Published in final edited form as:
Biochemistry. 2007 December 25; 46(51): 14751–14761. doi:10.1021/bi700970a.

Structure of the DNA-binding domain of the response regulator PhoP from *Mycobacterium tuberculosis*[†]

Shuishu Wang[‡,§,*], **Jean Engohang-Ndong**[‡], and **Issar Smith**[‡,||]

‡ *Public Health Research Institute, New Jersey Medical School – UMDNJ, Newark, New Jersey, USA*

§ *Department of Biochemistry and Molecular Biology, New Jersey Medical School – UMDNJ, Newark, New Jersey, USA*

|| *Department of Medicine, New Jersey Medical School – UMDNJ, Newark, New Jersey, USA*

Abstract

The PhoP-PhoR two-component signaling system from *Mycobacterium tuberculosis* is essential for the virulence of the tubercle bacillus. The response regulator, PhoP, regulates expression of over 110 genes. In order to elucidate the regulatory mechanism of PhoP, we determined the crystal structure of its DNA-binding domain (PhoPC). PhoPC exhibits a typical fold of the winged helix-turn-helix subfamily of response regulators. The structure starts with a four-stranded anti-parallel β-sheet, followed by a three-helical bundle of α-helices, and then a C-terminal β-hairpin, which together with a short β-strand between the first and second helices forms a three-stranded anti-parallel β-sheet. Structural elements are packed through a hydrophobic core, with the first helix providing a scaffold for the rest of the domain to pack. The second and third helices and the long, flexible loop between them form the helix-turn-helix motif, with the third helix being the recognition helix. The C-terminal β-hairpin turn forms the wing motif. The molecular surfaces around the recognition helix and the wing residues show strong positive electrostatic potential, consistent with their roles in DNA binding and nucleotide sequence recognition. The crystal packing of PhoPC gives a hexamer ring, with neighboring molecules interacting in a head-to-tail fashion. This packing interface suggests that PhoPC could bind DNA in a tandem association. However, this mode of DNA binding is likely to be non-specific because the recognition helix is partially blocked and would be prevented from inserting into the major groove of DNA. Detailed structural analysis and implications with respect to DNA binding are discussed.

Mycobacterium tuberculosis (MTB), the causative agent of tuberculosis, is one of the most successful human pathogens. MTB infects nearly one-third of the world's population and kills over 2 million people annually worldwide. Multi-drug resistant (MDR) and extreme drug resistant (XDR) strains of MTB are on the rise in the clinic (1), stressing the urgent need to develop novel anti-tuberculosis drugs.

Two-component systems (TCS) are major signaling systems in bacteria that mediate a variety of processes such as sporulation, transformation competence, membrane transport, inorganic nutrient uptake, chemotaxis, stress response, and virulence. In general, a TCS consists of a histidine kinase, which senses environmental signals, and a response regulator, which is

[†]This work was supported by NIH Grants GM079185 to S.W. and AI65987 to I.S. J.E.-N. was supported by a Heiser Program Fellowship from the New York Community Trust. Coordinates and observed structure factor amplitudes have been deposited in the Protein Data Bank (pdb code 2PMU).

*Corresponding address: Public Health Research Institute, New Jersey Medical School – UMDNJ, 225 Warren Street, Newark, NJ 07103, USA. Phone: (973) 854-3470. FAX: (973) 854-3101. E-mail: wang16@umdnj.edu.

phosphorylated by the cognate histidine kinase and in most cases functions as a transcription regulator to regulate expression of certain genes. Because they are absent in mammals, TCSs are potential targets for developing novel drugs. The MTB genome encodes 30 TCS proteins, comprising 11 systems in which the genes for the histidine kinase and response regulator are linked and 7 orphan histidine kinases or response regulators (2,3). Of these proteins, the PhoP-PhoR system shows the most severe effect on virulence among MTB strains in which TCSs are inactivated. The PhoP-PhoR system is essential for MTB virulence and is required for MTB to multiply in human and mouse bone marrow-derived macrophages and in mice (4,5). Global gene expression profiling indicates that 44 genes are up-regulated and another 70 genes are down-regulated by PhoP (5). Many of the up-regulated genes are involved in general or lipid metabolism and in substrate transport across the plasma membrane. One cluster of genes includes *msl3*, a polyketide β-ketoacyl synthase gene involved in the synthesis of polyacyltrehaloses, and *pks2* and *mmpL8*, genes implicated in the synthesis of sulfatides and in virulence. A PhoP knockout mutant MTB strain lacks sulfatides, diacyltrehaloses, and polyacyltrehaloses in the cell envelope, suggesting that the PhoP-PhoR system is involved in complex lipid biosynthesis (5,6).

The PhoR protein is a transmembrane histidine kinase that transmits signals from the environment by autophosphorylating a conserved histidine residue. The phosphoryl group is then transferred to a conserved aspartate residue of PhoP (7), a response regulator that regulates the expression of particular genes (5). In *Bacillus subtilis*, the PhoP-PhoR system senses phosphate, being activated under limited phosphate conditions, and it regulates the expression of 31 genes that are involved in phosphate utilization (8). The related PhoP-PhoQ system in *Salmonella enterica typhimurium*, however, senses Mg^{2+} (9), low pH (10,11), and antibacterial peptides (12,13); it regulates genes involved in Mg^{2+} transport and acquisition and also controls the expression of several virulence factors (14). Further support for the importance of PhoP in virulence and transmissibility is the observation that the *Mycobacterium bovis* B strain responsible for severe MDR TB outbreaks in Spain carries an *IS6110* insertion in the promoter region of *phoP*, which causes a strong up-regulation of the expression of *phoP* (15). Therefore, studying the signaling mechanism by PhoP-PhoR is an important step toward a better understanding of MTB pathogenicity.

PhoP of MTB belongs to the OmpR/PhoB subfamily, the largest subfamily of response regulators. Response regulators of this subfamily have two domains, an N-terminal regulatory domain and a C-terminal DNA-binding domain (also called the effector domain). Many members of this subfamily have been extensively studied, and most of the studies indicate that these response regulators bind DNA as a dimer. However, the type of dimer, i.e. tandem vs. symmetric, seems to vary among members of this subfamily. Most of currently known DNA recognition sites are tandem repeats, suggesting that these response regulators bind to DNA as tandem dimers. Currently, the only available DNA complex structure is the DNA-binding domain of PhoB from *E. coli* in complex with its *pho* box DNA (16), which shows a tandem dimer. The N-terminal domain of PhoP from *B. subtilis* also has a tandem association in the crystal structure (8). However, crystal structures of the N-terminal domain of *E. coli* ArcA, TorR, and KdpE show that they form a symmetric dimer mediated by the α4-β5-α5 face (17, 18). Biochemical studies of OmpR DNA binding also suggest that the response regulator binds DNA in a head-to-head orientation (19), arguing against tandem dimer formation.

The role of phosphorylation in transcription regulation is also unclear for the OmpR/PhoB subfamily of response regulators. Some members, such as PhoP from *S. enterica* (20) and *B. subtilis* (21), form dimers in solution and bind DNA regardless of the phosphorylation state. A recent paper by Gupta et al. (7) shows that phosphorylation has no effect on MTB PhoP binding to the *phoP* promoter. However, phosphorylation of PhoB from *E. coli* induces dimerization (22) and increases its affinity for DNA (23); AcrA from *E. coli* also dimerizes

upon phosphorylation (17). Crystal structures of full-length DrrB (24) and DrrD (25) from *Thermotoga maritima* indicate that in the unphosphorylated state the DNA recognition helix is freely exposed to the solvent, which would make it readily available for DNA binding. However, in the structure of PrrA from MTB, the recognition helix is involved in interactions with the regulatory domain, although in solution there may be an equilibrium with an open form (26).

In this paper, we determined the crystal structure of the DNA-binding domain of PhoP (referred to as PhoPC). PhoPC has the typical structure of the winged helix-turn-helix response regulators. In solution it exists predominantly as a monomer. However, it forms a hexamer ring in the crystal with tandem association between neighboring protomers. Two hexamers are related by the crystallographic two-fold symmetry to give a dodecamer. The structure of PhoPC is analyzed and compared to those of homologous response regulators, and the potential implications for its function in gene expression regulation are discussed.

MATERIALS AND METHODS

Cloning, Expression and Purification

The C-terminal domain of MTB *PhoP* gene (Rv0757) was amplified from the genomic DNA of MTB strain H37Rv by polymerase chain reactions (PCR) with KOD Hot Start DNA polymerase from Novagen. The following primers were used for PCR: 5′-primer GGCACATATGAAGGAACCACGTAATGTTCG, and 3′-primer CCAGAAGCTTTCGAGGCTCCCGCAGTACG, which generate *Nde*I and *Hin*dIII restriction sites (underlined). Amplified DNA was cleaved by restriction enzymes *Nde*I and *Hin*dIII, purified by agarose gel extraction, and ligated into a modified pET28a plasmid (Novagen) linearized with the same enzymes. The pET28a vector was modified by inserting DNA sequence encoding a TEV protease cleavage site before the *Nde*I restriction site and two stop codons, TAATAG, right after the *Hin*dIII site. The TEV protease cleavage site was generated by ligation of the *Nde*I-linearized pET28a with annealed oligos, TACGGGAGAAAATCTTTATTTTCAAGGTACCCA and TATGGGTACCTTGAAAATAAAGATTTTCTCCCG, and selection of the ligation product with the insert in the correct orientation by DNA sequencing. The resulting plasmid pET28-*PhoPC* encodes the C-terminal DNA-binding domain of PhoP (residues 144 – 247) and has a 6 × His tag at the N-terminus, which can be cleaved by the tobacco etch virus (TEV) protease, and two extra residues, LysLeu (amino acids are referred to as standard three-letter codes) encoded by the *Hin*dIII site, at the C-terminus. The DNA sequences were confirmed by sequencing.

The plasmid pET28-*PhoPC* was transformed into BL21(DE3) competent cells (Novagen) for protein overexpression. Cells containing the pET28-*PhoPC* plasmid were grown in LB medium containing 50 μg/ml Kanamycin at 37 °C to an OD_{600} of ~0.8 and induced with 50 μM IPTG for 3 hours at room temperature. Cells were collected by centrifugation, resuspended in 50 mM sodium phosphate, pH 7.4, 500 mM NaCl. PMSF was added to the cell suspension to a final concentration of 0.5 mM, and cells were lysed by passing twice through a French Press. The cell lysate was centrifuged at ~37,000 g for 30 min, and the supernatant was loaded onto a 5-mL HisTrap column (GE Healthcare) pre-equilibrated with 50 mM sodium phosphate, pH 7.4, 500 mM NaCl, and 20 mM imidazole (buffer A). The column was washed thoroughly with buffer A, and eluted with a linear gradient of imidazole from 20 to 300 mM in buffer A. The fractions containing the PhoPC protein were pooled, mixed with ~300 μg His-TEV (S219V) protease (27), and dialyzed against 2 L of buffer containing 50 mM sodium phosphate, pH 7.4, 200 mM NaCl, and 5 mM $MgCl_2$ overnight at 4 °C. The dialyzed mixture was loaded onto a HisTrap column equilibrated with buffer A to remove the His-TEV protease, cleaved His-tag, and the undigested protein. Impurities co-eluted from the first HisTrap column were

also removed at this second pass of the HisTrap column. Fractions containing PhoPC lacking the His-tag were concentrated to ~5 mg/mL, and the protein was further purified on a Superdex 75 column (GE Healthcare) equilibrated with 20 mM Bistris, pH 6.5, and 10 0 mM Li_2SO_4 The purified protein was analyzed by SDS-PAGE and MALDI mass spectrometry for purity.

Electrophoretic Mobility Shift Assays (EMSA)

The promoter DNA sequences, ~250–320 base pairs (bp) upstream of the genes and including 20 bp at 5′ end of the coding region, were amplified by PCR. Genes were selected based on global transcriptional profiling studies by Walters et al. (5) that indicate their expression is significantly affected by PhoP. The PCR-amplified fragments were labeled by phosphorylation with [γ-^{32}P]ATP using T4 polynucleotide kinase. The labeled promoter fragments were incubated with purified PhoPC protein for 10 min at room temperature in a 20 μl final volume containing 1 μg of poly[dA-dT] (Sigma), 20 mM Tris pH 8.0, 0.4 mM $MgCl_2$, 5 mM KCl, 0.2 mM DTT, 10% glycerol, and 0.01% NP-40 (BRL). The binding mixture was loaded onto a 6% non-denaturing polyacrylamide gel. After electrophoresis, the gel was vacuum dried and exposed overnight to autoradiographic film.

Crystallization and Data Collection

The purified protein was concentrated to above 10 mg/mL for crystallization. Crystallization experiments were carried out using the vapor-diffusion method. In each drop, 3 – 5 μL of protein solution was mixed with an equal volume of the well solution. The best crystals were obtained from drops set up with well solutions containing 1.2 – 1.4 M Na/K phosphate, pH 5.6, and 100 mM glycine.

Before data collection, crystals were soaked for 2–5 min in a cryogenic solution similar to well solutions, but with glycerol added to 30%, and then flash-frozen in a cryo stream of N_2 gas at 100 K on the goniometer. Diffraction data were collected at the beamline X6A of National Synchrotron Light Source, Brookhaven National Laboratory, with a Quantum-210 CCD detector. Data reduction and scaling were done with the programs DENZO and SCALEPACK (28). Data collection and processing statistics are listed in Table 1. The crystals belong to the space group C2, with 6 molecules per asymmetric unit and a V_M (Matthews volume) of 2.39 $Å^3$/Da. The data could also be processed in the F222 space group with good statistics (R_{sym} 0.054, three molecules per asymmetric unit); however, there would be steric clashes at the crystal packing interface along one of the two-fold axis. The data were converted to CCP4 format and structure factor amplitudes calculated by TRUNCATE and other programs in the CCP4 Suite (29).

Structural Determination, Model Building and Refinement

The structure of PhoPC was determined by molecular replacement with the program PHASER (30), using as a model the structure of the C-terminal domain of DrrD (PDB code 1kgs), an OmpR/PhoB homolog from *T. maritima* (25). The structural model was refined against the diffraction data using the programs CNS (31) and REFMAC (32), with the same subset of data for R_{free} calculation kept between the two programs. After each cycle of refinement, the model was manually adjusted with weighted 2Fo – Fc and Fo – Fc electron density maps using the program COOT (33). The initial R and R_{free} after one round of REFMAC rigid-body refinement of solutions from Phaser were 0.502 and 0.494, respectively. After manual building based on maps from the rigid-body-refined model and one round of maximum likelihood refinement with REFMAC, the R and R_{free} dropped to 0.370 and 0.446, respectively. Themodel was refined for a few cycles of REFMAC with TLS refinement (34) before convergence. Final refinement statistics for the refined coordinates are reported in Table 2.

There are six protein molecules in the asymmetric unit, forming a hexamer ring. The molecules are referred to as A to F. Molecules A, B, and C are nearly identical to D, E, and F, respectively, except at some flexible loops and some side chains. Therefore, non-crystallographic symmetry (NCS) restraints were applied to each pair of molecules throughout structural refinement. R factors, especially the R_{free} would go up if NCS restraints were not applied. Residues originating from the plasmid vector, i.e., GlyThrHisMet at the N-terminus after TEV protease cleavage and LysLeu at the C-terminus, as well as the last residue, Arg247, encoded by the *phoP* gene, are completely disordered in all six molecules and therefore were not modeled. In addition, residues 144–148 (residue numbers are according to the amino acid sequence of the *phoP* gene product) of molecule A, 144–147 and 204–210 of molecule C, 144–148 and 209–210 of molecule D, and 144–147 and 204–209 of molecule F were also not modeled due to lack of electron density.

Over 90.2% of protein residues fall in the most favored region of the Ramachandran plot. ProCheck (35) indicated that residues Trp203 of A and D and Ala154 of all 6 chains, are in the additional allowed or disallowed regions. These residues are well-ordered and have well-defined electron density. A more recent program for structure validation, MolProbity (36), which is based on new data from a large number of high resolution structures solved since the introduction of ProCheck, suggests that these residues are in the allowed region.

Three K^+ ions are modeled in the structure. Two of them are at the two-fold NCS interface between molecules A and D and are related by the two-fold NCS. Each ion is coordinated by the main chain carbonyl oxygen atoms of Asp200, His201, and Trp203 from one molecule and of His201 of the other molecule. A water molecule gives the 5^{th} coordination to both ions, with the 6^{th} coordination open in an otherwise octahedral coordination. The distances from the oxygen atoms to the ion range from 2.55 to 2.95 Å. Because K^+ and Na^+ are both present in the crystallization condition, these positions can be partially occupied by either ion. Modeling as Na^+ gave a B factor of ~26 $Å^2$, while modeling as K^+ gave a B factor of ~36 $Å^2$. B factors of the coordinating atoms range from 25.7 to 36.9 $Å^2$. The other K^+ ion sits on the two-fold NCS axis between molecules B and E. It has a B factor of ~32.7 $Å^2$ and mediates interactions between B and E to form a symmetric dimer (see below for more details of interactions). There are four unknown atoms modeled in the final refined structure. Two of them are ~2.06 Å from the main chain carbonyl oxygen of His201 of molecules B and E, respectively. When modeled as K^+, the B factors are 36.6 and 38.9 $Å^2$, which are lower than that of the corresponding carbonyl oxygen atoms that they are attached to (42.3 and 41.5 $Å^2$, respectively). The other two unknown atoms are 2.14 Å and 1.93 Å to the Oε1 atoms of Glu229 of chains A and D, respectively, and are within hydrogen-bonding distance to the Nη1 of Arg222 of chain C and F, respectively. A water molecule is also within a hydrogen-bonding distance to each unknown atoms. These two unknownatoms were modeled as Li^+ for refinement, which gave B factors of 24.5 and 29.0 $Å^2$. Li_2SO_4 is presentin the protein sample at 100 mM concentration. However, Li^+ is unlikely to be visible in the electrondensity at the data resolution.

There are four phosphate ions in the final refined model. One of them binds at a two-fold NCS axis between molecules A and D, with its 3-fold symmetry axis nearly coinciding with the two-fold NCS axis. As a result, this phosphate ion has two alternative positions related by the two-fold NCS and was modeled as two alternative conformations for the phosphorus atom and three oxygen atoms. The phosphate has hydrogen bonds to Oη of Tyr184, Nδ1 of His201, and Nζ of Lys197 of both molecules A and D, and it is well ordered with an average B factor of 26.5 $Å^2$, similar to that of nearby protein atoms. Another phosphate ion binds loosely near the side chains of Arg183 of molecules B and E in a large pocket at the two-fold NCS interface between these two molecules. This phosphate has an average B factor of 60.8 $Å^2$. Two oxygen atoms of the phosphate have a distance of 3.17 and 3.08 Å, respectively, each to an unknown atom bound to the carbonyl oxygen of His201 of molecules B and E, respectively. The other two

phosphate ions have an average B factor of ~47 $Å^2$ and interact with the side chains of Arg231 and His234 and the main chain carbonyl and NH groups of Thr235 of molecules A and D, respectively.

RESULTS

Expression and Purification of the C-terminal Domain of PhoP (PhoPC) and Activity Assays

An expression construct for producing PhoPC in *E. coli*, which encodes residues 144 to 247 of the PhoP protein from MTB (residues 144 to 147 are part of the linker between domains), was prepared as described in the Materials and Methods section. The PhoPC protein was purified by two passes through a Ni^{2+} column with the His-tag cleaved by the TEV protease before the second pass. Then the protein was passed through a gel filtration column. The purified protein showed no detectable impurity by either SDS-PAGE or MALDI mass spectrometry. Mass spectrometry gave a molecular weight of 12770 Da, identical to the calculated molecular weight from the amino acid sequence, which contained four residues, GlyThrHisMet, at the N-terminus, and two residues, LysLeu, at the C-terminus that originated from the plasmid vector. Gel filtration studies indicated that PhoPC is a monomer in solution. However, at high concentration (above 10 mg/mL) a small peak at about the dimer size can be seen, suggesting that PhoPC can form a weak dimer in solution.

The purified PhoPC protein is able to bind selectively promoter DNA sequences of certain genes. We have performed EMSA on the promoter sequence of 11 genes and found 4 of them bound PhoPC (Table 3), indicating that these genes are directly regulated by PhoP. Figure 1 shows the result ofEMSA with the promoter sequence of *phoP*. PhoP binds to its own gene promoter, suggesting auto-regulation of gene expression. A recent paper suggests that MTB PhoP down-regulates its own expression (7).

Overall Structure

PhoPC has the typical fold of the winged helix-turn-helix DNA-binding domain. The structure is composed of three α-helices flanked by two β-sheets (Figure 2). On the basis of known structures of homologous response regulators, the N-terminal domain of PhoP is expected to have a conserved structure consisting of an α/β fold with a five-stranded central β-sheet surrounded by five α-helices. For consistency, the α-helices and β-strands of PhoPC are therefore numbered starting from α6 and β6, respectively. At the N-terminus, there is a four-stranded antiparallel β-sheet with the last strand β9 connected to helix α6 by a short loop. The N-terminal β-sheet has hydrophobic interactions with helixα6 mediated by residues Leu151, Ile156, Leu158, His163, and Val165 from the β-sheet and residues Phe179, Leu182, Val186, and Ile187 from helix α6 (Figure 3). Helix α6 is mostly buried and forms the hydrophobic core, onto which the rest of the domain packs. Following helix α6 is a short strand β10, which pairs with strand β12 of the C-terminal β-hairpin to form a three-stranded antiparallel β-sheet. Helix α7 of the helix-turn-helix motif follows strand β10. This helix packs against α6 through a hydrophobic patch composed of residues Thr177, Thr180, Leu181, Tyr184, and Phe185 of α6 and residues Ile198, His201, and Val202 of α7 (Figure 3). A long loop between α7 and α8 is partially disordered. This loop is termed the transactivation loop because the corresponding loop in OmpR (37 and PhoB (38) of *E. coli* interacts with components of RNA polymerase to activate transcription, although this interaction for MTB PhoP is yet to be demonstrated. Helix α8 is the recognition helix, which is expected to play an important role in DNA sequence recognition by binding in the major groove of DNA. A 7-residue loop connects helix α8 to the C-terminal β-hairpin, which binds the minor groove of DNA with its hairpin turn in the PhoB-DNA complex (16). The hairpin turn is called the wing of the winged helix-turn-helix DNA-binding motif.

Crystal Packing Interactions

There are six protein molecules (designated as molecules A to F) in the asymmetric unit, forming a hexamer ring (Figure 4a). The interface between two neighboring subunits buries about 830 $Å^2$ to 1100 $Å^2$ of surface area (Table 4). Solvent accessible surface areas were calculated with the program AREAIMOL in the CCP4 Suite (29). The lower values of buried surface areas between molecules A and F and between C and D are due to the disordered loop between α7 and α8 (transactivation loop, residues 204 to 209) that was not modeled in molecules C and F. The interface, which is relatively flat, involves the N-terminal end of α6, the transactivation loop, and one side of the recognition helix α8 of the upstream molecule (e.g. molecule A of the interface between A and B) and one face of the N-terminal β-sheet of the downstream molecule. Interactions are mediated mostly by hydrogen bonds and π-electron interactions between aromatic groups and peptide planes. Tyr220 and Tyr217 from helix α8 of the upstream molecule each contribute a hydrogen bond from the Oη atom. Other residues of the upstream molecule at the interface are Ser175, Pro176, Thr177, Glu178, Val202, Phe207, and Lys224. Residues from the N-terminal β-sheet of the downstream molecule contributing to the intermolecular interface are Arg150, Thr152, Asp155, Glu157, Ala168, and Gly169. The side chain of Glu157 sits at the N-terminal end of the helix α6 of the upstream molecule; it has hydrogen bonds to Oγ1 and the main chain NH of Thr177, O γ of Ser175, and the Oη of Tyr217.

The two-fold crystallographic symmetry relates two hexamer rings to generate a dodecamer that stacks two rings face to face (Figure 4b). The interface between two hexamers buries ~5380 $Å^2$ of solvent accessible surface. On average a buried surface area of ~900 $Å^2$ exists between two neighboring protomers from different hexamers. Each protomer interacts with another protomer from the other hexamer ring mainly through a hydrogen bond from Oε2 of Glu229 of one molecule to Oγ of Arg222 of the neighboring molecule and charge-charge interactions between the two side chains. In 4 pairs of contacting protomers where Arg231 of one protomer interacts with a bound phosphate ion, the Arg231 side chain of the other protomer forms hydrogen bonds to the carbonyl oxygen atoms of Asp226 and Gly228. In comparison to interactions between tandem-associated protomers within a hexamer ring described above, these interactions between protomers from different hexamers are much weaker.

Electrostatic Potential Surface

The electrostatic potential surface of PhoPC shows that the molecular surface around the recognition helix (α8) and the wing (C-terminal β-hairpin) has a positive electrostatic potential (Figure 5), consistent with the role of these structural modules in DNA binding. Most of positive charges accumulate around Arg237 at the C-terminal β-hairpin turn in both front and back sides of the molecule, while most of the rest of the molecular surface is either negatively charged or neutral. This feature of the surface electrostatic potential may orient the molecule for initial binding to DNA. The C-terminal domain of PhoB of *E. coli* (16) has a similar electrostatic potential surface to that of PhoPC. However, OmpR of *E. coli* has a different feature, in which the exposed surface of the helix and the α-loop before the recognition helix bears most of the positive charges (39,40). This difference might allow OmpR to bind DNA in a different mode, in which the first helix of the helix -turn-helix motif, instead of the recognition helix, plays a major role in binding DNA, as suggested by cross-linking studies (19). In the hexameric form of PhoPC, the positively charged surfaces are located at the six vertices of the hexamer (Figure 5c). The positively charged surfaces are fully exposed, even though one side of the recognition helix is involved in hexamer formation (see the previous section for details of hexmeric interfaces).

Comparison among Protomers

Molecules A, B, and C are related to D, E, and F, respectively, by a two -fold NCS that is perpendicular to the ring. The rmsd between molecules A and D, excluding Gly209 and Asp210

that were not modeled in molecule D due to lack of electron density, is 0.714 Å for the Cα atoms. Most of the differences are at the transactivation loop (residues Phe207 to Val211), which is flexible and partially disordered in all molecules. Excluding this flexible loop, the rmsd between A and D is 0.04 Å for all other Cα atoms. The rmsd for 92 Cα atoms between molecules C and F is 0.125 Å. Major differences are at the N-terminal residues 148 – 150, residues 211 and 212 at the N-terminus of α8, and residues 235 – 239 that compose the wing.

Molecules B and E have an rmsd of 0.107 Å for all 103 Cα atoms. Most of the differences lie in the loop following the recognition helix and in the N- and C-termini, which also have high B factors. However, some significant differences exist at a crystal packing interface between these two molecules, where the side chains of Glu160 of B and E are near each other and related by a 2-fold NCS. These two side chains must adopt different conformations to avoid steric clashes. The side chains of Arg183 in both molecules B and E form salt bridges to the Glu160 side chains to neutralize the negative charges and thus dampen the charge-charge repulsion. This packing interface between molecules B and E generates a symmetric dimer. The major contributions to the dimer interface are the antiparallel β-strand hydrogen-bonds between residues 152 – 154 of strand β6 of one molecule and residues 146 – 148 of the other (these residues are disordered in molecules A, C, D, and F). In add ition, a K^+ (or Na^+) ion mediates interactions of the main chain oxygen of Arg150 and Oδ1 of Asn148 between B and E. When modeled as a Na^+, the B factor was ~23.7 $Å^2$, which is lower than the nearby protein atoms. However, when modeled as a K^+ ion, it gave a B factor of ~32.7 $Å^2$, similar to that of the coordinating protein atoms (~33.4 and ~32.5 $Å^2$ for the Oδ1 of Asn9 and ~28.9 and 29.1 $Å^2$ for the main chain carbonyl O of Arg150 of molecules B and E, respectively). The distances to the two main chain O atoms of Arg150 are 2.52 and 2.58 Å, while distances to the Oδ1 atoms of Asn148 are 2.85 and 2.87 Å. There are two water molecules with distances of ~3.2 Å to the ion to give a distorted octahedral coordination to the ion.

Molecules A, B, and C have a much larger difference with each other (Figure 6a), as do molecules D, E, and F. The largest difference is at the transactivation loop, which is disordered in molecules C and F but could be modeled although with high B factors for molecules A, B, and E. In molecule A, a half turn at the N-terminus of helix α8 is unwound relative to that in B, thus resulting in a different conformation of the loop. This reflects the flexibility of this transactivation loop to adopt various conformations for its interactions with the RNA polymerase and regulation of the DNA-binding activity. The N-terminal β-sheet rotates slightly relative to the rest of the domain. The N-terminal β-sheet interacts primarily with helix α6 through interactions among hydrophobic side chains, which allow a slight rotation in spite of the strong interactions. Molecules A and B have an rmsd of 1.7 Å for 98 Cα atoms from residues 149 – 246; most of the differences are at the transactivation loop, which has a different conformation with deviations of nearly 7 Å (Figure 6a). The wing residues also have a shift of ~0.7 Å in Cα positions. The side chain of Arg237 is partially disordered in all molecules and adopts different conformations. Molecules A and C have an rmsd of 0.87 Å over 91 Cα atoms. The lower value of rmsd is due to the transactivation loop (residues 204 – 210) in molecule C being disordered and therefore not included in rmsd calculation. However, molecules B and C are more similar with an rmsd for 91 Cα atoms of 0.36 Å.

DISCUSSION

Structural Comparison with other OmpR/PhoB Homolog Response Regulators

Among response regulators with known three-dimensional structures, the PhoPC sequence is the most similar to that of *T. maritima* DrrD with a ~30% sequence identity (Figure 3). The rmsd between the structures of PhoPC and the C-terminal domain of DrrD is ~1.28 Å for a core structure of 80 Cα atoms containing the β-sheets and α-helices, excluding residues at two termini, the transactivation loop, and the loop between helix α8 and strand β11, which have

large deviations and are either disordered or flexible with high B factors. Superposition of these 80 Cα atoms of structurally well-conserved residues gives an rmsd of 0.94, 1.14, 1.33, and 1.92 Å between the structure of PhoPC and that of apo PhoB (PDB code 1GXQ), PhoB -DNA complex (1GXP), OmpR (1OPC), and PrrA (1YS7), respectively. Figure 6b shows the structural alignment of MTB PhoPC with the C-terminal domains of PhoB and OmpR. The β-sheets and α-helices composing the core structure align well. The structure of the DrrD C-terminal domain was chosen as a model for molecular replacement because of its sequence identity to PhoPC and the availability of a high-resolution structure. However, the refined structure of PhoPC is the most similar to that of *E. coli* PhoB.

The helices superimpose very well, especially the recognition helix α8, consistent with its role in DNA-binding and sequence recognition for the PhoB/OmpR subfamily of response regulators. Strands of the C-terminal β-hairpin also superimpose well, but the wing residues, i.e. the hairpin turn, have a moderate shift of more than 3 Å in Cα positions. In the PhoPC structure, the wing residues have high B factors, indicating that this hairpin turn is relatively flexible. The hairpin turn in the structure of apo PhoB of *E. coli* also has relatively high B factors with the side chain of Arg219 disordered (16). In the PhoB-DNA complex, the Arg219 side chain is sandwiched between two sugar residues of the minor groove and interacts with an adenine base. The deviation of this β-hairpin turn among these known structures likely reflects its interactions with the minor groove of various DNA sequences. The wing of PhoPC superimposes very well with that of the *E. coli* PhoB (Figure 6b).

The largest deviation among the structures of PhoB/OmpR homologous proteins is at the transactivation loop, i.e. the loop between helices α7 and α8 (Figure 6b). This loop is called the α-loop in OmpR of *E. coli* because it interacts with the α-subunit of RNA polymerase (41, 42), while in PhoB this loop interacts with the σ70 subunit of RNA polymerase (43). The α-loop in OmpR has the largest deviation from the loop of PhoPC, while the transactivation loop of PhoB closely matches that of the PhoPC (Figure 6b). Relative to other structures, in OmpR (39) the α-loop has a unique structure with the first turn of helix α8 unwound. This conformation will result in steric clashes between the α-loop and the DNA if the recognition helix of OmpR binds DNA in a similar way as PhoB. However, it is possible that the α-loop changes conformation upon binding to DNA because this loop of OmpR is flexible with high B factors, similar to that of the MTB PhoPC structure, in which molecule A has a half turn unwound at the N-terminus of helix α8 and a large deviation in the conformation of the transactivation loop relative to other molecules as described above.

Another large deviation among the structures is at the loop following the recognition helix. This loop has a low sequence similarity with an insertion/deletion in the sequence alignment (Figure 3). Again the OmpR structure shows the largest deviation relative to other structures. In the PrrA structure, this loop is partially disordered (26). PhoPC is the most similar to DrrD in the structure of this loop among known structures the OmpR/PhoB subfamily members. Because of its variation in sequence and structure, this loop is likely to play different roles in different response regulators, such as dimer or oligomer interface, or interactions with other proteins. In the MTB PhoPC structure, the loop forms the major interface between the hexamer rings.

Similar to that observed among protomers in the MTB PhoPC hexamer but on a larger scale, the N-terminal β-sheet rotates relative to the rest of the domain among response regulators of the OmpR/PhoB subfamily (Figure 6b). The N -terminal β-sheet is proposed to play a role in interacting with the N-terminal domain, thus playing an important role in transmitting the signal of phosphorylation to the C-terminal domain for regulation of the DNA-binding activity (24). In the structure of PhoPC, the N-terminal β-sheet forms the hexamer interface by interacting with the N-terminal end of helix α6, the transactivation loop, and one side of α8 of the upstream

molecule. Strong interactions between the N-terminal β-sheet and the rest of the domain are necessary to transmit movements of the β-sheet to the DNA-binding elements, while relative deviations among different response regulators and some flexibility of small rotations allow variations in the regulation mechanisms of DNA-binding activity.

Possible DNA-binding Mechanism and Residues Involved

The above structural analysis indicates that PhoPC is similar to the DNA-binding domain of PhoB of *E. coli* in the structure of the core, transactivation loop, wing, and surface electrostatic potentials. Structural superposition between the PhoB-DNA complex and PhoPC indicates that PhoPC is likely to bind DNA in a similar way to the PhoB of *E. coli* (Figure 7). Helix α6, the first helix in the DNA-binding domain, can be N-capped by a phosphate group from DNA, with the N-terminal residues, Pro176 and Thr177, which are conserved between PhoP and PhoB, interacting with the sugar and phosphate backbone in the same way (16). Interactions of terminal residues of helix α7 with the sugar and phosphate of DNA can also be preserved, of which Trp203 is conserved and Lys195 substitutes for the Arg176 in PhoB. Arg222, whose corresponding residue Arg203 in PhoB interacts with DNA phosphate backbone, is conserved (Figure 3). Structural superposition suggests that the side chain of Arg223 can take the position of the Arg200 side chain of PhoB for interacting with a phosphate, even though these two residues do not align in sequence. The side chain of Lys224 is also capable of interacting with the backbone atoms of DNA based on the structural alignment. The wing residues are very well conserved in both sequence and three-dimensional structure (Figures 3 and 7). The side chain of Arg237 is at the same position as the Arg219 of PhoB that binds in the minor groove. Taken together, these favorable interactions make it likely that MTB PhoP binds DNA similarly to PhoB. In this binding mode, residues on the recognition helix, Asn212, Val213, Glu215, Ser216, and Tyr220, are likely to interact with the DNA bases, and thus play a role in DNA-binding specificity. The side chain of Ser219 is likely to have hydrogen bonds to sugar or phosphate residues of the DNA backbone. A SNP (single nucleotide polymorphism) was identified between the avirulent strain H37Ra and the virulent strain H37Rv of MTB, in which Ser219 is mutated to a Leu (B. Kreiswirth, personal communication). Mutation of the polar Ser219 side chain into a bulkier hydrophobic Leu would lose hydrogen bond interactions and introduce steric repulsions, thus affecting the DNA-binding affinity and/or the orientation and consequently the regulation of gene expression by this PhoP response regulator.

On the basis of the above DNA-binding model, it is unlikely that PhoP adopts the tandem association as found in the crystal when it binds DNA. As described above, the tandem association found in the crystal structure involves the N-terminal β-sheet, helices α6 and α8, and the transactivation loop. Therefore, several important DNA-binding interfaces, such as the N-terminal end of helix α6 and one side of the helix α8, are blocked. Although the wing residues, most of residues of helix α8, and the positively charged molecular surface are exposed for interacting with DNA, this tandem association would prevent the recognition helix from inserting into the major groove of DNA.

The molecular interactions found in the crystal, however, could be involved in the regulation of the activities of the response regulator, although this is yet to be tested experimentally. T he N-terminal β-sheet of the DNA-binding domain interacts with the regulatory domain in the DrrB structure, and it is proposed to transmit signals from the regulatory domain to the effector domain for regulation of the gene transcription. One model of this signaling mechanism can be postulated. Equilibrium exists among various forms of domain association through this β-sheet: associating with the N-terminal domain (active form), interacting with another effector domain in a tandem manner as observed in this crystal structure (inactive forms), or free in solution as a monomer (active form). Phosphorylation of the N-terminal domain makes the N-

terminal domain more favorable for binding of the β-sheet, thus shifting the equilibrium towards the active form.

From available structures of response regulators, it is noteworthy that the self association of the DNA-binding domain is not very strong, being mostly polar interactions and/or through a relatively small interface. This dynamic association allows better regulation of the DNA-binding activity. Dimerization of the DNA-binding domain is affected by DNA binding, dimerization of the regulatory domain, as well as phosphorylation. The crystal structure of PhoPC shows a significant amount of solvent-accessible surface area buried in the tandem association interface of the hexamer. However, the interactions are polar in nature, and PhoPC exists predominantly as a monomer in solution. The crystal packing also shows two other types of symmetric dimers. Although the dimeric interactions are not as strong, with associations of the regulatory domain or involvement of other proteins, these symmetric dimer interfaces can potentially participate in bringing different regions of DNA together for regulation of gene expression. Foot-printing assays indicate that PhoP protects more than one region on some gene promoters (Engohang-Ndong, unpublished observation).

Both full-length PhoP and its isolated C-terminal domain can bind promoter DNA. However, the C-terminal domain has reduced specificity relative to the full-length protein (Engohang-Ndong, unpublished observation). It is possible that the isolated DNA-binding domain can bind DNA in a different mode from the full-length PhoP. The C-terminal domain of PhoP is able to form a hexamer, in which the recognition helix is not fully exposed for inserting into the major groove of DNA. However, most of the positively charged surfaces are exposed in the hexamer (Figure 5), including the wing residues, which also play an important role in DNA binding by interacting with the minor groove of DNA. Multiple molecules of PhoPC are thus able to bind DNA through interactions with minor grooves. Because the recognition helix is not involved in DNA binding, specificity is lost in this DNA-binding mode. The presence of the N-terminal domain is likely to prevent this non-specific binding mode by interacting with the C-terminal domain. For example, if the N-terminal β-sheet of the DNA-binding domain prefers to interact with the N-terminal regulatory domain, the tandem association in the hexamer would not be formed, and the recognition helix would be exposed for productive DNA –binding in the presence of the N-terminal domain. In the structures of the DrrD (25) and DrrB (24) of *T. maritima*, the N-terminal β-sheet of the DNA-binding domain is found to interact with the N-terminal domain.

Acknowledgements

We thank Dr. Vivian Stojanoff, Dr. Fabiano Yokaichiya, and other staff members at the beamline X6A of the National Synchrotron Light Source (NSLS) for assistance in data collection. Helpful comments and revisions on the manuscript by Drs. Karl Drlica, David Dubnau, Leonard Mindich, Richard Pine, and David Wah are gratefully acknowledged.

Abbreviations

MTB
: *Mycobacterium tuberculosis*

PhoPC
: PhoP C-terminal domain

IPTG
: isopropyl β-D-1-thiogalactopyranoside

NCS
: non-crystallographic-symmetry

PMSF
: phenylmethanesulfonyl fluoride

PCR
: polymerase chain reaction

TEV
: tobacco etch virus

rmsd
: root mean square deviation

TCS
: two-component system

bp
: base pairs

EMSA
: electrophoretic mobility shift assay

References

1. Raviglione MC, Smith IM. XDR tuberculosis--implications for global public health. N Engl J Med 2007;356:656–659. [PubMed: 17301295]
2. Av-Gay, Y.; Deretic, V. Two-component systems, protein kinases, and signal transduction in Mycobacterium tuberculosis. In: Cole, ST., editor. Tuberculosis and the Tubercle Bacillus. ASM Press; Washington, D. C.: 2005. p. 359-367.
3. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE 3rd, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Barrell BG, et al. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature 1998;393:537–544. [PubMed: 9634230]
4. Perez E, Samper S, Bordas Y, Guilhot C, Gicquel B, Martin C. An essential role for phoP in Mycobacterium tuberculosis virulence. Mol Microbiol 2001;41:179–187. [PubMed: 11454210]
5. Walters SB, Dubnau E, Kolesnikova I, Laval F, Daffe M, Smith I. The Mycobacterium tuberculosis PhoPR two-component system regulates genes essential for virulence and complex lipid biosynthesis. Mol Microbiol 2006;60:312–330. [PubMed: 16573683]
6. Asensio JG, Maia C, Ferrer NL, Barilone N, Laval F, Soto CY, Winter N, Daffe M, Gicquel B, Martin C, Jackson M. The Virulence-associated Two-component PhoP-PhoR System Controls the Biosynthesis of Polyketide-derived Lipids in Mycobacterium tuberculosis. J Biol Chem 2006;281:1313–1316. [PubMed: 16326699]
7. Gupta S, Sinha A, Sarkar D. Transcriptional autoregulation by Mycobacterium tuberculosis PhoP involves recognition of novel direct repeat sequences in the regulatory region of the promoter. FEBS Lett 2006;580:5328–5338. [PubMed: 16979633]
8. Birck C, Chen Y, Hulett FM, Samama JP. The crystal structure of the phosphorylation domain in PhoP reveals a functional tandem association mediated by an asymmetric interface. J Bacteriol 2003;185:254–261. [PubMed: 12486062]
9. Groisman EA, Mouslim C. Sensing by bacterial regulatory systems in host and non-host environments. Nature reviews 2006;4:705–709.
10. Alpuche Aranda CM, Swanson JA, Loomis WP, Miller SI. Salmonella typhimurium activates virulence gene transcription within acidified macrophage phagosomes. Proc Natl Acad Sci U S A 1992;89:10079–10083. [PubMed: 1438196]
11. Martin-Orozco N, Touret N, Zaharik ML, Park E, Kopelman R, Miller S, Finlay BB, Gros P, Grinstein S. Visualization of Vacuolar Acidification-induced Transcription of Genes of Pathogens inside Macrophages. Mol Biol Cell 2006;17:498–510. [PubMed: 16251362]

12. Bader MW, Navarre WW, Shiau W, Nikaido H, Frye JG, McClelland M, Fang FC, Miller SI. Regulation of Salmonella typhimurium virulence gene expression by cationic antimicrobial peptides. Mol Microbiol 2003;50:219–230. [PubMed: 14507376]
13. Bader MW, Sanowar S, Daley ME, Schneider AR, Cho U, Xu W, Klevit RE, Le Moual H, Miller SI. Recognition of antimicrobial peptides by a bacterial sensor kinase. Cell 2005;122:461–472. [PubMed: 16096064]
14. Miller SI, Mekalanos JJ. Constitutive expression of the phoP regulon attenuates Salmonella virulence and survival within macrophages. J Bacteriol 1990;172:2485–2490. [PubMed: 2185222]
15. Soto CY, Menendez MC, Perez E, Samper S, Gomez AB, Garcia MJ, Martin C. IS6110 mediates increased transcription of the phoP virulence gene in a multidrug-resistant clinical isolate responsible for tuberculosis outbreaks. J Clin Microbiol 2004;42:212–219. [PubMed: 14715755]
16. Blanco AG, Sola M, Gomis-Ruth FX, Coll M. Tandem DNA recognition by PhoB, a two-component signal transduction transcriptional activator. Structure 2002;10:701–713. [PubMed: 12015152]
17. Toro-Roman A, Mack TR, Stock AM. Structural analysis and solution studies of the activated regulatory domain of the response regulator ArcA: a symmetric dimer mediated by the alpha4-beta5-alpha5 face. J Mol Biol 2005;349:11–26. [PubMed: 15876365]
18. Toro-Roman A, Wu T, Stock AM. A common dimerization interface in bacterial response regulators KdpE and TorR. Protein Sci 2005;14:3077–3088. [PubMed: 16322582]
19. Maris AE, Walthers D, Mattison K, Byers N, Kenney LJ. The response regulator OmpR oligomerizes via beta-sheets to form head-to-head dimers. J Mol Biol 2005;350:843–856. [PubMed: 15979641]
20. Perron-Savard P, De Crescenzo G, Le Moual H. Dimerization and DNA binding of the Salmonella enterica PhoP response regulator are phosphorylation independent. Microbiology 2005;151:3979–3987. [PubMed: 16339942]
21. Pragai Z, Allenby NE, O'Connor N, Dubrac S, Rapoport G, Msadek T, Harwood CR. Transcriptional regulation of the phoPR operon in Bacillus subtilis. J Bacteriol 2004;186:1182–1190. [PubMed: 14762014]
22. Fiedler U, Weiss V. A common switch in activation of the response regulators NtrC and PhoB: phosphorylation induces dimerization of the receiver modules. Embo J 1995;14:3696–3705. [PubMed: 7641688]
23. McCleary WR. The activation of PhoB by acetylphosphate. Mol Microbiol 1996;20:1155–1163. [PubMed: 8809768]
24. Robinson VL, Wu T, Stock AM. Structural analysis of the domain interface in DrrB, a response regulator of the OmpR/PhoB subfamily. J Bacteriol 2003;185:4186–4194. [PubMed: 12837793]
25. Buckler DR, Zhou Y, Stock AM. Evidence of intradomain and interdomain flexibility in an OmpR/PhoB homolog from Thermotoga maritima. Structure (Camb) 2002;10:153–164. [PubMed: 11839301]
26. Nowak E, Panjikar S, Konarev P, Svergun DI, Tucker PA. The structural basis of signal transduction for the response regulator PrrA from Mycobacterium tuberculosis. The Journal of biological chemistry 2006;281:9659–9666. [PubMed: 16434396]
27. Kapust RB, Tozser J, Fox JD, Anderson DE, Cherry S, Copeland TD, Waugh DS. Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. Protein Eng 2001;14:993–1000. [PubMed: 11809930]
28. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol 1996;276:307–326.
29. Collaborative Computational Project Number 4. The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 1994;50:760–763. [PubMed: 15299374]
30. McCoy AJ, Grosse-Kunstleve RW, Storoni LC, Read RJ. Likelihood-enhanced fast translation functions. Acta Crystallogr D Biol Crystallogr 2005;61:458–464. [PubMed: 15805601]
31. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr 1998;54(Pt 5):905–921. [PubMed: 9757107]

32. Murshudov GN, Vagin AA, Lebedev A, Wilson KS, Dodson EJ. Efficient anisotropic refinement of macromolecular structures using FFT. Acta Crystallogr D Biol Crystallogr 1999;55(Pt 1):247–255. [PubMed: 10089417]
33. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 2004;60:2126–2132. [PubMed: 15572765]
34. Winn MD, Isupov MN, Murshudov GN. Use of TLS parameters to model anisotropic displacements in macromolecular refinement. Acta Crystallogr D Biol Crystallogr 2001;57:122–133. [PubMed: 11134934]
35. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. ProCheck-a program to check the stereochemical quality of protein structures. J Appl Crystallopgr 1993;26:283–291.
36. Lovell SC, Davis IW, Arendall WB 3rd, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC. Structure validation by Calpha geometry: phi, psi and Cbeta deviation. Proteins 2003;50:437–450. [PubMed: 12557186]
37. Martinez-Hackert E, Stock AM. Structural relationships in the OmpR family of winged-helix transcription factors. J Mol Biol 1997;269:301–312. [PubMed: 9199401]
38. Makino K, Amemura M, Kawamoto T, Kimura S, Shinagawa H, Nakata A, Suzuki M. DNA binding of PhoB and its interaction with RNA polymerase. J Mol Biol 1996;259:15–26. [PubMed: 8648643]
39. Martinez-Hackert E, Stock AM. The DNA-binding domain of OmpR: crystal structures of a winged helix transcription factor. Structure 1997;5:109–124. [PubMed: 9016718]
40. Kondo H, Nakagawa A, Nishihira J, Nishimura Y, Mizuno T, Tanaka I. Escherichia coli positive regulator OmpR has a large loop structure at the putative RNA polymerase interaction site. Nat Struct Biol 1997;4:28–31. [PubMed: 8989318]
41. Pratt LA, Silhavy TJ. OmpR mutants specifically defective for transcriptional activation. J Mol Biol 1994;243:579–594. [PubMed: 7966283]
42. Kato N, Tsuzuki M, Aiba H, Mizuno T. Gene activation by the Escherichia coli positive regulator OmpR: a mutational study of the DNA-binding domain of OmpR. Mol Gen Genet 1995;248:399–406. [PubMed: 7565603]
43. Kumar A, Grimes B, Fujita N, Makino K, Malloch RA, Hayward RS, Ishihama A. Role of the sigma 70 subunit of Escherichia coli RNA polymerase in transcription activation. J Mol Biol 1994;235:405–413. [PubMed: 8289270]
44. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 1997;18:2714–2723. [PubMed: 9504803]
45. Kraulis PJ. MOLSCRIPT: A Program to Produce Both Detailed and Schematic Plots of Protein Structures. Journal of Applied Crystallography 1991;24:946–950.
46. 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]
47. Barton GJ. ALSCRIPT: a tool to format multiple sequence alignments. Protein Eng 1993;6:37–40. [PubMed: 8433969]
48. Merritt EA, Murphy ME. Raster3D Version 2.0. A program for photorealistic molecular graphics. Acta Crystallogr D Biol Crystallogr 1994;50:869–873. [PubMed: 15299354]

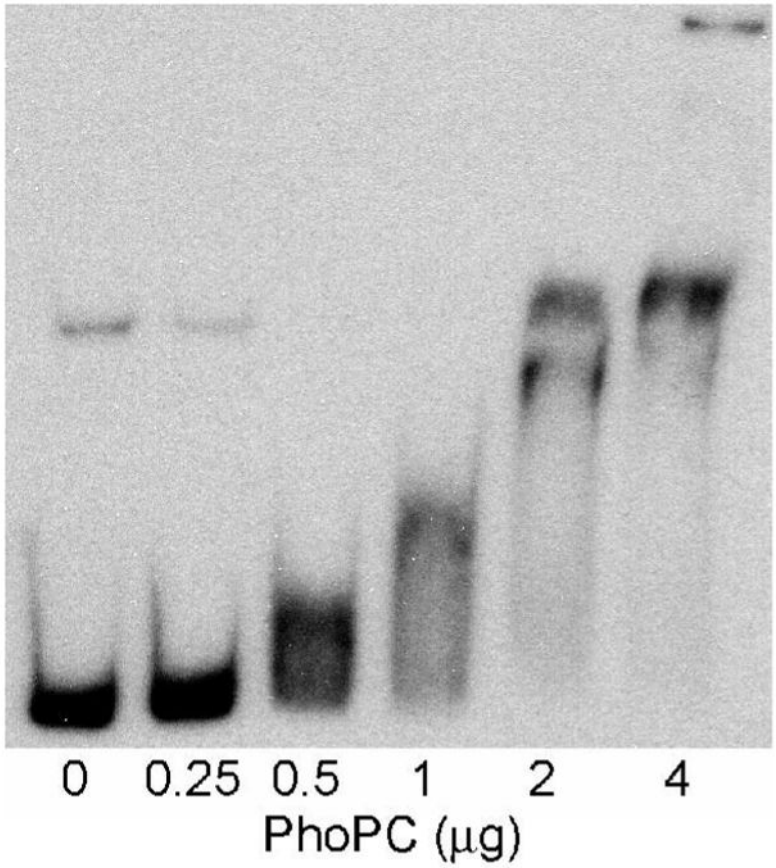

Figure 1.
Electrophoretic mobility shift assay of the binding of PhoPC with the promoter DNA of the *phoP* gene. Binding of PhoPC retards the electrophoretic mobility of DNA and thus causes a shift of the DNA band.

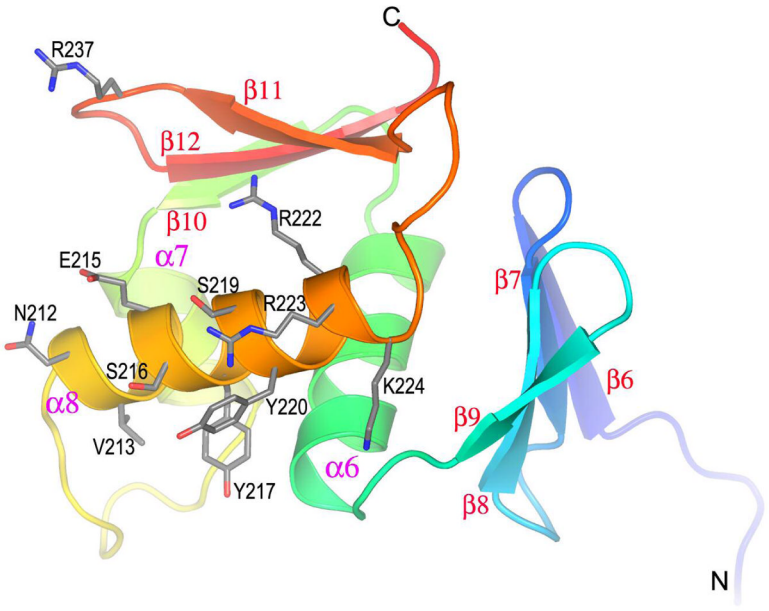

Figure 2.

Ribbon diagram of the structure of the C-terminal domain of MTB PhoP. Secondary structural elements are labeled. The β-strands and α-helices are numbered starting from β6 and α6, respectively, for consistency with the N-terminal regulatory domain structure. Side chains of residues on the recognition helix (α8) that are exposed and are likely to be involved in DNA binding and recognition are shown as sticks. The side chain of residue Arg237 at the wing is also shown. Amino acids are referred to as single letter codes in the figure for clarity. The figure was generated with the program PYMOL (http://www.pymol.org).

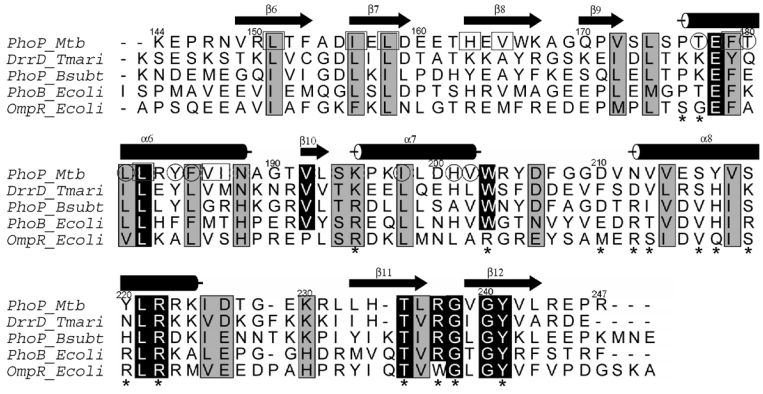

Figure 3.

Sequence alignment of the DNA-binding domain of response regulators of the OmpR/PhoB subfamily. Sequences were aligned with the program CLUSTALW (46), and the figure was generated with the program ALSCRIPT (47). The secondary structural elements shown are of the PhoPC structure. Residues labeled with a " * " are those found interacting with DNA in the structure of the PhoB-DNA complex (PDB code 1GXP) (16). Identical residues are highlighted in black boxes, while residues with high similarity as defined by CLUSTALW are highlighted in gray boxes. Among the sequences compared, OmpR of *E. coli* is the most distant from others. Several residues, including Val192, Trp203, and Arg237 (residue numbers refer to those of MTB PhoP sequence), are conserved among PhoP of MTB, DrrD of *T. maritima*, PhoP of *B. subtilis*, and PhoB of *E. coli*, but not in OmpR. Of these residues, Trp203 and Arg237 interact with DNA in the PhoB-DNA complex. Residues in the MTB PhoP sequence highlighted in rectangle boxes are involved in hydrophobic interactions between the N-terminal β-sheet and helix α6, and those in circles are involved in interactions between helices α6 and α7. These two hydrophobic clusters are connected by a cluster of side chains of residues Phe153 and Ala154 from the N-terminal β-sheet and residues Leu193, Leu232, Leu233, Tyr241, and Leu243 from the C-terminal β-sheet. Side chains of Val214, Val218, Leu221, and Ile225 from helix α8, as well as those of Leu174 and Trp203, also contribute to the hydrophobic core.

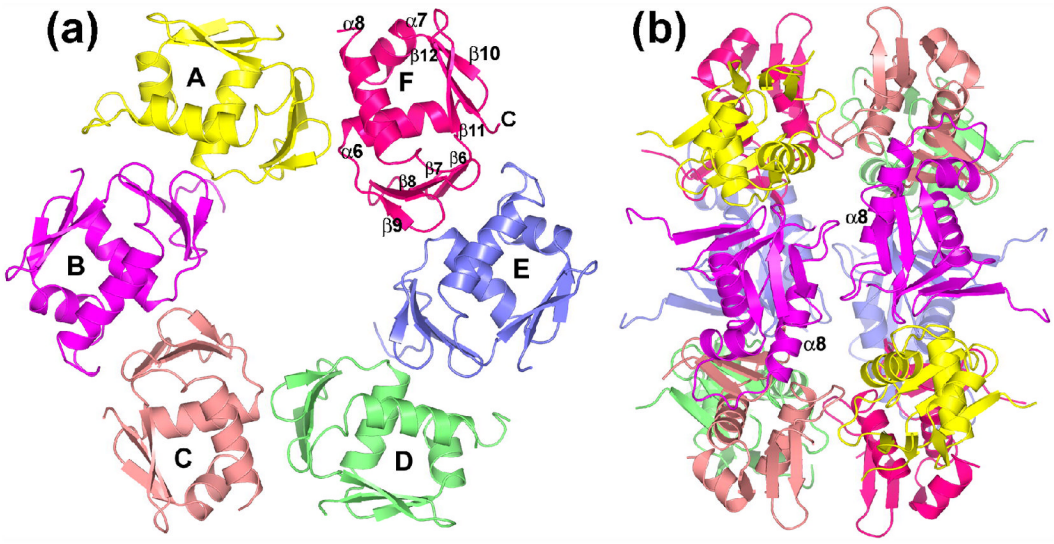

Figure 4.

Ribbon diagrams of hexamer and dodecamer in the crystal structure of PhoPC. (a) The six PhoPC molecules in the asymmetric unit form a hexamer with a loose six-fold non-crystallographic symmetry. The subunits are designated as A to F as described in text. Each subunit is colored differently. The secondary structure elements of subunit F are labeled. The recognition helix α8 is in the front while the N-terminus is in the back. The subunits interact with each other in a head-to-tail fashion. The molecular interface involves the N-terminal four-stranded β-sheet of one molecule and one side of the recognition helix α8, the N-terminal end of helix α6, and the transactivation loop of the other molecule. (b) The two-fold crystallographic symmetry relates two hexamers to give a double ring dodecamer, shown as a side-view of (a). Individual subunits in each ring are colored differently as in (a). Each subunit of one ring is related by a two-fold symmetry to the neighboring subunit of the other ring to give a symmetric dimer, as shown by the two magenta subunits (molecule B) in the front (recognition helix α8 is labeled). Major dimeric interface involves the loop following the recognition helix α8. The amino termini of all subunits point away from the dodecamer.

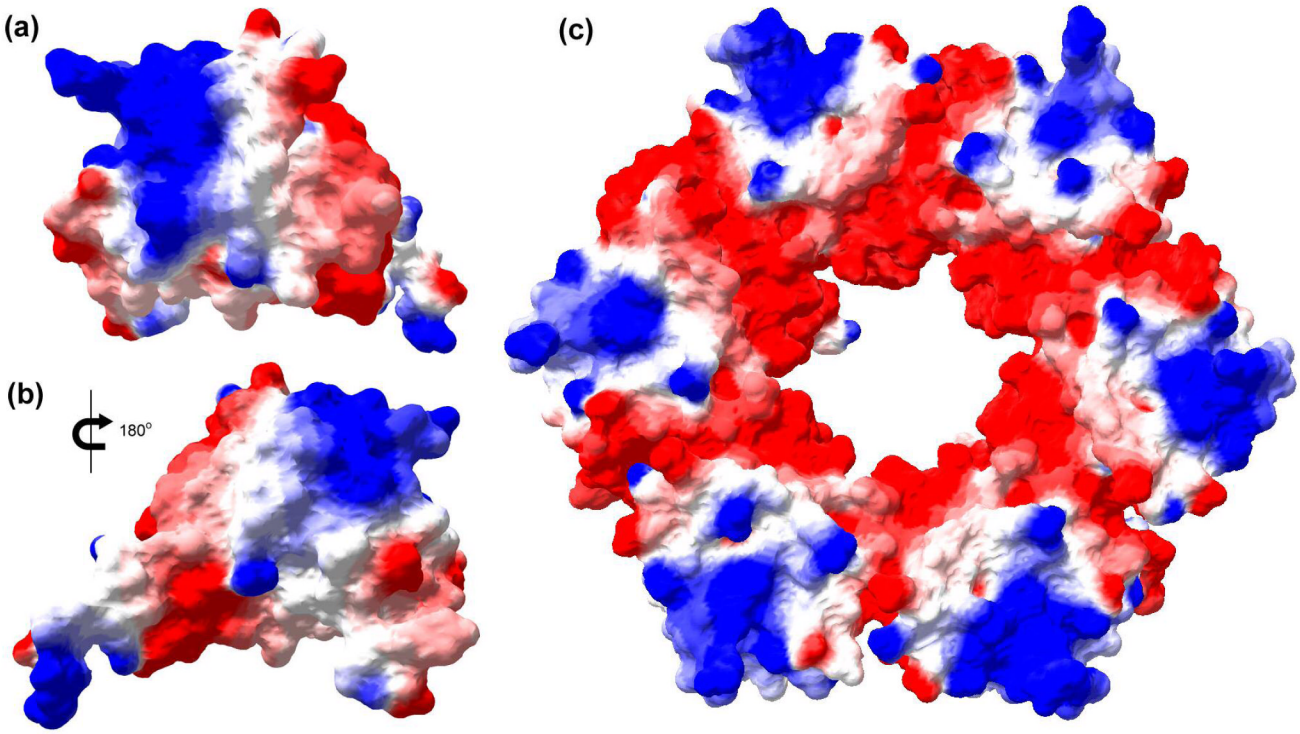

Figure 5.
Electrostatic potential surface diagrams of PhoPC. Positive charge is shown in blue, and negative charge is in red. The diagram in (a) shows the electrostatic potential surface in the same orientation as shown in the ribbon diagram in Figure 2. Shown in (b) is the backside of (a), i.e. approximately 180 ° rotation of (a) along a vertical axis. Panel (c) shows the surface potential of the hexamer in the same orientation as in Figure 4(a). Electrostatic potential was calculated and the figures were generated with the program Swiss-PdbViewer (44).

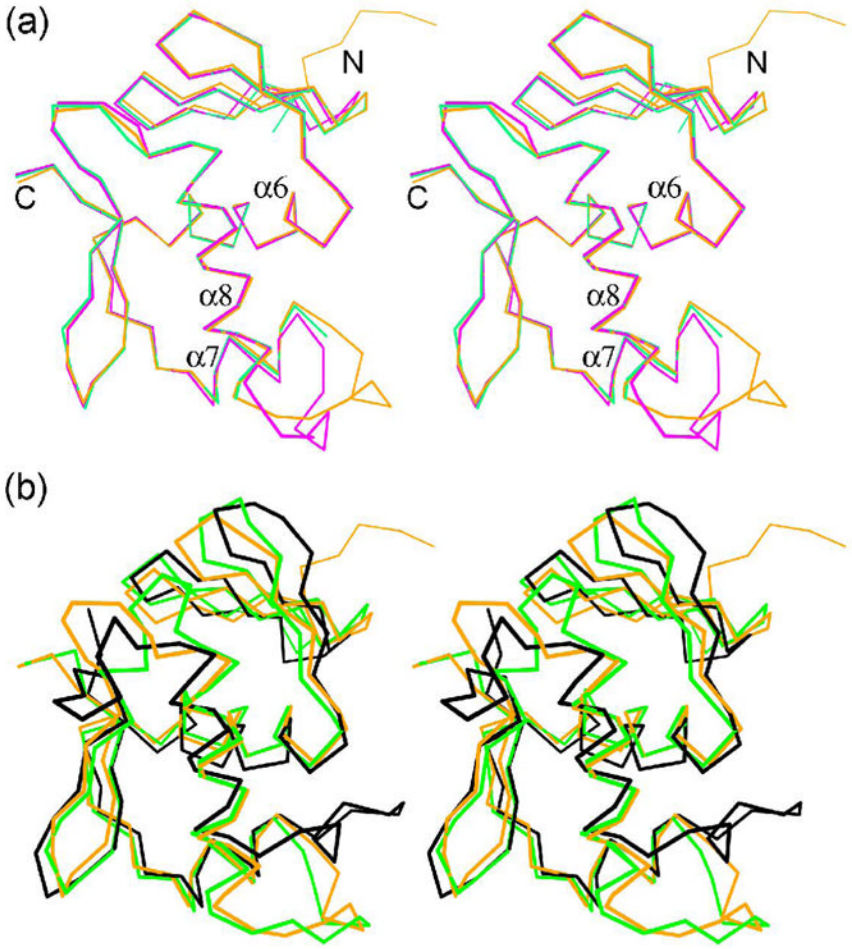

Figure 6.

Stereo views of structural superposition of molecules A, B and C in the PhoPC structure (a) and of PhoPC with structures of the C-terminal domain of OmpR and PhoB (b). The structures are shown as Cα traces. Molecules A, B, and C in (a) are shown in magenta, orange, and green, respectively. Helices and the two termini are labeled in (a). In (b), the molecule B of MTB PhoPC is colored in orange, OmpR of *E. coli* (PDB code 1OPC (39)) is colored in black, and PhoB of *E. coli*, (1GXQ (16)) is colored in green. PhoPC is more similar to PhoB in structure, while OmpR has the largest deviations, especially in the transactivation loop and the loop following the recognition helix. The figures were prepared with the program MOLSCRIPT (45).

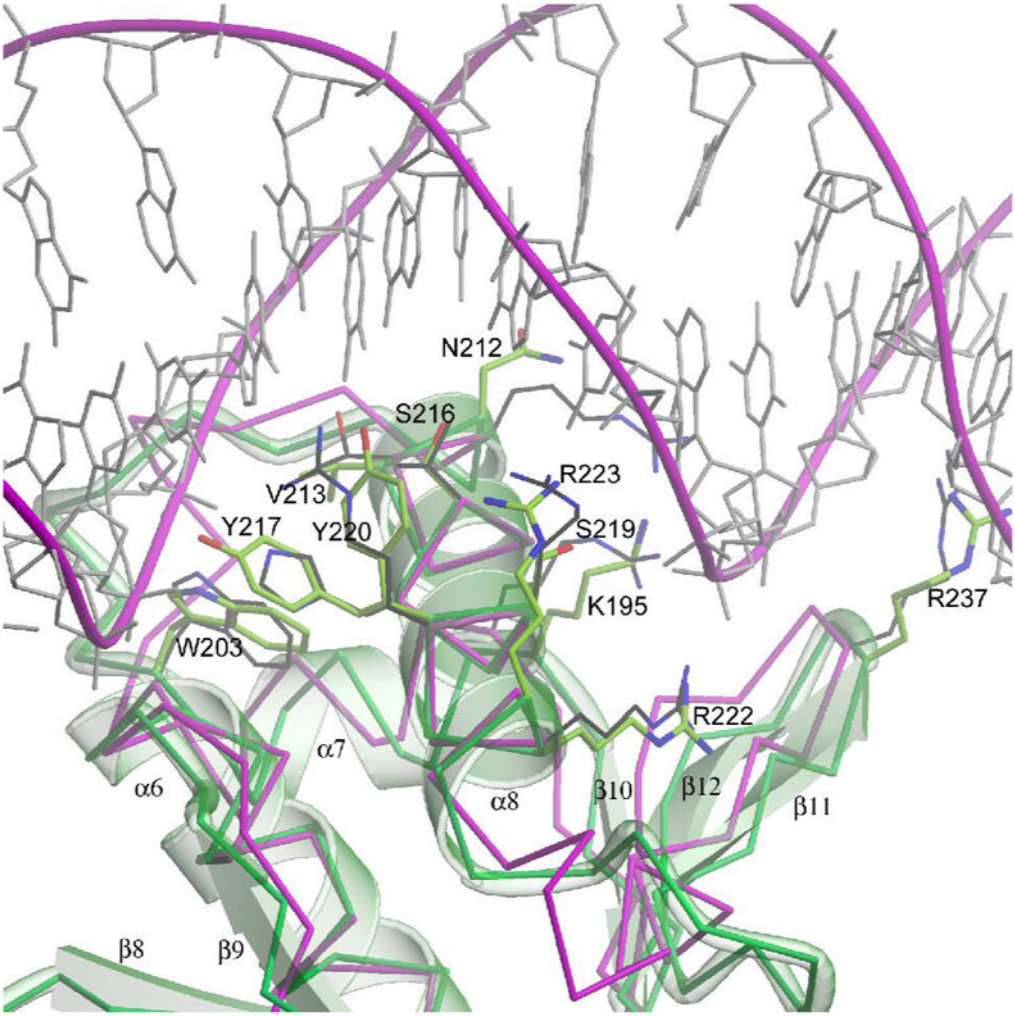

Figure 7.

Structural superposition of the molecule B of MTB PhoPC and the PhoB-DNA complex (PDB code 1GXP). The two structures were aligned based on the recognition helix. The PhoB structure is shown in magenta as a Cα trace, and the PhoPC structure is shown in green as a Cα trace and a ribbon diagram. The side chains shown are some of those found in the PhoB-DNA complex to interact with DNA and the corresponding residues in PhoP. Other side chains at the N-terminal end of helix α6 and at strands β11 and β12 that also interact with DNA in the PhoB-DNA complex are not shown for clarity. Also shown is the side chain of Arg223, which occupies the position of Arg200 in PhoB even though these two residues are not aligned in sequence. The residue labels are for those of the PhoP sequence with single letter codes for amino acids. Based on the structural superposition, PhoP is likely to be able to bind DNA in a similar way with most of the interactions with phosphate and sugar groups of DNA conserved. The figure was generated with the programs MOLSCRIPT (45) and RASTER3D (48).

Table 1

X-ray diffraction data for *M. tuberculosis* PhoPC[a].

Space Group	C2
Resolution (Å)	20-1.78
R_{merge}[c] (last bin)[b]	0.036 (0.582)
Completeness%[b]	94.8 (96.7)
I/σ[b]	26.4(2.0)
Redundancy[b]	2.9 (2.7)
Cell a (Å)	103.88
b (Å)	101.13
c (Å)	86.93
β (deg)	126.72

[a]. All data collected at cryogenic temperature of 100 K.

[b]. The numbers in parentheses are for the last bin of data, which is from 1.84 to 1.78 Å.

[c]. $R_{merge} = \Sigma|(I_{hkl} - \langle I_{hkl}\rangle)|/\Sigma I_{hkl}$, where $\langle I_{hkl}\rangle$ is the average of I_{hkl} over all symmetry equivalents.

Table 2

Atomic refinement statistics of PhoPC from *M. tuberculosis*.

Resolution (Å)	20-1.78
R_{work} (R_{free})[a]	0.196 (0.239)
Rmsd bonds (Å)	0.016
Rmsd angles (°)	1.55
No. protein atoms	4796
Ave B factors, protein atoms (Å^2)[b]	40.2
No. other molecules	3 glycine (56.0)
(Ave. B factors, Å^2)	4 phosphate (42.3)
	489 water (43.7)
	3 K^+ (35.0)[c]
	4 Cl^- (36.4)
	4 unknown atoms (37.8[d], 26.8[e])

[a]. R factors were calculated using data in the resolution range for refinement without a σ cutoff. R_{free} was calculated with a subset of data (5%) never used in the refinement. R_{work} was calculated against the data used in the refinement.

[b]. B factors reported for the protein atoms are total B factors calculated with TLSANL and BAVERAGE of the CCP4 Suite (29) after TLS refinements.

[c]. Some potassium ions might be sodium ions or partially occupied by sodium ions.

[d]. Two unknown atoms ~2.06 Å to the main chain carbonyl oxygen atoms are modeled as potassium ions for refinement.

[e]. Two other unknown atoms modeled as Li^+ atoms for refinement each is near a side chain oxygen atom of a Glu.

Table 3

Results of EMSA on selected genes regulated by PhoP

gene name	function	PhoPC binding
up-regulated genes		
PE_PGRS41	unknown, PE family	no
PPE19	unknown, PPE family	no
pks2	polyketide synthase	no
pks3	polyketide synthase	**yes**
down-regulated genes		
phoP	transcription regulator	**yes**
hsp	heat shock protein	**yes**
whiB6	putative transcription regulator	**yes**
pks5	polyketide synthase	no
Rv0042c	putative transcription regulator	no
Rv1816	putative transcription regulator	no
Rv2887	putative transcription regulator	no

Table 4

Solvent accessible surface areas that are buried at the molecular interfaces of the hexamer.

Molecules involved	AB	BC	CD	DE	EF	FA
Buried area (Å^2)	981	1055	833	922	1108	890