Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Paul Young April 10, 2018

Elucidating the Structure and Function of Heme Pocket Residues in Bacterial Diguanylate Cyclase-Containing Globin-Coupled Sensors

by

Paul Young

Dr. Emily Weinert, PhD
Adviser

Department of Biology

Dr. Emily Weinert, PhD

Adviser

Dr. Cassandra Quave, PhD

Committee Member

Dr. José Soria, PhD

Committee Member

2018

Elucidating the Structure and Function of Heme Pocket Residues in Bacterial Diguanylate Cyclase-Containing Globin-Coupled Sensors

By

Paul Young

Dr. Emily Weinert, PhD

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Sciences with Honors

Department of Biology

2018

Abstract

Elucidating the Structure and Function of Heme Pocket Residues in Bacterial Diguanylate Cyclase-Containing Globin-Coupled Sensors

By Paul Young

The ability to respond and adapt to the environment is a hallmark of biological systems. Globin-coupled sensors are one method bacteria have evolved to respond to the important environmental stimulus, molecular oxygen (O_2). To investigate the structure and signaling mechanisms of these proteins, the crystal structure of the globin domain from the diguanylate cyclase-containing globin-coupled sensor (DGC-GCS), *Bpe*GReg, has been solved in its O_2-bound state, making this the first bacterial sensor globin to be characterized in its native active state. Furthermore, key information about the role of heme pocket residues in ligand interaction has been obtained through Fourier-transform infrared spectroscopy (FTIR) on CO-bound *Bpe*GReg and the closely related *Pcc*GCS, which identified a tyrosine residue in the heme distal pocket as the primary hydrogen bond donor in both proteins. Additionally, FTIR data revealed an ordered water molecule within the heme pocket of *Bpe*GReg as an additional, previously unrecognized ligand interacting partner. Finally, the proximal histidine residue was identified as a major regulator of protein oligomeric state, as well as a contributor to ligand stability. These findings highlight the roles of heme pocket residues in modulating overall DGC-GCS structure and function.

Elucidating the Structure and Function of Heme Pocket Residues in Bacterial Diguanylate Cyclase-Containing Globin-Coupled Sensors

By

Paul Young

Dr. Emily Weinert, PhD

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Sciences with Honors

Department of Biology

2018

Acknowledgements

Shannon Rivera

Graduate Student Mentor

Benjamin Fontaine

Graduate Student Mentor

Weinert Lab Members

Dr. Stefan Lutz, PhD

Faculty Mentor

Lutz Lab Members

Table of Contents

Introduction

Bacteria have evolved a variety of strategies to translate environmental stimuli into phenotypic outputs, such as the formation of biofilms, which are communities of sessile bacteria structured in a self-synthesized polymeric matrix attached to living or inanimate surfaces.[1] The ability to form these communities evolved as a mode of protection for bacteria growing in inhospitable environments, as the dense structure of macromolecules provides a barrier to environmental stressors, such as antibiotics and host immune responses.[1] Thus, biofilm formation is a contributor to antibiotic resistance.[1-2] Indeed, biofilms are observed in 80% of chronic bacterial infections, which in addition to presenting individual health concerns also cost more than $18 billion in medical expenses each year.[2] Humans as well as plants are affected by these communities, as *Pectobacterium carotovorum* ssp. *carotovorum*, a pathogen responsible for soft rot in vegetables and fruit, incorporates biofilm formation as an essential step in its pathogenesis.[3-4] Consequently, a more comprehensive understanding of biofilms and the events initiating their formation is necessary, as a matter of both public health and cost-efficiency.

Under electron microscopy, biofilms appear to be branched structures attached to bacterial cell surfaces.[5] Analysis reveals that these branches are secreted exopolysaccharides (EPS), consisting primarily of mannose, galactose, and glucose, which serve as scaffolds for proteins, nucleic acids, and lipids.[5] While formation of EPS is not specific to a single type of bacterial biofilm, the specific carbohydrates present in a given biofilm can vary depending on the bacterial species.[5] For example, *Pseudomonas aeruginosa* biofilms are associated with EPS consisting of Pel, Psl, and alginate, which are mannose-, glucose-, and mannuronic acid-rich polysaccharides, respectively.[7] Alginate demonstrates the need for greater knowledge of biofilm chemistry, as it is critical to *P. aeruginosa* virulence and confers antibiotic- and immune-resistance.[5] Furthermore,

the durability of biofilms can be partially attributed to EPS-associated proteins, such as Fap amyloids in *Pseudomonads* and TasA fibers in *Bacillus subtilis*, which are able to tolerate harsh external forces as well as enhance biofilm growth.[5] In addition to providing antibiotic resistance through limiting the diffusion of molecules, biofilms grant genetic antibiotic resistance by enhancing conjugation of antibiotic resistance-carrying plasmids due to the increased effective concentration of plasmid donors and acceptors.[5]

c-di-GMP Production and Regulation

Integral to understanding biofilms and their formation is the small molecule bis-(3',5')-cyclic dimeric guanosine monophosphate (c-di-GMP), which is the second messenger responsible for regulating bacterial community behaviors, such as cell adhesion, secretion, motility, and biofilm formation.[4, 6-7] For example, cellulose—a common component of biofilms—is synthesized in a c-di-GMP-dependent manner in *Gluconacetobacter xylinus*, with c-di-GMP inducing cellulose synthase activity by binding to a PilZ domain on the glycosyltransferase protein of the cellulose synthase complex.[6] Furthermore, c-di-GMP can bind transcription factors, such as FleQ, a major regulator of flagellar proteins in *P. aeruginosa*, as well as RNA aptamers responsible for regulating biofilm-associated response elements.[6]

Given the importance of c-di-GMP in regulating bacterial lifestyles, it is unsurprising that mechanisms have evolved to keep its intracellular concentration tightly controlled. In general, intracellular levels of c-di-GMP are modulated in two ways: by a negative feedback loop and the opposing activities of diguanylate cyclases (DGCs) and specific phosphodiesterases (PDEs).[6, 8] DGCs are the enzymes responsible for the synthesis of c-di-GMP from two molecules of GTP, with the enzyme active site at the dimer interface of two DGC domains, each containing a GGDEF (Gly-Gly-Asp-Glu-Phe) active site motif.[6, 8] Through genomic analysis, this motif has been found

in 75% of bacterial species, demonstrating the conservation of c-di-GMP signaling.[9] The mechanism of DGCs is similar to other nucleotide cyclases, in which the 3'-OH of one GTP is deprotonated by the Glu residue to enable an intermolecular nucleophilic attack on the α-phosphate of the other GTP (Figure 1).[10] This reaction is made more favorable by the presence of a Mg^{2+}ion in each monomer, which is thought to coordinate to the β- and γ-phosphates of the bound GTP molecules, favoring nucleophilic attack from the 3'-OH.[10] While older studies suggested that all residues of the active site motif are necessary for DGC activity, more recent work has revealed that some bacteria have functional, though promiscuous DGCs with degenerate active site motifs.[8-10] Nonetheless, the importance of this general motif for c-di-GMP production has been demonstrated by site-directed mutagenesis studies.[10-11]

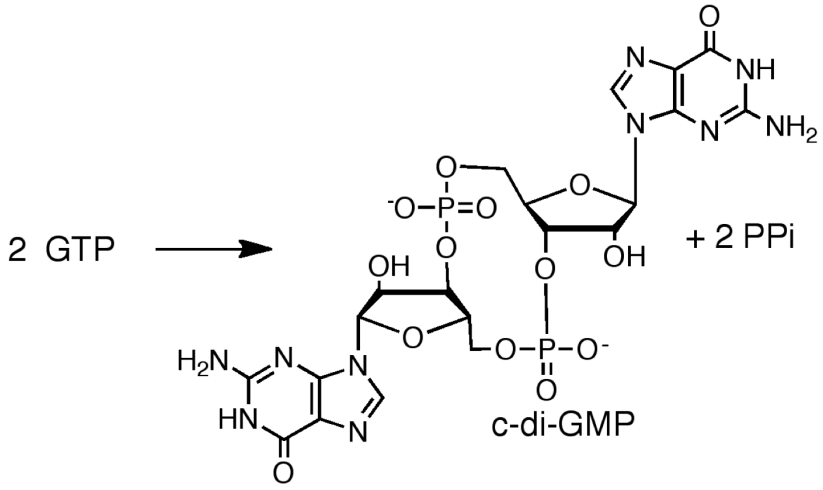

Figure 1 Following coordination of a GTP molecule to the Mg^{2+} of each DGC monomer, a glutamate residue on each monomer deprotonates the 3'-OH of the corresponding GTP, allowing nucleophilic attack from the resulting oxyanion to the α-phosphate of the other GTP and vice versa. This results in the formation of c-di-GMP and two pyrophosphate moieties.

Furthermore, DGCs often contain an allosteric site, termed the I-site, characterized by a RXXD (Arg-X-X-Asp) motif, where "X" can be any amino acid.[8] Studies on the DGC WspR have demonstrated that binding of c-di-GMP to the I-site induces a shift to a transient, partially inactive tetramer, followed by the formation of elongated dimers, which prevents the reaction from taking place across the dimer interface.[8, 12] This feedback inhibition prevents excessive GTP metabolism and sets a limit on c-di-GMP concentrations that acts as a buffer against stochastic variations in c-di-GMP levels.[8]

Along with the negative feedback loop caused by the I-site of DGCs, c-di-GMP levels are also controlled through the activity of PDEs that selectively hydrolyze c-di-GMP into linear pGpG.[6] These proteins are characterized by one of two domains, either the EAL (Glu-Ala-Leu) or HD-GYP (His-Asp-Gly-Tyr-Pro) domain, which are named for their respective active site motifs.[6] Both are often arranged in tandem with GGDEF domains as part of the mechanism that controls c-di-GMP levels.[6] The role of the EAL domain as a c-di-GMP-specific phosphodiesterase was elucidated by *in vitro* characterization of the VieA protein from *Vibrio cholerae*, which showed that it hydrolyzes c-di-GMP, as well as *in vivo* characterization, in which its elimination increases biofilm formation and its overexpression inhibits biofilm formation.[6, 13-14] Likewise, the HD-GYP domain PDEs exhibit hydrolytic activity toward c-di-GMP.[6] However, unlike EAL domain PDEs, HD-GYP PDEs form GMP as their end-product, not pGpG.[6] It is unclear if the enzyme releases pGpG as an intermediate and simply rebinds it, or if pGpG is not released and is instead immediately hydrolyzed further to GMP. Nevertheless, structural studies on the HD-GYP PDE from *Bdellovibrio bacteriovorus* have shown that the mechanism is likely similar to other PDEs, in that it activates the phosphodiester bond through bimetallic coordination, followed by nucleophilic attack from a metal-generated hydroxide.[6, 15]

Globin-Coupled Sensors

To be effective as a second messenger, c-di-GMP production must be enhanced in response to certain external stimuli, such as the presence or absence of specific molecules. One of these signals that induces c-di-GMP production and subsequent biofilm formation is gaseous oxygen in the environment.[16] In order to sense oxygen, bacteria have evolved DGC-linked protein sensors with non-covalently-bound heme cofactors, which may be held within a number of sensory domain folds, including PAS, GAF, and globin.[16] Bacterial globin-coupled oxygen sensors contain the heme cofactor within the eponymous globin fold, which is connected through a middle domain to an output domain, which may be one of several enzymes, including methyl-accepting chemotaxis proteins (MCP), phosphodiesterases, and diguanylate cyclases (DGC).[16-17] Previous work has shown that the diguanylate cyclase-containing globin-coupled sensors (DGC-GCS) from *E.* coli (*Ec*DosC),[18-20] *Desulfotalea psychrophila* (HemDGC),[21] *Shewanella putrefaciens* (HemDGC),[22] *Bordetella pertussis* (*Bpe*GReg),[4, 23] and *Pectobacterium carotovorum* ssp. *carotovorum* (*Pcc*GCS)[4] exhibit increased cyclase activity upon O_2 binding, confirming these proteins as bacterial oxygen sensors.[16] The existence of DGC-GCS proteins implies the importance of gaseous oxygen as a signal in mediating the transition in bacteria from the motile, virulence-associated planktonic form to the sessile biofilm form responsible for chronic infection. As virulence is costly for bacteria, DGC-GCS proteins may have evolved in part to signal conditions most conducive to this phenotype.[17, 24] While studies have revealed some of the intracellular effects these proteins have on factors such as motility,[6, 8, 17, 23] virulence,[6, 8, 17] and biofilm formation,[6, 10, 17, 22] the mechanism by which the oxygen-binding signal is propagated from the globin domain to the DGC output domain remains unknown.[17] A better understanding of this mechanism would allow insight into new methods of preventing biofilm formation and provide drug targets for the treatment of chronic disease.

The close relationship between structure and function in biology makes an understanding of DGC-GCS proteins structure vitally important to elucidating the signaling mechanism. To date, full-length crystal structures of GCS proteins have not been solved, but crystal structures of GCS globin domains have revealed homodimeric proteins exhibiting myoglobin-like folds with a four-helical bundle at the dimer interface.[20, 25-27] While DGC-GCS proteins show relatively low sequence similarity, some amino acid residues within the heme pocket are conserved.[25] One of these residues, the proximal histidine located on the F8 helix, is absolutely conserved across DGC-GCS proteins as the proximal ligand coordinated to the heme iron atom, tethering the heme cofactor to the protein.[25, 28] In addition to the proximal histidine, all characterized DGC-GCS proteins contain a distal tyrosine, which acts as a hydrogen bond donor to bound ligands.[25] Furthermore, while not as conserved as the distal tyrosine or proximal histidine, many DGC-GCS contain a second distal pocket residue capable of donating an hydrogen bond, such as serine or threonine (Figure 2).[25] While there is evidence that the heme pocket residues are involved in the signaling mechanism of DGC-GCS, their exact roles remain unclear.

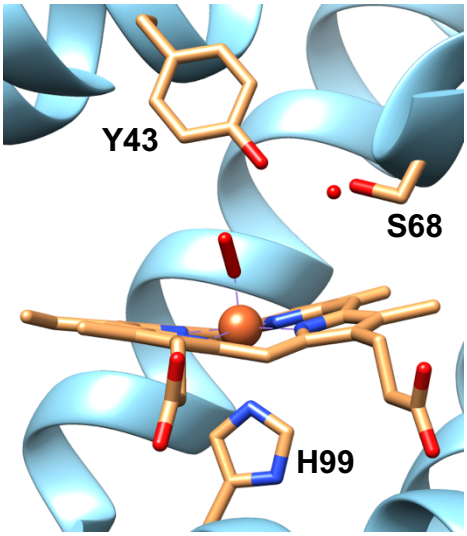

Figure 2 The heme pocket of *Bpe*GReg contains the two H-bond donors, tyrosine and serine on the heme distal side, as well as the conserved proximal histidine, which forms a connection between the heme and the protein scaffold by a coordination bond to the heme Fe atom.

One of the methods bacterial GCS proteins utilize to affect the activity of their output domains upon ligand binding is altering the protein oligomeric state.[4, 22, 29] Given that dimerization is necessary for cyclase activity, previous work on DGCs like WspR from *C. crescentus* and PleD from *Pseudomonas fluorescens* has demonstrated that activation of these proteins induces conformational changes that result in dimeric complexes.[25] While these are not examples of DGC-GCS proteins, crystal structures of isolated globin domains have suggested that DGC-GCSs undergo a similar mechanism mediated by the sensor globin domain.[25] Indeed, most of these globin crystal structures were isolated and examined in the Fe^{II} (ferrous unligated) and Fe^{III}-CN (ferric cyano) to yield important comparative information about the inactive and "active-like" conformations of these proteins.[25] The results of these studies suggest that ligand binding results in small heme pocket conformational rearrangements that are propagated to the globin-dimer interface, providing a potential mechanism for a ligand-mediated oligomeric equilibrium shift.[25] These findings have been corroborated by hydrogen/deuterium exchange coupled with mass spectrometry (HDX-MS) that found differences in solvent exposure of the dimer interface dependent on heme ligation.[29] Furthermore, Burns et al. demonstrated that addition of O_2 to *Pcc*GCS and *Bpe*GReg led to shifts in the oligomer equilibrium toward the more active tetrameric assembly, implying a direct link between oxygen binding, oligomeric state, and cyclase activity.[25] Taken together, these studies suggest that the heme pocket residues interact with the bound ligand in such a way that the three-dimensional structure of the protein is altered, thereby affecting the oligomeric states and output domain activity.

The potential importance of the heme distal pocket tyrosine in GCS signaling mechanisms has been demonstrated by a number of studies. Resonance Raman spectroscopy (RR) on O_2-bound HemAT-*Bs* and CO-bound HemDGC distal pocket mutants indicate that the tyrosine exists in two

states, an open or closed one, a characteristic that is thought to aid these proteins in ligand discrimination.[21, 30] This finding shows that this residue has a degree of mobility that allows it to respond to ligand binding, making it a candidate for a ligand-responsive signaling mechanism. Rivera et al. (2016) demonstrated the importance of the distal tyrosine residue as the primary hydrogen bond donor in both *Pcc*GCS and *Bpe*GReg.[31] In mutants with a Tyr→Phe mutation, O_2 binding resulted in rapid autoxidation, indicating that the distal tyrosine is necessary for stable ligand binding.[21] These findings are in agreement with work by Kitanishi et al. (2010), which showed enhanced O_2 dissociation and autoxidation of EcDosC (Y43F).[18] Furthermore, work on *Ec*DosC also showed shifts in the RR spectra of of O_2-, CO-, and NO-bound EcDosC (Y43F) when compared to the corresponding wild type spectra, indicating electrostatic interaction between the tyrosine and bound ligands, further implicating this residue in an O_2-dependent signaling mechanism.[18]

Many GCS proteins also have a second distal pocket residue capable of donating a hydrogen bond within the heme pocket of their sensor domains. In HemAT-*Bs*, RR studies on mutants lacking this secondary hydrogen bond donor, in this case the T95A variant, show a decrease in the number of isotope-sensitive Fe-O_2 stretching frequencies ($\nu_{Fe\text{-}O2}$) compared to the WT spectra, indicating that this residue forms hydrogen bonds with the O_2 ligand.[30] In this protein, it is believed that T95 and a water are responsible for maintaining the closed, ligand-bound state, while the distal tyrosine is necessary for distinguishing the ligation states for signaling purposes.[30] Similarly, studies on the sensor globin domains of *Bpe*GReg (*Bpe*Globin) and *Pcc*GCS (*Pcc*Globin) found that mutation of the distal serine residue, S68A and S82A, respectively, resulted in loss of biphasic dissociation kinetics observed in the dimerized wild type proteins.[31]

These important roles for the distal serine in ligand binding implicate it as another potential member of an intra-protein signaling mechanism.

In addition to distal pocket residues, the proximal pocket is also important for ligand binding and potentially for signal transduction. The heme iron atom can be hexa- or penta-coordinated, with four coordination bonds coming from the four pyrrole groups of the porphyrin, the fifth from the conserved proximal histidine residue, and the sixth from the bound ligand.[26] This proximal histidine serves as the covalent tether between the heme cofactor and the larger protein, but may also play a role in intra-protein signal transduction. Work on hemoproteins has shown that heme cofactors are not only often nonplanar, but that this non-planarity can influence protein activity.[32] Furthermore, studies on the PAS-fold hemoprotein FixL have determined a correlation between Fe-His bond distance and output domain activity.[29] Taken together, these findings imply a mechanism for DGC-GCS signaling by which ligand binding lengthens the Fe-His bond length, leading to heme distortion.[29] This distortion would in turn affect the surrounding heme pocket, resulting in steric repulsion from the heme propionates that could give rise to global protein rearrangements and potentially activate the cyclase domain.[29] Additionally, alterations to the Fe-His bond length could produce subtle rearrangements in the F8 helix, which is further connected to the G- and H-helices that comprise the dimer interface. Thus, perturbations to the F8 helix could then affect the dimer interface, driving the proteins to more active oligomers, such as dimers and tetramers.

PccGCS and BpeGReg

The DGC-GCS proteins of *Pectobacterium carotovorum* ssp. *carotovorum* and *Bordetella pertussis* are of particular interest, as these species are the causative agents of the plant disease soft rot and whooping cough, respectively.[4, 17] Notably, soft rot results in several billion dollars-worth

of crop loss in the world each year due to the lack of treatment for *P. carotovorum* infection.[4] Therefore, knowledge of the mechanisms by which these pathogens form biofilms would provide valuable drug targets for the prevention of disease (*B. pertussis*) and preservation of crops (*P. carotovorum*). These two proteins are often compared for their similar lengths, molecular weights, and sequence similarity, which is greater than other DGC-GCS pairs. Indeed, homology models suggest that the heme pockets of both *Pcc*GCS and *Bpe*GReg contain the proximal histidine, as well as tyrosine and serine residues as hydrogen bond donors on the heme distal side.[25] These models have been further validated by UV/visible spectroscopy, which shows spectra characteristic of histidyl-ligated heme proteins, as well as mutagenesis studies that reveal roles for the distal residues in ligand binding.[31] Furthermore, previous work has revealed that wild type *Pcc*GCS exists predominantly as a tetramer, but with octamer and dimer species present, as well.[4] *Bpe*GReg, on the other hand, has been found to be predominantly in a dimerized state.[4] For both proteins, the terameric state was found to be the most active.[4] The present work seeks to further characterize these two proteins through structural analysis and kinetic studies. Specifically, this thesis reports data obtained through crystallography, Fourier transform infrared spectroscopy (FTIR), UV/visible spectroscopy, and high-performance liquid chromatography (HPLC). Together, these approaches offer a powerful method of obtaining and relating structural and kinetic data to generate a more comprehensive signaling mechanism for these two bacterial oxygen sensors.

Methods

Site-Directed Mutagenesis

Codon-optimized genes encoding *Bpe*GReg and *Pcc*GCS in pET-20b[4] served as templates for site-directed mutagenesis. Mutagenic primer sequences can be found in the Supplmentary

Information Table S1. Isolated globin and full-length distal pocket mutants were produced by standard PCR protocol as previously described.[31] PCR mutagenesis conditions were as follows: 95ºC for 1:00 min; (95ºC for 0:50 min, 55ºC for 0:40 min, 68ºC for 6:00 min), repeated 16 times; 68ºC for 7:00 min.

Protein Expression

Proteins were expressed in *E. coli* Tuner (DE3) pLysS cells (Novagen) as previously described.[4, 25, 31] Briefly, codon-optimized pet20B plasmids encoding each protein were transformed via heat shock into *E. coli* Tuner (DE3) pLysS cells, which were grown overnight on LB agar plates inoculated with chloramphenicol (30 μg mL^{-1}) and ampicillin (100 μg mL^{-1}). Positive transformants grown on these plates were then grown in LB liquid culture with the appropriate antibiotics. These overnights were then used to inoculate yeast extract expression media and allowed to grow at 37ºC, shaking at 225 rpm until they achieved an OD_{600} of 0.6. Media for the expression of histidine mutants was supplemented with 10 mM imidazole. At this point, the temperature was dropped to 25ºC and ∂-aminolevulinic acid was added to give a final concentration of 50 mM in each expression mixture and left to shake. After 30 minutes, protein expression was induced by the addition of 100 mM IPTG. Cells were left to express each protein; full length proteins for 6 hrs and isolated globin proteins overnight. Cells were harvested by centrifugation (3500g, 4ºC, 20 min) and the resulting pellets were collected and stored at -80ºC.

Protein Purification

Cell pellets were resuspended in low imidazole buffer (50 mM Tris, 50 mM NaCl, 1 mM DTT, 20 mM imidazole, pH 7.0) and lysed using a homogenizer (Avestin). Proteins were purified from the lysate by centrifugation (40,000 rpm, 4ºC, 1 hr). The supernatants were loaded onto an equilibrated HisPur Ni-column (Fisher Scientific), washed with low imidazole buffer, and proteins

were eluted with high imidazole buffer (50 mM Tris, 50 mM NaCl, 1 mM DTT, 250 mM imidazole, pH 7.0). Purified proteins were further purified and desalted using a S200 gel filtration column into AGF buffer (50 mM Tris, 50 mM NaCl, 1 mM DTT, 5% glycerol (v/v), pH 7.0). All buffers used in the purification of histidine mutants were supplemented with 10 mM imidazole. Proteins were concentrated by ultrafiltration (10 kDa MWCO filter, Millipore), aliquoted, flash-frozen, and stored at -80ºC.

BpeGlobin Crystallization

Purified protein was reduced using sodium hydrosulfite under an anaerobic atmosphere (Coy Labs). *Bpe*Globin was then desalted, oxygenated by mixing with aerobic buffer, and concentrated to 5 mg/mL. Crystal trays were set up using a gradient of polyethylene glycol (PEG) 3350, 0.2 M calcium acetate, 10 mM tris, pH =7.0. Protein was mixed with the crystallization solution in equal and increasing amounts (i.e. 1+1, 2+2, etc.) in both hanging drop and sitting drop fashion. Crystals were grown at 20°C over several days before larger crystals were selected. Selected crystals were cryoprotected and flash frozen. All crystals were diffracted at Argonne National Laboratory in the insertional device synchrotron.

FTIR Characterization

Purified proteins were reduced using sodium hydrosulfite under anaerobic atmosphere (Coy Labs) and buffer exchanged into anaerobic buffer (50 mM Tris, 50 mM NaCl, 5% glycerol (v/v), pH 7.0). All proteins were then concentrated to >500 μM before being placed into sealed reactive vials and purging the headspace with CO gas. UV-Vis spectra of all proteins were taken before FTIR runs to ensure proteins were fully ligated. Protein and reference buffer were then placed in liquid FTIR cells with CaF_2 windows and run on either a Varian 660 FTIR (Thermo Fisher) or on Nicolet iS10 (Thermo Fisher).

The Nicolet iS10 was run with no additional modifications to the instrument. Sample and reference spectra were collected as an average of 2048 scans run with a 4 cm^{-1} resolution. The Varian 660 FTIR spectra were collected on a home built, dry air purged external box connected to the FTIR spectrometer and equipped with a liquid nitrogen cooled MCT detector. Sample and reference spectra were collected as an average of 2048 scans run with a 2 cm^{-1} resolution. For all spectra, the absorbance spectra ($-\log(I_{protein}/I_{buffer})$) were baseline corrected with a spline function.

$H_2{}^{18}O$ Buffer Exchange

1 mL of protein was brought into the anaerobic chamber and split into 2 equal aliquots. 50 µL of 500 mM sodium dithionite (DT) was added into both aliquots. 500 µL of 50 mM Na-P, 50 mM NaCl, and 1 mM DTT in $H_2{}^{18}O$ was added to 1 aliquot and allowed to buffer exchange overnight on a 4 °C block in the anaerobic glove bag. The $H_2{}^{18}O$ sample was then concentrated from 1000 µL to ~200 µL using a 5 kDa spin concentrator (Vivaspin) and additional $H_2{}^{18}O$ buffer was added to make a 500 µL aliquot. The aliquot was then allowed to sit for an additional 1-2 hours before repeating the concentrating-diluting one more time. After the second dilution, the protein was allowed to sit for 0.5-1 hr and then concentrated to ~50-100 µL. The control sample (protein in H_2O) was treated in the same manner except with 50 mM Tris, 50 mM NaCl, and 1 mM DTT in H_2O. Both samples were then placed in 1 mL sealed reactive vials, removed from the glove bag and exposed to CO gas by purging the headspace of the vials. UV-Vis spectra were taken before and after FTIR runs to ensure the protein samples were still Fe^{II}-CO.

Ultraviolet/Visible Spectroscopy

All spectra were recorded using an Agilent Cary 100. Prior to recording, all proteins were reduced with sodium hydrosulfite. For spectra taken from proteins in the Fe^{II}-O_2 or Fe^{II}-CO states,

reduced proteins were placed in reactive vials and the headspace was purged with the appropriate gas. For spectra taken from proteins in the Fe^{II}-NO state, proteins were exposed to 5 mM DEA NONOate.

High-performance Liquid Chromatography

Oligomeric states of *Bpe*GReg (H99G) and *Pcc*GCS (H112G) were detected via high-performance liquid chromatography using an Agilent 1200 infinity system with a Sepax SEC-300 (7.8 mm x 300 mm, 300 Å) and diode array detector (simultaneous detection at 214, 416, and 431 nm), as previously described.[4] Proteins were reduced in an anaerobic chamber and allowed to bind O_2 after mixture with aerobic buffer prior to injection onto the SEC-300 column. For all experiments, the mobile phase consisted of 150 mM sodium phosphate, pH 7.0 (AGF Buffer), with 1 mM DTT. Globular proteins (Sigma-Aldrich) consisting of thyroglobulin (669 kDa), ferritin (443 kDa), β-amylase (200 kDa), alcohol dehydrogenase (150 kDa), bovine serum albumin (66 kDa), and conalbumin (29 kDa) were used as molecular weight standards for calibration curves.

Results

BpeGlobin Crystal Structure

The isolated globin domain of *Bpe*GReg was crystallized at 5 mg/mL and solved to 2.3 Å (Table 1). The resulting crystal formed a C 1 2 1 space group in a parallelogram spaced crystal with three monomeric units of *Bpe*Globin held together by salt bridges along helices A, B and E (Table 1; Fig. 2 A). Interestingly, the trimer in the unit cell did not show interactions along helices G and H, the classic globin-dimer interface observed in all previously isolated sensor globin domains. Rather, the dimer interface of *Bpe*Globin can be found across the unit cell wall, a phenomenon that has not been observed in other sensor globin crystal structures (Fig. 2 B). Chains

A and B both contained density above the iron center of the heme, which was best fit to an O_2, while chain C did not contain any significant density above the iron atom (Figure 2C). Additionally, all three chains contained an ordered water within the heme pocket, which was located 3.5 ±0.4 Å away from the O_2 (Table 2; Figure 2B).

Table 1. Data collection and refinement statistics.

	***Bpe*Globin**
Wavelength	
Resolution range	35.63 - 2.3 (2.382 - 2.3)
Space group	C 1 2 1
Unit cell	145.16 49.64 68.14 90 110.585 90
Total reflections	169920 (18759)
Unique reflections	18912 (2041)
Multiplicity	9.0 (9.2)
Completeness (%)	92.16 (99.66)
Mean I/sigma(I)	26.08 (19.21)
Wilson B-factor	38.06
R-merge	0.07781 (0.1159)
R-meas	0.08293 (0.1227)
R-pim	0.02807 (0.03971)
CC1/2	0.993 (0.994)
CC*	0.998 (0.999)
Reflections used in refinement	18885 (2034)

Reflections used for R-free	1890 (205)
R-work	0.2094 (0.2580)
R-free	0.2527 (0.3143)
CC(work)	0.953 (0.862)
CC(free)	0.953 (0.800)
Number of non-hydrogen atoms	3877
macromolecules	3640
ligands	129
solvent	108
Protein residues	466
RMS(bonds)	0.005
RMS(angles)	1.06
Ramachandran favored (%)	97.39
Ramachandran allowed (%)	2.17
Ramachandran outliers (%)	0.43
Rotamer outliers (%)	0.80
Clashscore	7.67
Average B-factor	46.65
macromolecules	46.70
ligands	48.18
solvent	43.16

*Statistics for the highest-resolution shell are shown in parentheses.

Table 2. Heme pocket residue differences ± measurement error.

Protein	Residue	Distance (Å)
*Bpe*GReg (chain A/B/C)		
	Fe(II) - H99	2.37 ±0.42 2.35 ±0.49 2.44 ±0.47
	Fe(II) - Y43	4.62 ±0.39 4.70 ±0.47 4.69 ±0.45
	Fe(II) - S68	6.20 ±0.39 6.02 ±0.47 5.90 ±0.45
	Fe(II)- W72	8.07 ±0.41 8.03 ±0.45 8.34 ±0.42
	Y43 - S68	4.31 ±0.39 4.02 ±0.48 4.21 ±0.46
	O_2- Fe(II)	1.52 ±0.39 1.59 ±0.46 -
	O_2- Y43	2.26 ±0.41 2.43 ±0.48 -
	O_2- S68	5.46 ±0.41 4.66 ±0.48 -
	H_2O - Y43	2.78 ±0. 40 2.52 ±0.43 2.30 ± 0.43
	H_2O - S68	2.25 ±0.40 1.86 ±0.43 2.58 ±0.42

	Fe(II) - H_2O	4.50 ±0. 40 4.84 ±0.42 3.85 ±0.42
HemAT-*Bs* (chain A/B)		
(PDB=1or4)	Fe(III) - H180	2.013 2.004
	Fe(III) - Y70	5.650 5.578
	Fe(III) - T95	7.136 6.891
	Y70 - T95	5.700 5.690
	H_2O - Y70	4.017 3.610
	H_20 - T95	2. 658 2.839
	Fe(III)- H_2O	4.795 4.474
*Ec*DosC (chain A/B)		
(PDB= 4zvb)	Fe(II) - H99	2.161 2.157
	Fe(II) - Y43	6.548 6.692

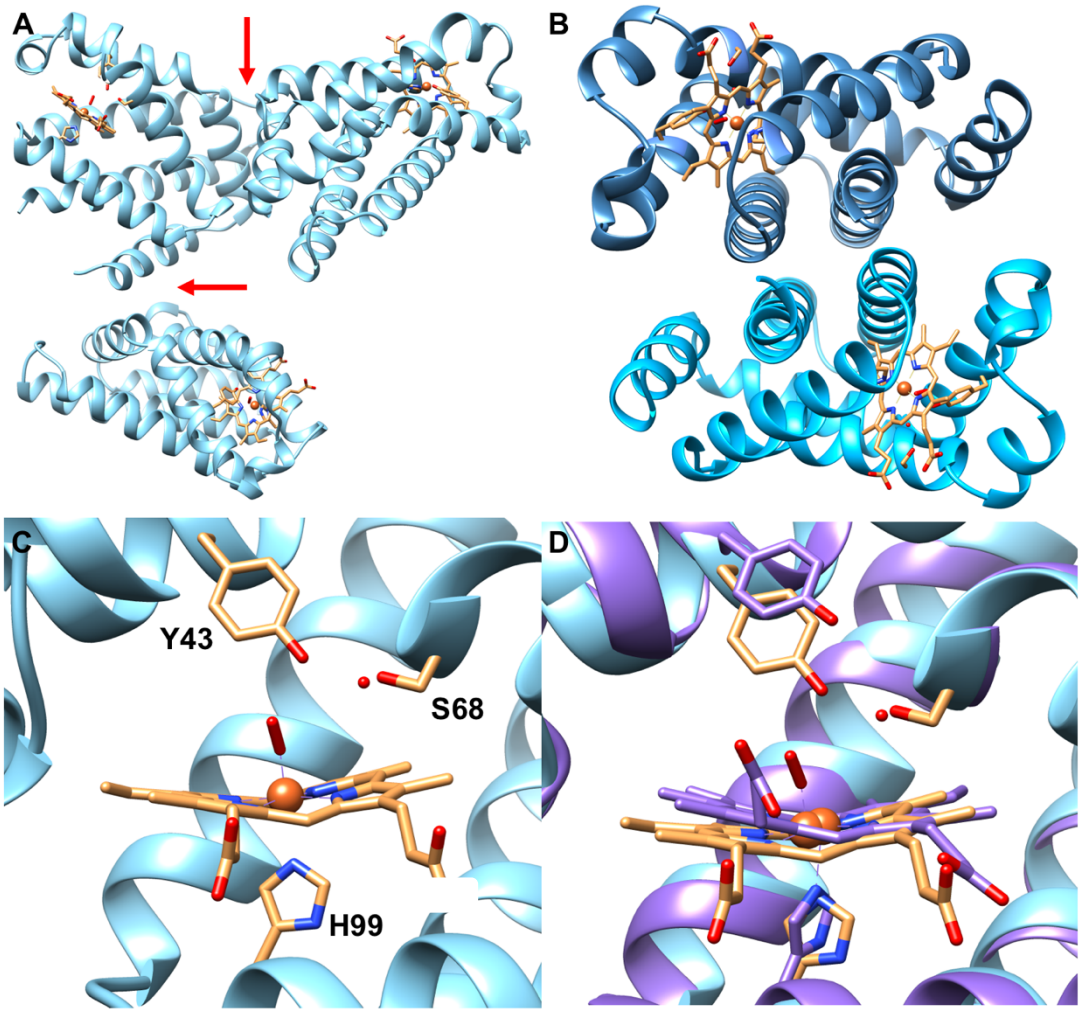

Figure 2 (A) The unit cell is made up of three monomeric units of *Bpe*Globin associated via salt bridges along helices A, B, and E, indicated by the red arrow. (B) The canonical sensor globin dimer interface comprised of helices G and H was found across monomers of different unit cells. (C) Depicted here, the heme pocket of chain B was one of two chains within the unit cell of *Bpe*Globin found with electron density above the heme iron consistent with an O_2 molecule. (D) Overlay of O_2-bound *Bpe*Globin and the ferrous GCS from *E. coli*, *Ec*DosC, reveals disparate deformations between the oxy- and unligated sensor globins.

Heme Distortion

Previous work has established the importance of heme deformation in the signaling mechanisms of heme-based sensor proteins.[32-33] To determine if heme distortion could possibly be involved in signal transduction within GCS proteins, the heme cofactors from all three chains of *Bpe*Globin were analyzed and compared to each other and other crystalized GCS hemes (Table S2).[34-35] All three chains of *Bpe*Globin have similar saddling (B2u) and propelling (A1u). However, there are significant differences in the Doop (overall distortion) of chain C, which contains the unligated heme. Although chains A and B have similar Doop and doming (A2u), chain B has significant ruffling (B1u) and waving (Eg(x and y)).

Heme distortion within *Bpe*Globin is likely due to interactions with heme pocket residues on the proximal side; a majority of the residues within 5 Å are hydrophobic, with most being isoleucines and valines. It was noted that the highest degree of difference for the homologous residues, in *Pcc*GCS, *Ec*DosC and HemAT-*Bs,* was seen for HemAT-*Bs* (Table S3). This further supports previous finding that heme distortions are most similar from proteins with the same function and that conserved heme pocket residues assist in signal transmission within DGC-GCS proteins. Additionally, these findings suggest that heme ruffling (B1u) transmits the ligation status of the heme-Fe into the globin fold.

Fourier-Transform Infrared Spectroscopy

In order to verify and extend the knowledge gained from the *Bpe*Globin crystal structure, we performed FTIR on CO-bound wild-type and mutant *Bpe*GReg/*Bpe*Globin, as well as the closely related GCS from *Pectobacterium carotovorum*, *Pcc*GCS. *Pcc*GCS exhibits a typical FTIR spectrum, with peaks at 1922 cm^{-1} and 1962 cm^{-1} corresponding to CO stretches with a hydrogen bonding interaction and apolar pocket, respectively. Interestingly, *Bpe*GReg exhibits three CO

stretching frequencies, 1925 cm^{-1}, 1963 cm^{-1}, and 1972 cm^{-1}, corresponding to a hydrogen bonding interaction, apolar interactions, and interaction with a negative charge. Binding of isotopically labeled ^{13}CO resulted in a downfield shift of all three peaks in the spectra of *Bpe*Globin (1925 cm^{-1}, 1963 cm^{-1}, and 1972 cm^{-1}) and two peaks in the spectra of *Pcc*Globin (1922 cm^{-1} and 1962 cm^{-1}), verifying that these stretching frequencies all correspond to the CO stretch (Figure 3A-B). In addition, FTIR spectra of full-length *Bpe*GReg and *Pcc*GCS were obtained and were consistent with the corresponding globin spectra, indicating that deletion of the middle and DGC domains does not substantially affect interactions between bound CO and the proteins (Figure 4).

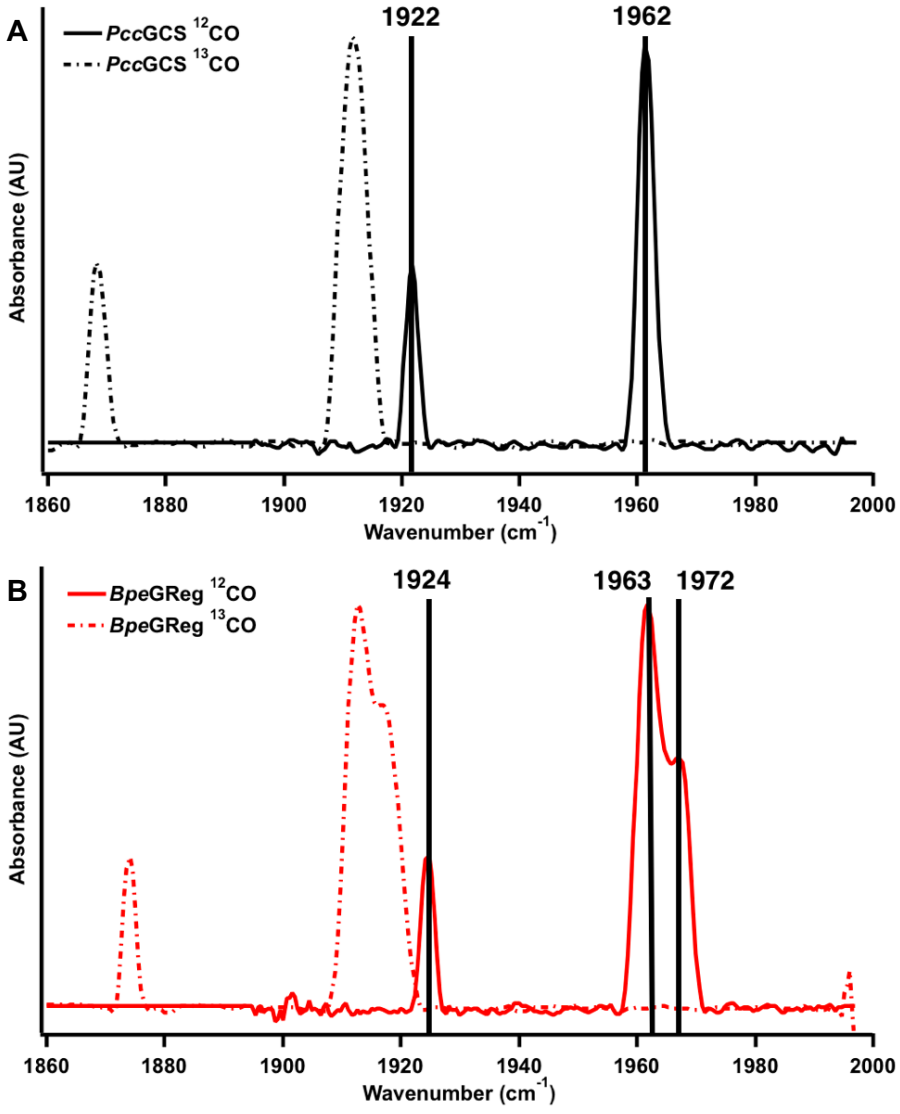

Figure 3 (A) Comparing the FTIR spectra of ^{12}CO- and ^{13}CO-bound *Pcc*GCS reveals two distinct stretches at 1922 cm^{-1} and 1962 cm^{-1} that can be attributed to the C-O triple bond stretch. These peaks indicate a hydrogen bond and an apolar stretch, respectively. (B) The same spectrum for *Bpe*GReg reveals three CO-dependent stretches at 1924 cm^{-1}, 1963 cm^{-1}, and 1972 cm^{-1}, indicating a similar hydrogen bonding interaction and apolar stretch to *Pcc*GCS, and an additional interaction with a negative charge, respectively.

To probe the roles of the heme distal pocket residues in interacting with bound CO, we obtained spectra of proteins with mutations to the distal hydrogen-bonding residues, (Y43F and S68A for *Bpe*GReg and *Bpe*Globin, as well as Y57F and S82A for *Pcc*Globin). For both proteins, only mutation of Y57 produced an appreciable change in FTIR spectrum. Spectra of both *Bpe*GReg/Globin (Y43F) and *Pcc*GCS/Globin (Y57F) lack the stretches at 1925 cm^{-1} and 1922 cm^{-1} (Figure 4A-B). Furthermore, the apolar stretch at 1963 cm^{-1} shifted in the spectra for both *Bpe*GReg/Globin (Y43F) and *Pcc*GCS/Globin (Y57F) (Figure 4).

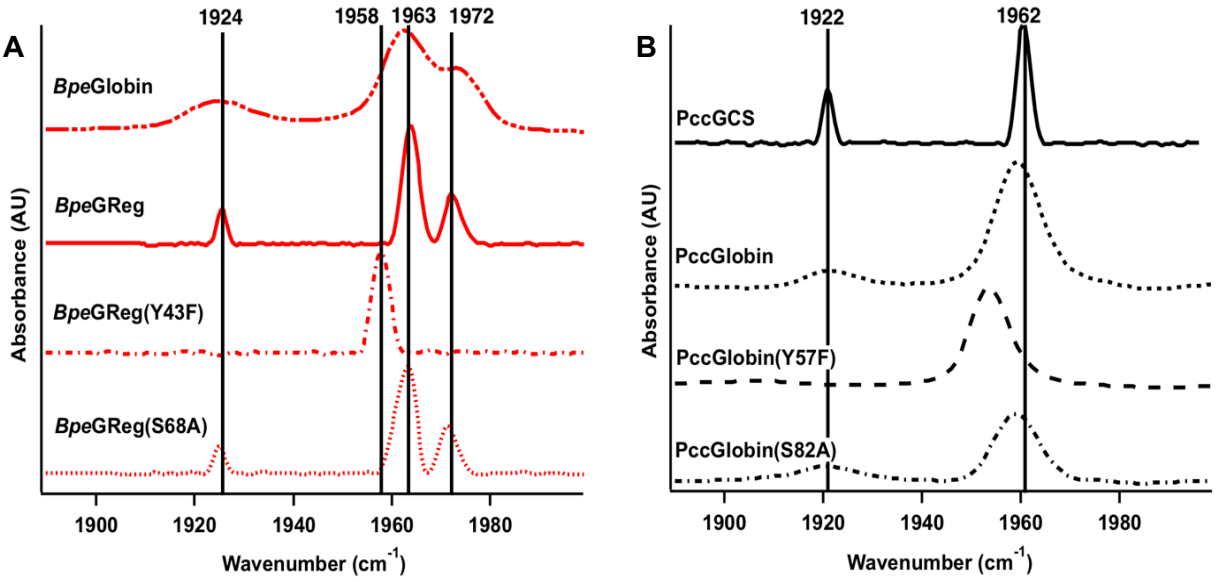

Figure 4 (A) The CO-bound spectra of *Bpe*Globin and full-length *Bpe*GReg show no difference. Mutation of the distal tyrosine (43) to a phenylalanine (Y43F) yields significant changes in the spectrum of the CO-bound protein. Namely, the stretches at both 1924 cm^{-1} and 1972 cm^{-1} disappear, indicating that the interactions responsible for these peaks are tyrosine-dependent. The spectrum of CO-bound protein with a mutation to the distal serine (S68) does not show any difference from the spectrum of WT *Bpe*GReg. (B) The spectra of *Bpe*Globin and full-length *Bpe*GReg are also identical. A similar shift to that observed in *Bpe*GReg (Y43F) is seen in *Pcc*GCS (Y57F). Like *Bpe*GReg, mutation of the distal serine (S82) in *Pcc*GCS had no effect on the FTIR spectrum.

Examination of Ordered Water

The observation of a CO stretching frequency at 1972 cm^{-1} suggested that the ordered water may be involved in direct interactions with bound CO. To explore the role of the water observed in the heme pocket of the *Bpe*Globin crystal structure, FTIR was performed on wild type *Bpe*Globin at pH values ranging from 5.0 to 8.0. The protein was stable at all pHs and did not form aggregates under any of the conditions. As the buffer pH was decreased, the peak at 1975 cm^{-1} decreased in intensity (Figure 5A). To further support the distal water as the moiety responsible for the stretch at 1972 cm^{-1}, FTIR spectra were taken in $H_2{}^{18}O$ labeled buffer. Within the *Bpe*Globin FTIR spectrum, the 1975 cm^{-1} peak shifts and thus loses intensity within the $H_2{}^{18}O$ buffer (Figure 6).

While *Pcc*GCS/*Pcc*Globin do not exhibit a peak in the 1970 cm^{-1} region like *Bpe*GReg/*Bpe*Globin, the two proteins contain a high enough sequence similarity (36% overall amino acid similarity, 61% globin domain amino acid similarity) and identity to suggest that *Pcc*GCS/*Pcc*Globin also contains a distal pocket water. Thus, in order to determine the presence

of a structured water in *Pcc*Globin (and *Pcc*GCS by extension), FTIR was performed on wild type *Pcc*Globin in a pH range of 7.0 to 9.0 (Figure 5B). Like *Bpe*Globin, *Pcc*Globin remained stable at all tested pH values, although it was noted that stability significantly decreased above pH 8.5/9. However, the pH range of 7.0-9.0 was sufficient to note the trend. The spectrum of *Pcc*Globin developed a peak at 1975 cm^{-1} as pH increased, as predicted by the results for *Bpe*Globin (Figure 5B).

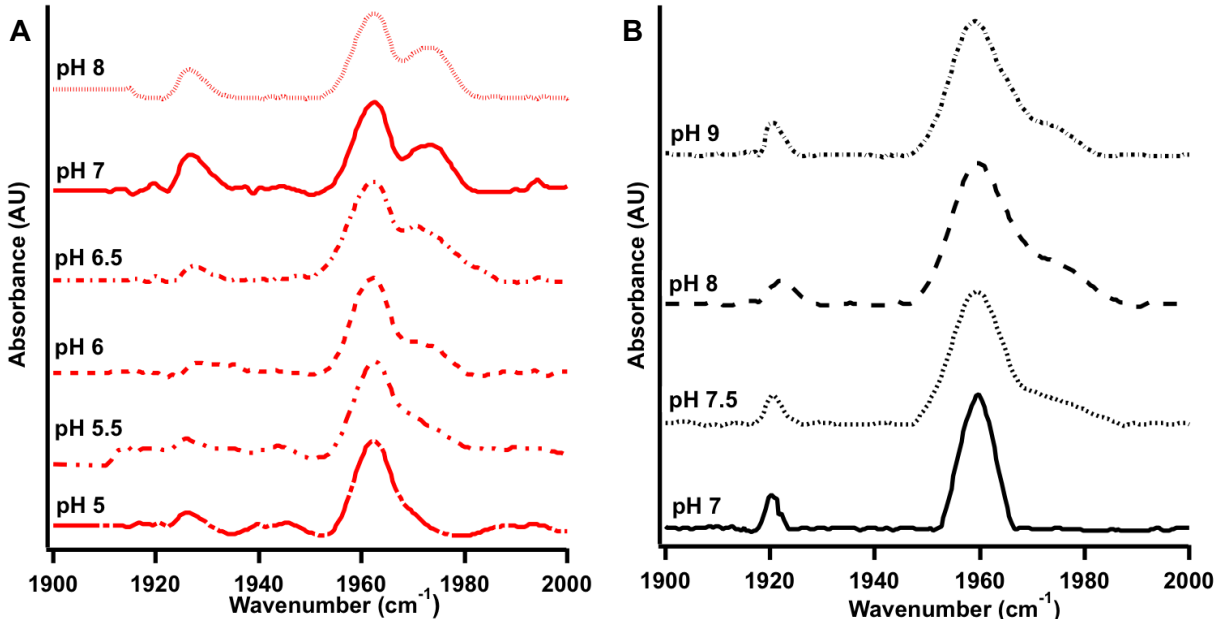

Figure 5 (A) FTIR spectra of CO-bound *Bpe*Globin at varying pH values were obtained. At more basic pH values, the CO stretch at 1972 cm^{-1} increases in intensity, indicating that the interaction that gives rise to this stretch is strengthened in more basic environments. (B) Similarly, for the CO-bound spectra of *Pcc*Globin, a stretch appears at 1975 cm^{-1} and increases in intensity as pH increases.

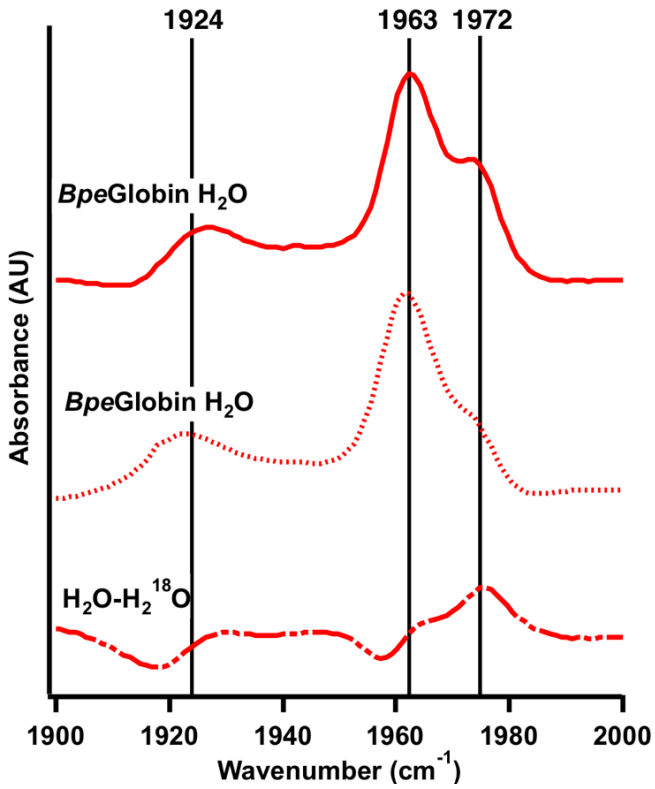

Figure 6 When exchanged into $H_2{}^{18}O$ buffer, the CO-bound spectrum of *Bpe*Globin shows a decrease in the intensity of the stretch at 1972 cm^{-1}, demonstrating that this peak is not only tyrosine-dependent, but also ordered water-dependent. A difference spectrum is shown to definitively demonstrate the shift of the 1972 cm^{-1} stretch.

Ultraviolet-Visible Spectroscopy

Ultraviolet-visible spectroscopy (UV/Vis) was used to examine the ligand binding capabilities of *Bpe*GReg (H99G) and *Pcc*GCS (H99G) (Figure 7 and Figure S2). *Bpe*GReg (H99G) was examined without the addition of ligand (Fe^{II}), bound to CO (Fe^{II}-CO), NO (Fe^{II}-NO), and O_2 (Fe^{II}-O_2). The protein was able to stably bind both CO and NO, but rapidly autoxidized when exposed to O_2, evidenced by the disappearance of the α/β bands in the 500-600 nm region (Figure 7). Results for the CO- and NO-bound state of *Pcc*GCS (H112G) are similar (Figure S2).

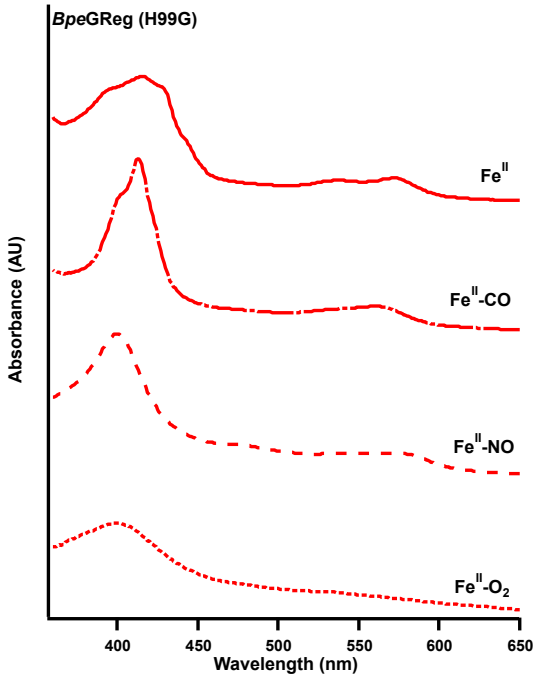

Figure 7 The UV/Vis spectra of *Bpe*GReg (H99G) in different ligation states demonstrates its ability to stably bind ligands. Fe^{II} (394, 416, 430, 443, 573 nm); Fe^{II}-CO (413, 560 nm); Fe^{II}-NO (400, 584 nm); Fe^{II}-O_2 (400).

High-Performance Liquid Chromatography

High-performance liquid chromatography (HPLC) was used to determine the oligomeric states of *Bpe*GReg (H99G) and *Pcc*GCS (H112G). Interestingly, although the wild type forms of the proteins exist in a mixture of oligomeric states, the mutant proteins were observed as only the tetramer state (Figure 8A-B).

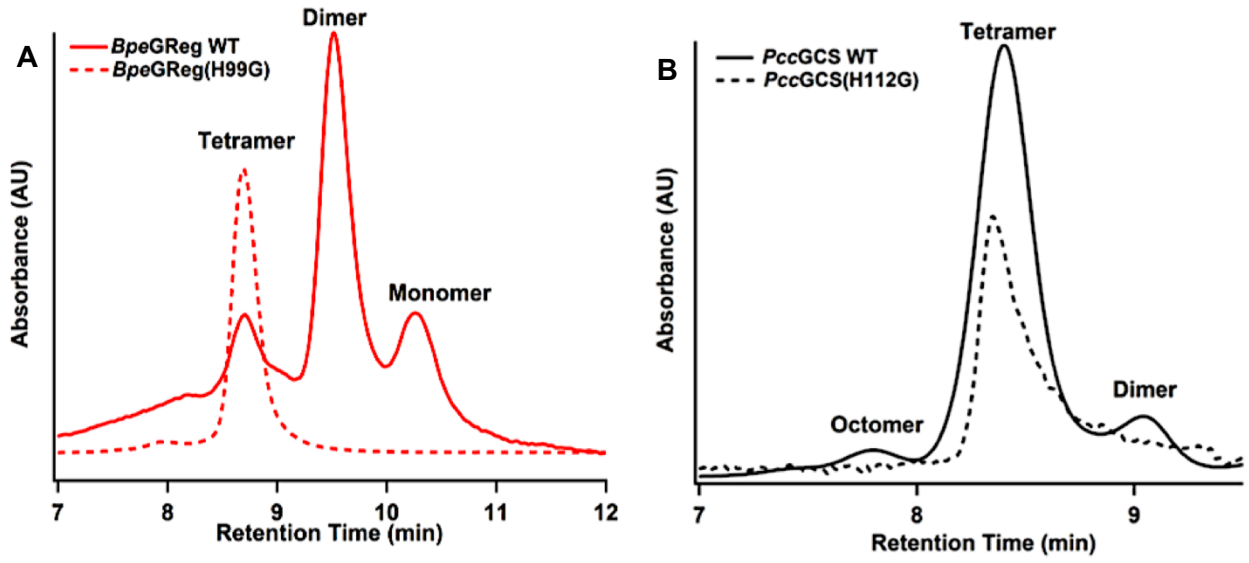

Figure 8 (A) Wild type *Bpe*GReg exists as a mixture of monomer, dimer, and tetramer, with the predominant form being dimer, while wild type *Pcc*GCS exists as a mixture of dimer, tetramer, and octamer, with the predominant form being tetramer. Mutation of the proximal histidine residue to a glycine (H99G and H112G, respectively) shifted both proteins completely to the tetramer state.

Discussion

Structural Insights

This is the first reported crystal structure of a bacterial sensor globin domain in an Fe^{II}-O_2 ligation state, offering a novel opportunity to examine a signaling domain in its native active state. Chains A and B were both found in the Fe^{II}-O_2 form, but C was found to be in the unligated Fe^{II} form. While chain C was found in a different ligation state than the other two, overlapping the three chains of *Bpe*Globin revealed no significant movement of the helices, suggesting that the three chains were likely in the same ligation state when the crystal lattice was established. Previous work has shown that Fe^{II}-O_2 *Bpe*GReg is stable for at least 18 hours with no significant oxidation, aggregation or oligomerization changes,[4] a characteristic that possibly enabled the initial crystal lattice to form in the Fe^{II}-O_2 state. The different ligation states in each monomer may be due to O_2 dissociation during the time required for sufficient crystal growth (~1 week) or photoreduction within the X-ray beam, providing a rationale for the similar overall protein structure of the monomeric units, despite different ligation states.

While overall protein structure remained similar among the monomeric units of the unit cell, there were differences in the heme distortions of the three units depending on the ligation state. As previous work has determined that heme distortions can indeed be a signaling mechanism for heme sensor proteins,[32-33] it is possible that the observed differences are relevant to the

signaling mechanism. The difference in the heme distortions of chains A and B despite identical ligation states is attributed to the fact that chain A dimerizes with chain A from another unit cell, while chain B dimerizes with chain C from another unit cell, respectively. As GCS activity is dimerization-dependent, and the ligation state of GCS proteins has been shown to influence oligomeric state,[25] it is possible that the A-A dimer represents the active, O_2-bound state, while B-C represents a transition between active and inactive states. Additionally, the large change in ruffling (B1u) values between chains A and B indicates that it may play a key role in signaling (Table S2).

A number of hydrophobic residues were found on the heme proximal side, within proximity for Van der Waals interactions (Figure 9). Due to the location of these residues on and around the F8 helix, which is itself further connected to the globin dimer interface, these residues may play a role in propagating ligand binding signals throughout the globin domain, thereby influencing the dimer interface (Figure 9). This mechanism might work by ligand binding changing the heme distortion, which could then interact sterically with the proximal residues and transmit the ligand binding signal to the F8 helix and the dimer interface.

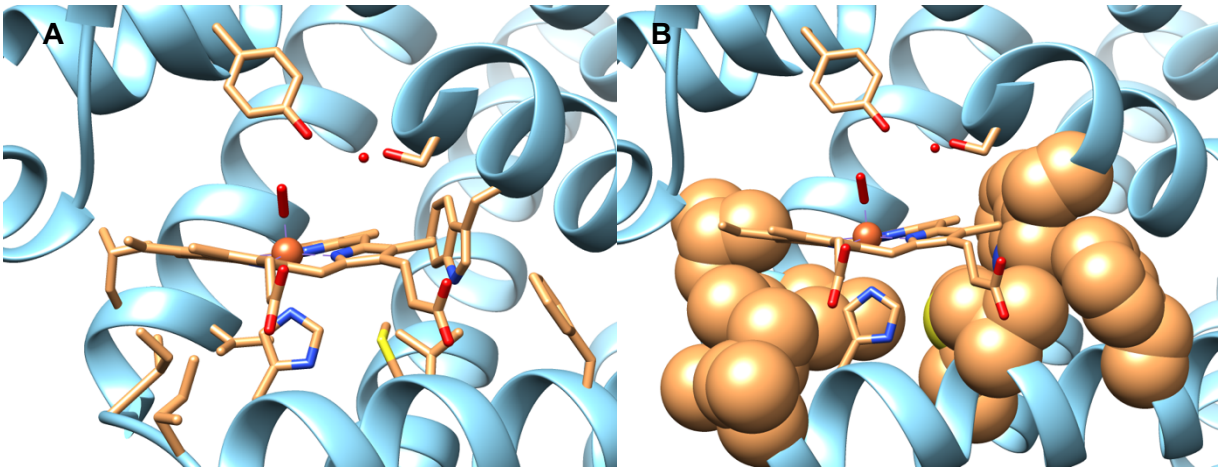

Figure 9 (A) All peripheral and proximal residues within 5 Å of the heme cofactor in the heme pocket of *Bpe*GReg are shown in stick form for identification. (B) All peripheral and

proximal residues within 5 Å of the heme cofactor are shown in the space-filling form to demonstrate the potential for steric interactions between these residues and the heme cofactor.

Furthermore, previous work has established that heme distortions among proteins of similar function are more similar than between proteins of disparate function. Thus, comparing the obtained heme distortion from *Bpe*Globin to those of the HemAT-*Bs* and *Ec*DosC globin domains offers insight into these trends. Indeed, the distortions of *Bpe*Globin and the globin domain from *Ec*DosC are more similar to each other than either are to the distortions found in the heme of the globin domain of HemAT-*Bs*. This is in keeping with the established pattern, as *Bpe*GReg and *Ec*DosC are both DGC-GCS proteins, whereas HemAT-*Bs* is an aerotaxis MCP-GCS.[20, 28] In particular, this trend holds true for the high Doop and B1u (ruffling) values of the heme cofactors of *Bpe*Globin and the globin domain of *Ec*DosC (Table S2). Indeed, overlaying the structures of unligated *Bpe*Globin with unligated *Ec*DosC or HemAT-*Bs* reveal these differences in heme pocket residue orientation and heme distortion (Figure 10). In addition to corroborating a previously noted trend, the similarity between such related proteins as *Bpe*GReg and *Ec*DosC implies a conserved intra-protein, heme distortion-mediated signaling mechanism of output activation.

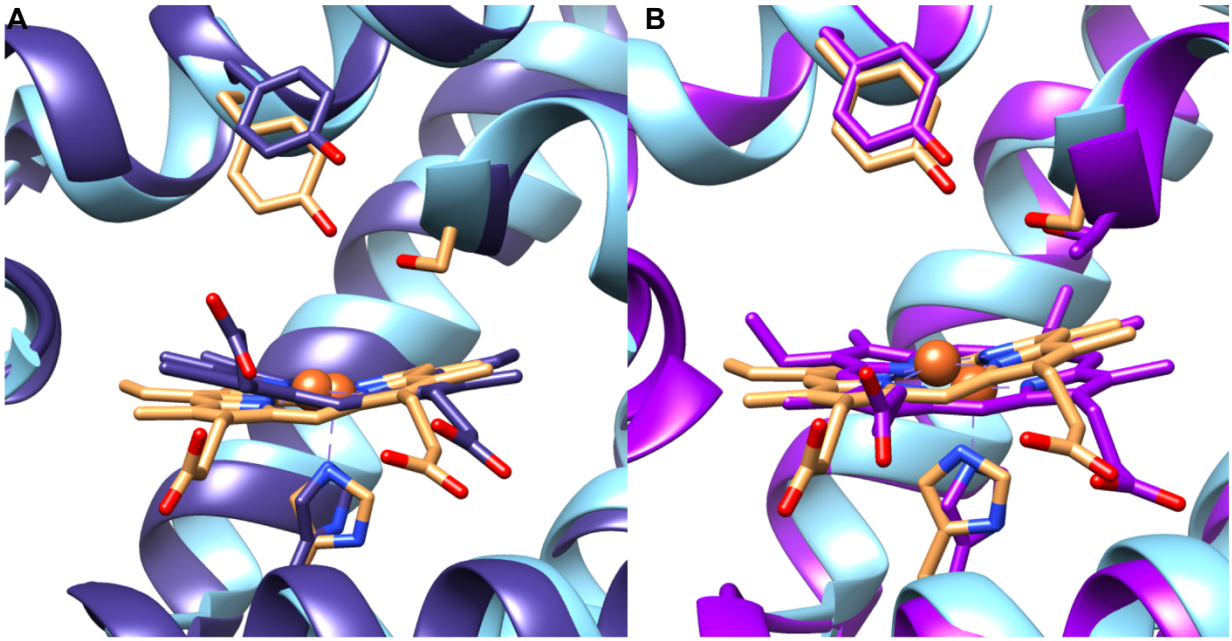

Figure 10 (A) Overlaying the heme pockets of *Bpe*GReg/Globin and the DGC-GCS from *E. coli* (*Ec*DosC) reveals minor differences in the heme distortion. Note that the heme cofactors are not perfectly aligned, resulting in the appearance of greater differences than there are. (B) In contrast, there are greater differences in the heme distortions of *Bpe*Globin and HemAT-*Bs*.

Interactions between heme distal residues and bound oxygen may also be a method by which the ligand binding signal is transmitted from the globin domain to the DGC output domain. Such a mode of intra-protein signal transduction has been hypothesized for GCS proteins in previous works, mediated by hydrogen-bonding interactions between the O_2 and distal residues.[18, 30] Measuring the distances between the distal residues and the bound oxygen in the *Bpe*Globin crystal structure reveals that both the distal tyrosine, as well as the distal serine are within the appropriate range to make hydrogen-bonding contacts with the ligand (Table 2). Thus, it is possible that the oxygen-binding event may be signaled to the DGC output domain via subtle

rearrangements within the heme pocket initiated by hydrogen bonding. Additionally, the distances between Y43 and the ordered water, as well as S68 and the ordered water are within range for potential hydrogen-bonding, implying a potential role for the water in ligand binding and signaling. The crystal structure of HemAT-*Bs* also contains a water molecule, which is 4.7 Å from the heme iron in the crystal structure, but has been demonstrated to interact with bound O_2 through resonance Raman studies. However, no water was found within the distal pocket of *Ec*DosC, suggesting that that the water may contribute to the differing behavior of the various GCS proteins.

Distal Interactions

While O_2 is the native ligand of *Bpe*Globin, the FTIR spectrum of O_2-bound heme is faint and easily lost in water vapor peaks, necessitating the examination of these proteins in their CO-bound states, as CO-bound heme gives sharp, pronounced FTIR peaks. For both *Bpe*GReg and *Pcc*GCS, FTIR spectra of CO-bound proteins resulted in two peaks, one in the 1922 cm^{-1} region and the other in the 1962 cm^{-1} region, indicating a hydrogen bonding interaction and the apolar C-O stretch, respectively (Figure 3A-B). The FTIR spectra of mutant proteins were examined to determine which distal residues were responsible for the observed interactions. Mutation of the distal tyrosine in both proteins caused the peak in the 1922cm^{-1} region to disappear, meaning that the distal tyrosine is responsible for making a hydrogen bond with the bound CO. Prior work has implicated the distal tyrosine as the major hydrogen bond donor in the heme pocket of both proteins, with a major role in ligand stability.[31] Furthermore, this result is in agreement with previous work, which implicates hydrogen bond-mediated structural shifts in the distal tyrosine as a means of ligand recognition and output domain activation.[30] Thus, hydrogen bonding between heme ligands and the distal tyrosine in *Bpe*Globin may play a role in DGC activation. It should be noted that although the spectra of CO-bound distal serine mutants did not show any difference

from wild type, this does not preclude interactions between this residue and other ligands. Indeed, past work has shown that the distal serine in *Pcc*GCS and *Bpe*GReg, as well as other GCS proteins, does play a role in O_2 binding.[30-31]

Notably, the CO-bound spectrum of the distal tyrosine mutants resulted in ~6 cm^{-1} downfield shifts of the 1960 cm^{-1}-region peak, indicating a change in the heme pocket environment from apolar to polar. While previous studies have recorded similar shifts in the FTIR spectra of CO-bound myoglobin, no explanation for this shift has been provided.[36] We hypothesize that this shift is due to a loss of steric interaction between the tyrosine hydroxyl group and the CO ligand, which would force the tyrosine phenyl ring out of the heme pocket in wild type proteins. This would allow the phenyl ring to remain in the heme pocket, thus permitting the partial positive charge in the center of the phenyl π-ring to interact with the C-O triple bond, drawing electrons through induction towards the oxygen atom and making the bond more polar.

In addition to the peaks at 1924 cm^{-1} and 1963 cm^{-1}, the spectrum of CO-bound *Bpe*GReg also contained a peak at 1972 cm^{-1}, which indicates an interaction between the bound CO and a negative charge within the heme pocket (Figure 3B). Such a pronounced discrepancy in the spectra of *Bpe*GReg and *Pcc*GCS was unexpected given the fact that the two proteins have homologous heme pockets. However, this difference does suggest a potential explanation for the different ligand binding kinetics of these two proteins.[4, 31] As the heme pocket of *Bpe*Globin does not contain any charged amino acids, we hypothesized that the negative charge interaction could be due to the ordered water within the heme pocket observed in the crystal structure. The pH-dependent increase in the intensity of the peak at 1972 cm^{-1} for the CO-bound spectrum of *Bpe*Globin indicates an increase in the strength of the interaction between the negative charge and bound CO (Figure 5A). This positive correlation between deprotonation and strength of interaction

suggests that the ordered water in the pocket of *Bpe*Globin exists in a partially deprotonated state. Thus, under more basic conditions, the strength of the negative charge on the water would increase, and the interaction between the CO and that negative charge would intensify accordingly. Dependence of the 1972 cm^{-1} peak on the ordered water was corroborated by spectra taken in $H_2{}^{18}O$ buffer, which showed a shift in the 1972 cm^{-1} peak in the $H_2{}^{18}O$ buffer (Figure 6A). Taken together, these results show that the peak at 1972 cm^{-1} in the CO-bound spectrum of *Bpe*GReg is likely the result of interactions between the CO, a partially deprotonated ordered water, and the distal tyrosine.

While the lack of a peak in the 1970 cm^{-1} region of the CO-bound spectrum of *Pcc*GCS/Globin prevented us from obtaining meaningful results through $H_2{}^{18}O$ labelling (Figure S1), the spectra of CO-bound *Pcc*Globin taken at varying pH values are telling. The appearance and rise in intensity of a peak at 1975 cm^{-1} suggests the presence of an ordered water within the heme pocket of *Pcc*Globin, similar to that of *Bpe*GReg (Figure 5B). However, the fact that this peak only appears at elevated pH indicates that the ordered water within the pocket of *Pcc*GCS is fully protonated at pH 7.0, and therefore cannot donate a negative charge to the CO. This difference in the ordered water protonation state between *Bpe*GReg and *Pcc*GCS implies a discrepancy in the heme pocket environments, resulting in different local pH values. Thus, either the different protonation states, or the different heme pocket environments that give rise to the different protonation states, may provide a rationale for the different O_2 binding kinetics observed in these two proteins, despite homologous heme pockets.[4, 31]

Roles of the Proximal Histidine

Additionally, UV/Vis data obtained from *Pcc*GCS (H112G) and *Bpe*GReg (H99G) implicate the proximal histidine in ligand binding and stability. While the spectra of *Pcc*GCS

(H112G) and *Bpe*GReg (H99G) are similar to those of wild type in the ferrous (Fe^{II}) form and when ligated to nitric oxide (Fe^{II}-NO), there are notable differences in the spectra of both histidine mutants when in CO- and O_2-bound states (Fe^{II}-CO and Fe^{II}-O_2, respectively; Figure 7 and Figure S2). Unlike the wild types, the spectra of *Bpe*GReg (H99G) in the Fe^{II}-CO state showed two Soret peaks, one at 400 nm and one at 413 nm (Figure 7). While the soret at 413 nm indicates the protein in its CO-bound state, it is unclear what exactly the peak at 400 nm represents, but it does indicate that the protein is not fully CO-bound. Full CO binding could not be achieved even in CO-saturated buffer or by leaving the protein in a CO atmosphere for >30 minutes. We hypothesize that the soret at 400 nm represents the protein in a bis-imidazole ligation state, as similar absorbance values have previously been reported in the literature and the constant presence of imidazole these proteins require is likely to result in a bis-ligated state.[37-38] This would explain the difficulty we experienced with binding CO to these proteins.

Although the soret peaks for O_2-bound *Bpe*GReg (H99G) were similar to previously reported maxima for the O_2-bound wild type proteins,[4, 23] the disappearance of the α/β bands around 550 nm and 580 nm indicate rapid autoxidation of the heme iron. As this same phenomenon has been observed in the tyrosine mutants of both proteins, and tyrosine was thus established as a major contributor to ligand stability, it follows that the proximal histidine also contributes to ligand stability.[31] The rapid autoxidation and aggregation of *Pcc*GCS (H112G) under aerobic conditions prevented its observation.

The oligomeric state of DGC-GCS proteins has been shown to have a direct effect on the activity of the DGC output domain.[4, 25, 29, 31] Thus, the relationship between oligomeric state and the proximal histidine residue was explored due to its placement on the F8 helix (Figure 11). It is hypothesized that the proximal histidine residue may modulate the oligomeric state of DGC-GCS

proteins by introducing subtle shifts in the F8 helix upon ligand binding, which would then lead to shifts in helices G and H (Figure 11). Indeed, previous work has shown that mutation to the proximal histidine of the histidine kinase-containing globin-coupled sensor from *Anaeromyxobacter* sp. Fw109-5, *Af*GcHK, results in solvent exposure for residues along the dimer interface different from that of wild type.[29] For both *Pcc*GCS and *Bpe*GReg, mutation of the proximal histidine to a glycine (H112G and H99G, respectively) resulted in alterations to the HPLC spectra, meaning that mutation of this residue caused shifts in the oligomeric state equilibrium of these proteins (Figure 8A-B). Specifically, for both proteins, mutation to the proximal histidine led to a complete shift in the oligomeric states to the tetramer, which is surprising given that the tetramer is the most active form of both proteins (Figure 8A-B).[4] This implies that the proximal histidine does have a role in translating the ligand binding signal into oligomeric shifts, though discerning the exact mechanism, as well as how this change affects activity, will require further work.

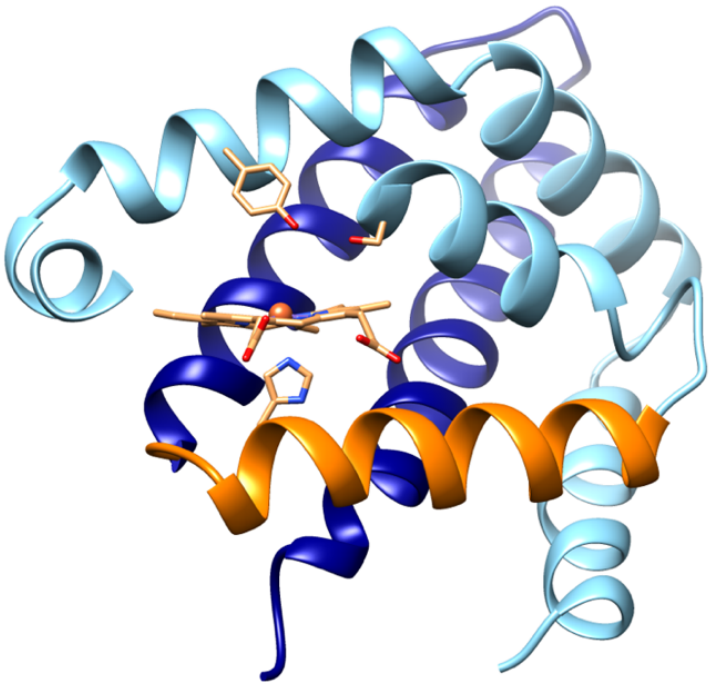

Figure 11 The proximal histidine residue is located on the F8 helix (orange), which is further connected to the G and H helices (dark blue) comprising the dimer interface. This presents a potential pathway by which changes in the Fe-His bond length could be transmitted to affect global protein structure and DGC activity.

Conclusions

The crystal structure of the globin domain from *Bpe*GReg has been solved in its O_2-bound state, making it the first structure of a bacterial sensor globin to be solved in its native active form. The structure is similar to previously characterized sensor globin structures. This structure revealed few differences in overall protein structure, but implicates differential, ligand-dependent heme deformation as a mode of signaling. Furthermore, FTIR was used in conjunction with crystallization to determine the roles of heme distal pocket moieties in ligand interaction in *Bpe*GReg and the closely related DGC-GCS, *Pcc*GCS. These explorations support previous data that implicate the distal tyrosine as the major hydrogen bond donor within the heme pocket of these proteins. Furthermore, this work provides the first thorough characterization of the ordered water within the heme pocket of a bacterial sensor globin. FTIR studies show that this water interacts with both the bound ligand through the donation of a negative charge, as well as with the distal tyrosine. Thus, we propose that this water is partially deprotonated and therefore partially negatively charged by donating a hydrogen bond to the distal tyrosine. Finally, data show that the heme pocket proximal histidine likely plays a role in determining the oligomeric state, and therefore the activity of the output domain, of *Pcc*GCS and *Bpe*GReg, as well as in ligand binding stability. Together, these data provide a thorough exploration of the structure-function relationship of heme pocket residues for two bacterial DGC-GCS proteins, which will be expanded on through kinetic assays in future work.

References

1. J.W. Costerton, P. S. S., E.P. Greenberg, *Science* **1999,** ***284***, 1318-1322.
2. A.E.L. Roberts, K. N. K., T. Bjarnsholt, S.P. Diggle, *J. Mol. Biol.* **2015,** ***427***, 3646-3661.
3. D.H. Lee, J.-A. L., J. Lee, E. Roh, K. Jung, M. Choi, C. Oh, S. Ryu, J. Yun, S. Heu, *Microbiology* **2013,** ***159***, 1487-1496.
4. J.L. Burns, D. D. D., E.E. Weinert, *Mol. BioSyst.* **2014,** ***10***, 2823-2826.
5. N. Rabin, Y. Z., C. Opuku-Temeng, Y. Du, E. Bonsu, H.O. Sintim, *Future Med. Chem.* **2015,** ***7*** (4), 493-512.
6. U. Römling, M. Y. G., M. Gomelsky, *Microbiol. Mol. Biol. R.* **2013,** ***77*** (1), 1-52.
7. S.M. Vorobiev, H. N., B. Yu, J. Seetharaman, R. Xiao, T.B. Acton, G.T. Montelione, J.F. Hunt, *J. Struct. Funct. Genomics* **2012,** ***13*** (3), 177-183.
8. Hengge, R., *Nat. Rev. Microbiol.* **2009,** ***7***, 263-273.
9. Z.F. Hallberg, X. C. W., T.A. Wright, B. Nan, O. Ad, J. Yeo, M.C. Hammond, *PNAS* **2015,** ***113*** (7), 1790-1795.
10. C. Chan, R. P., D. Samoray, N.C. Amiot, B. Giese, U. Jenal, T. Schirmer, *PNAS* **2004,** ***101*** (49), 17084-17089.
11. J.G. Malone, R. W., M. Christen, U. Jenal, A.J. Spiers, P.B. Rainey, *Microbiol.* **2007,** ***153***, 980-994.
12. N. De, M. P., P.V. Krasteva, R.V. Raghavan, H. Sondermann, *PLOS Biol.* **2008,** ***6*** (3), e67.
13. R. Tamayo, A. D. T., A. Camilli, *J. Biol. Chem.* **2005,** ***280*** (39), 33324-33330.
14. Camilli, A. D. T. a. A., *Mol. Microbiol.* **2004,** ***53*** (3), 857-869.
15. A.L. Lovering, M. J. C., C. Lambert, L. Hobley, R.E. Sockett, *MBio* **2011,** ***2*** (5), e00163-11.
16. T. Shimizu, D. H., F. Yan, M. Stranava, M. Bartosova, V. Fojtíková, M. Martínková, *Chem. Rev.* **2015,** ***115***, 6491-6533.
17. J.L. Burns, P. B. J., S. Rivera, B.M. Fontaine, L. Briggs, E.E. Weinert, *ACS Chem. Biol.* **2017,** ***12*** (8), 2070-2077.
18. K. Kitanishi, K. K., Y. Kawamura, I. Ishigami, T. Ogura, K. Nakajima, J. Igarashi, A Tanaka, T. Shimizu, *Biochemistry* **2010,** ***49***, 10381-10393.
19. J.R. Tuckerman, G. G., E.H.S. Sousa, X. Wan, J.A. Saito, M. Alam, M.-A. Gilles-Gonzalez, *Biochemistry* **2009,** ***48***, 9764-9774.
20. M. Tarnawski, T. R. M. B., I. Schlichting, *Acta Cryst.* **2015,** ***D71***, 2158-2177.
21. H.Sawai, S. Y., T. Uchida, M. Hyodo, Y. Hayakawa, K. Ishimori, S. Aono, *BBA-Proteins and Proteom.* **2010,** ***1804***, 166-172.
22. C. Wu, Y.-Y. C., H. Yin, X.-N. Song, W.-W. Li, X.-X. Zhou, L.-P. Zhao, L.-J. Tian, J.-C. Han, H.-Q. Yu, *Sci. Rep.* **2013,** ***3***, 1945.
23. X. Wan, J. R. T., J.A. Saito, T.A.K. Freitas, J.S. Newhouse, J.R. Denery, M.Y. Galperin, G. Gonzalez, M.-A. Gilles-Gonzalez, M. Alam, *J. Mol. Biol.* **2009,** ***388***, 262-270.
24. A. Sturm, M. H., M. Arnoldini, A. Benecke, M. Ackerman, M. Benz, J. Dormann, W.-D. Hardt, *PLOS Pathog.* **2011,** ***7*** (7), e1002143.
25. J.L. Burns, S. R., D.D. Deer, S.C. Joynt, D. Dvorak, E.E. Weinert, *Biochemistry* **2016,** ***55***, 6642-6651.
26. Jr., W. Z. a. G. N. P., *Structure* **2003,** ***11*** (9), 1097-1110.

27. A. Pesce, L. T., M. Nardini, F. Desmet, L. Sisinni, L. Gourlay, A. Bolli, M. Coletta, S. Van Doorslaer, X. Wan, M. Alam, P. Ascenzi, L. Moens, M. Bolognesi, S. Dewilde, J., *J. Mol. Biol.* **2009,** ***386*** (1), 246-260.
28. S. Hou, T. F., R.W. Larsen, M. Piatibratov, V. Sivozhelezov, A. Yammamoto, E.A. Meleshkevitch, M. Zimmer, G.W. Ordal, M. Alam, *PNAS* **2001,** ***98*** (16), 9353-9358.
29. M. Stranava, P. M., T. Skálova, P. Kolenko, J. Blaha, V. Fojtikova, V. Martínek, J. Dohnálek, A. Lengalova, M. Rosulek, T. Shimizu, M. Martínková, *J. Biol. Chem.* **2017,** ***292*** (51), 20921-20935.
30. T. Ohta, H. Y., S. Yoshioka, S. Aono, T. Kitigawa, *J. Am. Chem. Soc.* **2004,** ***126***, 15000-15001.
31. S. Rivera, J. L. B., G.E. Vansuch, B. Chica, E.E. Weinert, *J. Inorg. Biochem.* **2016**.
32. J.A. Shelnutt, X.-Z. S., J.-Guo Ma, S.L. Jia, W. Jentzen, C.J. Medforth, *Chem. Soc. Rev.* **1998,** ***27***, 31-41.
33. J.G. Ma, J. Z., R. Franco, S.L. Jia, I. Moura, J.J. Moura, P.M. Kroneck, J.A. Shelnutt, *Biochemistry* **1998,** ***37***, 12431-12442.
34. W. Jentzen, X.-Z. S., J.A. Shelnutt, *J. Phys. Chem. B* **1997,** ***101*** (9), 1684-1699.
35. A.B. Graves, M. T. G., M.D. Liptak, *J. Phys. Chem. B* **2016,** ***120*** (16), 3844-3853.
36. T. Li, M. L. Q., G.N. Phillips, Jr., J.S. Olson, *Biochemistry* **1994,** ***33***, 1433-1446.
37. S. Hay, T. W., *Biochemistry* **2005,** ***44*** (431-439).
38. E. Karnaukhova, S. R., M. Rajabi, L.W. Rosenlöf, A.I. Alayash, B. Åkerström, *Front. Physiol.* **2014,** ***5*** (465), 1-11.

Supporting Information

Table S1. Primers for Site-Directed Mutagenesis

Mutation	Primer Sequences
*Bpe*Globin	5'-CGAAATGATGTGTCATTAATATTCAGTGTCGCAC-3' 5'-GTGCGACACTGAATATTAATGACACATCATTTCG-3'
*Bpe*GReg (S68A)	5'-CAAGCTGCACGCAGCGATGCAAGATT-3' 5'-GCCAATCTTGCATCGCTGCGTGCAGC-3'
*Bpe*GReg (Y43F)	5'- GGCCGATTATTTTTTTGAATGCATGC-3' 5'-GCCAGCATGCATTCAAAAAAATAATC-3'
*Bpe*GReg (H99G)	5'-CAAAGTCGGCGAAGTGGGCGCGCGT-5' 5'-GATATCAATACGCGCGCCCACTTCGC-3'
*Pcc*Globin	5'-CCAGTCTCATGACCTGTAAGCCAAAAATGAAGAATC-3' 5'-GATTCTTCATTTTTGGCTTACAGGTCATGAGACTGG-3'
*Pcc*GCS (S82A)	5'- CTGCACGGCGCGATGTCCAAATGGATC-3' 5'- GATCCATTTGGACATCGCGCCGTGCAG-3'
*Pcc*GCS (Y57F)	5'-GCTGACAAGTTCTTTAGTTACATGCTG-3' 5'CAGCATGTAACTAAAGAACTTGTCAGC-3'
*Pcc*GCS (H112G)	5'-GAAAAAGATCGGCCAAATTGGAGCGCGTATTGGTATCCC G-3' 5'-CGGGATACCAATACGCGCTCCAATTTGGCCGATCTTTTT C-3'

Table S2. Heme Distortion Measurements for *Bpe*Globin, HemAT-*Bs*, and *Ec*DosC

Protein	basis	Doop (Over All)	B2u (Saddle)	B1u (Ruffle)	A2u (Dome)	Eg(x) (Wave)	Eg(y) (Wave)	A1u (Propeller)
BpeGlobin Chain A Fe(II)-O2	min.	1.2097	-0.2166	1.1334	-0.2677	-0.2420	-0.0449	0.0078
BpeGlobin Chain B Fe(II)-O2	min.	1.1112	-0.1265	-1.0331	-0.3227	0.0863	-0.2027	0.0042
BpeGlobin Chain C Fe(II)	min.	0.8975	-0.1820	0.8521	0.1451	-0.0832	0.1360	0.0019
HemAT-Bs 1or4 A Fe(III)-CN	min.	0.2709	0.2261	-0.1294	-0.0453	0.0058	0.0715	-0.0182
HemAT-Bs 1or4 B Fe(III)-CN	min.	0.4348	0.3655	-0.1064	-0.1169	0.1215	0.0992	-0.0797

HemAT-Bs 1or6 A Fe(III)	min.	0.5925	-0.2538	-0.4783	-0.2091	0.0097	0.0810	0.0855
HemAT-Bs 1or6 B Fe(III)	min.	0.5777	-0.0525	0.4509	0.3537	0.0188	0.0064	0.0275
EcDosC 4zva A Fe(III)	min.	0.8753	-0.0195	0.6838	0.5309	0.1071	-0.0962	0.0097
EcDosC 4zva B Fe(III)	min.	0.9615	-0.0467	0.7192	0.6272	0.1301	-0.0370	0.0092
EcDosC 4zvb A Fe(II)	min.	0.7853	-0.1611	-0.5816	-0.4979	-0.0278	-0.0921	-0.0069
EcDosC 4zvb B Fe(II)	min.	0.9832	-0.0143	-0.7557	-0.6335	0.0073	0.0310	-0.0087
EcDosC 4zvb C Fe(II)	min.	0.9618	-0.1107	-0.7576	-0.5773	-0.0152	-0.1054	-0.0107
EcDosC 4zvb D Fe(II)	min.	0.9594	-0.0491	0.7510	0.5946	-0.0062	-0.0808	0.0073

Figure S1. FTIR Spectrum of CO-Bound *Pcc*Globin in $H_2{}^{18}O$-Containing Buffer

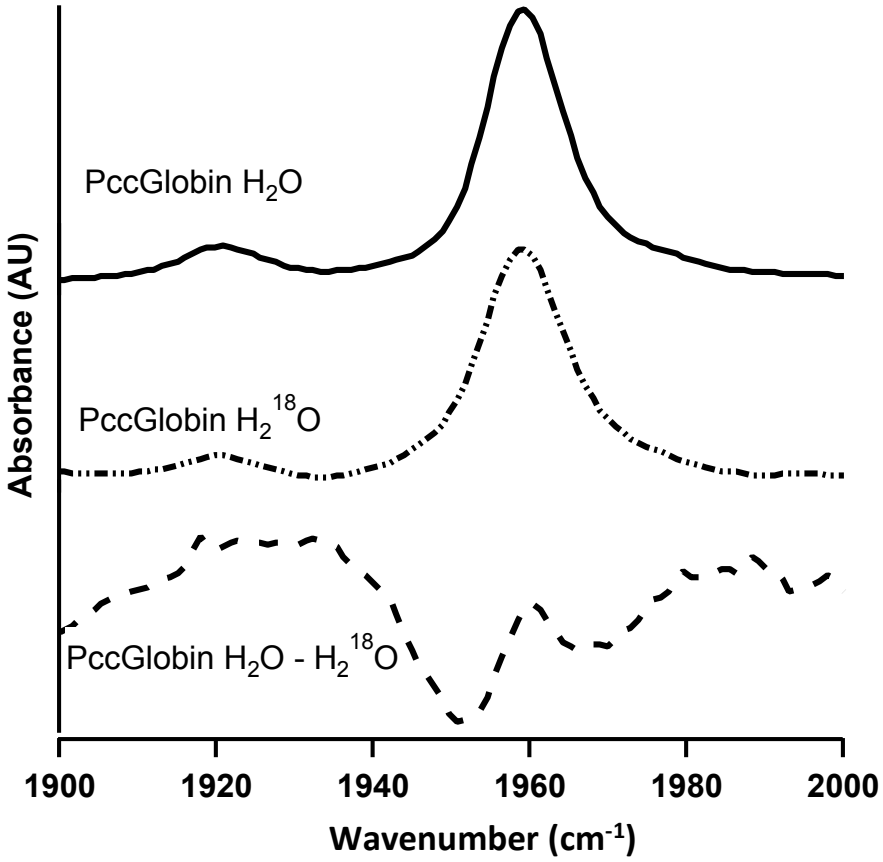

Figure S2. UV/Vis Spectra of *Pcc*GCS (H112G) in Different Ligation States