Published in final edited form as:

J Mol Biol. 2009 July 24; 390(4): 686–698. doi:10.1016/j.jmb.2009.05.023.

Crystal Structure of Histidine Phosphotransfer Protein ShpA, an Essential Regulator of Stalk Biogenesis in *Caulobacter crescentus*

Qingping Xu^{1,2}, Dennis Carlton^{1,3}, Mitchell D. Miller^{1,2}, Marc-André Elsliger^{1,3}, S. Sri Krishna^{1,4,5}, Polat Abdubek^{1,6}, Tamara Astakhova^{1,4}, Prasad Burra^{1,5}, Hsiu-Ju Chiu^{1,2}, Thomas Clayton^{1,3}, Marc C. Deller^{1,3}, Lian Duan^{1,4}, Ylva Elias^{1,3}, Julie Feuerhelm^{1,6}, Joanna C. Grant^{1,6}, Anna Grzechnik^{1,3}, Slawomir K. Grzechnik^{1,4}, Gye Won Han^{1,3}, Lukasz Jaroszewski^{1,4,5}, Kevin K. Jin^{1,2}, Heath E. Klock^{1,6}, Mark W. Knuth^{1,6}, Piotr Kozbial^{1,5}, Abhinav Kumar^{1,2}, David Marciano^{1,3}, Daniel McMullan^{1,6}, Andrew T. Morse^{1,4}, Edward Nigoghossian^{1,6}, Linda Okach^{1,6}, Silvy Oommachen^{1,2}, Jessica Paulsen^{1,6}, Ron Reyes^{1,2}, Christopher L. Rife^{1,2}, Natasha Sefcovic^{1,5}, Christine Trame^{1,2}, Christina V. Trout^{1,3}, Henry van den Bedem^{1,2}, Dana Weekes^{1,5}, Keith O. Hodgson^{1,7}, John Wooley^{1,4}, Ashley M. Deacon^{1,2}, Adam Godzik^{1,4,5}, Scott A. Lesley^{1,3,6}, and Ian A. Wilson^{1,3,*}

¹Joint Center for Structural Genomics, <http://www.jcsg.org>

²Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Stanford University, Menlo Park, CA, USA

³The Scripps Research Institute, La Jolla, CA, USA

⁴Center for Research in Biological Systems, University of California, San Diego, La Jolla, CA, USA

⁵Burnham Institute for Medical Research, La Jolla, CA, USA

⁶Genomics Institute of the Novartis Research Foundation, San Diego, CA, USA

⁷Photon Science, SLAC National Accelerator Laboratory, Menlo Park, CA, USA

Abstract

Cell cycle regulated stalk biogenesis in *Caulobacter crescentus* is controlled by a multi-step phosphorelay system consisting of the hybrid histidine kinase ShkA, the histidine-phosphotransfer protein ShpA and the response regulator TacA. ShpA shuttles phosphoryl groups between ShkA and TacA. When phosphorylated, TacA triggers a downstream transcription cascade for stalk synthesis in an RpoN-dependent manner. The crystal structure of ShpA was determined to 1.52 Å resolution. ShpA belongs to a family of monomeric histidine phosphotransfer (HPT) proteins, which feature a highly conserved four-helix bundle. The phosphorylatable histidine, His56, is located on the surface of the helix bundle and is fully solvent exposed. One end of the four-helix bundle in ShpA is shorter compared to other characterized histidine phosphotransfer proteins, whereas the face that potentially interacts with the response regulators is structurally conserved. Similarities of the interaction surface around the phosphorylation site suggest that ShpA is likely to share a common mechanism for

*Corresponding author. Dr. Ian Wilson, JCSG, The Scripps Research Institute, BCC206, 10550 North Torrey Pines Road, La Jolla, CA 92037. wilson@scripps.edu.

Accession number

Atomic coordinates and experimental structure factors for ShpA at 1.52 Å resolution have been deposited in the PDB (<http://www.wwpdb.org>) with accession number 2ooc.

molecular recognition and phosphotransfer with yeast phosphotransfer protein YPD1 despite low overall sequence similarity.

Keywords

Stalk biogenesis; phosphorelay; two-component signal transduction; Histidine phosphotransfer protein (HPt)

Introduction

Two-component signaling systems, the predominant signal transduction pathways in bacteria, are essential for bacteria survival, growth and development by enabling them to adapt to the environment ^{1,2}. In a canonical two-component system, the translation of extracellular environmental signals into cellular responses is achieved by two proteins. A cytoplasmic histidine kinase (HK), often fused to an extracellular sensor domain through a transmembrane helix (sensor HK), autophosphorylates at a histidine residue using ATP upon ligand binding to the sensor domain. The phosphoryl group is then transferred to an aspartate residue of a response regulator (RR) which becomes activated upon phosphorylation. The activated response regulator triggers appropriate downstream responses (*e.g.* gene expression). Bacteria, lower eukaryotes and plants have also evolved a more intricate phosphorelay system in which an additional histidine phosphotransfer protein accepts the phosphoryl group from the first response regulator domain (R1) and transfers it to the second response regulator domain (R2). This arrangement coordinates integration of multiple signals from different sources, propagation of one signal to multiple targets, as well as more complex regulation scenarios ³.

Two-component systems are critical for controlling cell division and development in *Caulobacter crescentus*, which has become an important model system for studying the regulation of the cell cycle and cellular differentiation ^{4–6}. *Caulobacter crescentus* divides asymmetrically, producing two cells with different structures and functions: a stalk daughter cell and a motile swarmer daughter cell with a single polar flagellum and pili. The swarmer cell cannot initiate DNA replication, but can differentiate into a stalk cell by shedding its polar flagellum, retracting its pili, and synthesizing a stalk at the former flagellum site after 30–45min of swimming. The adhesive polysaccharide holdfast is then synthesized, in parallel with initiation of a single round of DNA replication. The stalk of the pre-divisional cell is then extended while a new polar flagellum and pili are built at the opposite pole of the stalk, followed by a new round of cell division (Fig. 1a). The master cell cycle regulator CtrA (an RR) controls several key cell cycle events ⁷. Phosphorylated CtrA directly activates or represses the transcription of 95 genes which are essential for cell cycle processes, such as DNA methylation, morphogenesis and cell division. CtrA is responsible for activation of the genes involved in the biosynthesis of pili, holdfast and flagellum ^{6,8}. Phosphorylated CtrA also inhibits the initiation of DNA replication in the swarmer cell by binding to the replication origin.

Bacterial stalks are likely nutrient-scavenging antennae that allow stalked bacteria to survive in nutrient-limited environments ^{9,10}. Stalk synthesis is regulated by the cell cycle and the stalk length is controlled by both cell cycle and environmental cues. The cell cycle regulated stalk biogenesis is controlled by a ShkA-ShpA-TacA multi-step phosphorelay system (Fig. 1b) in a sigma factor $\sigma 54$ (RpoN)-dependent manner ¹¹. ShkA (stalk biogenesis histidine kinase A) is a cytoplasmic hybrid HK that contains an HK domain and a C-terminal RR domain. ShpA (stalk biogenesis histidine phosphotransferase A) is a small monomeric histidine phosphotransfer (HPt) protein (112 residues) that transfers the phosphoryl group from ShkA to TacA. TacA is an RpoN-dependent, NtrC-like transcription activator with an RR domain at its N-terminus ^{12,13}.

DNA microarray analysis has identified 30 genes involved in stalk biogenesis controlled by this phosphorelay system¹¹. This pathway was shown to be essential for stalk biogenesis as $\Delta shkA$, $\Delta shpA$ or $\Delta tacA$ mutants are stalkless, and $\Delta ShkA$ or $\Delta ShpA$ deletions can be rescued by the TacA D54E mutant which functionally mimics phosphorylated TacA. Phosphorylated TacA collaborates with RpoN to activate the expression of genes involved in the cell cycle regulated stalk biogenesis. Details about the upstream signal(s) for ShkA, and how the target genes of TacA function together, have yet to be fully explored. Several proteins involved in stalk synthesis, such as penicillin-binding protein 2 (PBP2), RodA, and MreB, are also required for cell elongation, suggesting that stalk synthesis is a special type of cell elongation¹⁴.

At least two types of histidine phosphotransfer proteins are found in two-component systems¹. Spo0B-like proteins assemble the four-helix bundle in the dimer¹⁵, whereas in ArcB^c or YPD1-like proteins, the four-helix bundle is present in the monomer^{16–18}. Despite their equivalence in functionality, it is likely these two types of histidine phosphotransfer proteins have evolved from different origins due to significant differences in sequence and structure. The monomeric form (hereafter referred as HPT) is more prevalent, as demonstrated by the PF01627 family (Pfam 23.0), which contains 2669 putative HPT domains¹⁹. A monomeric HPT exists either as a domain in a multi-domain protein (e.g. hybrid HKs) or as a separate protein. HPT domains located at the N-termini of hybrid HKs (e.g. CheA) often have HK-dependent autophosphorylation activity in addition to phosphotransfer activity, while C-terminal HPTs (e.g. ArcB^c) or monomeric HPTs (e.g. YPD1 or ShpA) are not able to undergo autophosphorylation^{1,2}. HPTs are found predominantly in bacteria, but are also present in archaea, lower eukaryotes, fungi, viruses, plants and probably amphibians, such as *Xenopus tropicalis* (based on transcript sequence homology)^{19,20}. The structures of HPTs are highly conserved, despite significant sequence diversity. Further study of their structure/function and evolutionary relationships is required to understand how specificity is achieved among cognate partners in two-component systems²¹.

Here, we report the crystal structure of ShpA from *Caulobacter crescentus*, determined using the high-throughput pipeline of the Joint Center for Structural Genomics (JCSG, <http://www.jcsg.org>)²², a large-scale Production Center of the National Institute of General Medical Sciences' Protein Structure Initiative (<http://www.nigms.nih.gov/Initiatives/PSI/>). The high resolution crystal structure shows unambiguously that ShpA is a prototypical HPT protein with surface properties similar to other HPTs despite low sequence similarity. The structure also suggests that flexible regions in HPT proteins may play a role in defining their specificity to RRs.

Results and Discussion

Overall structure

The selenomethionine form of full length ShpA (corresponding to residues 1–112) was expressed in *Escherichia coli* with an N-terminal TEV cleavable His-tag and purified by metal affinity chromatography (see Material & Methods). The crystal structure of ShpA was determined in space group P3₂21 at 1.52 Å resolution, with two monomers (A, B) per asymmetric unit (ASU), using the MAD method. The RMSD between monomers is 1.7 Å (104 C α). The model has good geometry with an all-atom clash score of 3.86, only 1.4% rotamer outliers, and all residues in favored regions in the Ramachandran plot according to MolProbity²³. The final model of ShpA contains residues 8–111 in monomer A, 8–112 in monomer B, two glycerols, 3 polyethylene glycols (PEG) and 219 waters. Electron density for residues A0 and B0 (N-terminal glycines remaining after cleavage of the expression and purification tag), A1-7, B1-7, A112, as well as the side chains of A72, A75, and A108, were poorly defined and were not built into the model. The Matthews' coefficient is 2.69 Å³/Da, corresponding to an

estimated solvent content of 54.2 %²⁴. Data collection, model, and refinement statistics are summarized in Table 1.

Analytical size exclusion chromatography, in combination with static light scattering, indicated that ShpA exists as a monomer in solution. ShpA has an elongated shape with molecular dimensions of 25.4 Å × 23.9 Å × 49.5 Å (Fig. 2), and is an all- α protein with 75.2% of its residues in a helical conformation (H1–H5). Four of these helices, H2–H5, form an up-and-down four-helix bundle that encloses a hydrophobic core. H3 (residues 51–65) and H4 (68–78) on one side, are shorter than H2 (24–43) and H5 (84–111) on the other side. The hydrophobic interior between the two long helices (H2 and H5) is further protected by an additional N-terminal “capping” helix H1 (residues 12–16), which sits atop the H3–H4 helices and oriented at 131.5° to helix H2. The connecting loops between helices are short in ShpA. The H2–H3 loop (Pro42–Pro49) contains three prolines, whereas H3 and H4 form an anti-parallel helical hairpin connected by a γ turn so that the H3–H4 distance is only 7.2 Å, and significantly closer than other helical pairs in the bundle [H2–H3 (9.8 Å), H4–H5 (9.3 Å) and H2–H5 (9.9 Å)]. The fold of ShpA is typical of monomeric HPts, such as ArcB^c 16 and YPD1¹⁸. The site of phosphorylation, His56, is located in the second turn of the four-turn helix H3 on the surface of the protein.

Conservation of monomeric HPts

Several monomeric HPt domains have been structurally characterized, including the C-terminal domain of ArcB (ArcB^c) from *Escherichia coli* 16, YPD1 from *Saccharomyces cerevisiae* 17,18, ZmHP2 from maize²⁵ and AK104879 from rice (87% sequence identity to ZmHP2, Center for Eukaryotic Structural Genomics), bacteria P1 domains of CheA (CheA-P1)^{26–28}, LuxU from *Vibrio harveyi* 29, and YojN from *E. coli* 30 (Fig. 3a). Structural comparisons of ShpA to other HPts are outlined in Table 2. ShpA is structurally similar to other HPts with RMSDs between 2.3–3.3 Å despite low overall sequence identities between 7–24%. The overall structure of ShpA is most similar to ZmHP2, YPD1, and ArcB^c.

One unique structural feature of ShpA is that the H3 helix upstream of His56 is one turn shorter, with 4 turns compared to the usual five turns in other HPts, such as YPD1 (Fig. 3a). The other helices in the bundle (H2, H4 and H5) are also shorter (2–3 turns) on this face of the molecule (bottom). As a result, His56 is closer to the bottom of the bundle (Fig. 2). On the other hand, the spacing between His56 and helix H1, the likely binding interface for RRs, does not change significantly.

The N-terminal helix (*i.e.* H1 of ShpA) is also present in all HPts, except for the CheA P1 domains. One of the roles for this helix is to provide structural stability to the four-helix bundle¹⁸. Additionally, the equivalent helix in YPD1 is directly involved in the interaction with RRs³¹. The placement of these N-terminal helices varies significantly in HPt structures. The protrusion of these helices with respect to the plane perpendicular to the phosphorylatable histidine (P-His) is correlated with the length and level of bending of the second and last helices in the four-helix bundle. The ArcB^c helix is the most protruding among HPts due to severely kinked helices, while the corresponding YojN and LuxU helices protrude less. The N-terminal helices of ShpA and YPD1 occupy similar spatial locations (Fig. 3a).

The four-helix bundle, which usually contains two shorter helices at the front and two longer helices at the back, is conserved in all HPts. The conserved core common to all known HPts consists of a compact four-helix bundle (~60 residues) which can be mapped to helices H2 (residues 27–41), H3–H4 (residues 50–78), and H5 (residues 83–98) of ShpA (Fig. 3b). The anti-parallel helical hairpin H3–H4 is the most conserved region in all known HPts¹⁸. This region of ShpA is highly similar in both sequence and structure to the equivalent regions of other HPts, with RMSDs of 1.1 Å or less, with the exception of LuxU (Table 2, Figs. 3c–d).

The helical hairpin region of LuxU has an RMSD of 3.2 Å compared to ShpA, similar to the overall RMSD between ShpA and LuxU. A closer inspection indicates that P-His of LuxU is located approximately one residue down from the conserved histidine position in other HPTs (*i.e.* equivalent to position 57 of ShpA). In addition to the H3–H4 hairpin region, H2 of ShpA is also highly conserved structurally among HPTs (FIG. 3C).

The site of phosphorylation

The H56A ShpA mutant disrupts phosphotransfer *in vitro* and stalk formation *in vivo*, suggesting that His56 is the only site of phosphorylation in ShpA¹¹. His56 is located on the surface of helix H3, and adopts a low-energy rotamer ($\chi_1=170^\circ$, $\chi_2=80^\circ$). It is fully exposed to the solvent with well-defined density. Residues on the solvent exposed surface of the C-terminal portion of H3 (upstream from His56) either have small or no side chains (Gly60, Ala61 and Gly64) or point away from His56 (Lys59 and Arg63) (Fig. 4a). Phosphorylation of YPD1 and ArcB^c is severely affected by mutations of the highly conserved glycine to bulkier side chains next to the phosphorylation site (equivalent to Gly60 of ShpA)^{32–34}. Thus, conservation of this feature in ShpA is consistent with the finding that substantial solvent exposure of the P-His is critical for its interaction with RRs (Fig. 3c).

Inter-helical hydrogen bonds are commonly observed within the helical hairpin in HPTs, and may help stabilize the conformation of the H3–H4 hairpin. In ShpA, three hydrogen bonds are observed between Lys59 and Glu78, and Arg63 with Glu72 and Glu75. One particular hydrogen bond, between P-His (designated as position 0) and a nearby glutamate/glutamine residue from the second helix of the helix-hairpin (position +22), is highly conserved among HPTs (Fig. 3d). This hydrogen bond may orient the imidazole side chain or increase its nucleophilicity during the phosphotransfer reaction³⁴. Absence of a hydrogen bond to P-His may affect the side-chain conformation of P-His. For example, the His80 imidazole in the rice ZmHP2 homolog AK104879 (PDB 1yvi), which instead contains an arginine at +22, is flipped 180° with respect to the commonly observed rotamer (likely due to crystal packing). In addition, a glutamate at +22 can form an additional hydrogen bond with the highly conserved lysine at position +3, such as in CheA P1 domains^{26,27} (Fig. 4c). ShpA contains the equivalent three residues as the CheA P1 domains: His56, Lys59 and Glu78. However, only one hydrogen bond between Glu78 and Lys59 was observed in ShpA (Fig. 4a). Glu78O ϵ and His56N δ are too far apart (4.1–4.2 Å) to form a hydrogen bond. Instead, a water-mediated bridge is observed between these side chains. The orientation of the His56 side chain remains similar to those in other HPTs. The conserved hydrogen bond to the P-His is likely to have a similar functional role in all HPTs. Mutational analysis of the glutamate in the CheA P1 domain indicated that the loss of the glutamate had an adverse effect on the ATP-dependent autophosphorylation of the P1 domain, but little effect on the interaction with its cognate response regulator CheY²⁷, in agreement with similar observations with YPD1³³ and ArcB^c³⁵. Although it has been proposed that this hydrogen bond may not be essential for phosphotransfer between HPTs and RRs²⁷, a more recent kinetic study of YPD1 suggested that it may affect the rate of phosphotransfer³⁴.

Lys59 and Arg63 form a small positive patch on one side of the P-His (Fig. 4b). In monomer A, Lys52 further contributes to expansion of this basic surface. Mutational studies of YPD1 and structural studies of the YPD1/SLN1 complex have indicated that this lysine (Lys59 in ShpA) plays a role in binding the conserved glutamate/aspartate residue in the active site of RRs^{31,34,36}. Positive residues on one side of the P-His are also observed in other HPTs, although they are not conserved in the sequence (except for Lys59). For example, Arg90 of YPD1 is also located near P-His (Fig. 4c) and affects the half-life of the phosphorylated histidine, as well as the phosphotransfer rate^{33,34}.

Interactions with RRs

A triangular hydrophobic surface, formed by H1, H2 and the C-terminal portion of H3, is located above His56 (Fig. 5a). Similar hydrophobic surfaces are also present in YPD1 and ArcB^c 17,¹⁸. Yeast two-hybrid assays of YPD1 with its cognate RRs (SLN1, SKN7 and SSK1), as well as the crystal structure of the YPD1/SLN1-RR complex, revealed that YPD1 interacts with RRs through this common hydrophobic surface^{31,36–38}. Therefore, it is likely that ShpA interacts with TacA and ShkA using a common mechanism³¹.

Many residues involved in the YPD1-SLN1 interaction are conserved in ShpA (Fig. 5b). The most conserved regions include helix H1 (Tyr14 and Phe18), the N-terminal portion of H2 (Val25, Val29 and Leu32) and the C-terminal portion of H3 (Gly60, Ala61 and Gly64). The primary interaction between SLN1-RR and YPD1 involves docking of the first helix (α 1) of SLN1-RR between helices H2 and H3 of YPD1³¹. The designation of this region of ShpA as a *bona fide* helix docking site is supported by the crystal packing interactions which mimic the interaction observed in the YPD1/SLN1 complex (Fig. 5c). Furthermore, the hydrophobic residues on SLN1-RR, which interact with YPD1, are also conserved in cognate RRs TacA and ShkA of ShpA (Fig. 5d).

A BLAST sequence similarity search against a non-redundant database³⁹, using ShpA as a probe, uncovered 30 closely related HPt proteins, that are primarily distributed in α -proteobacteria. Additional ShpA-like domains are also present in the C-terminal region of hybrid HKs. Sequence conservation among these homologous HPts is found in the H1–H4 region (residues 8–78). The most highly conserved residues (Asp11, Leu15, Val25, Glu28, Val29, Leu30, Leu32, Phe33, Gln36, His56, Val58, Lys59 and Gly60, Ala61, Ala62, Gly66 and Glu78) correlate well with those identified through comparisons of YPD1 and ShpA. Most of these residues are important for interaction with RRs (Fig. 4d). Glu28 (Q/E) is highly conserved in ShpA homologs, but not YPD1, suggesting that a specific role in the interaction with TacA or ShkA. Asp11 forms an “N-cap” (with Tyr14 main-chain NH) to stabilize the H1 helix.

Structural flexibility in ShpA

Beyond the structurally well-conserved helix hairpin and the N-terminal portion of H2, the other helices in HPts are generally less conserved. Helices H1 and H5 of ShpA display higher RMSDs in comparison to other HPts. This structural variation distant from the hairpin appears to be a general feature of HPts. Structural differences in the H1 and H2 regions are likely to be functionally relevant since these two helices are important for RR interactions.

Interestingly, ShpA itself also displays structural flexibility in this region. A comparison of the two monomers in the ASU indicates that H1, the H1–H2 loop, H2 and C-terminal portion of H5 display the largest structural differences (Figs. 6a–b). These differences are likely induced by crystal packing as these structural elements have different packing environments. Significant structural rearrangements are observed in the H1–H2 loop in order to optimize inter-molecular contacts (Fig. 6c). Helix H1 in the B monomer is shorter by one turn at the C-terminus. The large structural differences in this region, when the two monomers are compared, correlate with much higher B-values in monomer B (Fig. 6a). As the H1 helix and the H1–H2 loop of the monomer B make few crystal contacts, their conformations are likely more representative of the solution state. Thus, the higher thermal motion and the propensity for the H1–H2 region to undergo structural adaptation in the crystal lattice suggest that the observed flexibility in this region is likely an inherent property in solution. Flexibility of the H1 and H1–H2 loops are also observed in other HPts, such as YPD1 and ZmHP2^{18,25}. A second region with large atomic displacements occurs at the C-terminal end of H2, due to Trp40 adopting

two distinct conformations. Its side chain is completely buried between H2 and H3 in monomer A, but is solvent exposed in monomer B.

The CheA HPT domain also displays different conformational states²⁶. Protein dynamics are important for the function of the RRs⁴⁰. However, there are currently no direct studies on the effect of structural heterogeneity on the function of an HPT. Studies of YPD1 with different RRs indicate that residues in these flexible regions are involved in binding of the response regulators³⁷. Thus, we postulate that some flexibility may facilitate the binding of HPTs to multiple partners. The induced fit of a flexible region to different binding partners could be one of the determinants for specificity or affinity between HPTs and RRs.

Specificity between ShpA and RRs

Since HPTs must interact with at least two RRs which are homologous in structure, they usually possess features which allow them to bind RRs promiscuously. It has been observed that HPT proteins can interact with non-cognate RRs. For example, yeast YPD1 can accept the phosphoryl group from bacterial CheY *in vitro*⁴¹ and complement the phosphorelay protein RdeA from *Dictyostelium discoideum*⁴². These observations are not surprising since the general features of HPT-RR interaction are conserved in both HPTs and RRs. For example, response regulators CtrA and CpdR also contain conserved hydrophobic residues, which SLN1-RR enlist to bind YPD1 (Fig. 5d). ShpA can transfer the phosphoryl group to other RRs in *Caulobacter crescentus* (e.g. CC3302 and CC0921); however, it clearly has significantly higher specificity for its cognate RRs (ShkA and TacA)¹¹. Additionally, it has also been observed that the HPTs often have different affinities among multiple cognate RRs. For example, YPD1 has a higher affinity for SSK1 compared to SLN1 and SKN7⁴³. Thus, the use of common interaction interfaces raises the question of how fidelity between HPTs and RRs are ensured and how cross-talk between similar pathways is avoided. The molecular basis by which such specificity is achieved is not well understood²¹. Nevertheless, interactions between YPD1 and all of its cognate RRs have been systematically studied by yeast two-hybrid assays^{37,38}. These studies suggest that, although a conserved recognition mechanism is utilized for HPT interaction with all of the RRs, specificity between HPT and particular RRs may come from specific localized interactions, mostly at the perimeter of the core recognition area³⁷. Possible factors that may contribute to the strength of HPT-RR interaction could include surface complementarity, size of the contact area, and the balance between the favorable (e.g. hydrophobic contact and hydrogen bonds) and unfavorable (e.g. steric hindrance or charge repulsion) interactions. The known HPT structures seem to support the hypothesis above about the specificity determinants. Although many residues on the potential RR interaction surfaces are conserved to preserve the hydrophobic core (Fig. 3c), differences arise in the shape of these surfaces. Additionally, conserved residues specific to each HPT subfamily, such as Glu28 in ShpA, are likely involved.

The number of two-component HK/RR genes in *Caulobacter crescentus* (34 HK genes, 44 RR genes and 27 hybrid HK/RR genes)⁴⁴ is higher than in typical bacteria, such as *Escherichia coli* (23 HKs, 32 RRs, 5 hybrid HK/RR)⁴⁵, potentially increasing the risk of cross-talk. One possible strategy of achieving signaling efficiency and accuracy is to maintain spatial proximity between cognate partners, including most of HPTs, since they are frequently fused together in two-component systems. Moreover, the population of potential promiscuous components is small. Only four YPD1-like HPTs have been identified through sequence analysis in *Caulobacter crescentus*, and occur in monomeric HPTs ShpA and CC1220 (192 residues, gi|16125470)¹¹, as well as at the N-terminal domain of two CheAs (gi|16124848 and gi|16124688)²⁰. CC1220 and ShpA only have limited sequence similarity around the phosphorylation site. Additional novel HPTs may exist in *Caulobacter crescentus*, such as ChpT in the CckA-ChpT-CtrA phosphorelay⁴⁶. However, ChpT lacks global or local sequence

similarity to ShpA. Therefore, the limited number of HPts and the significant sequence/structure diversity among them could be another strategy to avoid cross-talk. *Caulobacter crescentus* may have additional layers of protection against unwanted cross-talk. The regulatory proteins of *Caulobacter crescentus*, such as TacA, may also be subject to temporal and spatial control^{11,47,48}.

Materials and Methods

Protein production and crystallization

Clones were generated using the PIPE cloning method⁴⁹. The gene encoding ShpA (GenBank: NP_419930, GI:16125366, Swiss-Prot: Q9A980) was amplified by polymerase chain reaction (PCR) from *Caulobacter crescentus* ATCC 19089 genomic DNA using *Pfu*Turbo DNA polymerase (Stratagene) and I-PIPE (insert) primers that included sequences for the predicted 5' and 3' ends. The expression vector, pSpeedET, which encodes an amino-terminal tobacco etch virus (TEV) protease-cleavable expression and purification tag (MGSDKIHSHHHHHENLYFQG), was PCR amplified with V-PIPE (vector) primers. V-PIPE and I-PIPE PCR products were mixed to anneal the amplified DNA fragments. *Escherichia coli* GeneHogs (Invitrogen) competent cells were transformed with the V-PIPE / I-PIPE mixture and dispensed on selective LB-agar plates. The cloning junctions were confirmed by DNA sequencing. Expression was performed in a selenomethionine-containing medium. At the end of fermentation, lysozyme was added to the culture to a final concentration of 250 µg/ml, the cells were harvested and frozen. After one freeze/thaw cycle the cells were sonicated in the lysis buffer [50mM HEPES pH 8.0, 50mM NaCl, 10mM imidazole, 1mM Tris(2-carboxyethyl)phosphine-HCL (TCEP)] and the lysate was clarified by centrifugation at 32,500 × g for 30min. The soluble fraction was passed over nickel-chelating resin (GE Healthcare) pre-equilibrated with lysis buffer, the resin washed with wash buffer [50mM HEPES pH 8.0, 300mM NaCl, 40mM imidazole, 10% (v/v) glycerol, 1mM TCEP], and the protein eluted with elution buffer [20mM HEPES pH 8.0, 300mM imidazole, 10% (v/v) glycerol, 1mM TCEP]. The eluate was buffer exchanged with TEV buffer [20mM HEPES pH 8.0, 200mM NaCl, 40mM imidazole, 1mM TCEP] using a PD-10 column (GE Healthcare) and incubated with 1mg of TEV protease per 15 mg of eluted protein. The protease-treated eluate was passed over nickel-chelating resin (GE Healthcare) pre-equilibrated with TEV buffer and the resin was washed with the same buffer. The flow-through and wash fractions were combined, concentrated, and further purified on a 1.6 × 60 cm HiLoad Superdex 200 column (GE Healthcare) with isocratic elution in crystallization buffer [20mM Tris pH 7.9, 150mM NaCl, 0.25mM TCEP]. The protein was concentrated to 15.6 mg/ml by centrifugal ultrafiltration (Millipore) for crystallization trials. ShpA was crystallized using the nanodroplet vapor diffusion method⁵⁰ with standard JCSG crystallization protocols²². The crystallization reagent contained 1.5% polyethylene glycol 400, 15% (v/v) glycerol, 1.9M ammonium sulfate, and 0.1M HEPES pH 6.9. No additional cryoprotectant was added to the crystals. Initial screening for diffraction was carried out using the Stanford Automated Mounting system (SAM)⁵¹ at the Stanford Synchrotron Radiation Lightsource (SSRL, Menlo Park, CA). The data were indexed in space group P3₂21. The molecular weight and oligomeric state of ShpA were determined using a 1 × 30cm Superdex 200 column (GE Healthcare) in combination with static light scattering (Wyatt Technology). The mobile phase consisted of 20mM Tris pH 8.0, 150mM NaCl, and 0.02% (w/v) sodium azide.

Data Collection, Structure Solution, and Refinement

Multi-wavelength anomalous diffraction (MAD) data were collected from a crystal at SSRL beamline 1–5 with an ADSC Quantum 4 detector at wavelengths corresponding to the high energy remote (λ_2) and inflection (λ_3) of a selenium MAD experiment. Another data set (λ_1 Se) at slightly higher resolution was collected at a wavelength corresponding to the high

energy remote from a second crystal at SSRL beamline 11-1 using a MarCCD 325 detector (MarResearch, USA). Both data sets were collected at 100 K using the BLU-ICE data collection environment⁵². The data processing and structure solution were carried using Xsolve, an automatic structure solution pipeline developed at the JCSG. The MAD data were integrated and reduced using XDS and then scaled with the program XSCALE⁵³. Selenium sites were located with SHELXD⁵⁴. Phase refinement and automatic model building were performed using autoSHARP⁵⁵ and ARP/wARP⁵⁶. The data for refinement (λ_1 Se) were processed with MOSFLM AND SCALA^{57,58}. Model completion and refinement were performed with COOT⁵⁹ and REFMAC⁶⁰. the CCP4 suite of programs were used for data conversion and other calculations⁶¹. Data and refinement statistics are summarized in Table 1. Analysis of the stereochemical quality of the model was accomplished using MolProbity²³. All molecular graphs were prepared with PyMOL with the monomer B (DeLano Scientific) unless specifically stated. Structural comparisons were performed by DALI⁶² and LSQAB⁶³. The electrostatic potential was calculated with APBS using default parameters⁶⁴. Multiple sequence alignment of the response regulator domains in Fig. 5d was calculated by Clustal W⁶⁵.

Abbreviations used

HPT, histidine phosphotransfer protein; HK, histidine kinase; RR, response regulator; P-His, phosphorylatable histidine; RMSD, root mean square deviation; SeMet, selenomethionine-labeled; ASU, asymmetric unit.

Acknowledgements

We greatly appreciate valuable comments on the manuscript by Professor Ann West of the University of Oklahoma. The project is sponsored by National Institutes of General Medical Sciences Protein Structure Initiative (U54 GM074898). Portions of this research were carried out at the Stanford Synchrotron Radiation Lightsource (SSRL) and the Advanced Light Source (ALS). The SSRL is a national user facility operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research, and by the National Institutes of Health (National Center for Research Resources, Biomedical Technology Program, and the National Institute of General Medical Sciences). Genomic DNA from *Caulobacter crescentus* ATCC 19089 (ATCC Number: 19089D) was obtained from the American Type Culture Collection (ATCC). The content is solely the responsibility of the authors and does not necessarily represent the official views of NIGMS or NIH.

References

1. Stock AM, Robinson VL, Goudreau PN. Two-component signal transduction. Annu. Rev. Biochem 2000;69:183–215. [PubMed: 10966457]
2. West AH, Stock AM. Histidine kinases and response regulator proteins in two-component signaling systems. Trends Biochem. Sci 2001;26:369–376. [PubMed: 11406410]
3. Appleby JL, Parkinson JS, Bourret RB. Signal transduction via the multi-step phosphorelay: not necessarily a road less traveled. Cell 1996;86:845–848. [PubMed: 8808618]
4. Jacobs-Wagner C. Regulatory proteins with a sense of direction: cell cycle signalling network in *Caulobacter*. Mol. Microbiol 2004;51:7–13. [PubMed: 14651607]
5. Laub MT, Shapiro L, McAdams HH. Systems biology of *Caulobacter*. Annu. Rev. Genet 2007;41:429–441. [PubMed: 18076330]
6. Skerker JM, Laub MT. Cell-cycle progression and the generation of asymmetry in *Caulobacter crescentus*. Nat. Rev. Microbiol 2004;2:325–337. [PubMed: 15031731]
7. Quon KC, Marczyński GT, Shapiro L. Cell cycle control by an essential bacterial two-component signal transduction protein. Cell 1996;84:83–93. [PubMed: 8548829]
8. Laub MT, Chen SL, Shapiro L, McAdams HH. Genes directly controlled by CtrA, a master regulator of the *Caulobacter* cell cycle. Proc. Natl. Acad. Sci. USA 2002;99:4632–4637. [PubMed: 11930012]

9. Lawler ML, Brun YV. Advantages and mechanisms of polarity and cell shape determination in *Caulobacter crescentus*. *Curr. Opin. Microbiol* 2007;10:630–637. [PubMed: 17997127]
10. Wagner JK, Setayeshgar S, Sharon LA, Reilly JP, Brun YV. A nutrient uptake role for bacterial cell envelope extensions. *Proc. Natl. Acad. Sci. USA* 2006;103:11772–11777. [PubMed: 16861302]
11. Biondi EG, Skerker JM, Arif M, Prasol MS, Perchuk BS, Laub MT. A phosphorelay system controls stalk biogenesis during cell cycle progression in *Caulobacter crescentus*. *Mol. Microbiol* 2006;59:386–401. [PubMed: 16390437]
12. Marques MV, Gomes SL, Gober JW. A gene coding for a putative sigma 54 activator is developmentally regulated in *Caulobacter crescentus*. *J. Bacteriol* 1997;179:5502–5510. [PubMed: 9287006]
13. Brun YV, Shapiro L. A temporally controlled sigma-factor is required for polar morphogenesis and normal cell division in *Caulobacter*. *Genes Dev* 1992;6:2395–2408. [PubMed: 1459461]
14. Wagner JK, Brun YV. Out on a limb: how the *Caulobacter* stalk can boost the study of bacterial cell shape. *Mol. Microbiol* 2007;64:28–33. [PubMed: 17376069]
15. Varughese KI, Madhusudan, Zhou XZ, Whiteley JM, Hoch JA. Formation of a novel four-helix bundle and molecular recognition sites by dimerization of a response regulator phosphotransferase. *Mol. Cell* 1998;2:485–493. [PubMed: 9809070]
16. Kato M, Mizuno T, Shimizu T, Hakoshima T. Insights into multistep phosphorelay from the crystal structure of the C-terminal HPt domain of ArcB. *Cell* 1997;88:717–723. [PubMed: 9054511]
17. Song HK, Lee JY, Lee MG, Moon J, Min K, Yang JK, et al. Insights into eukaryotic multistep phosphorelay signal transduction revealed by the crystal structure of Ypd1p from *Saccharomyces cerevisiae*. *J. Mol. Biol* 1999;293:753–761. [PubMed: 10543964]
18. Xu Q, West AH. Conservation of structure and function among histidine-containing phosphotransfer (HPt) domains as revealed by the crystal structure of YPD1. *J. Mol. Biol* 1999;292:1039–1050. [PubMed: 10512701]
19. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, et al. The Pfam protein families database. *Nucleic Acids Res* 2004;32:D138–D141. [PubMed: 14681378]
20. Gough J, Karplus K, Hughey R, Chothia C. Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *J. Mol. Biol* 2001;313:903–919. [PubMed: 11697912]
21. Laub MT, Goulian M. Specificity in two-component signal transduction pathways. *Annu. Rev. Genet* 2007;41:121–145. [PubMed: 18076326]
22. Lesley SA, Kuhn P, Godzik A, Deacon AM, Mathews I, Kreusch A, et al. Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline. *Proc. Natl. Acad. Sci. USA* 2002;99:11664–11669. [PubMed: 12193646]
23. Davis IW, Murray LW, Richardson JS, Richardson DC. MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res* 2004;32:W615–W619. [PubMed: 15215462]
24. Matthews BW. Solvent content of protein crystals. *J. Mol. Biol* 1968;33:491–497. [PubMed: 5700707]
25. Sugawara H, Kawano Y, Hatakeyama T, Yamaya T, Kamiya N, Sakakibara H. Crystal structure of the histidine-containing phosphotransfer protein ZmHP2 from maize. *Protein Sci* 2005;14:202–208. [PubMed: 15576555]
26. Quezada CM, Gradinaru C, Simon MI, Bilwes AM, Crane BR. Helical shifts generate two distinct conformers in the atomic resolution structure of the CheA phosphotransferase domain from *Thermotoga maritima*. *J. Mol. Biol* 2004;341:1283–1294. [PubMed: 15321722]
27. Mourey L, Da Re S, Pedelacq JD, Tolstykh T, Faurie C, Guillet V, et al. Crystal structure of the CheA histidine phosphotransfer domain that mediates response regulator phosphorylation in bacterial chemotaxis. *J. Biol. Chem* 2001;276:31074–31082. [PubMed: 11387324]
28. Zhou H, Dahlquist FW. Phosphotransfer site of the chemotaxis-specific protein kinase CheA as revealed by NMR. *Biochemistry* 1997;36:699–710. [PubMed: 9020767]
29. Ulrich DL, Kojetin D, Bassler BL, Cavanagh J, Loria JP. Solution structure and dynamics of LuxU from *Vibrio harveyi*, a phosphotransferase protein involved in bacterial quorum sensing. *J. Mol. Biol* 2005;347:297–307. [PubMed: 15740742]

30. Rogov VV, Bernhard F, Lohr F, Dotsch V. Solution structure of the *Escherichia coli* YojN histidine-phosphotransferase domain and its interaction with cognate phosphoryl receiver domains. *J. Mol. Biol* 2004;343:1035–1048. [PubMed: 15476819]
31. Xu Q, Porter SW, West AH. The yeast YPD1/SLN1 complex: insights into molecular recognition in two-component signaling systems. *Structure* 2003;11:1569–1581. [PubMed: 14656441]
32. Matsushika A, Mizuno T. The structure and function of the histidine-containing phosphotransfer (HPt) signaling domain of the *Escherichia coli* ArcB sensor. *J. Biochem* 1998;124:440–445. [PubMed: 9685739]
33. Janiak-Spens F, West AH. Functional roles of conserved amino acid residues surrounding the phosphorylatable histidine of the yeast phosphorelay protein YPD1. *Mol. Microbiol* 2000;37:136–144. [PubMed: 10931311]
34. Janiak-Spens F, Cook PF, West AH. Kinetic analysis of YPD1-dependent phosphotransfer reactions in the yeast osmoregulatory phosphorelay system. *Biochemistry* 2005;44:377–386. [PubMed: 15628880]
35. Matsushika A, Mizuno T. Mutational analysis of the histidine-containing phosphotransfer (HPt) signaling domain of the ArcB sensor in *Escherichia coli*. *Biosci. Biotechnol. Biochem* 1998;62:2236–2238. [PubMed: 9972245]
36. Zhao X, Copeland DM, Soares AS, West AH. Crystal structure of a complex between the phosphorelay protein YPD1 and the response regulator domain of SLN1 bound to a phosphoryl analog. *J. Mol. Biol* 2008;375:1141–1151. [PubMed: 18076904]
37. Porter SW, West AH. A common docking site for response regulators on the yeast phosphorelay protein YPD1. *Biochim. Biophys. Acta* 2005;1748:138–145. [PubMed: 15769590]
38. Porter SW, Xu Q, West AH. Ssk1p response regulator binding surface on histidine-containing phosphotransfer protein Ypd1p. *Eukary. Cell* 2003;2:27–33.
39. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;25:3389–3402. [PubMed: 9254694]
40. Stock AM, Guhaniyogi J. A new perspective on response regulator activation. *J. Bacteriol* 2006;188:7328–7330. [PubMed: 17050920]
41. Janiak-Spens F, Sparling JM, Gurfinkel M, West AH. Differential stabilities of phosphorylated response regulator domains reflect functional roles of the yeast osmoregulatory SLN1 and SSK1 proteins. *J. Bacteriol* 1999;181:411–417. [PubMed: 9882653]
42. Chang WT, Thomason PA, Gross JD, Neweil PC. Evidence that the RdeA protein is a component of a multistep phosphorelay modulating rate of development in *Dictyostelium*. *Embo J* 1998;17:2809–2816. [PubMed: 9582274]
43. Janiak-Spens F, Sparling DP, West AH. Novel role for an HPt domain in stabilizing the phosphorylated state of a response regulator domain. *J. Bacteriol* 2000;182:6673–6678. [PubMed: 11073911]
44. Nierman WC, Feldblyum TV, Laub MT, Paulsen IT, Nelson KE, Eisen JA, et al. Complete genome sequence of *Caulobacter crescentus*. *Proc. Natl. Acad. Sci. USA* 2001;98:4136–4141. [PubMed: 11259647]
45. Mizuno T. Compilation of all genes encoding two-component phosphotransfer signal transducers in the genome of *Escherichia coli*. *DNA Res* 1997;4:161–168. [PubMed: 9205844]
46. Biondi EG, Reisinger SJ, Skerker JM, Arif M, Perchuk BS, Ryan KR, et al. Regulation of the bacterial cell cycle by an integrated genetic circuit. *Nature* 2006;444:899–904. [PubMed: 17136100]
47. Jenal U, Stephens C. The *Caulobacter* cell cycle: timing, spatial organization and checkpoints. *Curr. Opin. Microbiol* 2002;5:558–563. [PubMed: 12457698]
48. Ausmees N, Jacobs-Wagner C. Spatial and temporal control of differentiation and cell cycle progression in *Caulobacter crescentus*. *Annu. Rev. Microbiol* 2003;57:225–247. [PubMed: 14527278]
49. Klock HE, Koesema EJ, Knuth MW, Lesley SA. Combining the polymerase incomplete primer extension method for cloning and mutagenesis with microscreening to accelerate structural genomics efforts. *Proteins* 2008;71:982–994. [PubMed: 18004753]

50. Santarsiero BD, Yegian DT, Lee CC, Spraggon G, Gu J, Scheibe D, et al. An approach to rapid protein crystallization using nanodroplets. *J. Appl. Cryst* 2002;35:278–281.
51. Cohen AE, Ellis PJ, Miller MD, Deacon AM, Phizackerley RP. An automated system to mount cryo-cooled protein crystals on a synchrotron beamline, using compact samples cassettes and a small-scale robot. *J. Appl. Cryst* 2002;35:720–726.
52. McPhillips TM, McPhillips SE, Chiu HJ, Cohen AE, Deacon AM, Ellis PJ, et al. Blu-Ice and the Distributed Control System: software for data acquisition and instrument control at macromolecular crystallography beamlines. *J. Synchrotron Radiat* 2002;9:401–406. [PubMed: 12409628]
53. Kabsch W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst* 1993;26:795–800.
54. Schneider TR, Sheldrick GM. Substructure solution with SHELXD. *Acta Cryst. D* 2002;58:1772–1779. [PubMed: 12351820]
55. Vonrhein C, Blanc E, Roversi P, Bricogne G. Automated structure solution with autoSHARP. *Methods Mol. Biol* 2007;364:215–230. [PubMed: 17172768]
56. Cohen SX, Morris RJ, Fernandez FJ, Ben Jelloul M, Kakaris M, Parthasarathy V, et al. Towards complete validated models in the next generation of ARP/wARP. *Acta Cryst. D* 2004;60:2222–2229. [PubMed: 15572775]
57. Leslie AGW. Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4+ESF-EAMCB Newsletter on Protein Crystallography*. 1992
58. Evans P. Scaling and assessment of data quality. *Acta Cryst. D* 2006;62:72–82. [PubMed: 16369096]
59. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Cryst. D* 2004;60:2126–2132. [PubMed: 15572765]
60. Winn MD, Murshudov GN, Papiz MZ. Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol* 2003;374:300–321. [PubMed: 14696379]
61. CCP4. The CCP4 suite: programs for protein crystallography. *Acta Cryst. D* 1994;50:760–763. [PubMed: 15299374]
62. Holm L, Sander C. Dali: a network tool for protein structure comparison. *Trends Biochem. Sci* 1995;20:478–480. [PubMed: 8578593]
63. Kabsch W. A solution for the best rotation to relate two sets of vectors. *Acta Cryst. A* 1976;32:922–923.
64. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA* 2001;98:10037–10041. [PubMed: 11517324]
65. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994;22:4673–4680. [PubMed: 7984417]

Fig. 1.

Cell cycle of *Caulobacter crescentus* and the ShkA-ShpA-TacA phosphorelay system controlling the stalk biogenesis. (a) Progression of *Caulobacter crescentus* cells through three cell cycle stages (adapted from reference 3) : G1 (mobility swarmer phase), S (DNA replication) and G2 (cell division). The timing of the three main morphogenic events is indicated at the bottom. (b) Schematic representation of the ShkA-ShpA-TacA phosphorelay system controlling the stalk biogenesis in *Caulobacter crescentus*. The domains are labeled as follows: histidine kinase domain (HK), response regulator domains (R1 and R2), histidine phosphotransfer protein (HPt), AAA+ ATPase domain (AAA+), and DnaA binding domain (HTH). The length of each protein is also shown.

Fig. 2. Crystal structure of the histidine phosphotransfer protein ShpA. Stereo ribbon diagram of a ShpA monomer color-coded from N-terminus (blue) to C-terminus (red). Helices are labeled sequentially as H1 to H5. The site of phosphorylation (His56) is shown as a stick model.

Fig. 3.

Structural comparisons between ShpA and monomeric HPTs. (a) Structures superimposition of known HPTs: ShpA (PDB 2ooc, orange), ZmHP2 (PDB 1wn0, cyan), ArcB^c (PDB 2a0b, purple), CheA P1 domain (PDB 1tqg, blue), YojN (PDB 1sr2, red), and YPD1 (PDB 1qsp, green). LuxU (PDB 1y6d) is not shown due to its large RMSD with the others. P-His residues are shown in sticks. (b) The conserved structural core of HPTs (red) mapped onto ShpA, is shown in cartoon (white). P-His are shown in sticks. (c) A structure-based sequence alignment of helix H2 and the anti-parallel H3–H4 hairpin region of ShpA with those of structurally known monomeric HPTs. The P-His, and two highly conserved residues in close proximity are highlighted in boxes. The additional conserved residues are highlighted in yellow. The columns

marked with asterisks (*) correspond to the YPD1 residues involved in the interaction with SLN1. **(d)** Sequence logo representation of the sequence conservation of the H3–H4 region in HPts domains in Pfam (PF01627) with P-His assigned as position 0. A logo consists of stacks of symbols, one stack for each position in the sequence. The overall height of the stack indicates the sequence conservation at that position, while the height of symbols within the stack indicates the relative frequency of each amino or nucleic acid at that position.

Fig. 4.

The site of phosphorylation in ShpA. (a) Stereo view of the phosphorylation site of ShpA. The hydrogen bonds are shown in red dashed lines. The distances for each HB bonds are also shown. (b) The electrostatic surface of ShpA. Positively charged residues are in blue, and negatively charged residues in red. Three basic residues (Lys52, Lys59 and Arg63) and the phosphorylation site (His56) are labeled. (c) The phosphorylation sites for the CheA P1 domain, YPD1 and ZmHP2. All of the models are in the same orientation as (a).

Fig. 5. Interaction between ShpA and RRs. (a) The molecular surface of ShpA colored by a hydrophobicity gradient, from hydrophilic (green) to hydrophobic (white). The three helices (H1–H3) that contribute the hydrophobic patch and the phosphorylation site (His56) are labeled. (b) Mapping of the RR interaction surface of YPD1 onto ShpA. The YPD1 residues involving in the SLN1-RR binding are shown on the left. The corresponding ShpA residues (right) were obtained by structural superimposition of ShpA onto YPD1. The YPD1 residues which are primary contributors to the SLN1-YPD1 interaction are shown in bold on a yellow background. The conserved residues between YPD1 and ShpA are highlighted in red. The equivalent residues for the first 20 residues of ShpA and those of YPD1 are approximate. Two additional highly conserved residues (Leu15 and Glu28) present in closely related ShpA homologs are labeled in blue on ShpA. (c) Crystal packing of the two ShpA monomers in the asymmetric unit mimics the inter-helical interaction of SLN1-YPD1. Monomer B is shown as a cartoon (orange). The section of helix H5 (95–111) from monomer A that interacts with monomer B is shown in green. The SLN1-RR $\alpha 1$ helix which interacts with corresponding region of YPD1 is shown in red. (d) Conservation of hydrophobic residues on the solvent exposed surface of the helix $\alpha 1$ of selected *Caulobacter crescentus* RRs. Residues of SLN1-

RR that are involved in contact with YPD1 are shown in bold and highlighted in yellow. Conserved residues (mostly hydrophobic) are highlighted in boxes.

Fig. 6. Structural flexibility in ShpA. **(a)** Plot of the C α atomic displacements between the two monomers in the ASU and average B-factors (per residue) for both monomers (A- green, B- orange). The corresponding secondary structure of monomer B is shown on the top. **(b)** Structural comparison of monomer A (green) and monomer B (orange). **(c).** Close-up view of the 14–25 region where large structural differences are observed.

Table 1
Data collection and refinement statistics for ShpA (PDB: 2ooc).

Data collection	λ_1 Se	λ_2 MADSe	λ_3 MADSe
Cell dimensions (Å)	$a=b=62.37$, $c=115.57$	$a=b=62.60$, $c=115.28$	
Wavelength (Å)	0.9184	0.9184	0.9786
Resolution range (Å)	28.9-1.52	39.5-1.57	39.5-1.61
Number of observations	284,312	200,540	178,291
Number of unique reflections	40,951	37,307	34,209
Completeness (%)	100 (100) ^a	100 (100)	99.3 (95.7)
Mean I/σ (I)	16.2 (1.6) ^a	13.6 (1.8)	17.6 (2.0)
Rsym on I (%)	0.07 (0.81) ^a	0.07 (0.89)	0.05(0.66)
Highest resolution shell (Å)	1.56-1.52	1.65-1.57	1.70-1.61
Model and refinement statistics			
Resolution range (Å)	28.9-1.52	Data set used in refinement	λ_1 Se
No. of reflections (total)	40,897 ^b	Cutoff criteria	$ F >0$
No. of reflections (test)	2,051	R _{cryst}	0.166
Completeness (% total)	99.9 ^b	R _{free}	0.202
Stereochemical parameters			
Restraints (RMS observed)			
Bond angle (°)	1.70		
Bond length (Å)	0.018		
Average isotropic B-value (Å ²)	32.8		
ESU based on R _{free} (Å)	0.07		
Protein residues / atoms	209 / 1,624		

^aHighest resolution shell in parentheses. The high resolution cutoff was chosen such that the mean I/σ (I) in the highest resolution shell is around 2.

^bTypically, the number of unique reflections used in refinement is less than the total number that were integrated and scaled. Reflections are excluded due to systematic absences, negative intensities, and rounding errors in the resolution limits and cell parameters.

ESU = Estimated overall coordinate error.

$R_{sym} = \sum |I_i - \langle I_i \rangle| / \sum I_i$, where I_i is the scaled intensity of the i^{th} measurement and $\langle I_i \rangle$ is the mean intensity for that reflection.

$R_{cryst} = \sum |F_{obs} - F_{calc}| / \sum F_{obs}$, where F_{calc} and F_{obs} are the calculated and observed structure factor amplitudes, respectively.

R_{free} = as for R_{cryst} , but for 5% of the total reflections chosen at random and omitted from refinement.

Table 2
Structural comparisons of ShpA with other monomeric HPT proteins^a

Protein	PDB code	RMSD (Å)	Aligned length (aa)	No. residues (aa)	Z score	Seq. id (%)	P-His	Organism	Function
ShpA	2ooc-B	1.7 (0.2) ^b	104	104	18.6	100	56	<i>C. crescentus</i>	Stalk biogenesis
AK104879	1yvi-A	2.5 (0.6)	100	140	10.9	22	80	<i>Rice</i>	Unknown
ZmHP2	1wn0-A	2.6 (0.9)	97	131	10.2	24	80	<i>Maize</i>	Hormone signaling/stress response
ArcB ^c	2a0b	3.2 (1.1)	101	118	9.6	20	715	<i>E. coli</i>	Aerobic respiration
YPD1	1qsp-A	2.5 (1.1)	97	165	8.7	14	64	<i>Yeast</i>	Osmoregulation/oxidativ e stress
CheA-P1	1tqg	2.3 (0.8)	83	105	8.4	7	45	<i>T. maritima</i>	Chemotaxis
LuxU	1y6d	3.3 (3.2)	83	114	5.1	7	58	<i>V. harveyi</i>	Quorum sensing
YojN	1sr2	2.6 (0.9)	78	116	6.4	15	842	<i>E. coli</i>	Capsule synthesis/cell division/motility

^aThe alignment of the full length proteins is performed by the DALI⁶² structural comparison server using monomer A of ShpA as a search probe; monomer B gives similar results (not shown). For structures with multiple copies in the asymmetric unit, only results for the one with highest Z-score are shown. For proteins with multiple structures (such as YPD1 and CheA P1 domains), only one representative is shown. The value for ShpA are a comparison between the two monomers in the ASU, this value is relatively high due to high structural flexibility in residues 19–22 and 103–112. The RMSD is 0.59 Å for 93 C α atoms if these two regions are excluded. No. residues is the number of residues present in the model used for comparison. P-His is the residue number of the phosphorylatable histidine.

^bThe value in parentheses is the RMSD of the C α atoms between ShpA and each protein listed for the antiparallel α -helix hairpin region (H3–H4 of ShpA), calculated by LSQKAB⁶³. The range of the residues is defined as (P-His-6 to P-His+22) which corresponds to residues 50–78 in ShpA.