The full-length, two-domain response regulator RegX3 from *Mycobacterium tuberculosis* is a dimer stabilized by three-dimensional domain swapping. Dimerization is known to occur in the OmpR/PhoB subfamily of response regulators upon activation but has previously only been structurally characterized for isolated receiver domains. The RegX3 dimer has a bipartite intermolecular interface, which buries 2357 Å² per monomer.

The two parts of the interface are between the two receiver domains (dimerization interface) and between a composite receiver domain and the effector domain of the second molecule (interdomain interface). The structure provides support for the importance of threonine and tyrosine residues in the signal transduction mechanism. These residues occur in an active-like conformation stabilized by lanthanum ions. In solution, RegX3 exists as both a monomer and a dimer in a concentration-dependent equilibrium. The dimer in solution differs from the active form observed in the crystal, resembling instead the model of the inactive full-length response regulator PhoB.

Tuberculosis is a global threat to human health, with one-third of the world’s population infected with the causative agent of tuberculosis, *Mycobacterium tuberculosis*. With the emergence of multidrug-resistant strains of *M. tuberculosis*, novel therapeutic agents are urgently required. This will be best achieved by developing a better understanding of the molecular biology of this pathogen. In this context, signaling systems play an important role in the continued survival of the organism and are central to the way *M. tuberculosis* responds to environmental stress, especially that generated by the host immune system.

The predominant signaling system for adaptive gene expression changes in bacteria is called a two-component system (TCS).³ TCS are found in eubacteria, archaea, and eukarya; however, they are rare in eukarya. As the name implies, the system typically consists of two proteins: a sensor histidine kinase (SK) and a response regulator (RR) (1). SKs are usually membrane-anchored proteins with a characteristic core consisting of a histidine-containing dimerization domain and a catalytic domain with ATPase activity. In response to extracellular conditions, and in some cases intracellular conditions, the SK autophosphorylates at a histidine residue within the dimerization domain. This autophosphorylation occurs upon dimerization through a phosphate transfer from the catalytic domain (annotated as a HATPase_c domain by the SMART server (2, 3)) of the adjacent protomer to the conserved histidine in the dimerization domain (HisKa domain). The activated SK may then function as a phosphate donor to a universally conserved aspartic acid residue in the RR. RRs typically consist of two domains, a receiver and an effector domain, although a large family of σ²⁴ activators also contain a central AAA+ domain. The N-terminal receiver domain acts as the phosphoacceptor, whereas the C-terminal effector domain is usually a DNA-binding domain involved in the transcriptional regulation of genes required to respond to the sensed environment. Phosphorylation of the receiver domain results in the activation of the effector domain in a manner that is not yet completely understood, in part due to the lack of structural information on full-length RRs in the activated state. Dimerization is a proposed method of regulation (4). This is based upon structures of the receiver domain alone in both active and inactive states (4).

Some TCS appear to be essential for the survival of bacteria, and to date homologous TCS have not been identified in the animal kingdom. TCS are therefore considered attractive targets for the development of novel agents that could alter the response of the bacterium to its detriment. More than 180 different TCS have been identified in bacteria (5); however, the number varies enormously between species. *M. tuberculosis* contains 30 genes encoding TCS proteins (6), 12 complete TCS, and several orphan TCS proteins possibly belonging to the SK and RR families (6, 7). Of these 12 complete TCS, all but five have been found deleted or exist as pseudogenes in the closely related intracellular pathogen *Mycobacterium leprae* (8). The TCS SenX3-RegX3 is one of the five conserved TCS, emphasizing its importance for the survival of mycobacteria (not just the

* This work was supported by the Bundesministerium für Bildung und Forschung (German Science Ministry)-funded X-MTB consortium. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (code 2OQR) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

1 Present address: Argonne National Laboratory, 9700 S. Cass Ave., Argonne, IL 60439.

2 To whom correspondence should be addressed. Tel.: 49-40-89901129; Fax: 49-40-89902149; E-mail: tucker@embl-hamburg.de.

3 The abbreviations used are: TCS, two-component system(s); SK, sensor kinase; RR, response regulator; SAXS, small angle X-ray scattering; MM, molecular mass; wHTH, winged-helix-turn-helix; r.m.s., root mean square.
The Crystal Structure of RegX3

virulent strains) in hostile environments. Previous studies carried out on this TCS have revealed that both SenX3 and RegX3 communicate in the normal fashion, with His\textsuperscript{167} of SenX3 being the active histidine residue that is phosphorylated and Asp\textsuperscript{52} being the corresponding phosphorylation target in RegX3, the cognate RR (9). The production of *M. tuberculosis* strains with a compromised *senX3-regX3* operon (through the removal of the 3' end of the *senX3* gene, the intergenic repeats, and the 5' end of the *regX3* gene) has shown that an intact SenX3-RegX3 TCS is required for a progressive infection (10). Microarray experiments comparing *M. tuberculosis* strains with and without the *senX3-regX3* operon revealed 50 possible regulon members (10). One operon in particular, Rv0096-Rv0101, appears to be directly, negatively regulated with expression found to be 5–17-fold increased in the absence of the SenX3-RegX3 TCS (10). Further experiments using *M. tuberculosis* strains with a compromised *senX3-regX3* operon displayed significant attenuation in active and resting macrophages as well as in immunocompromised and immunocompetent mice, thus demonstrating that the *senX3-regX3* operon is involved in the virulence of *M. tuberculosis* (10). Recently, experiments in *Mycobacterium smegmatis* revealed that the SenX3-RegX3 system responds to limiting environmental phosphate concentrations by inducing the *phoA* and *pstS* genes (11). In *E. coli*, these genes are regulated by another TCS, PhoR-PhoB (12, 13). Through DNA-binding experiments, it was shown that phosphorylated RegX3 directly regulates *phoA* and *pstS*, in *M. smegmatis*, by binding to their promoter regions. Previously, RegX3 has only been demonstrated to have DNA-binding properties with the promoter region of *senX3* (9). A putative “pho-box” for RR binding has been derived for *M. smegmatis* consisting of an inverted repeat (GTGAAC) separated by seven nonconserved nucleotides (11).

Most RRs are multidomain proteins. All RRs contain structurally similar receiver domains with a characteristic B1-α1-β2-α2-β3-α3-β4-α4-β5-α5 topology and a conserved aspartic acid that is phosphorylated. They can be subdivided into families based upon similarities in their (output) effector domains. The three largest families are named after their founding members: OmpR/PhoB, FixJ/NarL, and NtrC/DctD. The RR studied here belongs to the OmpR/PhoB family. This family is characterized by a winged helix-turn-helix (wHTH) DNA-binding domain (14–16). This domain has been experimentally shown to bind direct tandem repeat half-sites (17–19) and inverted repeats (11) of DNA. Currently, four full-length structures of OmpR/PhoB type response regulators exist in the Protein Data Bank. A comparison of these structures suggests that the DNA-binding domain can exist in two types of conformations, closed and open. MtrA (20) and PrrA (21) exist as very compact structures with the recognition helix completely inaccessible to DNA. By comparison, the structures of DrrD (22) and DrrB (23) exhibit very open conformations with the DNA recognition helix of the effector domain completely accessible. In other RR families, the recognition helix is occluded in the unphosphorylated state, becoming accessible only upon phosphorylation (24, 25). This was recently suggested to be the effect of signal transduction by phosphorylation in PrrA (21). PrrA is one of the closed, compact structures of the OmpR/PhoB subfamily. Dimerization has been suggested as important for signal transduction in other OmpR/PhoB members. This is based upon several biochemical experiments and the two full-length crystal structures of DrrD (22) and DrrB (23), which are both open structures with accessible recognition helices. The α4-β5-α5 face of the receiver domain of a RR is the region most affected by phosphorylation and is the proposed site of dimerization in the OmpR/PhoB family. This site in this family displays high residue conservation as compared with other RR families (4), consistent with the importance of this region as a dimerization interface. However, dimerization is not only restricted to activated RRs. ArcA, for instance, exists as a dimer in the inactive state and forms an octamer upon activation (26). PmrA is also known to exist as a dimer in both states (27), and PhoB has been shown to undergo a transition that involves changing the dimerization interface from the α1-α5 region in the inactive form to the α4-β5-α5 region in the active form (28, 29).

Three-dimensional domain swapping is becoming increasingly more evident in the Protein Data Bank as more structures are deposited. Three-dimensional domain swapping describes the formation of a symmetry interaction between two molecules via the exchange of identical regions. The region that is swapped can be a single element of secondary structure or an entire globular domain (30). Although many proteins display this phenomenon, the functional and physiological relevance is often not understood. Some potential advantages attained through domain swapping include higher local concentrations of active sites, larger binding surfaces, new active sites at subunit interfaces, the possibility of allosteric control, and an economic way to produce a large protein interaction network (31). Here we present the structure of the full-length RR, RegX3, stabilized in what we believe to be an active form that exhibits three-dimensional domain swapping. As such, it would represent the first full-length RR to be captured in an activated dimeric state and to display three-dimensional domain swapping. We also present an analysis of the solution structure of RegX3 in an inactive form as both a monomer and dimer.

**EXPERIMENTAL PROCEDURES**

**Cloning and Expression**—The gene sequence encoding RegX3 was cloned by PCR from genomic *Mycobacterium tuberculosis* (H37Rv strain) DNA. Flanking DNA primers 5'-TATACCATG-GCAACCGATGTTGATTGAGG-3' and 5'-CAGAA-GCTTATTACTAGCCCTCGAGTTTGTAGCCCAGCCCG-CGC-3' were engineered to provide melting temperatures of ~65 °C with extensions encoding NcoI and HindIII (New England Biolabs, Frankfurt, Germany) restriction sites. The corresponding PCR product was ligated with T4 DNA ligase (New England Biolabs, Frankfurt, Germany) into a pETM11 (EMBL) expression vector with an N-terminal His\textsubscript{6} tag. For the forward primers inserted a GCA codon immediately following the start codon to preserve the reading frame. Following sequence confirmation, the plasmid was transformed into Rosetta(DE3)pLysS cells. Cells were induced with 1 mM isopropyl 1-thio-β-D-galactopyranoside at *A\textsubscript{600} = 0.7* and grown for 5 h at 25 °C. The cells were harvested by centrifugation and lysed by sonication in a buffer containing 50 mM Tris-HCl, pH 8.0, 200 mM NaCl, 5 mM β-mercaptoethanol (buffer A), and two tablets of protease inhibitor mixture (Roche Applied Science).
The lysate was spun down and loaded onto a 5-ml nickel-nitrilotriacetic acid affinity column (Qiagen) equilibrated in buffer A. The protein was eluted with a 100-ml linear gradient of 10–250 mM imidazole in buffer A. The fractions containing RegX3 were pooled and diluted to a final imidazole concentration of 50 mM. The protein was then mixed with (His-tagged) tobacco etch virus protease in a 1:20 ratio and incubated at 4 °C overnight in order to cleave the His6 tag. The digested protein was applied to the nickel-nitrilotriacetic acid affinity column previously equilibrated in buffer A containing 50 mM imidazole. The flow-through fractions were collected, concentrated to a final volume of 5 ml, and further purified on a HiLoad 16/60 Superdex 75 column (Amersham Biosciences) equilibrated with 20 mM Tris-HCl, pH 8.0, 100 mM NaCl, and 2 mM dithiothreitol (buffer B). The peak fractions were concentrated using a 10-kDa cut-off Centricon (VIVASPIN) to a final concentration of 10 mg/ml as measured by Bradford’s assay (Bio-Rad) (32).

Crystallization, Data Collection, and Processing—Crystallization of RegX3 was carried out at 19 °C using the hanging drop vapor diffusion method. The initial conditions (0.2 M calcium acetate, 18% polyethylene glycol 8000, 0.1 M sodium cacodylate at pH 6.0) were identified using the Crystal Screen from Hampton Research.

Following optimization, the best crystals were obtained after 5 days in a mixture containing 2 μl of protein solution with 2 μl of reservoir buffer. The reservoir buffer consisted of 0.2 M calcium acetate, 2–4% polyethylene glycol 4000, and 0.1 M sodium cacodylate at pH 6.5. These crystals diffracted to a maximum resolution of 6.0 Å. To obtain better quality crystals, it was necessary to mix the protein drop with LaCl3 to obtain a final concentration of 0.01 M La3+. For data collection, the crystals were transferred to a drop-containing reservoir solution mixed with 2-methyl-2,4-pentendiol to a final concentration of 20% and then flash-cooled to 100 K in a cold nitrogen gas stream. Two data sets were collected: one at 1.5 Å, to a resolution of 6.0 Å, and the other at 2 Å to a resolution of 2.6 Å, the solvent content (63%) was relatively high, and model building was performed using the program ARP/wARP (39). The resultant model contained 184 residues of 228. In the last cycle of the ARP/wARP, 144 residues were docked with the correct sequence. The model was completed with manual building using COOT (40) alternating with additional REFMAC5 (41, 42) refinement cycles. The refinement cycles included a bulk solvent correction and anisotropic scaling and used the entire molecule as a single rigid body for TLS refinement. The resultant refined coordinates were then used to refine against the 2.03 Å data set using REFMAC5. The overall geometric quality of the model was assessed using the program PROCHECK (43). The refined coordinates and the structure factors for RegX3 have been deposited in the Protein Data Bank (Protein Data Bank code 2OQR). Data collection and refinement statistics are shown in Table 1. The superimpositions and r.m.s. deviation calculations were carried out using the programs COOT and SSM (44). All molecular images were generated using PYMOL (45).

Small Angle Scattering—Synchrotron x-ray scattering data from solutions of RegX3 were collected at the X33 EMBL beamline using an MAR345 image plate detector (46). The scattering patterns were measured with a 3-min exposure time for multiple solute concentrations ranging from 2.7 to 15.3 mg/ml. No indication of radiation damage was detected. This was checked by the comparison of duplicate scattering patterns from 2-min exposures. Using a sample-detector distance of 2.7 m, the momentum transfer range, 0.009 < q < 0.5 Å−1, was covered (where s = 4πsin(θ)/λ, 2θ is the scattering angle, and λ = 1.5 Å is the x-ray wavelength).

The data were processed using standard procedures by the program package PRIMUS (47). The forward scattering I(0) and the radii of gyration Rg were evaluated using the Guinier approximation (48), assuming that at very small angles (s < 1.3/Rg), the intensity is represented as I(s) = I(0)exp(−(sRg)2/3). The maximum particle dimensions (Dmax) and the interatomic distance distribution function p(r) were computed using the program GNOM (49). The molecular mass (MM) of the solute at various concentrations was evaluated by a comparison of the forward scattering with that from a reference solution of bovine serum albumin (MM = 66 kDa). Particle shape at low resolution was reconstructed ab initio using the program DAMMIN (50), which represents the particle as a collection of M = 1 densely packed beads inside a sphere with the diameter Dmax. Each bead belongs either to the particle or to the solvent, and the shape is described by a binary string of length M. Starting from a random string, simulated annealing is employed to search for a compact model that fits the experimental data Iexp(s) to minimize the discrepancy χ:

$$\chi^2 = \frac{1}{N-1} \sum \left( \frac{I_{\text{exp}}(s_j) - I_{\text{calc}}(s_j)}{\sigma(s_j)} \right)^2$$ (Eq. 1)

where N is the number of experimental points, c is a scaling factor, Iexp(sj) and Icalc(sj) are the experimentally determined and calculated intensities, respectively, and σ(sj) is the experimental error of Iexp(sj) at the momentum transfer sj. The program GASBOR (51) was also used to create ab initio
The Crystal Structure of RegX3

### TABLE 1
Data collection and refinement statistics

| Parameters | Beamline BW7A EMBL-Hamburg, DESY | Beamline X11 EMBL-Hamburg, DESY |
|------------|----------------------------------|----------------------------------|
| Data collection* | | |
| Wavelength (Å) | 1.508 | 0.813 |
| Temperature (K) | 100 | 100 |
| Space group | P422,2 | P422,2 |
| Cell parameters (Å) | a = b = 124.2, c = 44.7 | a = b = 124.3, c = 44.8 |
| Maximum resolution (Å) | 2.60 | 2.03 |
| No. of unique reflections | 11,221 | 23,225 |
| Completeness (%) | 99.7 (100.0) | 99.9 (97.7) |
| Multiplicity | 7.7 (7.3) | 12.1 (8.8) |
| R(f0) | 13.6 (2.3) | 31.9 (3.9) |
| Rmerge (%) | 12.2 (76.7) | 7.6 (66.4) |
| Rfree (%) | 4.5 (29.9) | 2.3 (23.4) |

**Refinement**

| Parameters | Values |
|------------|--------|
| Resolution (Å) | 30.0-2.03 |
| Rmerge/"Rfree" (%) | 17.8/21.6 |
| No. of residues | 226 |
| Missing residues | 4 |
| No. of water molecules | 181 |
| No. of acetate ions | 1 |
| No. of β-mercaptoethanol molecules | 1 |
| No. of La(III) ions | 5 |
| r.m.s. deviation bonds (Å) | 0.016 |
| r.m.s. deviation angles (degrees) | 1.5 |

**Ramachandran plot**

| Parameters | Values |
|------------|--------|
| Most favored (%) | 90.4 |
| Additional allowed (%) | 8.6 |
| Generously allowed (%) | 0.5 |
| Disallowed (%) | 0.5 |

* Numbers given in parentheses are from the last resolution shell.

### Data Collection

Table 1 presents data collection and refinement statistics for the crystal structure of RegX3, including wavelength, temperature, space group, cell parameters, maximum resolution, number of unique reflections, and completeness. The data were collected at two beamlines, BW7A and X11, with resolutions ranging from 30.0 Å to 2.03 Å. The statistics include Rmerge, Rfree, and Rwork, indicating the quality of the data.

### Refinement

The refinement process involved several steps, starting with preliminary phases derived from single wavelength data and progressing to multiple runs using the program CRYSOL. The resulting model was refined with REFMAC5 to a resolution of 2.6 Å. The overall structure was refined against 2.0 Å data collected at λ = 0.813 Å from a second crystal co-crystallized with lanthanum chloride. Phasing information from the five lanthanum ions in the asymmetric unit was used to calculate the initial experimental electron density map. Following manual building of the initial model, the structure was refined to 2.6 Å. The resulting coordinates were then refined and used to refine against 2.0 Å data collected at λ = 1.5 Å.

### Models

Models consisting of dummy residues instead of beads simulated annealing were employed to construct a model with protein-like distribution of beads providing the best fit to the experimental data. For the *ab initio* and rigid body analyses, multiple runs were performed to verify the stability of the solution, and typical three-dimensional reconstructions are presented below.

### Overall Structure

The overall structure, shown in Fig. 1, consists of two domains: the receiver domain, which at first sight has the expected secondary structural elements although in a totally new spatial arrangement, and a wHTH DNA binding domain characteristic of the OmpR/PhoB RR subfamily (14–16). The structure contains five lanthanum ions that were identified and distinguished from other possible metal ions by their anomalous signal. Of the five ions, one occupies the active site, two interact between the domain-swapped receiver domain and the effector domain, and the remaining two stabilize the crystal lattice (see below). The largest contact area between molecules is mediated by the α4-β5-α5 face and forms a two-fold symmetric crystallographic dimer burying a total surface area of 2357 Å²/monomer (Fig. 2). An examination of the crys-
tal packing reveals three other contact areas that stabilize the crystal lattice. One contact occurs between the C-terminal end of α2, the α2β3-loop, the N-terminal end of α4, and the α3β4-loop of the protomer in the asymmetric unit with the N-terminal end of α8 and the β11β12-loop of a symmetry-related molecule (−x + 1/2, y − 1/2, −z) burying an area of 591 Å²/monomer. The next interface occurs between the N-terminal region of α1 and the linker region between α5 and β6 of a symmetry-related molecule (−y, −x, −z + 1). The smallest interface, burying 208 Å², occurs between the β11β12-loop and the C-terminal region of α5 from a symmetry-related molecule (y + 1/2, x − 1/2, z).

A typical receiver domain has an α/β topology that consists of a central five-stranded β-sheet (β2-β1-β3-β4-β5), which forms a hydrophobic core surrounded by two helices (α1 and α5) on one side and three (α2–α4) on the other side. The monomer of RegX3 in the asymmetric unit has an apparently incomplete receiver domain with only a four-stranded parallel β-sheet (β2-β1-β3-β4) and three α-helices, of which α1 is positioned on one side of the β-sheet and α2–α3 are positioned on the other side (Fig. 1). The linker region connecting the two domains is made up of α4 and β5, which, as stated above, form a strong dimerization interface. The three-dimensional domain-swapped element, namely the incomplete receiver domain (β1, α1, β2, α2, β3, α3, and β4) from molecule 2 of the crystallographic dimer, interacts with α4, β5, and α5 from molecule 1. Strand β5 of molecule 1 aligns parallel to the domain-swapped, four-stranded β-sheet of molecule 2, forming a composite five-stranded β-sheet. Helix α4 of molecule 1 lies just outside of β5 on one side of the composite β-sheet with α1 and α2 from the domain-swapped element, whereas α5 from molecule 1 lies on the other side of the composite β-sheet (Fig. 2). The new composite receiver domain is now complete, with three α-helices on the other side of a five-stranded β-sheet. When the composite receiver domain is superimposed with other previously characterized receiver domains from the OmpR/PhoB family (Fig. 3A), such as the active PhoB receiver domain (29), the inactive PhoB receiver domain (28), DrrD (22), DrrB (23), PrrA (21), and MtrA (20), the secondary structures occupy similar positions with an overall r.m.s. deviation between Cα atoms ranging from 1.45 to 2.19 Å (Table 2). Not surprisingly, the region of highest variance is the dimerization interface composed of α4β5α5.

The effector domain of RegX3 has a typical wHTH fold, consisting of a four-stranded antiparallel β-sheet followed by a three-helix bundle and a C-terminal β-hairpin (Fig. 2). The DNA recognition helix (α8) is fully exposed to the solvent, as is also the case for the OmpR/PhoB family members DrrB (23) and DrrD (22) but unlike the other structurally characterized full-length family members PrrA (21) and MtrA (20).

The Receiver Domain—The active site of all RRs is in a crevice in the receiver domain. In the case of RegX3, this is formed by the β5α5-loop from molecule 1 and by the
The Crystal Structure of RegX3

![Structure diagram](image)

**TABLE 2**

r.m.s. deviation (Å) comparison of RegX3 against existing OmpR/PhoB subfamily member structures of individual domains

|        | RegX3       | Receiver domain | Effector domain |
|--------|-------------|-----------------|-----------------|
| PhoB (active) | 1.68 (101 to 101 Cα atoms) | 1.54 (83 to 83 Cα atoms) |
| PhoB (inactive) | 1.99 (105 to 105 Cα atoms) | 2.19 (99 to 99 Cα atoms) |
| OmpR   | 1.93 (105 to 105 Cα atoms) | 1.56 (84 to 84 Cα atoms) |
| DrdB   | 1.93 (105 to 105 Cα atoms) | 1.78 (58 to 58 Cα atoms) |
| PrrA   | 1.75 (100 to 100 Cα atoms) | 1.12 (78 to 78 Cα atoms) |
| MtrA   | 1.45 (102 to 102 Cα atoms) | 1.33 (85 to 85 Cα atoms) |

β1α1-loop, β3α3-loop, and β4α4-loop from the domain-swapped element of molecule 2. The active site of the complete, but composite, receiver domain contains all of the residues essential for phosphorylation. These are Glu8, Asp9, Glu10, Asp52, Leu53, Met56, Thr79, Thr81, Lys103, Ile55, Asn56, and Gly82 have been shown, in the structure of the activated RR ArcA, to be important in the stabilization of the phosphate mimic BeF3⁻ (4). In RegX3, the active site residues Asp52, Thr79, Lys101, Leu53, and Glu84 are required for the formation and stabilization of the phosphorylated species. Of particular interest is the residue Lys101 on β5, since it is the only residue that is not from the three-dimensional domain-swapped element. It forms a hydrogen bond (2.84 Å) with a carboxylate oxygen atom from Asp52. In the phosphorylated molecule, this bond would be broken, and an alternative hydrogen bond would be formed between the phosphate moiety and Lys101, as has been observed in ArcA (4). The signature switch residues known to be crucial in signal transduction by the rearrangements resulting from phosphorylation are serine/threonine and phenylalanine/tyrosine located in the β4α4-loop and in β5, respectively (54, 55). In inactivated RRys the tyrosine or phenylalanine occupy positions facing away from the active site in gauche⁻ (OH group points outward from the domain, exposed to solvent) and gauche⁺ (OH group points toward the C terminus of α4) positions. This is believed to be due to the outward orientation of the threonine/serine residue on the β4α4-loop. In RegX3, these residues are Thr79 and Tyr98, respectively. Thr79 is from the three-dimensional domain-swapped element, whereas Tyr98, like Lys101, is not. In the RegX3 structure, Thr79 is oriented toward the active site in a position that would coordinate phosphate, if it were there, whereas Tyr98 adopts an inward transposition, occupying the space freed by the inward rotation of Thr79 (Fig. 5). La-1 is coordinated by the carboxyl oxygen atom of Tyr98. This La³⁺ ion is positioned between α4 and β5. It also displays nonadentate coordination, interacting directly with both carboxylate oxygen atoms of Glu86, a single carboxylate oxygen atom from Asp57, and the carbonyl oxygen atom from
The Crystal Structure of RegX3

The nonadentate coordination spheres for the five La$^{3+}$ ions involved in the crystallization and structure determination of RegX3. A, La-1 is positioned proximal to the ‘switch’ residue Tyr$^{98}$. La-1 is coordinated by the carbonyl oxygen atom of Tyr$^{98}$, forcing the residue into an active conformation. La-1 exhibits nonadentate coordination. B, La-2 is positioned at the C terminus of α1 between the receiver domain and the effector domain of the domain-swapped partner. It stabilizes the α7α8-loop region. C, La-3 and La-4 are positioned on a crystallographic intermolecular contact point at the N terminus of α1, forming an interface with the α5-β6 linker region of the symmetry-related molecule. D, La-5 is positioned in the active site of the receiver domain and is directly coordinated by the highly conserved Asp$^{32}$, which is phosphorylated in vivo. Other highly conserved residues complete the active site, including Lys$^{101}$ from the dimeric partner, which also hydrogen-bonds with Asp$^{32}$. The black dashed lines show the metal coordination with distances in Å. The electron density (weighted 2Fo − Fc) in these regions is shown in a wire frame representation contoured at 1.5σ.

FIGURE 4. The nonadentate coordination spheres for the five La$^{3+}$ ions involved in the crystallization and structure determination of RegX3. A, La-1 is positioned proximal to the ‘switch’ residue Tyr$^{98}$. La-1 is coordinated by the carbonyl oxygen atom of Tyr$^{98}$, forcing the residue into an active conformation. La-1 exhibits nonadentate coordination. B, La-2 is positioned at the C terminus of α1 between the receiver domain and the effector domain of the domain-swapped partner. It stabilizes the α7α8-loop region. C, La-3 and La-4 are positioned on a crystallographic intermolecular contact point at the N terminus of α1, forming an interface with the α5-β6 linker region of the symmetry-related molecule. D, La-5 is positioned in the active site of the receiver domain and is directly coordinated by the highly conserved Asp$^{32}$, which is phosphorylated in vivo. Other highly conserved residues complete the active site, including Lys$^{101}$ from the dimeric partner, which also hydrogen-bonds with Asp$^{32}$. The black dashed lines show the metal coordination with distances in Å. The electron density (weighted 2Fo − Fc) in these regions is shown in a wire frame representation contoured at 1.5σ.

FIGURE 5. The key switch residues, Tyr$^{98}$ and Thr$^{79}$, adopt an inward position in the active state and an outward position in the inactive state. RegX3 (three-dimensional domain-swapped elements in red, the remainder in blue) displays a geometry representing the activated state and is similar to the activated structure of PhoB (magenta). The inactive conformation of PhoB (green) differs in the orientation of the threonine and tyrosine residues, which adopt a conformation pointing away from the active site indicated by the conserved aspartate (Asp$^{32}$ in RegX3). For better visualization, the domain-swapped element has been clipped from residue 82 onward.

Tyr$^{98}$, the switch residue. Five water molecules complete the coordination (Fig. 4A). The coordination of the carbonyl oxygen atom of Tyr$^{98}$ forces the aromatic ring to be orientated inward in the active state position.

Dimerization Interface—The total buried surface area created as a result of dimerization is 2357 Å$^2$/monomer shared equally between two distinct subinterfaces. The first, which we refer to as the dimerization interface, buries 1169 Å$^2$/monomer and is formed between the two composite receiver domains. It is centered on the helices that pack about the crystallographic two-fold axis (Fig. 6A). The second interface is formed between the composite receiver domain (consisting of the N-terminal, three-dimensional domain-swapped element, α8, β5, and α5) and the effector domain (Fig. 6B). The dimerization interface contains four salt bridges per monomer with a forked salt bridge within the active site between E8 (β1α1-loop) and D52 (C terminus of β3) from the domain-swapped element of molecule 2 and Lys$^{101}$ (α4β5-loop) from molecule 1. A salt bridge is also formed between Asp$^{68}$ (α4β5-loop) from molecule 2 and Asp$^{79}$ (α3β4-loop) from molecule 1. The extended five-stranded β-sheet created by the three-dimensional domain swapping is stabilized by an extensive network of hydrogen bonds, which further strengthen the interface. Asp$^{68}$ (C terminus of α3) from the domain-swapped fragment of molecule 2, already participating in a salt bridge, is further stabilized by Gly$^{94}$ (α4β4-loop) from molecule 1. Ile$^{76}$ (β4), also from the domain-swapped element, stabilizes Asp$^{96}$ (β5) from molecule 1 by forming two hydrogen bonds. One hydrogen bond is formed between the carbonyl oxygen atom of Ile$^{76}$ and the main-chain nitrogen atom of Asp$^{96}$. The other bond is formed between the main-chain nitrogen atom of Ile$^{76}$ and one of the carboxylate oxygen atoms of Asp$^{96}$ that is already involved in a salt bridge. Two hydrogen bonds are formed between both the carbonyl oxygen atom and main-chain nitrogen atom of Val$^{79}$ (β4) from molecule 2 with Asp$^{97}$ and Val$^{99}$ (β5) from molecule 1, respectively. The last two hydrogen bonds are formed between the main-chain nitrogen atom and carbonyl oxygen atom of Ala$^{90}$ (β4) from molecule 2 and the carbonyl oxygen atom of Val$^{99}$ (β5) from molecule 1, respectively. The last salt bridge involved in the dimerization interface is formed between Asp$^{97}$ (β5) and Arg$^{111}$ (α5) both from molecule 1, ensuring the two secondary structure elements maintain close contact as they complete the composite receiver domain. The hydrophobic surface formed by Leu$^{108}$, Ile$^{109}$, Ile$^{112}$, Val$^{115}$, and Leu$^{116}$ of α5 protects an otherwise exposed hydrophobic surface of the receiver domain formed by α1 and the central β-sheets. In other RRs, this hydrophobic surface is protected by both α4 and α5; however, in RegX3, α4 acts as a linker packing against the α4 of the dimeric partner. The two helices pack against each other on a hydrophobic interface made by Ile$^{85}$, Val$^{88}$, Val$^{98}$, and Leu$^{93}$. The N-terminal resi-
The Crystal Structure of RegX3

FIGURE 6. A schematic diagram of the crystallographic three-dimensional domain-swapped dimer and the resulting functional unit. A, the three-dimensional domain-swapped dimer, one molecule being darker than the other. The position of the two-fold crystallographic axis is shown. Tyr98 in both molecules is displayed in a stick representation. B, the hybrid RR produced as a result of the three-dimensional domain swapping. The domain-swapped element consists of β1, α1, β2, α2, β3, α3, and β4 from the monomer in gray. The remaining elements α4, β5, α5, and the effector domain are in block.

dues of each α4 helix play important roles in completing the active site of their own molecule.

Interdomain Interface—We suggest that the model of RegX3 provides the first high resolution structural data of the interdomain contacts occurring between the regulatory and the effector domains for a member of the OmpR/PhoB subfamily in an activated state. The interaction is formed by α5 from the composite receiver domain with the β7/β8-loop and C terminus of α7 from the effector domain. This interface is the second sub-interface of the three-dimensional domain-swapped dimer and buries a total surface area of 1188 Å²/monomer. This is significantly larger than that in previously characterized full-length RRs, PrrA (820 Å²), MtrA (738 Å²), DrrB (751 Å²), and DrrD (245 Å²).

The RegX3 interdomain interface is most similar to the interface of DrrD in that it involves the face of α5. However, unlike DrrD and the other structures, RegX3 exhibits a large linker region of 11 residues between α5 and β6, which is involved in many interactions that strengthen the interface. Two hydrogen bonds are formed from the carboxyl oxygen atoms of Asp¹²² and Ser¹²⁴ to the N⁴⁴ atoms of Arg¹⁶⁷ (α6). The other two hydrogen bonds occur between the carboxyl oxygen atoms of Glu¹²⁵ and the N⁴⁴ atoms of Arg¹⁸¹ (α7). This extended linker region may function as a “clasp” to fix the orientation of the effector domain by coordinating with two of the three helices that form the DNA-binding domain. However, in the RegX3 crystal structure, the “clasp” (residues 123–129) is distorted by two La³⁺ ions at a crystal contact (Fig. 4C). The ions La-3 and La-4 are located as a bimetallic cluster proximal to α1, which interacts with the “clasp” region from the effector domain of a symmetry-related monomer (−y, −x, −z + 1). Both ions display nonadentate geometry and are coordinated by a single residue each from α1. La-3 is coordinated by a carboxylate oxygen atom from Asp¹⁵, and La-4 is coordinated by both carboxylate oxygen atoms from Glu¹¹. Five residues from the “clasp” region between α5 and β-sheets 6, 7, and 8 of the symmetry-related monomer complete the coordination. Two residues, Asp¹²² and Asp¹²³, bridge the two ions with both carboxylate oxygen atoms coordinating individually with each ion. Asp¹²⁸ and three waters complete the coordination sphere for La-3. La-4 is additionally coordinated by a carboxylate oxygen atom from both Asp¹²¹ and Glu¹⁴¹ and three water molecules (Fig. 4C).

Another contact point between the two domains occurs at the position of La-2, which lies between α7 from the effector domain of one molecule and the end of α1 from the receiver domain of the second molecule of the dimer. La-2 also displays nonadentate coordination geometry. It is coordinated by a single carboxylate oxygen atom from Asp¹⁸⁶ of the α7α8-loop, both carboxylate oxygen atoms from Asp¹⁸⁰ at the N terminus of α8, and both carboxylate oxygen atoms from Glu²⁴ in the C terminus of α1 from the symmetry-related molecule. Four water molecules complete the coordination sphere (Fig. 4B). As a result of the stabilization of the α7α8-loop, a forked salt bridge is formed between Gly¹⁸⁴ from the loop and Arg¹¹³ from α5.

The Effector Domain—The wHTH fold contains the structural and sequence elements responsible for DNA recognition. OmpR and PhoB are the founding members of this family. The isolated structures of the OmpR (15) and PhoB (18) effector domains complexed with DNA as well as the uncomplexed but full-length structures of DrrB (22), DrrD (23), and PrrA (21) have all been described in the literature. The RR subfamily differs from other wHTH non-RR proteins in that they have an N-terminal, four-stranded, antiparallel β-sheet platform and a large loop between the positioning helix (α7) and the recognition helix (α8) called the transactivation loop (14–16). In RegX3, the β-sheet platform is positioned to the side, and the recognition helix is completely accessible (Fig. 7). The isolated effector domains from all of the above structures superimpose quite well on the isolated effector domain from a single monomer of RegX3 (Table 2). The largest backbone deviations occur in the transactivation loop regions, the α7α8-loop and the α8β11-loop (Fig. 3B).

A single residue in the effector domain (Asn¹⁴⁸) is found in the disallowed region of the Ramachandran plot (Table 1). This residue is located at a β-hairpin on the β8β9-loop and is involved in main-chain hydrogen bonding but in a region with poor electron density.

Small Angle X-ray Scattering—We believe that the crystal structure of RegX3 reflects an activated RR dimer generated by the presence of lanthanum ions rather than through phosphorylation. Small angle x-ray scattering (SAXS) was used to analyze the protein in solution in the inactive state without the
The Crystal Structure of RegX3

Concentration measured using different programs (DAMMIN and GASBOR) with and without two-fold symmetry imposed. The overall shape in solution, derived by GASBOR with two-fold symmetry applied, appears linear in a “trans-shaped” dimer (Fig. 8B) compared with the “cis-shaped” crystal dimer (Fig. 8A). The elongated trans-shape is strikingly similar to the proposed model of the inactive full-length PhoB structure (29), where the positioning of the effector domains in opposite directions provides an additional means of inhibition. We then used, as prior knowledge, the inactive PhoB dimerization interface contacts and the full-length RR DrrD for rigid body modeling against the 15.3 mg ml\(^{-1}\) RegX3 SAXS data. Several runs of the program SASREF, assuming two-fold symmetry, produced consistent results and a good fit (\(\chi = 2.05\)) to the scattering data (Fig. 9). This model is superimposed with the \textit{ab initio} envelope in Fig. 8B.

The crystal dimer has two-fold symmetry, which is inconsistent with binding to both parts of a DNA direct repeat, but could conceivably bind to both elements of a DNA inverted repeat. This agrees with recent work on the SenX3-RegX3 system in \textit{M. smegmatis} (11). Note that the crystal dimer fails to provide a reasonable fit to the SAXS data (\(\chi = 9.6\)) (Fig. 9A).

The scattering data from 4.8 mg ml\(^{-1}\) solutions of RegX3 yield a molecular mass of 22.2 ± 2.2 kDa, consistent with a monomer (24.8 kDa). Moreover, the scattering curve at this concentration agrees well (\(\chi = 0.748\)) with the computed scattering pattern of DrrD (22) (Fig. 9A). DrrD is a monomeric RR but has a more compact shape when compared with the structure of a single protomer from the crystallographic RegX3 dimer, which does not fit the SAXS data. RegX3, although existing as a monomer at low concentrations, is driven into a dimeric state with increasing concentration. No higher oligomers are present, and the scattering at intermediate concentrations can be fitted well by linear combinations of monomeric and dimeric species (Table 3). The overall shape determined from the SAXS measurements is dominated by the more accurately measured lower resolution data, and the deviations at higher angles (\(s > 0.3 \text{ Å}^{-1}\), corresponding to resolutions better than 20 Å) between the experimental data and some of the fits in Fig. 9A may be explained by small scale conformational flexibility or possibly by detailed differences between the monomeric DrrD

The Crystal Structure of RegX3

Concentration measured using different programs (DAMMIN and GASBOR) with and without two-fold symmetry imposed. The overall shape in solution, derived by GASBOR with two-fold symmetry applied, appears linear in a “trans-shaped” dimer (Fig. 8B) compared with the “cis-shaped” crystal dimer (Fig. 8A). The elongated trans-shape is strikingly similar to the proposed model of the inactive full-length PhoB structure (29), where the positioning of the effector domains in opposite directions provides an additional means of inhibition. We then used, as prior knowledge, the inactive PhoB dimerization interface contacts and the full-length RR DrrD for rigid body modeling against the 15.3 mg ml\(^{-1}\) RegX3 SAXS data. Several runs of the program SASREF, assuming two-fold symmetry, produced consistent results and a good fit (\(\chi = 2.05\)) to the scattering data (Fig. 9). This model is superimposed with the \textit{ab initio} envelope in Fig. 8B.

The crystal dimer has two-fold symmetry, which is inconsistent with binding to both parts of a DNA direct repeat, but could conceivably bind to both elements of a DNA inverted repeat. This agrees with recent work on the SenX3-RegX3 system in \textit{M. smegmatis} (11). Note that the crystal dimer fails to provide a reasonable fit to the SAXS data (\(\chi = 9.6\)) (Fig. 9A).

The scattering data from 4.8 mg ml\(^{-1}\) solutions of RegX3 yield a molecular mass of 22.2 ± 2.2 kDa, consistent with a monomer (24.8 kDa). Moreover, the scattering curve at this concentration agrees well (\(\chi = 0.748\)) with the computed scattering pattern of DrrD (22) (Fig. 9A). DrrD is a monomeric RR but has a more compact shape when compared with the structure of a single protomer from the crystallographic RegX3 dimer, which does not fit the SAXS data. RegX3, although existing as a monomer at low concentrations, is driven into a dimeric state with increasing concentration. No higher oligomers are present, and the scattering at intermediate concentrations can be fitted well by linear combinations of monomeric and dimeric species (Table 3). The overall shape determined from the SAXS measurements is dominated by the more accurately measured lower resolution data, and the deviations at higher angles (\(s > 0.3 \text{ Å}^{-1}\), corresponding to resolutions better than 20 Å) between the experimental data and some of the fits in Fig. 9A may be explained by small scale conformational flexibility or possibly by detailed differences between the monomeric DrrD

The Crystal Structure of RegX3

Concentration measured using different programs (DAMMIN and GASBOR) with and without two-fold symmetry imposed. The overall shape in solution, derived by GASBOR with two-fold symmetry applied, appears linear in a “trans-shaped” dimer (Fig. 8B) compared with the “cis-shaped” crystal dimer (Fig. 8A). The elongated trans-shape is strikingly similar to the proposed model of the inactive full-length PhoB structure (29), where the positioning of the effector domains in opposite directions provides an additional means of inhibition. We then used, as prior knowledge, the inactive PhoB dimerization interface contacts and the full-length RR DrrD for rigid body modeling against the 15.3 mg ml\(^{-1}\) RegX3 SAXS data. Several runs of the program SASREF, assuming two-fold symmetry, produced consistent results and a good fit (\(\chi = 2.05\)) to the scattering data (Fig. 9). This model is superimposed with the \textit{ab initio} envelope in Fig. 8B.

The crystal dimer has two-fold symmetry, which is inconsistent with binding to both parts of a DNA direct repeat, but could conceivably bind to both elements of a DNA inverted repeat. This agrees with recent work on the SenX3-RegX3 system in \textit{M. smegmatis} (11). Note that the crystal dimer fails to provide a reasonable fit to the SAXS data (\(\chi = 9.6\)) (Fig. 9A).

The scattering data from 4.8 mg ml\(^{-1}\) solutions of RegX3 yield a molecular mass of 22.2 ± 2.2 kDa, consistent with a monomer (24.8 kDa). Moreover, the scattering curve at this concentration agrees well (\(\chi = 0.748\)) with the computed scattering pattern of DrrD (22) (Fig. 9A). DrrD is a monomeric RR but has a more compact shape when compared with the structure of a single protomer from the crystallographic RegX3 dimer, which does not fit the SAXS data. RegX3, although existing as a monomer at low concentrations, is driven into a dimeric state with increasing concentration. No higher oligomers are present, and the scattering at intermediate concentrations can be fitted well by linear combinations of monomeric and dimeric species (Table 3). The overall shape determined from the SAXS measurements is dominated by the more accurately measured lower resolution data, and the deviations at higher angles (\(s > 0.3 \text{ Å}^{-1}\), corresponding to resolutions better than 20 Å) between the experimental data and some of the fits in Fig. 9A may be explained by small scale conformational flexibility or possibly by detailed differences between the monomeric DrrD
The Crystal Structure of RegX3

crystal structure and the RegX3 solution structure. In Table 3, the concentration-dependent equilibrium is quantitatively characterized in terms of the volume fractions of monomers and dimers using the program OLIGOMER. We note that the observed monomer-dimer equilibrium is fully reversible. If a high concentration sample is diluted, the volume fraction of monomers increases accordingly.

DISCUSSION

The crystallization of multidomain RRs has proven difficult, probably due to conformation variability. Based on information from the full-length structures that have been determined, it is evident that (at least) two distinct structural subclasses exist. These subclasses are based upon the accessibility of the DNA recognition helix. The structures of DrrD, DrrB, and now RegX3 all exhibit accessible recognition helices and hence belong to the “open” structural subclass (Fig. 7A). PrrA and MtrA, on the other hand, are very compact structures with inaccessible recognition helices and therefore belong to the “closed” subclass (Fig. 7B). To date, all of the known full-length RRs have been structurally characterized in an inactive state, and the only structural information about the activation mechanism has come from isolated activated receiver domains. In this paper, we have presented the first full-length structure of a response regulator that not only supports the dimerization hypothesis proposed by Toro-Roman et al. (56) but also exhibits three-dimensional domain swapping. Although crystals were not grown in the presence of any activating substances, such as BeF3, we believe, based upon the position of a number of important residues surrounding the active site, the length of helix α4, and the existence of a dimer, that the structure we observe represents RegX3 driven to an active state by the presence of lanthanum ions.

Three-dimensional Domain Swapping—The formation of a three-dimensional domain-swapped dimer requires the breaking of many interactions within the monomer in order for the domains to be swapped. In the case of RegX3, it also requires the partial unfolding of the receiver domain, which must be a process with a relatively high energy of activation. This leads to speculation as to what conditions would be conducive to the formation of a three-dimensional domain-swapped dimer. The majority of three-dimensional domain-swapped dimers in the past have been formed in vitro at high protein concentrations under environmental conditions that favor unfolding. These conditions exist in many crystallization experiments and are a likely reason for the observation of this and other three-dimensional domain-swapped dimers. In RegX3, the partial unfolding may be due to the LaCl3 present in the crystallization conditions, which is subsequently stabilized in the crystal lattice. The in vivo conditions that lead to three-dimensional domain swapping are still largely unknown.
Changes in the length of α4 have previously been associated with the activation of the receiver domain. In the active form of PhoB, for instance, the helix extends from Glu⁸⁷ to Thr⁹⁷, as opposed to Gly⁶⁶ to Gly⁹⁸ in the inactive form. Consequently, it adopts a position closer to the other activated receiver domains (29).

This helix in all receiver domain structures, apart from DrrD and RegX3, is positioned parallel to the central five-stranded β-sheets. It is proposed that the helix packs alongside α5 in the inactive state, thereby shielding the hydrophobic face of the β-sheets. In the active state, the helix extends, generating a rotational shift that exposes a hydrophobic face. This newly exposed hydrophobic face forms the active dimerization interface.

In RegX3, the role of α4 is slightly different. Although the helix has the same length as an active helix (12 residues), it is mainly involved as a linker between the two receiver domains formed about the crystallographic two-fold axis. Its hydrophobic face packs against the hydrophobic face of the other protomer’s α4. These differences lead to an alternative dimerization interface, which is evident when the composite receiver domain is superimposed upon other “activated” receiver domains (Fig. 7). The dimerization interface is significantly larger than in other (isolated) active receiver domains due to the existence of two subinterfaces. The α4 helix from RegX3 is positioned at an angle of ~70° from the face of the central β-sheets of the same protomer such that the N-terminus of this helix donates a residue (Glu⁸⁹) to the active site of that protomer (Fig. 2). The only other RR with an α4 helix oriented differently is DrrD, and this is reported to be due to the crystal packing (22).

Switch Residues—The key switch residues in RegX3, as mentioned previously, are Thr⁷⁵ and Tyr⁹⁶. Although structures of receiver domains alone in an activated state exist, such as those of PhoB (29), ArcA (4), KdpE (56), TorR (56), and MicA (60), RegX3 represents the first full-length RR structure from the OmpR/PhoB family in what we believe represents an activated state. RegX3 is forced into an active state through its interaction with one of the La³⁺ ions. La-1 is coordinated by the carbonyl oxygen atom of Tyr⁹⁶, twisting the side chain of the residue into the active state position. The repositioning of Tyr⁹⁶ forces Thr⁷⁸ to shift inward also into an active state position. This mechanism of a geometric switch is in the reverse order to the normally proposed activation mechanism, where the threonine residue shifts in response to phosphorylation followed by a shift in the orientation of the tyrosine residue.

In other active receiver domains, it has been noted that the tyrosine, in its active position, coordinates through the side chain’s hydroxyl group to a carbonyl oxygen atom of an arginine or lysine residue. This coordination is proposed to stabilize the active site (29). In RegX3, Tyr⁹⁶ is oriented so that the side-chain hydroxy group points toward the carbonyl oxygen atom of Arg⁸¹ from the dimeric partner, located toward the N-terminus of α4. However, the distance (3.9 Å) between the two residues is too large for direct hydrogen bonding. Instead, the hydroxyl group hydrogen bonds a water molecule, which in turn hydrogen bonds with Thr⁷⁸ from β5 of the same protomer.

**TABLE 3**
The oligomeric ratio for the intermediate concentrations of RegX3 in solution measured by SAXS and the discrepancy (χ) of the Oligomer fits to the data

| Concentration | χ    | Monomer fraction | Dimer fraction |
|---------------|------|-----------------|---------------|
| mg ml⁻¹       | %    | %               | %             |
| 4.8           | 0.79 | 58 ± 2          | 42 ± 1        |
| 8.0           | 1.16 | 53 ± 2          | 47 ± 1        |
| 11.3          | 1.79 | 30 ± 1          | 70 ± 1        |

RRs are DNA-binding proteins, and it is possible that the domain-swapped dimer represents a form able to recognize alternative regulons. Three-dimensional domain swapping has previously been observed in the RR Spo0A (Protein Data Bank code 1DZ3) (57), albeit only for the isolated receiver domain. The result of the domain swapping, however, was a nonfunctional receiver domain with an incomplete active site. In contrast, RegX3 appears to be structurally complete. This observation is based upon the formation of a complete composite receiver domain, a complete active site with all of the residues necessary to stabilize a phosphate group, the formation of the dimer on a highly conserved interface that is strongly linked to dimer formation, and the presentation of the DNA recognition helices in an orientation capable of interacting with DNA.

As a result of the three-dimensional domain swapping, two new functional units are created, each consisting of a partly domain-swapped, composite receiver domain and an effector domain. The buried surface area per monomer consists of two parts of roughly equal size and is much greater than that seen in other full-length RRs or in isolated receiver domains. One interface region, the dimerization interface, consists of the elements α4 and β5, which are exchanged about the two-fold crystallographic axis facilitated by the orientation of α4. Strand β5 is tightly associated with the domain-swapped four-stranded β-sheet from the other protomer, forming an extended five-stranded β-sheet. The second interface region, mediated by α5, forms the interdomain interface between the composite receiver domain and the effector domain.

Dimerization Interface—Although three-dimensional domain swapping upon activation may not occur in other OmpR/PhoB subfamily members, the fact that the same structural elements (α4, β5, and α5) are also involved in the formation of dimeric interfaces in non-domain-swapped active receiver domains reflects a similar primary structure within the subfamily.

Multiple sequence alignments of members of the OmpR/PhoB subfamily as well as other RR families reveal that the key residues involved in the α4β5α5 interface interactions are highly conserved only within this subfamily (4). This suggests that the α4, β5, and α5 are of specific importance to the OmpR/PhoB subfamily.

The largest variation seen when aligning the composite receiver domain with other characterized receiver domains is in the orientation of α4. Its unique orientation positions β5 and α5 near the protomer’s incomplete receiver domain, allowing the formation of the composite receiver domain.

**Helix α4**—Helix α4 has previously been shown to be partially unfolded in some RRs, such as CheY (58) and NtrC (59).
The Crystal Structure of RegX3

In the inactivated full-length RR structures, PrrA, MtrA, and DrrB, the switch tyrosine is oriented away from the receiver domain in an inactive state and forms hydrogen bonds with the effector domain. In the “closed” models (PrrA and MtrA), these bonds occur with the loop region between α7 and α8 (the DNA recognition helix). In the “open” model, represented by DrrB, the interaction occurs with the β7β8-loop from the β-scaffold. DrrD is an exception, since the orientation of the switch tyrosine is apparently influenced by the effect of crystal packing on α4 (22). It has been proposed that the inward orientation of the tyrosine would destabilize the interdomain interface between the receiver and effector domains, thereby activating the RR by exposing the DNA recognition helix in preparation for DNA binding (21). The idea of dimerization is strongly attached to RRs of the “open” subclass. This type of RR does not have the option of opening up upon activation, since they are apparently already open and in the case of RegX3 already bound to DNA in an inactive state. Dimerization may allow the recognition of other DNA consensus sequences, and although this speculation still requires investigation, it is consistent with the report of an inverted repeat present in three RegX3-regulated operons in M. smegmatis (11). The RegX3 structure would support this idea, because the switch residue is distant from the interdomain interface and is positioned on the dimerization interface (Fig. 6A). This suggests that the reorientation of the switch tyrosine is associated with a destabilization of the interdomain interface, leading to a reorientation of the effector domain and the creation of the dimerization interface (where the tyrosine is now positioned).

Interdomain Interface—All of the current structures of inactive full-length OmpR/PhoB members exhibit different interdomain interfaces formed by some of the secondary structure elements α4, β5, and α5. The type of interdomain interface appears to be dependent on the subclass of RR. For the “closed” subclass, the interdomain interface involves all three components of the α4β5α5 surface. However, members of the “open” class consisting of RegX3, DrrD, and DrrB form interfaces on only a single component. In DrrB, the interface involves α4, whereas in DrrD and RegX3 the interface involves only α5 (Fig. 7A). Interestingly, although either domain from all of the known full-length OmpR/PhoB RRs can be superimposed individually on one another with small r.m.s. deviations, when the full-length proteins are superimposed, it becomes apparent that the two domains have very different spatial orientations (Fig. 7A). In some instances, of course, this could result from crystal packing, although in the case of (inactivated) PrrA, the crystal packing does not determine the interdomain interface (21).

The most noticeable difference between RegX3 and the other “open” inactive RRs is the orientation of the effector domain. Although all of them have an accessible α8 DNA recognition helix, DrrD and DrrB mediate the interdomain interface through the β-scaffold. In the effector domains of the active, three-dimensional domain-swapped RegX3 dimer, the β-scaffold does not interact with the interdomain interface but instead is orientated toward the side in a similar position to the orientation of the scaffold in the “closed” subclass, where the whole effector domain is rotated so as to bury α8. The different orientation of the β-scaffold in relation to the interdomain interface in the RegX3 structure may represent the active state for the “open” subclass of RR, allowing the two α8 helices from the RR dimer to correctly interact with DNA.

The Inactive Dimer—Inactive dimers of OmpR/PhoB response regulators have been reported in the past. Although their function remains largely unknown, it is assumed that they serve an inhibitory function. The dimer in solution was detected by SAXS at a high concentration (15.3 mg/ml), suggesting that it is a low affinity dimer. Although this concentration may appear unrealistic inside a cell, it may be influenced through intracellular macromolecular crowding. As suggested by Bachhawat et al. (29), the inactive dimer may play a role in regulation. Initially, RegX3 exists as an inactive monomer. When activated, it autoregulates itself, and expression increases until there is an abundance of RegX3 in vivo, whereupon it dimerizes, allowing other regulatory functions. It is possible that when the stress signal (possibly low environmental phosphate concentration (11)) that the SenX3-RegX3 system is responding to subsides and the active kinase is no longer present, the active RegX3 dimers, which are at high concentration, switch to an inactive dimeric form, thus providing a quick attenuation of the response.

The formation of an inactive dimer may inhibit the RR activities in many ways, such as disruption of DNA binding by orienting the effector domains in opposite directions. The dimer may also serve to disrupt or enhance the interaction interface with the SK, or the new dimer may facilitate rapid dephosphorylation by further exposing the active site containing the unstable phosphoaspartate.

We also tried to fit the experimental scattering data from the dimeric species by keeping the three-dimensional domain-swapped interface and allowing for the rigid body movements of the peripheral portions of the effector domain. This modeling was able to give extended trans-shapes, which could almost fit the experimental data as well as the inactive PhoB-like dimer shown in Fig. 8B. The latter dimer, however, seems more likely, since the monomer dimer equilibrium is reversible.

Conclusions—Based upon the position of the switch residues and the dimerization interface, we believe that the RegX3 structure represents the RR in an active state. The cis-dimer seen in the crystal structure is the first full-length OmpR/PhoB member RR to be crystallized as a dimer in an apparently active state. The RR dimer formed by three-dimensional domain swapping retains a complete active site as is necessary for phosphorylation. This structure provides the first detailed analysis of the interdomain interface for an active dimeric RR. However, SAXS measurements show that the orientation of the effector domains may not be representative of the dimer in vivo in an inactive state. The SAXS studies revealed that the inactive protein exists as a monomer and a dimer in solution. The inactive trans-dimer formed in solution provides the first structural evidence for a previously hypothesized regulatory mechanism. A structural comparison of the active and inactive RegX3 bound to DNA would provide answers regarding conformational changes of the effector domain and interdomain interface.
