures (Hegyi, 1999). The arm-in-arm BRR structures presented in a previous chapter allow the prediction of a sequence motif that corresponds to that structural arrangement (Baker, 2016). The motif acts as a handle to search all bacterial genomes for that coding sequence, and assign likely arm-in-arm BRRs to over a hundred bacterial species. This allows questions about the earliest common ancestor bearing the arm-in-arm BRR, and what if any common lifestyles or environments are associated with the arm-in-arm BRR to be addressed. This analysis reveals that the arm-in-arm is prevalent in certain classes of the Proteobacteria that share a soil-dwelling and/or plant-associated lifestyle.

Biochemical studies of BphP signal transduction has shown that in many bacterial systems, BphP histidine kinase genes and BRR genes, which are typically encoded adjacent to one another on the genome interact to initiate phosphotransfer (Bhoo, 2001; Karniol, 2003; Giraud, 2005; Psakis, 2011; Baker 2016). This linear signal transduction relay fits the classic model of a two-component signal transduction system. Any downstream BRR-interacting molecules are presently unknown, and thus the mechanism for phenotypic adjustment to the red light environment is poorly understood. In select cases, red light signaling is more complex than

a simple linear relay. In *R. palustris* two BphPs, RpBphP2 and RpBphP3 can both phosphorylate a single domain response regulator, rpa3017 (Giraud, 2005). In *Rhizobium* sp. NT-26, the BphP can phosphorylate its adjacently encoded BRR and can also phosphorylate the N-terminal receiver domain of a hybrid histidine kinase encoded downstream of the BRR (Wojnowska, 2013). Thus, convergent or branched signaling pathways possibly exist for other BphP-BRR systems. Importantly, the partners which interact in these more complex red light phosphorelays are encoded in the same genomic neighborhood in their respective species. In order to discover additional signal transduction partners and possible mechanistic effects of red light sensing, an extensive survey of gene products present in BphP gene neighborhoods was undertaken. Further, the study sought to determine if similar or different common gene neighbors were present near BphPs with arm-in-arm BRRs and BphPs with other BRR types. The findings identified hybrid histidine kinase genes in the majority of BphP gene neighborhoods that also contain a BRR and suggest the branched phosphorelay model of signal transduction may be more common than previously known.

Although signal transduction partners are often encoded in the same gene neighborhoods, the previous approach would miss potential signal transduction partners encoded elsewhere on the genome that are part of diverse cellular systems. BphP sensing and signaling have been connected to diverse phenotypes in several characterized bacterial systems. These physiological responses include peripheral light harvesting system in phototropic bacteria (Hubschmann, 2005; Giraud 2005, Jaubert 2008), carotenoid dependent (Davis, 1999) and independent stress responses (Kumar 2012; Barkovitz 2011, 2008), and motility (Oberpichler, 2008; Wu 2013) among others. The genes that control these varied responses, and other undiscovered responses, cannot plausibly be localized with the BphP gene cassette in every species. The goal in this study was to consider the entire genomes of nearly 100 different BphP-encoding species and to discover commonly present and absent

genes related to BphP-BRR signaling. By inputting a large number of carefully selected bacterial genomes into a gene clustering computational pipeline (Altschul, 1990; Dongen, 2000), we were able to address the question what gene products are present in arm-in-arm BRR containing species, but lacking in species without these proteins? Any hits which are often present alongside arm-in-arm BRRs represent potential signaling partners that could be encoded anywhere on the genome. This powerful technique yielded several interesting and novel gene products which can be studied in the laboratory for their role in a red light signaling pathway. Several stress response proteins were identified in agreement with other studies. In addition, two conserved uncharacterized proteins were also identified as potential BRR interaction partners.

Mining the wealth of information contained in the thousands of sequenced and publicly available bacterial genomes has been productive in generating rich, testable hypotheses for future work in BphP-BRR signal transduction studies. Taken together with the literature and the biochemical and structural data presented in the previous chapters, the *in silico* studies presented in this chapter allow new models of red light signal transduction networks to be constructed and tested in future work.

Results:

The arm-in-arm dimer motif

The previously identified arm-in-arm motifs consist of LxN (near the N terminus) and DLGhFWAhLNEPPP (near the C terminus). The C-terminal motif shares a common conserved FW with inverted 4-5-5 dimer BRRs but the hydropathic character of the surrounding residues differs between the two types. Arm-in-arm BRRs have hydrophobic residues (h = A, I, V, L, M) flanking FW and inverted 4-5-5 dimer BRRs have charged and polar residues (S, E, D, N) in those positions. Additionally, inverted 4-5-5 dimer BRRs lack the LxN motif near the N-terminus. Using the ExPasy ProtParam webserver (Gasteiger, 2003), the sequence flanking FW can be numerically scored by hydropathicity to predict the BRR dimer type (Baker, 2016). Using this method, the BRRs following several well studied BphPs can be predicted (Table 1). BRRs that form the arm-in-arm dimer in the crystal structure (*R. tataouinensis* and *A. tumefaciens*) have high Grand Average of Hydropathicity (GRAVY) scores between 2.6-2.7. BRRs whose crystal structure reveal they form the inverted 4-5-5 dimer have low GRAVY scores between -0.6-0.12. Searching for the presence of the additional LxN motif in arm-in-arm BRR sequences and the absence of LxN in inverted 4-5-5 BRR sequences gives an additional layer of predictive power. Of the BphP-encoding species which have been prominent in the literature, several are predicted to have arm-in-arm BRRs: *Pseudomonas syringae*, *Rhizobium* sp. NT-26, and *Azospirillum brasilense*. Other species studied for BphP signaling are predicted to have inverted 4-5-5 BRRs: *Stigmatella aurantiaca* and *Deinococcus radiodurans*. It is worth noting that *R. palustris* encodes both an SDRR (classified in Table 1) and a RR with a HTH domain in the BphP gene neighborhood.

Table 1. BRR dimer type prediction for BphP study species from the literature based on hydropathicity scoring.

species	strain	FW flanking sequence	hydropathicity score*	LxN present?	Experimental dimer type	Predicted dimer type
Ramlibacter tataouinensis	TTB310	IFWAVL	2.70	Y	arm-in-arm	arm-in-arm
Pseudomonas syringae	CFII64	IFWAVL	2.70	Y	ND	arm-in-arm
Rhizobium sp. NT-26		VFWALL	2.58	Y	ND	arm-in-arm
Agrobacterium tumefaciens	F2	VFWALL	2.58	Y	arm-in-arm	arm-in-arm
Stigmatella aurantiaca	DW4/3-1	EFWLDF	0.25	N	ND	inverted 4-5-5
Deinococcus radiodurans	ATCC13939	AYWFGT	0.22	N	ND	inverted 4-5-5
Synechocystis sp. PCC 6803		SFWLET	0.12	N	inverted 4-5-5	inverted 4-5-5
Tolypothrix sp. PCC 7601		KYWLDI	-0.22	N	inverted 4-5-5	inverted 4-5-5
Rhodopseudomonas palustris	CGA009	HFWMNT	-0.60	N	inverted 4-5-5	inverted 4-5-5

*Grand average of hydropathicity derived from the ExPasy ProtParam server (Gasteiger, 2003)

Taxonomy and lifestyle of bacterial species encoding arm-in-arm BRRs

A taxonomic tree based on derived from the NCBI Taxonomy Database (Federhen, 2012) was generated for 102 bacterial species with the previously identified arm-in-arm motifs (Figure 1). Additionally, the isolation source for each strain was recorded (Table 2) in order to identify common environments in which arm-in-arm BRRs may be required for signaling. All 102 species with predicted arm-in-arm BRRs are from non-photosynthetic bacteria. Overwhelmingly, BRR species are found in the Proteobacteria and strictly within the alpha, beta, and gamma Proteobacteria. No isolates with arm-in-arm BRR motifs were identified in the delta or epsilon Proteobacteria. Only two species with predicted arm-in-arm BRRs were identified outside the Proteobacteria: *Acidobacterium* sp. PMMR2 Acidobacteria phylum and *Zavarzinella formosa* within the Planctomycetes phylum. Since both have the sole arm-in-arm predicted thus far in their entire phyla it is possible that the arm-in-arm BRR genes were acquired by lateral gene transfer in these two species. The most populous genus in the study was Pseudomonas with 31 species (30.4% of the total species) and second most populous was the genus Burkholderia with 21 species (20.6% of the total species). Initially it was suspected that the high proportion of species from these two genera may reflect the overrepresentation of these genera in databases of sequenced bacterial genomes. However, consulting NCBI's microbial genome sequencing database (Markowitz, 2006) revealed that Pseudomonas species make up 1.1% of all bacterial genomes and Burkholderia make up 0.3%, therefore these genera are genuinely enriched with arm-in-arm BRRs. Further, the majority (87.5%) of sequenced Burkholderia species encode a predicted arm-in-arm BRR, underscoring the potential signaling importance of these proteins in that genus.

Categorizing the species by isolation source (Figure 1, Table 2) reveals an important clue to the lifestyle of species with arm-in-arm BRRs. 79.4% of species were isolated from soil or plant sources. Of these, many were isolated from agriculturally important crops such as wheat, rice, sugarcane, corn, tomato, potato, and peanut. Very few of the species identified are

known plant pathogens (*P. syringae* and *B. glumae* infect a variety of crops and rice, respectively) and many are plant commensal species. The bulk of arm-in-arm species live in the soil, where BphP signaling through that unique dimer may confer a survival advantage, or may have become fixed in these species. The rest of the arm-in-arm species were isolated from aquatic sources (11.8%), from insects (4.4%), and from human clinical samples (4.4%).

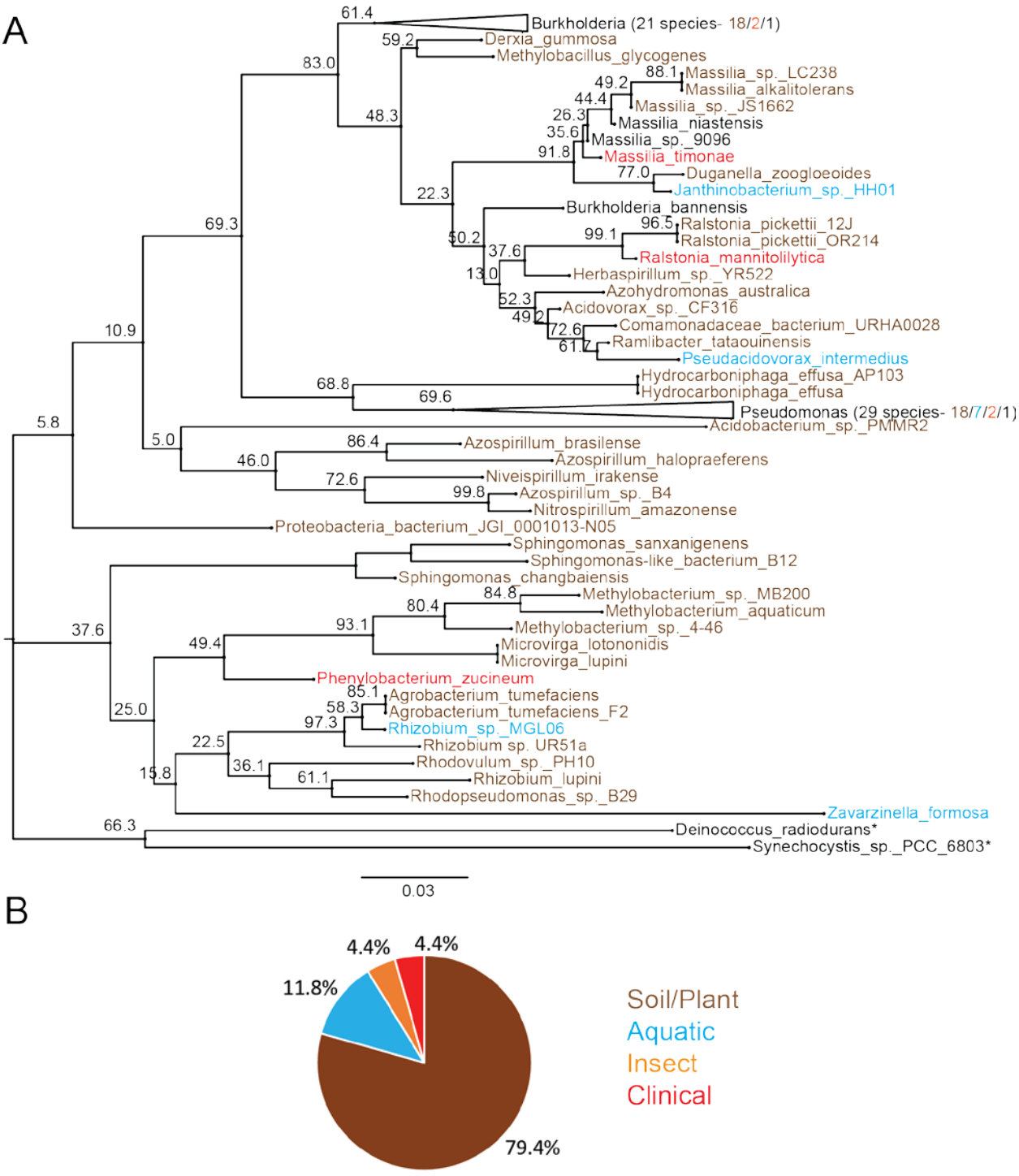


Figure 1. Evolutionary relationships of predicted arm-in-arm encoding bacterial species and their isolation sources. A. Neighbor-joining tree based on the 16S RNA sequences of 102 bacterial species with the arm-in-arm motif in their BRR amino acid sequences,classified by their isolation source. Outgroups are indicated with an asterisk. Nodes are labeled with the percentage of bootstrap replicates (out of 1000 replicates). B. Pie chart of all species according to their isolation source: soil isolates (brown), aquatic isolates (blue), insect isolates (orange), and human clinical isolates (red). Outgroups (*D. radiodurans* and Synechocystis 6803) and isolates from unknown sources are depicted in black.

Table 2. Species with predicted arm-in-arm BRR used in taxonomic tree and their isolation sources.

species	**isolation source**
Acidobacterium sp. PMMR2	biofilm of dead tree near volcano
Acidovorax sp. CF316	roots of populus deltoides
Agrobacterium tumefaciens	zinc lead mine tailing, soil, fungus
Agrobacterium tumefaciens F2	soil near Harbin Institute of Technology
Azohydromonas australica	soil
Azospirillum brasilense	wheat rhizosphere
Azospirillum halopraeferens	roots of kallar grass
Azospirillum sp. B4	stems of rice plants
Burkholderia andropogonis	sugarcane/ cultured habitat
Burkholderia bannensis	torpedo grass
Burkholderia bryophila	moss associated
Burkholderia caledonica	rhizosphere soil
Burkholderia gladioli	corn/ sinus of human
Burkholderia glathei	soil, sediment
Burkholderia glumae	most found in rice
Burkholderia grimmiae	xerophilous moss
Burkholderia jiangsuensis	soil
Burkholderia sordidicola	white rot fungus from plants
Burkholderia sp. 9120	missing
Burkholderia sp. A1	stag beetle gut
Burkholderia sp. BT03	*Populus deltoides* endosphere (tree)
Burkholderia sp. CCGE1003	rhizosphere
Burkholderia sp. MR1	simulated Florida golf course soil

Burkholderia sp. RPE64	bacterial symbiont of the Bean Bug *Riptortus pedestris*
Burkholderia sp. UYPR1.413	isolation source not given, host is tree Parapiptadenia rigida
Burkholderia sp. WSM2232	host is *Nemcia capitata* (plant)
Burkholderia sp. YI23	golf course - degrades fenitrothion (insecticide)
Burkholderia terrae	forest soil
Burkholderia terrae BS001	soil underneath mushroom foot
Candidatus *Burkholderia kirkii* UZHbot1	*Psychotria* leaf (coffee plant)
Comamonadaceae bacterium URHA0028	rhizosphere, in soil
Derxia gummosa	soil
Duganella zoogloeoides	activated sludge
gamma proteobacterium L18	freshwater lake in michigan
Herbaspirillum sp. YR522	roots of populus deltoides
Hydrocarboniphaga effusa	oil contaminated soil
Hydrocarboniphaga effusa AP103	soil contaminated with heavy fuel oil
Janthinobacterium sp. HH01	water from a watering pot
Massilia alkalitolerans	soil
Massilia niastensis	air sample
Massilia sp. 9096	missing
Massilia sp. BSC265	biological soil crust
Massilia sp. JS1662	peanut rhizosphere
Massilia sp. LC238	cave surface layer
Massilia timonae	human blood
Methylobacillus glycogenes	decayed tomato

Methylobacterium aquaticum	*Racomitrium japonicum* moss
Methylobacterium sp. 4-46	*Lotononis bainesi* (plant) nodules
Methylobacterium sp. MB200	agricultural digester remainder
Microvirga lotononidis	root nodules
Microvirga lupini	roots of *L. texensis* (plant)
Microvirga sp. BSC39	soil crust
Nitrospirillum amazonense	forage grass and plants from palmaceae family
Niveispirillum irakense	rice roots
Oxalobacteraceae bacterium JGI 0001012-C15	rhizosphere of arabidopsis thaliana
Phenylobacterium zucineum	intracellular isolation from erythroleukemia cell
Proteobacteria bacterium JGI 0001013-N05	rhizosphere from *Arabidopsis thaliana*
Pseudacidovorax intermedius	south china sea
Pseudomonas alcaligenes	oyster mantle tissue
Pseudomonas azotifigens	hyperthermal compost material
Pseudomonas fulva	leaf from *Arabidopsis thaliana*
Pseudomonas lutea	rhizosphere of grasses
Pseudomonas luteola	soil
Pseudomonas monteilii	clinical
Pseudomonas mosselii	medical specimen/none
Pseudomonas nitroreducens	sewage sludge/sediment from river
Pseudomonas parafulva	rice
Pseudomonas plecoglossicida	fish pathogen
Pseudomonas putida	surface sterilized seeds, antartic sediment, soil from a basin in India, river water, sewage sludge, rhizosphere soil

Pseudomonas rhizosphaerae	rhizosphere of grasses
Pseudomonas savastanoi pv. phaseolicola	leaves of bean plant
Pseudomonas sp. 10B238	deep sea sediment
Pseudomonas sp. 20_BN	Saudi Arabian bank notes
Pseudomonas sp. 313	kelp
Pseudomonas sp. CB1	coal mine soil
Pseudomonas sp. CCOS191	water in Zurich, Switzerland
Pseudomonas sp. CFII64	hyporheic zone of river
Pseudomonas sp. ES3-33	soil
Pseudomonas sp. FeS53a	iron stressed rice soil
Pseudomonas sp. GM84	roots of populus deltoides
Pseudomonas sp. HPB0071	human gut
Pseudomonas sp. LAIL14HWK12:I7	plant associated
Pseudomonas sp. M47T1	trails made from *B. xylophilis (*nematode)
Pseudomonas sp. MT-1	Mariana Trench mud
Pseudomonas sp. StFLB209	potato leaf
Pseudomonas sp. URMO17WK12:I8	plant associated
Pseudomonas sp. WCS358	potato root
Pseudomonas stutzeri	rice roots, chemical hydrocarbon sludge, sediment from west mediterrenean sea, roots, soil near durum wheat
Pseudomonas syringae	irrigation canal, biofilm on rocks in strem, river water, barley
Ralstonia mannitolilytica	bone marrow
Ralstonia pickettii 12J	copper contaminated sediment
Ralstonia pickettii OR214	sediment

Ramlibacter tataouinensis	sub-desert soil
Rhizobium lupini	saline dessert soil
Rhizobium sp. MGL06	sea water
Rhizobium sp. UR51a	rice roots
Rhodopseudomonas sp. B29	rice shoots
Rhodovulum sp. PH10	mangrove forest soil
Sphingomonas changbaiensis	forest soil
Sphingomonas sanxanigenens	soil
Sphingomonas-like bacterium B12	rice shoots
Zavarzinella formosa	Siberian peat bog

Survey of genes encoded near BphPs

The proximity of genes on chromosomes and their conserved order often correlates to the proteins they code for interacting *in vivo* (Aravind, 2000; Snel, 2000; Valencia, 2002; Yanai, 2002; Ciria, 2004; Rogozin, 2004). In the case of BphPs and BRRs, this is experimentally supported. From species studied in the literature for their BphPs, the classic order of genes in a BphP operon is: heme oxygenase / bacteriophytochrome / BRR (Bhoo, 2001). However, there are often other genes flanking this triad encoded in the same orientation and these are largely unstudied for their potential contributions to red light signaling pathways. In order to address this knowledge gap, 155 bacterial species with putative BphPs were surveyed for the genes surrounding the BphP gene. Gene neighborhoods were identified as all genes flanking the BphP gene, in the same sense direction, with 50 noncoding bp or less between genes. Of the 155 species, 32 have predicted arm-in-arm BRRs, 76 have another type of SDRR such as a predicted inverted 4-5-5 dimer BRR, and 47 have no BRR encoded near the BphP. We examined the prevalence of genes in the entire study group and searched for trends that differed between the BRR type groups (Figure 2).

For the entire 155 species (Figure 2A), the most common genes encoded near BphPs were BRRs directly following the BphP (67% of species), non-BphP histidine kinase proteins (51.2%), heme oxygenase (33.5%), transferase enzymes (12.2%), and other RR proteins not encoded directly downstream of the BphP (11.6%). The next most common genes (found in ≤10% of neighborhoods) include transcriptional regulators, domain of unknown function proteins, and peptidase/proteases. These data suggest non BphP HKs may play a part in red light signal transduction based on their prevalence. Next, the different BRR groups were analyzed for gene prevalence surrounding the BphP gene. Histidine kinase genes were found in 81.2% of the arm-in-arm species, 64.5% of the other BRR type species, and only 8.5% of non-BRR encoding species. The correlation between BRR genes and HK genes in BphP regions suggest that a potential signaling pathway would involve these proteins interacting together,

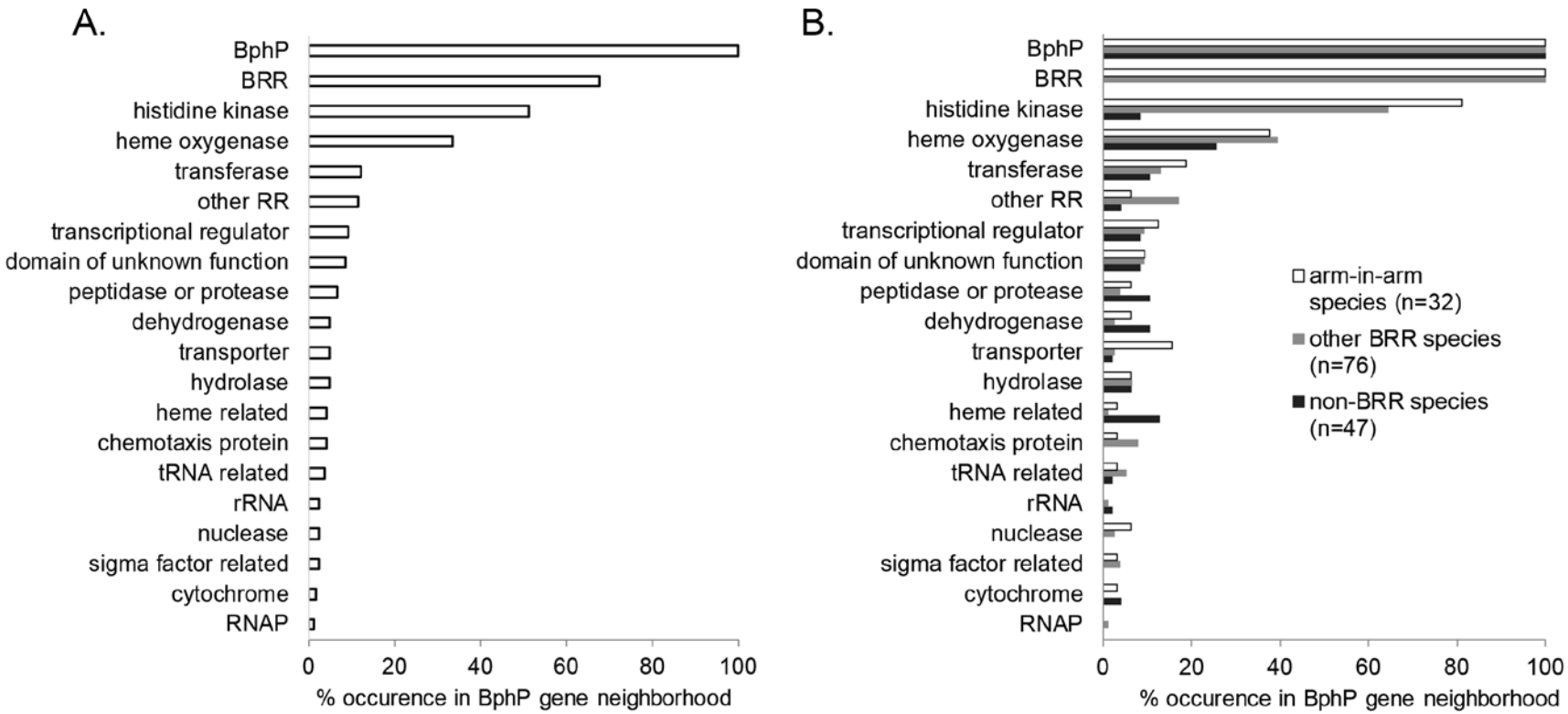


Figure 2. Top 20 occurring genes in the BphP gene neighborhood survey of 155 bacterial genomes. A. Occurance of genes for all species. B. Occurence of genes for each BRR type group. Experimentally verified BphP interaction partners BRR and heme oxygenase are joined by other histidine kinase genes as the top three occurring genes. Species are grouped by the type of BRR encoded in the BphP gene neighborhood: arm-in-arm BRRs (white); other BRRs (gray) or no BRR present (black).

that BphP can interact with either of the two. The difference between prevalence of HKs in the arm-in-arm BRRs and other BRRs (Figure 2B) was not statistically significant ($p = 0.084$).

Since conserved gene order can correlate to protein-protein interaction (Dandekar, 1998; Overbeek, 1999), the most common genes from all species were analyzed for the frequency of following the classic BphP operon order with the addition of HK following the BRR. Figure 3 demonstrates the frequency that heme oxygenase, BphP, BRR, and HK genes appear in order where the BphP position is 0, heme oxygenase is -1, BRR is +1, and HK is +2. The data reveal that for species with BphPs and BRRs, the BphP/BRR/HK order of genes is more common than the classic HO/BphP/BRR order. Additionally, the frequency of HK appearing +2 genes downstream of BphP and immediately following the BRR is more common in arm-in-arm species than in species with other BRR types ($p = 0.015$).

Other histidine kinase proteins besides BphPs are commonly encoded near the phytochrome and response regulator genes based on this survey. But what role in signaling might these HKs play? In order to predict a functional role, the amino acid sequences of the arm-in-arm species HKs were annotated for any common conserved domains (Bateman, 2002) and analyzed for the frequency of those domains occurring (Figure 4A). These data allow rudimentary structural and functional modeling of an average HK from the BphP operon. In order, starting from the amino terminus, the most common domains are: a receiver domain; a PAS 3 domain; a second PAS domain of type 3, 4, or 9; a HisKA domain, an HATPase_C domain, and a final receiver domain (Figure 4B). These domains were recognized in at least 75% of the amino acid sequences. Therefore, the typical HK from a BphP operon is a hybrid histidine kinase, that contains both an N-terminal and C-terminal aspartic acid phosphoacceptor domain, with one or more PAS domains and a histidine kinase histidine phosphoacceptor/phosphotransfer and ATPase domains between the two receivers. PAS domains are known dimerization domains with the ability to sense environmental signals, often

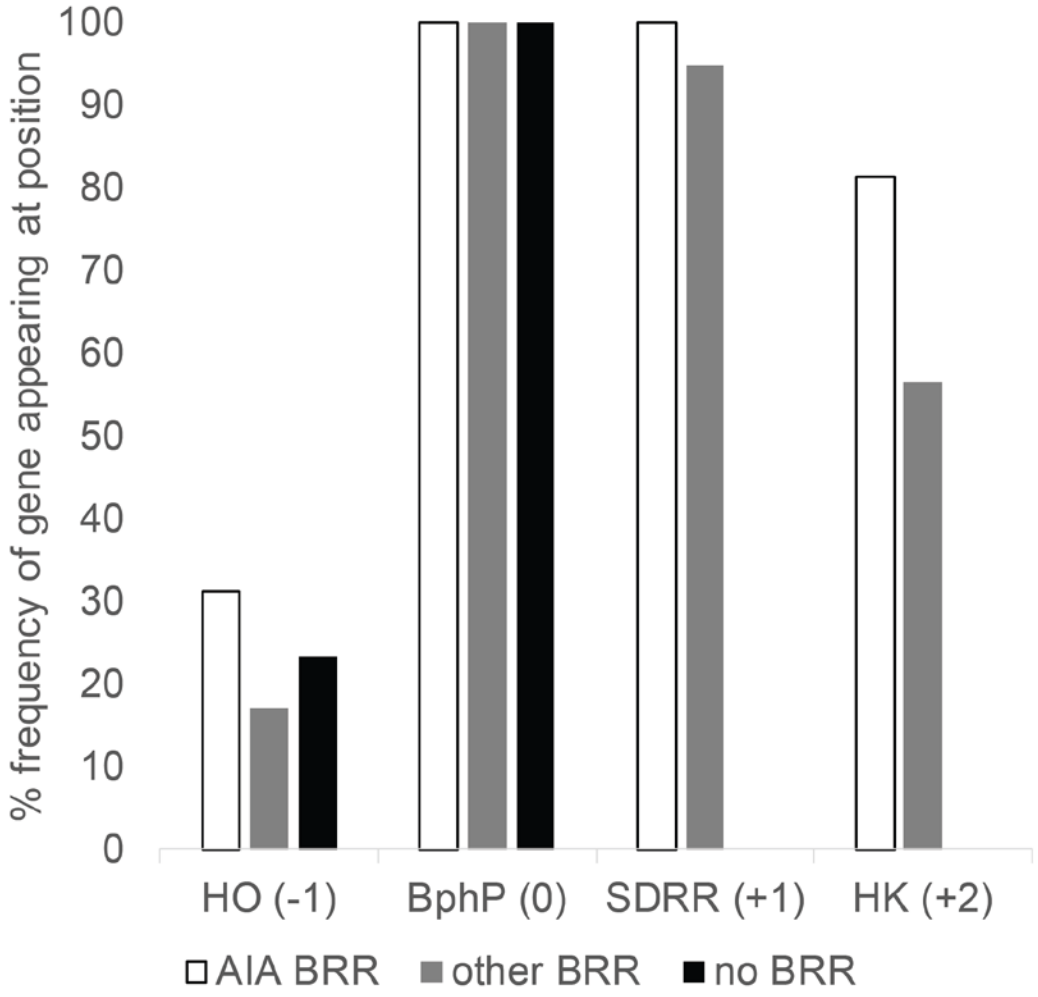


Figure 3. Order of top occurring genes from gene neighborhood survey, relative to the BphP gene (position 0). Heme oxygenase genes were scored for their frequency occurring directly before the BphP, BRR genes were scored for frequency of occurring directly after the BphP gene, and HK genes were scored for frequency of occurring 2 positions after the BphP gene. Colored bard indicate the BRR type groups: arm-in-arm BRRs (white); other BRRs (gray) or no BRR present (black).

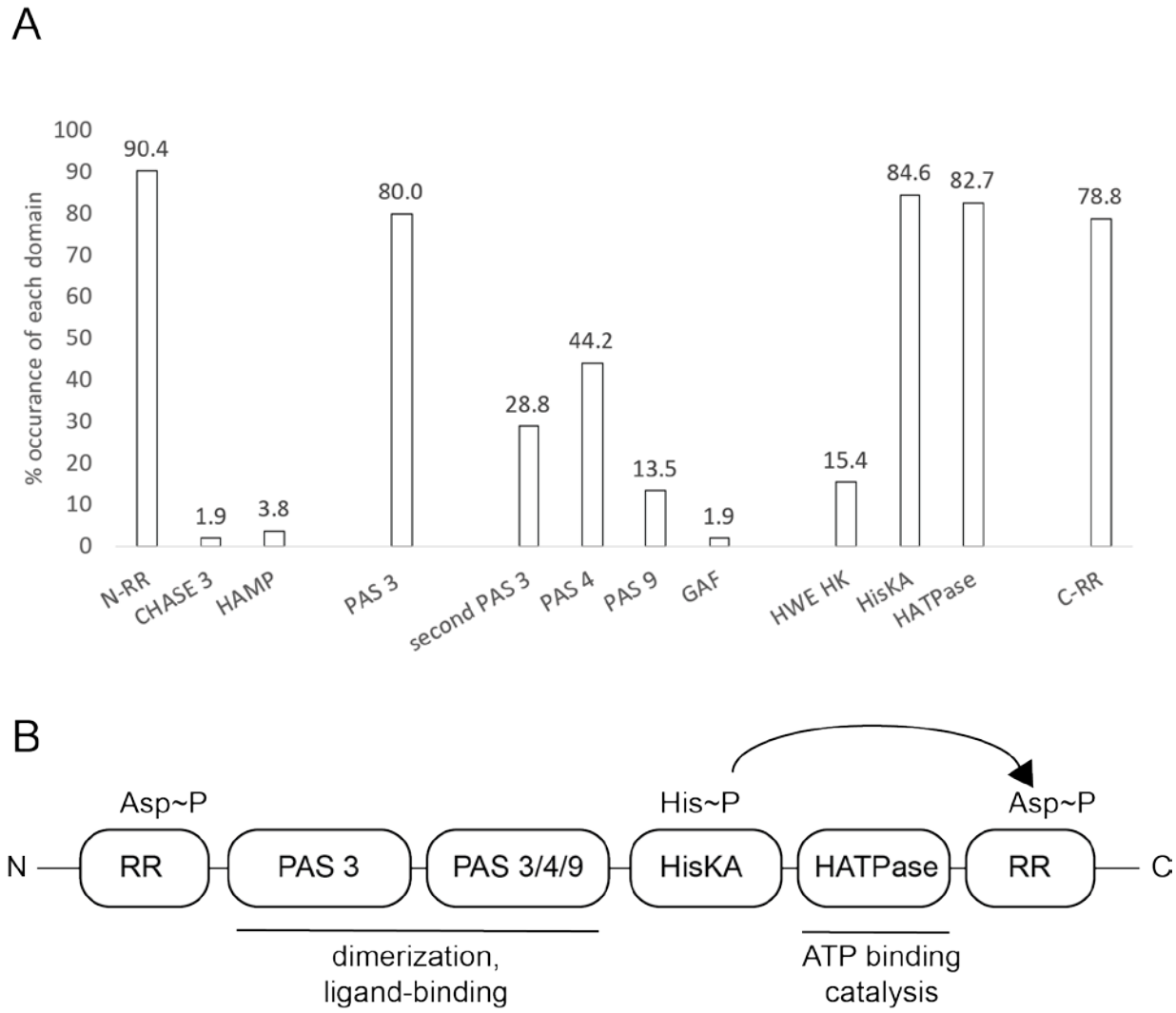


Figure 4. Typical domain architecture of HHK proteins encoded in BphP with arm-in-arm BRR gene neighborhoods. A. Pfam domain assignments of 32 HHK amino acid sequences, in order from N terminus to C terminus. B. Domains occurring in more than 75% of sequences were compiled to create a domain architecture model for a typical HHK associated with BphPs and arm-in-arm BRRs.

through bound ligands (Taylor, 1999). A model of how such a hybrid histidine kinase may function in BphP-BRR signal transduction pathways is presented in the discussion.

Gene Cluster Search for arm-in-arm BRR signal transduction partners

In order to perform a more exhaustive, genome-wide search for undiscovered signal transduction partners, a collaborative study with Sarah Stevens of the McMahon Lab at UW-Madison was undertaken. The study was jointly designed by Anna Baker and Sarah Stevens. Briefly, Sarah Stevens performed gene clustering by BLAST of all ORFs in the study genomes and Anna Baker performed the analysis to find genes differentially encoded between arm-in-arm BRR species and outgroup species. The top hits can be considered for a potential role in red light signal transduction.

Gene clusters which resulted from a 60% identity cutoff were used for the following analysis, since of the three cutoffs studied (50%, 60%, and 70%) the 60% data set output arm-in-arm BRR proteins and histidine kinase-domain containing proteins as the #1 and #3 hits respectively. Given that arm-in-arm BRRs are known to be found in the ingroup species and not the outgroups (by study design), and given that BphP histidine kinases are proven interaction partners with BRRs we reasoned that any gene clusters with similar ingroup%-outgroup% scores to these positive control clusters could be considered valid hits. All clusters with a differential score of >50% are listed in Table 3.

There were several gene clusters that are essential and universally conserved genes in bacteria, and therefore cannot be considered as valid differential hits in this analysis: these include DNA polymerase, DNA topoisomerase, and RNA polymerase subunit omega. Possible reasons for the leakage of essential genes into the top clusters will be discussed.

Table 3. Top Gene clusters with 60% identity cutoff, ranked by differential occurrence in arm-in-arm species vs. outgroups (ingroup%-outgroup%). Presence of clusters in total outgroup % and individual outgroups (BphP species with no arm-in-arm BRR, Proteobacteria with no BphPs, and distant phyla) are listed separately.

gene product	ingroup %	outgroup %	ingroup%-outgroup%	BphP outgroup %	Proteobacteria outgroup %	other phyla outgroup %
response regulator receiver protein	93.8	5.9	**87.9**	20	0.0	0
DUF971 family protein	81.5	17.6	**63.8**	40	14.3	0
His Kinase A (phospho-acceptor) domain-containing protein	90.1	29.4	**60.7**	60	28.6	0
putative DNA modification/repair radical SAM protein	75.3	17.6	**57.7**	40	14.3	0
peptide-methionine (R)-S-oxide reductase	98.8	41.2	**57.6**	80	42.9	0
error-prone DNA polymerase, DnaE-like	86.4	29.4	**57.0**	60	28.6	0
ferredoxin	91.4	35.3	**56.1**	60	42.9	0
peroxiredoxin, Ohr subfamily	79.0	23.5	**55.5**	60	14.3	0
Glucose/arabinose dehydrogenase, beta-propeller fold	72.8	17.6	**55.2**	40	14.3	0
DNA repair photolyase	84.0	29.4	**54.5**	60	28.6	0
Uncharacterized conserved protein, circularly permuted ATPgrasp superfamily	77.8	23.5	**54.2**	60	14.3	0
DNA-directed RNA polymerase subunit omega	87.7	35.3	**52.4**	60	42.9	0
adenosylhomocysteinase	92.6	41.2	**51.4**	60	42.9	0
undecaprenyl diphosphate synthase	86.4	35.3	**51.1**	60	42.9	0
3-oxoacyl-[acyl-carrier-protein] reductase	74.1	23.5	**50.5**	40	28.6	0

2-keto-4-pentenoate hydratase/2-oxohepta-3-ene-1,7-dioic acid hydratase	67.9	17.6	**50.3**	40	14.3	0
polyhydroxyalkanoate synthase	67.9	17.6	**50.3**	40	14.3	0

Discussion:

The arm-in-arm BRR dimer motif was identified as consisting of conserved LxN near the N-terminus and an extended DLGhFWAhLNEPPP near the C-terminus. The second motif can be abbreviated to hFWAhL in the case of hydropathicity scoring. The resultant hydropathicity scores can be used to bin BRR proteins into arm-in-arm type or inverted 4-5-5 type based on this short sequence alone. Scores 2.5 or higher appear to correspond to arm-in-arm BRR dimers, and scores 0.25 and lower appear to correspond to inverted 4-5-5 BRR dimers. Of the BRRs from literature cited BphP-bearing species, *P. syringae*, *Rhizobium* NT-26, and *A. brasilense* are predicted to be arm-in-arm dimers by this method. BRRs from *S. aurantiaca* and *D. radiodurans* are predicted to be inverted 4-5-5 dimers. As more BRR biochemical characterizations and structures appear in the literature, this method can be validated by comparing experimental stoichiometries to these predictions.

The phylogeny of bacterial species with putative arm-in-arm BRRs suggests that the conservation of FW as dimerization residues can be traced back to a more distant ancestor than the FW + hydrophobic motif. BRRs associated with the phytochrome TCSs from cyanobacteria species do not form arm-in-arm dimers, yet retain the FW surrounded by more polar and charged residues (Im, 2001; Benda, 2004). Species predicted to have arm-in-arm BRR dimers are mainly within the α, β, and ϒ-proteobacteria and even within these groups there are notable species which lack the motif in their BRR. It is possible that arm-in-arm dimers evolved in a proteobacterial ancestor by mutations resulting in hydrophobic residues near FW, likely in a soil-dwelling species, and that descendant species that no longer inhabit such an environment have lost the adaptation and reverted to polar/charged residues. Given the common soil and plant-associated lifestyle exhibited by most of the arm-in-arm species, arm-in-arm BRR dimers may confer a red light signaling advantage to bacteria inhabiting subsurface soil, or may confer an

advantage to bacteria associated with plants, whose cellular processes are heavily regulated by light cycles.

The targeted survey of BphP gene neighborhoods discovered a high prevalence of other histidine kinase protein coding sequences, especially in neighborhoods that also contained BRRs. BphP and BRR are typically considered members of two component systems, in which BphP autophosphorylates and transfers phosphate with fidelity to BRR. Experimental evidence presented in this work and by others has demonstrated that BphPs transfer phosphate specifically to cognate BRRs (Bhoo, 2001; Karniol, 2003; Giraud, 2005; Psakis, 2011). The presence of a hybrid histidine kinase with possible sensory domains raises the possibility that a BphP-BRR-HHK trio might form a branched signal transduction pathway (Figure 5) (Laub, 2007). How might such a branched pathway function? As previously demonstrated, a BphP can hydrolyze ATP and transfer a phosphate group onto histidine more efficiently in the dark. In the presence of BRR proteins, phosphate is transferred from histidine on the HK to aspartate on the RR. However, the average HK in a BphP gene neighborhood also bears an N-terminal RR domain. It remains to be tested whether BphP can also transfer phosphate to this RR in the *R. tataouinensis* system, but work in the *Rhizobium* NT-26 system has shown specific phosphotransfer from a BphP to either a BRR (predicted in this work as an arm-in-arm BRR) or to a HHK (Wojnowska, 2013). Once the HHK accepts the input from BphP indicating that the environment is dark, the PAS domains encoded in the protein's core might sense additional environmental signals and integrate light with that second input to control the final histidine kinase activity. PAS domain sensors are varied beyond their role in BphPs and are known to sense redox state, oxygen, voltage, and bind small ligands such as heme, hydroxycinnamic acid, flavin adenine nucleotide, and 2Fe-2S clusters (Taylor, 1999). Any of these may be a second signal integrated with the light signal from BphP by the HHK proteins encoded in BphP gene neighborhoods. Biochemical characterization of one or more of these HHKs is needed to

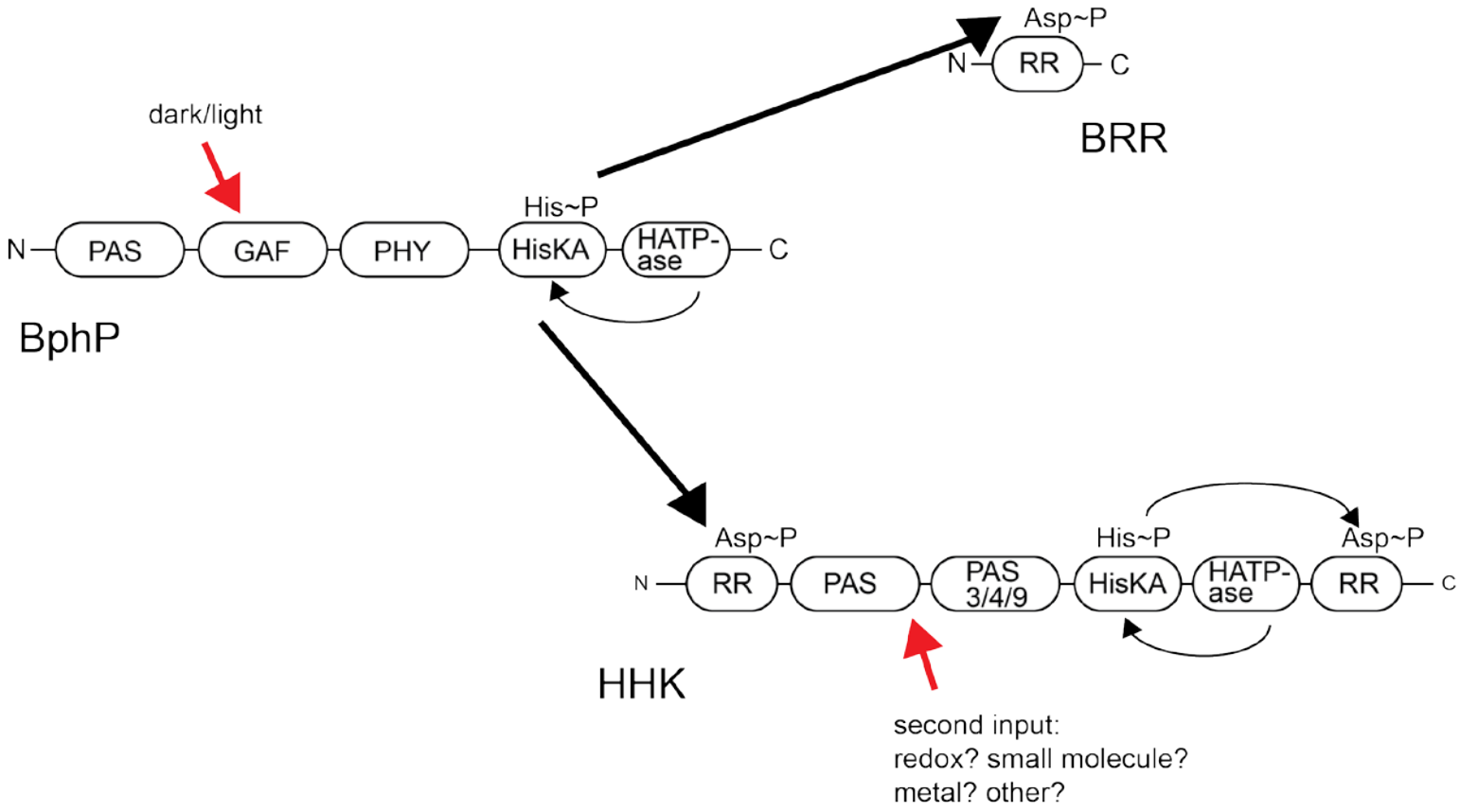


Figure 5. Model of a branched BphP-BRR-HHK signal transduction network regulated by red light and a second, unknown input. Red arrows represent sensory input. Thin black arrows represent intra-protein phosphotransfer events. Thick black arrows represent protein-protein interaction and inter-protein phosphotransfer events.

verify a role in red light signal transduction. Additionally, signal transduction partners beyond the HHK would need to be identified to determine how integrated/branched signaling through a BphP/BRR/HHK system results in cellular adjustment to the environment.

Gene clustering searching and analysis was designed to identify potential signal transduction partners beyond arm-in-arm BRRs which might be missed by the gene neighborhood survey. Although the arm-in-arm species (ingroups) and outgroups were carefully chosen to find genes which are prevalent in only in ingroups, some essential bacterial proteins were identified as top differential hits in the 60% homology dataset. Proteins such as DNA topoisomerase and RNA polymerase subunit omega must be present in bacterial genomes for those bacteria to reproduce, and typically such evolutionarily conserved proteins would bear significant homology between different groups of bacteria (Jordan, 2002; Zhang, 2009). In the case of RNAP subunit omega, the gene was identified as homologous at the 60% identity cutoff in 87.7% of the arm-in-arm ingroup species, but in only 35.3% of the combined outgroups. Examining systematic presences and absences among the different outgroups can explain, in part, why these essential genes appeared as differential hits. RNAP subunit omega was identified as a homolog with the others in the cluster in 60% of the outgroup with a BphP but no BRR, 42.8% in the alpha/beta/gamma proteobacteria without a BphP outgroup, and 0% in the outgroup species from distant phyla. The absence of homology in the most phylogenetically distant groups points to the essential genes being less similar in these species because all genes would be less similar the more distantly related the species. Further analysis of all the top gene cluster hits (ingroup-outgroup ≥50%) in the different outgroups shows an overall trend that none of the top genes were identified as homologous in the distant phyla outgroup, and more were identified as homologous in the sub-outgroup species that encode a BphP. Compared to the BphP-only outgroup fewer top hits were identified as homologous in the outgroup from the Proteobacteria that lack BphPs. This suggests that the top hits may indeed be related to red

light signal transduction, since several of the BphP-only outgroup species are outside the Proteobacteria and less homology would be expected for unrelated proteins from more distant groups. When the top gene clusters are cutoff at 45% ingroup%-outgroup% or lower, hits start to appear that are more homologous in the Proteobacteria outgroup than in the BphP-only outgroup.

The second highest scoring cluster, found in 81.5% of arm-in-arm species and 17.6% of outgroups corresponds to a domain of unknown function (DUF) protein known as DUF971. This is a functionally uncharacterized protein with sequence homology to the histidine biosynthesis pathway protein HisA (Tars, 2010). Fortunately, a DUF971 has been structurally characterized by NMR as part of a structural genomics initiative (Figure 6) (PDB: 2L6P). Although the function of this small protein remains unknown, surface electrostatics modeling makes this an attractive candidate for a direct protein-protein interaction with the arm-in-arm BRR. A long positively charged stripe originally noted as a potential interaction surface on RtBRR is complemented by a negatively charged stripe on the surface of DUF971 (Figure 6). Future work could involve cloning and expression of the *R. tataouinensis* DUF971 protein and assaying complex formation with RtBRR.

Several oxidative stress and DNA repair proteins were also found as top hits: a putative DNA repair radical S-adenyl-methionine protein (75.3% ingroup / 17.6% outgroup); a peptide-methionine (R)-S-oxide reductase (98.8% ingroup / 41.2% outgroup); peroxiredoxin (79.0% ingroup / 23.5% outgroup); and DNA repair photolyase (84.0% ingroup / 29.4% outgroup). Stress response to light is a common theme in photoreceptor research in bacteria. In a proteomic study of red light phenotypes from a predicted arm-in-arm species, *Azospirillum brasilense*, a general stress response was found (Kumar, 2012). However, that response involved a completely different set of proteins being upregulated in response to light than those identified in this work. General stress responses to blue light have also been identified in *B.*

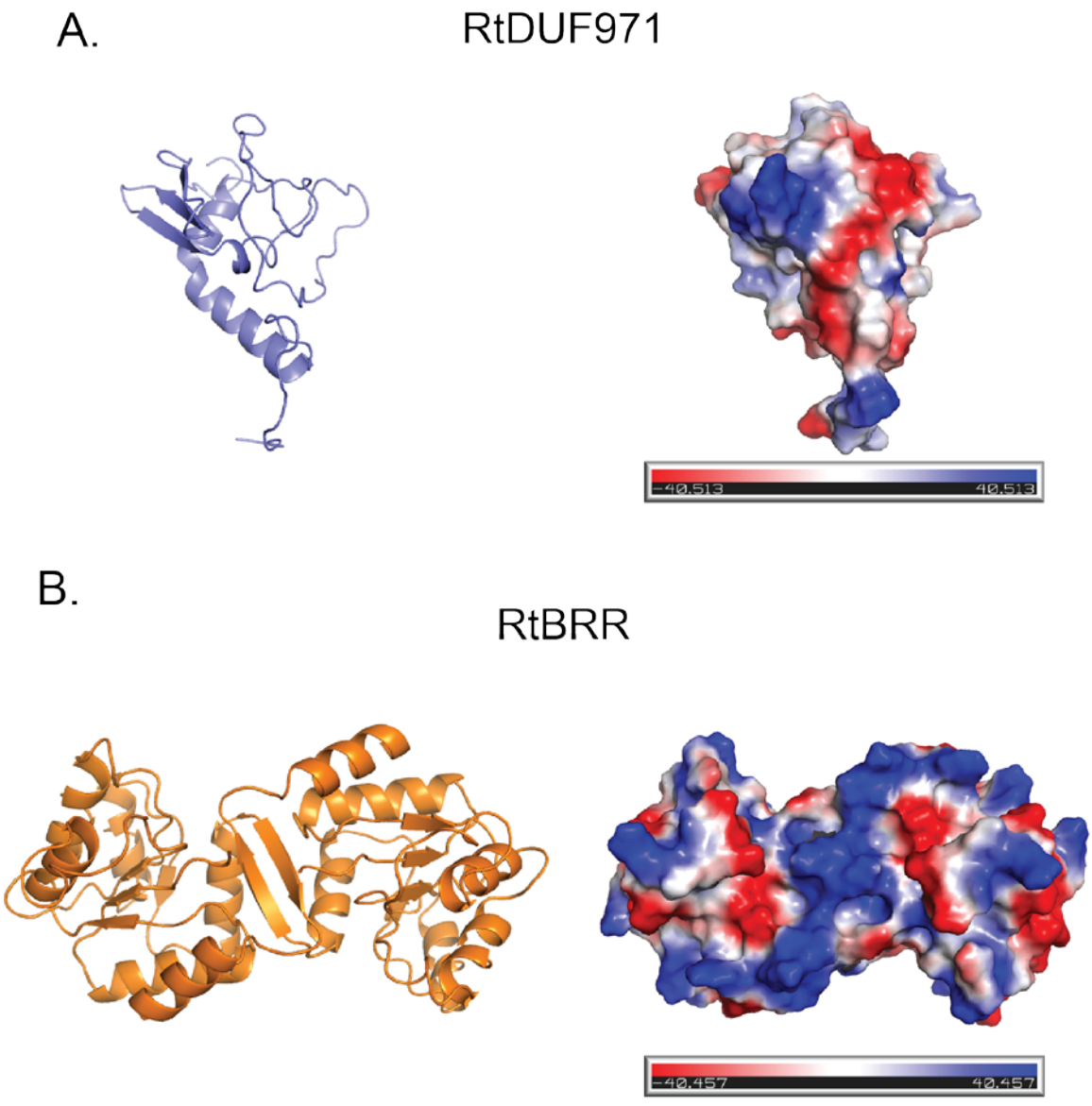


Figure 6. Cartoon and surface electrostatics models of DUF971 and RtBRR proteins, which are predicted protein interaction partners discovered by gene clustering. A. Homology model of the *R. tataouinensis* DUF971 protein based on PDB 2L6P. B. X-ray crystal structure of an RtBRR arm-in-arm dimer (taken from Baker, 2016).

subtilis, mediated by a LOV-HK protein which works through yet another mechanism, by regulating the sigma factor σ^B (Avila-Perez, 2006; Gaidenko, 2006). Perhaps the redox response proteins we identified in silico are part of the response detected in *A. brasilense*, yet were not upregulated or highly expressed enough to be detected in the 2D-gel/mass spectrometry experiment conducted by Kumar et al. Another possibility is that these oxidative stress and DNA-repair proteins represent a separate stress response pathway to red light.

In the context of the *R. tataouinensis* genome and phenotypes, two more clusters stand out. The uncharacterized circularly permuted ATPgrasp superfamily protein (77.8% ingroup / 23.5% outgroup) is encoded in the RtBphP1 gene neighborhood, three genes downstream. Proteins with this conserved fold have ATP binding capability and diverse functions including fatty acid biosynthesis, peptidoglycan biosynthesis, ribosome modification, and microtubule assembly to name a few (Fawaz, 2011). Polyhydroxyalkanoate (PHA) synthase (67.9% ingroup / 17.6% outgroup) has a potential link to the large intracellular granules of PHA that form within *R. tataouinensis* cells (Heulin, 2003). These granules are found in both rod and cyst-like cell types. How PHA content of *R. tataouinensis* cells may be affected by red light exposure remains to be tested.

Finally, by combining the results of the biochemical characterizations done here and by others on BphP-BRR signal transduction systems with these new bioinformatics leads a red light signaling interaction map can be constructed (Figure 7). This map illustrates the experimentally verified and newly predicted protein-protein interactions and the interrelatedness of the proteins in the pathway. Central to the pathway is red light sensing by BphP, which can directly interact with three partners: heme oxygenase, to receive the biliverdin chromophore (Shah, 2012); BRR, to transfer phosphate in a light-regulated fashion; and also to the N-terminal receiver domain of HHK, as evidenced by Wojnowska, et al. All four of these are linked as gene neighbors, and the four in sequence share conserved gene order. Also depicted are some of the

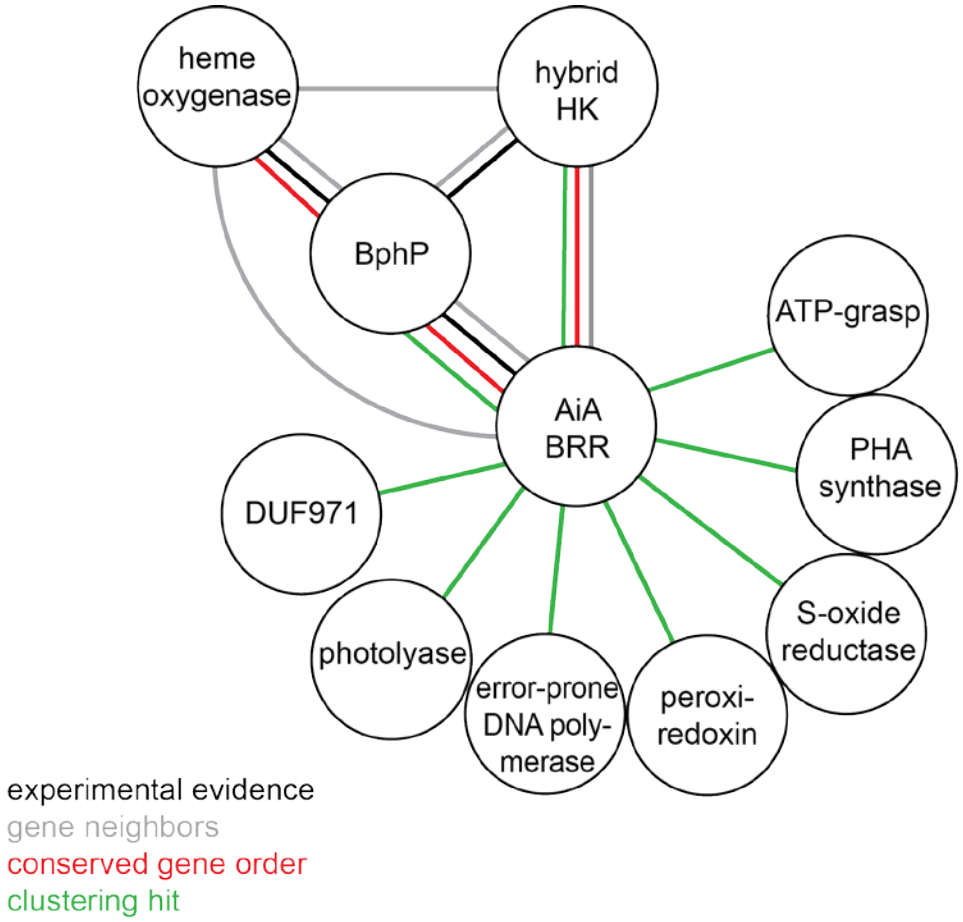


Figure 7. Interaction map summarizing previous experimental findings and new bioinformatics predictions about red light signal transduction networks. Each circle depicts a protein in the network. Links between proteins are colored according to interaction evidence type: experimental evidence (black), gene neighbors (grey), conserved gene order (red), gene clustering hit (green).

gene cluster hits which are interaction partners for the arm-in-arm yet to be tested. Experiments with these new potential interaction partners for BRR revealed by bioinformatics, particularly HHK and DUF971, represent an exciting new direction for the Forest Lab.

Methods

Identifying the arm-in-arm motif and predicting arm-in-arm BRRs

The arm-in-arm dimer motifs were identified by manual analysis of a multiple amino acid sequence alignment generated with ClustalOmega (Sievers, 2011) which included the sequences for BRR proteins from *Ramlibacter tataouinensis*, *Agrobacterium tumefaciens*, *Pseudomonas syringae*, *Burkholderia glumae*, *Rhizobium sp. NT-26*, *Synechocystis* sp. PCC 6803, *Tolypothrix* sp. PCC 7601, *Rhodopseudomonas palustris*, *Deinococcus radiodurans* and the CheY sequence from *Escherichia coli.* Sequence features which were common in the structurally characterized arm-in-arm BRRs from *R. tataouinensis* and *A. tumefaciens* and absent in inverted 4-5-5 BRR dimers were identified as critical arm-in-arm sequence motifs.

Grand average of hydropathicity (GRAVY) scores were computed for the six-amino acid sequences corresponding to the abbreviated C-terminal arm-in-arm motif (IFWAVL in *R. tataouinensis*). Abbreviated sequences from *Ramlibacter tataouinensis, Pseudomonas syringae, Rhizobium* sp. NT-26, *Agrobacterium tumefaciens*, *Stigmatella aurantiaca, Deinococcus radiodurans, Synechocystis* sp. PCC 6803, *Tolypothrix* sp. PCC 7601, and *Rhodopseudomonas palustris* were scored using ExPasy ProtParam (Gasteiger, 2003).

Evolutionary relationships and lifestyle of BRR species

A pBLAST search (Altschul, 1990) was carried out for the sequence DLGxFWAxLNEP within the Bacteria and hits were filtered by the following criteria: expect value cutoff < 1; only putative SDRRs with the search sequence near the C-terminus; x = hydrophobic amino acids. Then, pBLAST using the RtBphP1 sequence was run on the 104 candidate BRR-containing genomes, only those with a putative PAS-GAF-PHY-HK BphP with cysteine in the conserved biliverdin binding position were included (102/104). 16S rRNA sequences of the 102 species passing all filters were downloaded from the SILVA database (Quast, 2013) and aligned with ClustalOmega

(Sievers, 2011). A neighbor-joining tree was generated using MEGA7 (Kumar, 2016), using the Maximum Composite Likelihood Method with 1000 bootstrap replicates. The resulting tree was visualized with FigTree (Rambaut, 2009). The isolation sources of bacterial species were found in the corresponding IMG genome record for each species (Markowitz, 2006).

BphP gene neighborhood survey and hybrid histidine kinase domain assignment

BLASTP was performed against the RtBphP1 amino acid sequence, and yielded 1734 hits for BphP homologs (Altschul, 1990). Results were filtered with an expect cutoff of 1e-50 or less and sequence coverage cutoff of 90% or greater. Since certain genera, such as *Pseudomonas*, dominated the list, these were further filtered to include only one representative species for each unique genus, with preference going to those species which have previously been studied in the BphP literature. This yielded 155 species/genus representatives to annotate the BphP gene neighborhood. The 155 species were further split into three BRR type groups: those with a predicted arm-in-arm BRR (n=32), those with a BRR of another predicted type such as inverted 4-5-5 dimer (n=76), and those with no BRR in the gene neighborhood (n=47). For each species, the IMG database (Markowitz, 2006) was used to BLAST for and identify the BphP with highest homology to RtBphP1 (in cases of multiple BphPs in one species) and the gene neighborhood was manually annotated in a newly created database. Gene neighborhoods were identified as all genes flanking the BphP gene, in the same sense direction, with 50 noncoding bp or less between genes. After neighborhoods were annotated for all species, genes were counted by their type and listed from most to least commonly occurring in the neighborhoods. This analysis was done for the 155 species and for each BRR type group for comparison.

Common domains of hybrid histidine kinase proteins adjacent to BRR coding regions were identified by entering the amino acid sequences of the 32 arm-in-arm species from the gene neighborhood survey into the Pfam server (Bateman, 2002). The output domain predictions

were arranged from N to C terminus and plotted for the frequency of occurrence. This yielded the domain architecture of an average arm-in-arm BRR-associated hybrid histidine kinase.

Gene Cluster Search

In order to discover gene products which are present in arm-in-arm BRR species but absent in other species, 81 ingroup species were chosen which bear a predicted arm-in-arm BRR in the same gene neighborhood as a BphP. Three separate outgroups were chosen to contrast the ingroup: species with a BphP but no arm-in-arm BRR (n=6); species in the alpha, beta, and gamma proteobacteria but which lack a BphP or BRR (n=7); and species from more distantly related phyla (n=5) for a total of 18 total outgroup species.

Pairwise BLAST searches were run through the UW-Madison Center for High Throughput Computing, comparing every predicted protein coding sequence from the genomic records of all 99 study species against all the others, to determine the homology of every pair (Altschul, 1990). These data were filtered to include pairs with a query coverage cutoff of 75% or greater. These results were further filtered with sequence identity cutoffs of 50%, 60%, and 70%. Each of the 50%, 60%, and 70% data sets were sorted into gene clusters using the MCL clustering algorithm (Dongen, 2000).

The presence or absence of each gene cluster in the ingroup and outgroups were recorded, and then clusters were ranked by the ingroup %-outgroup %. At this stage, the 50%, 60%, and 70% identity cutoff cluster lists were manually compared. The 60% identity cutoff data set was chosen for the final analysis because at this level of homology, the arm-in-arm BRR and histidine kinase proteins were the #1 and #3 gene clusters identified as differential between the ingroup and the outgroups, meaning that if these known red light signal transduction partners were successfully identified, other hits could be considered as valid potential interaction partners. The list of hits was further filtered to remove: clusters not present in the study species

R. tataouinensis; clusters present in greater than 50% of outgroup species; and clusters not present in at least 65% of ingroup species. The top cluster hits are reported at 50% ingroup-outgroup or higher.

References

1. **Altschul S**, **Gish W**, **Miller W**. 1990. Basic local alignment search tool. J Mol … **215**:403–410.

2. **Aravind L**. 2000. Guilt by association: Contextual information in genome analysis. Genome Res **10**:1074–1077.

3. **Ávila-Pérez M**, **Hellingwerf KJ**, **Kort R**. 2006. Blue light activates the σB-dependent stress response of Bacillus subtilis via YtvA. J Bacteriol **188**:6411–6414.

4. **Baker AW**, **Satyshur KA**, **Moreno Morales N**, **Forest KT**. 2016. Arm-in-arm response regulator dimers promote intermolecular signal transduction. J Bacteriol **198**:1218–1229.

5. **Barkovits K**, **Harms A**, **Benkartek C**, **Smart JL**, **Frankenberg-Dinkel N**. 2008. Expression of the phytochrome operon in Pseudomonas aeruginosa is dependent on the alternative sigma factor RpoS. FEMS Microbiol Lett **280**:160–168.

6. **Barkovits K**, **Schubert B**, **Heine S**, **Scheer M**, **Frankenberg-Dinkel N**. 2011. Function of the bacteriophytochrome BphP in the RpoS/Las-quorum sensing network of Pseudomonas aeruginosa. Microbiology **157**:1651–1664.

7. **Bateman A**, **Birney E**, **Durbin R**, **Eddy SR**, **Howe KL**, **Sonnhammer EL**, **Cerruti L**, **Durbin R**, **Etwiller L**, **Eddy SR**, **Griffiths-Jones S**, **Howe KL**, **Marshall M**, **Sonnhammer EL**. 2002. The Pfam protein families database. Nucleic Acids Res **28**:276–280.

8. **Benda C**, **Scheufler C**, **Tandeau de Marsac N**, **Gärtner W**. 2004. Crystal structures of two cyanobacterial response regulators in apo- and phosphorylated form reveal a novel dimerization motif of phytochrome-associated response regulators. Biophys J **87**:476–487.

9. **Bender CL**, **Alarcón-Chaidez F**, **Gross DC**. 1999. Pseudomonas syringae phytotoxins: mode of action, regulation, and biosynthesis by peptide and polyketide synthetases. Microbiol Mol Biol Rev **63**:266–92.

10. **Bhoo SH**, **Davis SJ**, **Walker J**, **Karniol B**, **Vierstra RD**. 2001. Bacteriophytochromes are photochromic histidine kinases using a biliverdin chromophore. Nature **414**:776–779.

11. **Ciria R**, **Abreu-Goodger C**, **Morett E**, **Merino E**. 2004. GeConT: Gene context analysis. Bioinformatics **20**:2307–2308.

12. **Dandekar T**, **Snel B**, **Huynen M**, **Bork P**. 1998. Conservation of gene order: A fingerprint of proteins that physically interact. Trends Biochem Sci **23**:324–328.

13. **Davis SJ**. 1999. Bacteriophytochromes: Phytochrome-Like Photoreceptors from Nonphotosynthetic Eubacteria. Science (80-) **286**:2517–2520.

14. **Fawaz M V.**, **Topper ME**, **Firestine SM**. 2011. The ATP-grasp enzymes. Bioorg Chem **39**:185–191.

15. **Federhen S**. 2012. The NCBI Taxonomy. Nucleic Acids Res **40**:D136–D143.

16. **Gaidenko TA**, **Kim TJ**, **Weigel AL**, **Brody MS**, **Price CW**. 2006. The blue-light receptor YtvA acts in the environmental stress signaling pathway of Bacillus subtilis. J Bacteriol **188**:6387–6395.

17. **Gasteiger E**, **Gattiker A**, **Hoogland C**, **Ivanyi I**, **Appel RD**, **Bairoch A**. 2003. ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res **31**:3784–3788.

18. **Giraud E**, **Zappa S**, **Vuillet L**, **Adriano JM**, **Hannibal L**, **Fardoux J**, **Berthomieu C**, **Bouyer P**, **Pignol D**, **Vereglio A**. 2005. A new type of bacteriophytochrome acts in tandem with a classical bacteriophytochrome to control the antennae synthesis in Rhodopseudomonas palustris. J Biol Chem **280**:32389–32397.

19. **Hegyi H**, **Gerstein M**. 1999. The relationship between protein structure and function: a comprehensive survey with application to the yeast genome. J Mol Biol **288**:147–164.

20. **Heulin T**, **Barakat M**, **Christen R**, **Lesourd M**, **Sutra L**, **De Luca G**, **Achouak W**. 2003. Ramlibacter tataouinensis gen. nov., sp. nov., and Ramlibacter henchirensis sp. nov., cyst-producing bacteria isolated from subdesert soil in Tunisia. Int J Syst Evol Microbiol **53**:589–594.

21. **Hubschmann T**, **Jorissen HJMM**, **Borner T**, **Gartner W**, **De Marsac NT**. 2001. Phosphorylation of proteins in the light-dependent signalling pathway of a filamentous cyanobacterium. Eur J Biochem **268**:3383–3389.

22. **Im YJ**, **Rho SH**, **Park CM**, **Yang SS**, **Kang JG**, **Lee JY**, **Song PS**, **Eom SH**. 2002. Crystal structure of a cyanobacterial phytochrome response regulator. Protein Sci **11**:614–624.

23. **Jaubert M**, **Vuillet L**, **Hannibal L**, **Adriano JM**, **Fardoux J**, **Bouyer P**, **Bonaldi K**, **Fleischman D**, **Giraud E**, **Vermeglio A**. 2008. Control of peripheral light-harvesting complex synthesis by a bacteriophytochrome in the aerobic photosynthetic bacterium Bradyrhizobium strain BTAi1. J Bacteriol **190**:5824–5831.

24. **Jordan IK**, **Rogozin IB**, **Wolf YI**, **Koonin E V**. 2002. Essential Genes Are More Evolutionarily Conserved Than Are Nonessential Genes in Bacteria Essential Genes Are More Evolutionarily Conserved Than Are Nonessential Genes in Bacteria. Genome Res 962–968.

25. **Karniol B**, **Vierstra RD**. 2003. The pair of bacteriophytochromes from *Agrobacterium tumefaciens* are histidine kinases with opposing photobiological properties. Proc Natl Acad Sci USA **100**:2807–2812.

26. **Kumar S**, **Kateriya S**, **Singh VS**, **Tanwar M**, **Agarwal S**, **Singh H**, **Khurana JP**, **Amla DV**, **Tripathi AK**. 2012. Bacteriophytochrome controls carotenoid-independent response to photodynamic stress in a non-photosynthetic rhizobacterium, Azospirillum brasilense Sp7. Sci Rep **2**:872.

27. **Kumar S**, **Stecher G**, **Tamura K**. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol Biol Evol **33**:1870–1874.

28. **Laub MT**, **Goulian M**. 2007. Specificity in two-component signal transduction pathways. Annu Rev Genet **41**:121–145.

29. **Markowitz VM**, **Korzeniewski F**, **Palaniappan K**, **Szeto E**, **Werner G**, **Padki A**, **Zhao X**, **Dubchak I**, **Hugenholtz P**, **Anderson I**, **Lykidis A**, **Mavromatis K**, **Ivanova N**, **Kyrpides NC**. 2006. The integrated microbial genomes (IMG) system. Nucleic Acids Res **34**:D344–D348.

30. **Oberpichler I**, **Rosen R**, **Rasouly A**, **Vugman M**, **Ron EZ**, **Lamparter T**. 2008. Light affects motility and infectivity of Agrobacterium tumefaciens. Environ Microbiol **10**:2020–2029.

31. **Overbeek R**, **Fonstein M**, **Souza MD**, **Pusch GD**. 2014. The use of gene clusters to infer functional coupling. Proc Natl Acad Sci USA **96**:2896–2901.

32. **Psakis G**, **Mailliet J**, **Lang C**, **Teufel L**, **Essen LO**, **Hughes J**. 2011. Signaling kinetics of cyanobacterial phytochrome Cph1, a light regulated histidine kinase. Biochemistry **50**:6178–6188.

33. **Quast C**, **Pruesse E**, **Yilmaz P**, **Gerken J**, **Schweer T**, **Yarza P**, **Peplies J**, **Glöckner FO**. 2013. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. Nucleic Acids Res **41**:590–596.

34. **Rambaut A**. 2009. FigTree.

35. **Rogozin IB**, **Makarova KS**, **Wolf YI**, **Koonin E V**. 2004. Computational approaches for the analysis of gene neighbourhoods in prokaryotic genomes. Br Bioinform **5**:131–149.

36. **Schaeffer SE**. 2007. Graph clustering. Comput Sci Rev. University of Utrecht.

37. **Shah R**, **Schwach J**, **Frankenberg-Dinkel N**, **Gärtner W**. 2012. Complex formation between heme oxygenase and phytochrome during biosynthesis in Pseudomonas syringae pv. tomato. Photochem Photobiol Sci **11**:1026–31.

38. **Sievers F**, **Wilm A**, **Dineen D**, **Gibson TJ**, **Karplus K**, **Li W**, **Lopez R**, **McWilliam H**, **Remmert M**, **Söding J**, **Thompson JD**, **Higgins DG**. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol **7**:539.

39. **Snel B**, **Lehmann G**, **Bork P**, **Huynen MA**. 2000. STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. Nucleic Acids Res **28**:3442–4.

40. **Tars K**, **Rumnieks J**, **Zeltins A**, **Kazaks A**, **Kotelovica S**, **Leonciks A**, **Sharipo J**, **Viksna A**, **Kuka J**, **Liepinsh E**, **Dambrova M**. 2010. Crystal structure of human gamma-butyrobetaine hydroxylase. Biochem Biophys Res Commun **398**:634–639.

41. **Valencia A**, **Pazos F**. 2002. Computational methods for the prediction of protein interactions. Curr Opin Struct Biol **12**:368–373.

42. **Wojnowska M**, **Yan J**, **Sivalingam GN**, **Cryar A**, **Gor J**, **Thalassinos K**, **Djordjevic S**. 2013. Autophosphorylation activity of a soluble hexameric histidine kinase correlates with the shift in protein conformational equilibrium. Chem Biol **20**:1411–1420.

43. **Wu L**, **McGrane RS**, **Beattie GA**. 2013. Light regulation of swarming motility in Pseudomonas syringae integrates signaling pathways mediated by a bacteriophytochrome and a LOV protein. MBio **4**:e00334–13.

44. **Yanai I**, **Mellor JC**, **DeLisi C**. 2002. Identifying functional links between genes using conserved chromosomal proximity. Trends Genet **18**:176–179.

45. **Zhang R**, **Lin Y**. 2009. DEG 5.0, a database of essential genes in both prokaryotes and eukaryotes. Nucleic Acids Res **37**:455–458.

Chapter V:

Conclusions and Future Directions

The intention of this work was to elucidate BphP-BRR two component signal transduction and understand the signaling outcomes of the red light regulated TCSs in non-photosynthetic bacteria. Prior to this work, the atomic details of light-stimulated conformational regulation in photosensory portions of BphPs from phototrophs and chemotrophs had been well studied thanks in large part to crystallographic studies. However, the catalytic histidine kinase domains of BphPs were and remain structurally uncharacterized. The structures of BphP-associated response regulator proteins from phototrophs were also known, yet no BRRs from chemotrophic bacteria had been characterized structurally. Biochemical evidence for cognate BphP-BRR phosphotransfer existed for several phototrophic and chemotrophic bacterial systems. Phenotypes controlled by BphP-BRR systems in photosynthetic bacteria were consistent, with the expression of light harvesting machinery regulated by red light in those species. BphP-BRR associated phenotypes found in chemotrophic species were known but highly varied. This work has characterized the *R. tataouinensis* BphP-BRR system as a new model for the chemotrophic bacteria with a red light sensing TCSs. The previous chapters looked beyond signal transduction within the BphP molecule itself to understand the atomic details of the unique arm-in-arm BRR dimer, investigated novel post translational modifications of the BphP, and identified potential interaction partners for the BRR as well as potential red light controlled phenotypes for arm-in-arm BRR encoding species.

Conclusions of Chapter II

The RtBphP1 and RtBRR genes are adjacent and encoded in overlapping reading frames on the *R. tataouinensis* chromosome and were expected to form a TCS. RtBphP1 was characterized as a dimeric PAS-GAF-PHY-HK BphP which binds biliverdin as a chromophore and acts as an autokinase, catalyzing the hydrolysis of ATP and initial formation of His~P. The initial rate of autophosphorylation is slower, by twofold, when RtBphP1 is illuminated with red light. Dark-adapted RtBphP1 can transfer a phosphate group specifically to RtBRR to form

Asp~P. The crystal structures of RtBRR and AtBRR (from another chemotrophic bacterium) revealed these SDRRs are topologically linked into dimers which relies on a conserved hFWAhLNEPPP motif near the C-terminus. Truncation of RtBRR to exclude the motif and terminus reduced the efficiency of phosphotransfer between RtBphP1 and RtBRR, therefore the arm-in-arm dimer promotes phosphotransfer in this red light regulated TCS.

Conclusions of Chapter III

Beyond the histidine/aspartate phosphorylation expected and observed in TCS proteins, additional post translational modifications were detected on several full length BphPs including RtBphP1 and Agp1, each purified from *E. coli* protein expression cell lines. Serine phosphorylation was detected in RtBphP1 in two locations in the GAF domain flanking the A and D rings of the biliverdin binding site. Lysine acetylation was observed across all domains of BphPs, including the dynamic PHY domain tongue region and the histidine kinase catalytic core. Both serine phosphorylation and lysine acetylation were previously unknown for BphPs.

Conclusions of Chapter IV

The previously identified arm-in-arm dimer motif was used to predict the presence of similar dimeric BRRs in over a hundred bacterial species with putative BphPs, indicating that these signal transduction promoting oligomers are common players in red light signal transduction. This group of arm-in-arm bearing species are primarily α, β, and γ Proteobacteria and the majority are soil-dwelling chemotrophs. Gene neighborhood analysis of a larger group of BphP-encoding species including a subgroup of arm-in-arm species found that hybrid histidine kinase genes are frequently encoded in the vicinity of the classic TCS partners and suggest a role for branched signaling by BphPs. A gene clustering study identified potential interaction partners for the arm-in-arm BRR protein, particularly the uncharacterized DUF971.

Phenotypes indicated in the search are related to an oxidative stress response, DNA repair, and PHA synthesis.

The work shown here illuminates the red light TCS from non-photosynthetic bacteria biochemically and structurally and uses bioinformatics to generate hypothesis about the next steps in cellular signal transduction after BRR phosphorylation and red light controlled phenotypes. These findings here have also raised a number of new questions for researchers including but not limited to: the role of the arm-in-arm interface before, during, and after phosphotransfer; the effects of alternative PTM on BphP activities; and possible protein interactions with the arm-in-arm BRR.

Future directions: the arm-in-arm RtBRR dimer in signal transduction

Truncation of RtBRR to eliminate dimerization via the arm-in-arm motif resulted in reduced phosphotransfer efficiency between RtBphP1 and RtBRR. One possibility for this effect is that the local pool of BRR monomers in proximity of a BphP histidine kinase domain is enriched when each BRR molecule brings along a second molecule as part of the dimer. If this were the case, engineering RtBRR or another RR to dimerize by a different topological arrangement would also promote phosphotransfer and monomerization of natively dimeric RRs would decrease phosphotransfer efficiency. Another possibility is that the relative positions of the phosphoacceptor sites on each arm-in-arm protomer encourage simultaneous, and therefore more rapid phosphorylation of the BRR by multiple BphP molecules. A third possibility is that the arm-in-arm dimer, and not the monomer, is recognized specifically by some region of the BphP protein, and this specific interaction contributes to promotion of phosphotransfer. A crystal structure of RtBphP1 and RtBRR in complex would be extremely valuable in understanding the relationship between the arm-in-arm interface and signal transduction. How phosphorylation affects the arm-in-arm BRR in terms of conformational signaling is also an open

question. The protein interface statistics generated from the crystal structure of RtBRR did not suggest phosphorylation would break the highly hydrophobic interface between the two protomers and preliminary gel filtration results show that RtBRR incubated with RtBphP1 and ATP remains a dimer. Therefore, subtler effects on the protein's conformation are suggested to result from phosphorylation. As discussed in Chapter 1, the "switch" tyrosine which adopts two different conformations in CheY in response to phosphorylation is not occluded by the dimer interaction in the arm-in-arm BRR and has not been ruled out as a possible signaling mode for this protein. Crystals of RtBRR grown for this study were soaked with the phosphate analog BeF_3 in order to produce a phosphorylated-like conformation, but density for BeF_3 was not detected in the resulting electron density map. New crystal growth conditions for BeF_3-treated RtBRR would need to be screened for and optimized in order to generate an informative structure to answer questions about conformational changes in response to phosphorylation.

Future directions: novel posttranslational modifications of BphPs

Chapter three described the discovery of phosphorylated serine and acetylated lysine residues on BphPs purified from *E. coli*. These PTMs are new information for the BphP field, which up to now has focused on histidine and aspartate phosphorylation as the only known PTMs in BphP-BRR systems. Novel PTMs were detected on the full-length BphPs RtBphP1 and Agp1 by all methods (Phos-tag gels, Western Blot, and MS/MS). Novel PTMS were also indicated by one or more method on RtBphP2 and DrBphP. The finding that several BphPs bear these PTMS, coupled with the dominance of using *E. coli* protein expression systems in the field, raises the concern that full-length BphPs which have been spectrally and enzymatically characterized to date also bear phospho-serine and acetyl-lysine modifications. Depending on the results of experiments that determine the effect of PTMs on enzyme activity, new strategies may be recommended for BphP expression and purification that eliminate these PTMs. One strategy is to express BphPs in acetyltransferase and Ser/Thr kinase gene knockout strains of

E. coli, which are already constructed and available (Starai, 2002; Baba, 2006). This strategy may not be effective if small molecule donors such as acetyl phosphate are responsible for phosphorylation and acetylation of BphPs in the *E. coli* cytoplasm. Therefore, production of modified BphPs in wild type *E. coli* followed by enzymatic dephosphorylation by lambda phosphatase and deacetylation by sirtuin may be preferable and work in the Forest lab has begun optimizing this method. A third strategy is the production of BphPs in a cell-free system such as wheat germ *in vitro* translation, although producing proteins at large scale by this method is limited by resources and cost. These methods would need empirical testing of the produced BphPs by Western Blot and/or MS/MS to verify the absence of serine phosphorylation and lysine acetylation before proceeding to enzymatic characterization.

After production of experimentally verified unmodified full length BphPs, experiments to compare the enzymatic activities of modified and unmodified BphPs can be conducted to determine what, if any, effect novel PTMs have. Based on the mapping of discovered PTMs to the homology model of full-length RtBphP1 presented in Chapter 3, the following activities are potentially affected and should be characterized: biliverdin binding; photoconversion; refolding of the PHY tongue region; ATP binding; autophosphorylation; and phosphotransfer to BRR. Methods for evaluating each activity include: zinc-fluorescence gels; UV-Vis spectrophotometry; circular dichroism; DRaCALA with a fluorescently labeled or radiolabeled ATP molecule; ^{32}P labeling and autoradiography; and Phos-tag gel analysis respectively. The Forest lab is experienced in most of these methods or has access to collaborations with labs with experience. Experiments characterizing RtBphP1 in Chapter 2 found that biliverdin bound to only 78% of protein molecules, and phosphorylation of serine at two sites in the biliverdin binding site were detected in Chapter 3, therefore a hypothesis that unmodified RtBphP1 will bind biliverdin to a greater degree is proposed. Additionally, the kinetics of autophosphorylation indicated that RtBphP1 autophosphorylation is just twofold suppressed by red light, and still

becomes actively phosphorylated when illuminated. This could indicate a defect in dark-state autophosphorylation activity, or an increase in illuminated state activity attributed to lysine acetylation in the histidine kinase region of BphP. Therefore, a hypothesis that the difference in red light regulation of autokinase activity will be greater in unmodified full length BphPs is proposed.

A final question about novel PTMs in full-length BphPs remains: does serine phosphorylation and lysine acetylation also occur in the source organisms for these proteins? Specifically, does RtBphP1 or Agp1 expressed natively in *R. tataouinensis* or *A. tumefaciens* also display the same unexpected PTMs? Is there a possible regulatory role for serine phosphorylation and/or lysine acetylation *in vivo*? In order to address the first question, work has begun in the Forest lab to generate custom polyclonal antibodies to full length RtBphP1. Briefly, purified RtBphP1 recombinantly expressed in *E. coli* was submitted to GeneTel Laboratories, Madison, WI, for immunization of egg-laying hens to produce chicken polyclonal antibodies specific to RtBphP1. Soluble whole cell lysates from *R. tataouinensis* cultures can then be assayed by standard SDS-PAGE separation, Phos-tag gel separation, and Western blot with anti-RtBphP1, anti-phosphoserine, and anti-acetyl lysine probes to determine if RtBphP1 bands also bear novel PTMs. If BphP expression from whole cells is too low to detect, the same custom antibodies may be used to enrich *R. tataouinensis* lysates for BphPs using a co-immunoprecipitation strategy.

Future directions: BphP and BRR protein-protein interaction studies

Chapter 4 utilized complementary bioinformatic techniques to identify putative interaction partners that may be part of the greater red light signal transduction network. The next steps involve experimental evaluation of protein-protein interactions between known the TCS players BphP and BRR and these potential partners. Hybrid histidine kinase genes were found in the

majority of gene neighborhoods with BphP-BRR pairs, and most of these HHKs encode an N-terminal receiver domain with a putative phosphoacceptor aspartate. The HHK from the RtBphP1 gene cassette should be cloned, expressed, and purified both as the full length protein and as the isolated N-terminal and C-terminal receiver domains. These constructs can then be subjected to phosphotransfer assays with RtBphP1. If RtBphP1 can phosphorylate the N-terminus of the HHK an addition to its known phosphotransfer to RtBRR, competition assays can be conducted to determine which RR and under which conditions (dark, light) is preferentially phosphorylated by the BphP. Depending on the outcome of the RtBphP1-HHK interaction studies, biochemical characterization can proceed to determine what signal may be sensed by the HHK protein to contribute to integrated signaling by light and a second message.

Protein-protein interaction experiments between RtBRR and RtDUF971 should be conducted, as this gene was the top hit (other than known BRR interaction partners) in the gene clustering analysis. The observation that RtBRR and RtDUF971 have complementary electrostatic surface potential suggests these two may form a stable interaction. If an interaction is verified, the problem remains that the function of DUF971 proteins is unknown in any system. However, this would still represent the first finding of a signal transduction partner beyond the BphP-BRR pair for any non-photosynthetic red light TCS and therefore remains a valuable experimental target. Similar protein-protein interaction studies may also be conducted between RtBRR and the redox stress proteins and PHA synthase identified in the gene cluster analysis.

General observations and advice

A common theme to much of the work presented in this dissertation would be: expect the unexpected when investigating something seemingly as simple as a two component signal transduction system. For example, the structure of RtBRR was not expected, based on its primary amino acid sequence, to reveal a quaternary fold that is both new to the BphP field and

new to science. Serine phosphorylation and lysine acetylation were not expected to be part of the BphP story by us or by other researchers, but these PTMs may have enormous impact on the field. Hybrid histidine kinase genes were not by any means hidden in the proximity of BphP and BRR genes, but it took a careful analysis to reveal their prevalence and possible role in red light signaling. To researchers who undertake the next steps in these lines of research, follow the data where they lead despite any preconceived expectations while maintaining a hypothesis driven approach and scientific rigor.

References

1. **Baba T**, **Ara T**, **Hasegawa M**, **Takai Y**, **Okumura Y**, **Baba M**, **Datsenko KA**, **Tomita M**, **Wanner BL**, **Mori H**. 2006. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol **2**:2006.0008.

2. **Starai VJ**, **Celic I**, **Cole RN**, **Boeke JD**, **Escalante-semerena J c**. 2002. Sir2-dependent activation of Acetyl-CoA Synthetase by deacetylation of active lysine. Science **298**:2390–2392.