JOURNAL OF BACTERIOLOGY, July 2003, p. 4087–4098 Vol. 185, No. 14
0021-9193/03/$08.00+0 DOI: 10.1128/JB.185.14.4087–4098.2003

The Purine Repressor of *Bacillus subtilis*: a Novel Combination of Domains Adapted for Transcription Regulation

Sangita C. Sinha,[1]† Joseph Krahn,[1]‡ Byung Sik Shin,[2] Diana R. Tomchick,[1]§ Howard Zalkin,[2] and Janet L. Smith[1]*

Departments of Biological Sciences[1] and Biochemistry,[2] Purdue University, West Lafayette, Indiana 47907

Received 31 January 2003/Accepted 29 April 2003

The purine repressor from *Bacillus subtilis*, PurR, represses transcription from a number of genes with functions in the synthesis, transport, and metabolism of purines. The 2.2-Å crystal structure of PurR reveals a two-domain protein organized as a dimer. The larger C-terminal domain belongs to the PRT structural family, in accord with a sequence motif for binding the inducer phosphoribosylpyrophosphate (PRPP). The PRT domain is fused to a smaller N-terminal domain that belongs to the winged-helix family of DNA binding proteins. A positively charged surface on the winged-helix domain likely binds specific DNA sequences in the recognition site. A second positively charged surface surrounds the PRPP site at the opposite end of the PurR dimer. Conserved amino acids in the sequences of PurR homologs in 21 gram-positive bacteria cluster on the proposed recognition surface of the winged-helix domain and around the PRPP binding site at the opposite end of the molecule, supporting a common function of DNA and PRPP binding for all of the proteins. The structure supports a binding mechanism in which extended regions of DNA interact with extensive protein surface. Unlike most PRT proteins, which are phosphoribosyltransferases (PRTases), PurR lacks catalytic activity. This is explained by a tyrosine side chain that blocks the site for a nucleophile cosubstrate in PRTases. Thus, *B. subtilis* has adapted an enzyme fold to serve as an effector-binding domain and has used it in a novel combination with the DNA-binding winged-helix domain as a repressor of purine genes.

Regulation of transcription initiation of several *Bacillus subtilis* genes with functions in purine synthesis, transport, and metabolism requires the purine repressor, PurR, and specific DNA sequences, PurBoxes, in the upstream control regions of affected genes. The PurR-PurBox system regulates transcription of all genes encoding enzymes for synthesis of IMP from the starting material α-D-5-phosphoribosyl-1-pyrophosphate (PRPP) and for synthesis of AMP from IMP (10, 26, 38). The two-step pathway from IMP to GMP is separately regulated. The PurR-PurBox system also regulates transcription of the *purR* operon (*purR* and *yabJ*) (38) and of several genes involved in cofactor biosynthesis (*glyA* and *folD*), purine salvage (*guaC* and *xpt*), and purine transport (*pbuG*, *pbuO*, and *pbuX*) (28).

B. subtilis PurR is a 285-residue, homodimeric protein (38). The PurR sequence includes a 13-residue PRPP-binding motif (Fig. 1) that is characteristic of the PRT protein family (32, 38). Most PRT family members are phosphoribosyltransferases (PRTases) with a PRPP substrate, but no catalytic activity has been detected for PurR (H. Zalkin, unpublished data). Precedent for adaptation of the PRT fold to a regulatory function occurs in PyrR, a transcription attenuator for pyrimidine biosynthetic genes in *B. subtilis* (35, 36). PurR is about 70 amino acids longer at the N terminus than most PRT proteins. This region is a proposed DNA interaction domain (38), although at the sequence level it is not obviously related to any other DNA-binding protein.

PurR binds specifically to an extended region of DNA upstream of the transcription start sites of *purA* and of the *pur*, *purR*, and *pyr* operons (4, 31, 38). The control site for PurR regulation is a 14-nucleotide PurBox with the sequence 5′-AWWWCCGAACWWTW-3′ (where W represents A or T) (15, 28). In *B. subtilis*, PurR regulates transcription from genes whose upstream regions contain an imperfect, inverted repeat of PurBoxes separated by 16 to 17 nucleotides (28, 31). Extensive biochemical studies with purified PurR demonstrated that DNase I footprints exhibit an alternating pattern of protected and hypersensitive sites extending over more than 60 base pairs, that PurR introduces right-handed supercoils in the DNA, and that two to six PurR dimers bind cooperatively to *pur* control site DNA. These results led to the suggestion that an extended region of control site DNA wraps around PurR (31). A GAAC sequence in the conserved region of the PurBox, 75 base pairs upstream of the *pur* operon, is essential for PurR binding, but the same sequence in the inverted repeat is not essential.

PRPP appears to be the inducer of genes regulated by PurR because it is the only molecule among many nucleobases, nucleosides, and nucleotides known to affect PurR-DNA binding in vitro (38). This is consistent with the PRPP-binding sequence motif in PurR. In vivo, adenine affects PurR-mediated repression of transcription initiation (9). Excess adenine signals are thought to be transmitted to PurR via the cellular PRPP pool (38).

* Corresponding author. Mailing address: Department of Biological Sciences, Purdue University, 915 W. State St., West Lafayette, IN 47907. Phone: (765) 494-9246. Fax: (765) 496-1189. E-mail: smithj@purdue.edu.

† Present address: Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, Dallas, TX 75390.

‡ Present address: National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709.

§ Present address: Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, TX 75390.

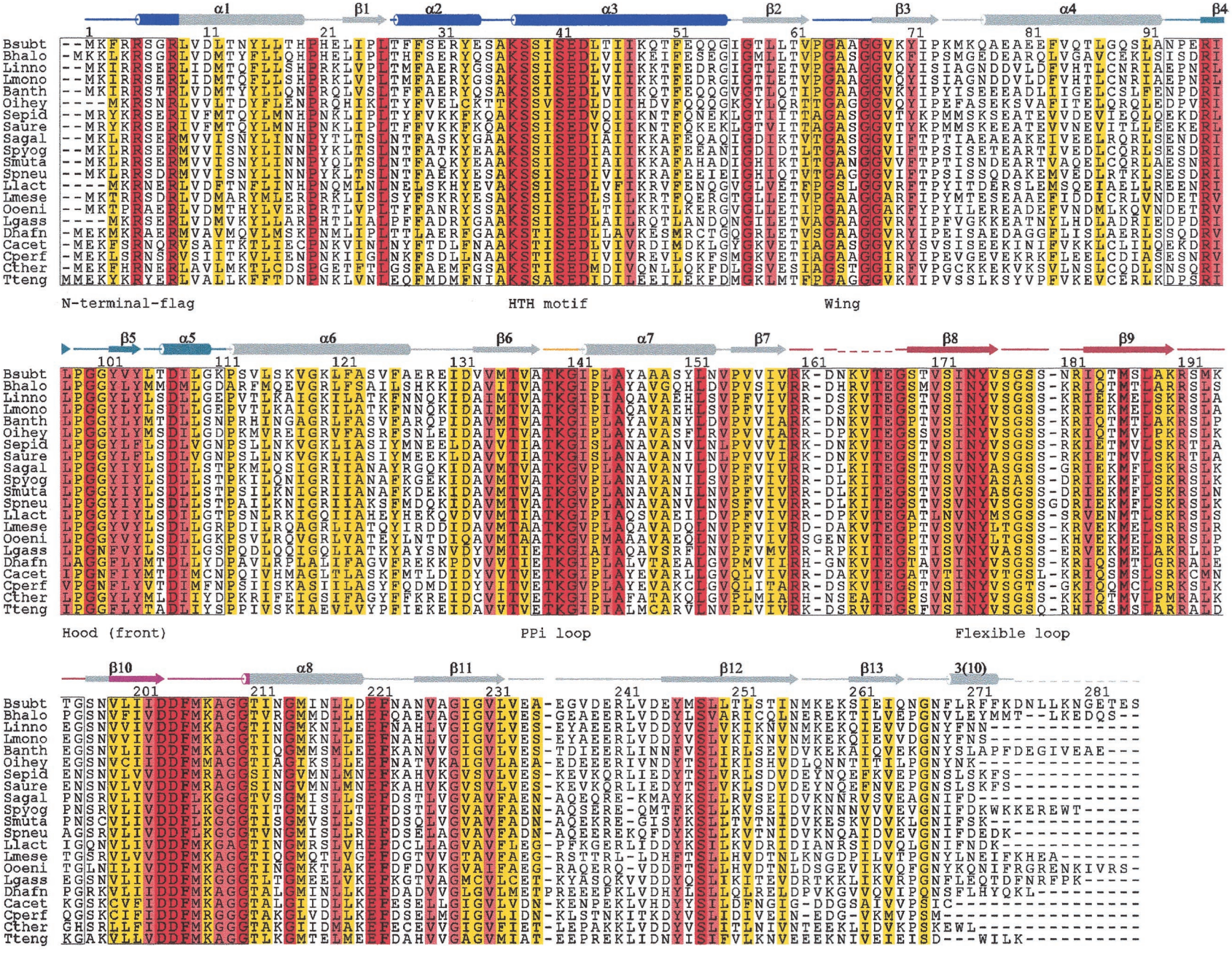

α1 β1 α2 α3 β2 β3 α4 β4
N-terminal-flag
HTH motif
Wing
β5 α5 α6 β6 α7 β7 β8 β9
Hood (front)
PPi loop
Flexible loop
β10 α8 β11 β12 β13 3(10)
Bsubt
Bhalo
Linno
Lmono
Banth
Oihey
Sepid
Saure
Sagal
Spyog
Smuta
Spneu
Llact
Lmese
Ooeni
Lgass
Dhafn
Cacet
Cperf
Cther
Tteng

FIG. 1. Alignment of sequences with homology to both domains of *B. subtilis* PurR. Residues numbers above the alignment correspond to *B. subtilis* PurR. Biological sources and accession codes for the sequences are: Bsubt, *B. subtilis* (gi586880); Bhalo, *B. halodurans* (gi15612625); Linno, *Listeria innocua* (gi16799308); Lmono, *Listeria monocytogenes* (gi16802230); Banth, *B. anthracis* (gi21398009); Oihey, *Oceanobacillus iheyensis* (gi23097511); Sepid, *Staphylococcus epidermidis* (gi16024897); Saure, *Staphylococcus aureus* (gi15923486); Sagal, *Streptococcus agalactiae* (gi22537911); Spyog, *Streptococcus pyogenes* (gi15674448); Smuta, *Streptococcus mutans* (gi24378853); Spneu, *Streptococcus pneumoniae* (gi15901802); Llact, *Lactococcus lactis* (gi15674241); Lmese, *Leuconostoc mesenteroides* (gi23024251); Ooeni, *Oenococcus oeni* (gi23037595); Lgass, *Lactobacillus gasseri* (gi23003426); Dhafn, *Desulfitobacterium hafniense* (gi23111814); Cacet, *Clostridium acetobutylicum* (gi15896471); Cperf, *C. perfringens* (gi18311474); Cther, *C. thermocellum* (gi23020712); and Tteng, *Thermoanaerobacter tengcongensis* (gi20808923). Secondary structures are indicated above the sequence alignment (cylinders for helices, arrows for β strands, lines for coils, and dashed lines for disordered regions missing from the structure). Background colors reflect increasing similarity among the sequences, with similar residues highlighted in yellow, sites of conservative substitution in pink, and invariant residues in red. Regions of functional importance are labeled, and the secondary structures are colored as in Fig. 2.

Homologs having 40 to 65% sequence identity with *B. subtilis* PurR occur in 20 other gram-positive bacteria (Fig. 1). However, functional studies have been reported for only the *Lactococcus lactis* homolog, which also regulates transcription of purine genes. *L. lactis* PurR activates transcription of *purC* and *purDEK* (15). PurBoxes located 76 nucleotides upstream of *purC* and *purD* are required for transcription activation, with the conserved G in the central GAAC sequence being essential. A second, tandem PurBox is located 93 nucleotides upstream of the *purC* start site but is not required for transcription regulation. PurR appears to autorepress transcription of *purR* in *L. lactis* (16).

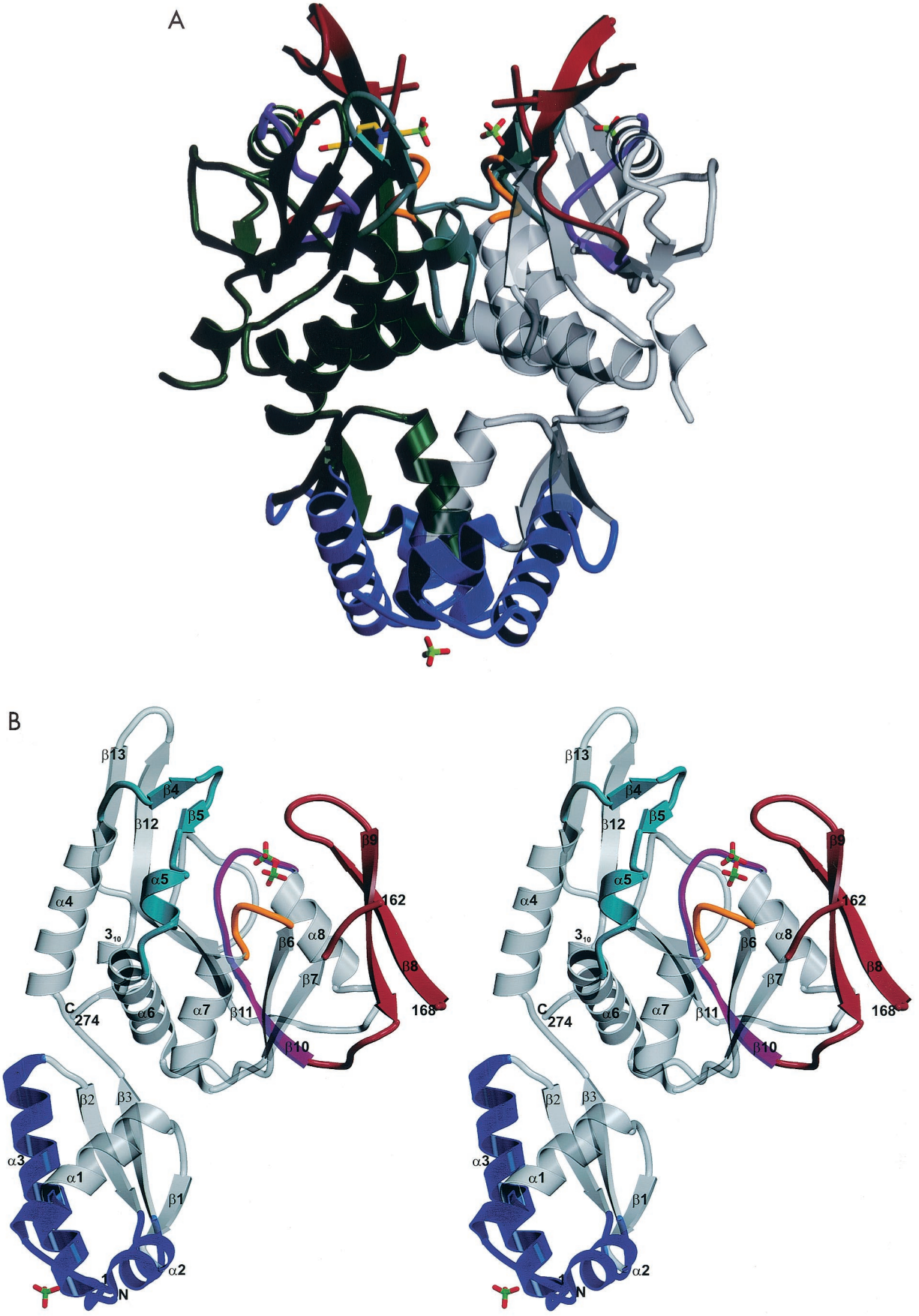
B. subtilis PurR is unrelated to the more familiar purine repressor from *Escherichia coli*, also called PurR. *E. coli* PurR is a member of the LacI family and binds to a well-defined 16-base-pair DNA palindrome via an N-terminal helix-turn-helix (HTH) domain. Hypoxanthine and guanine are corepressors that increase the DNA affinity of *E. coli* PurR. *B. subtilis* and *E. coli* PurR proteins belong to different protein families, bind different DNA sequences, and regulate transcription by different mechanisms. With its PRT domain, *B. subtilis* PurR represents a new protein family among bacterial repressors that is thus far limited to gram-positive eubacteria.

Here we report the 2.2-Å crystal structure of *B. subtilis* PurR. Surfaces involved in DNA and PRPP binding are apparent in the dimer of winged-helix and PRT domains. The structure provides a foundation for understanding the mechanism of transcription regulation by PurR (4).

MATERIALS AND METHODS

Crystallization. Purification of His-tagged selenomethionyl (SeMet) PurR from *B. subtilis* by Ni-chelate affinity chromatography was identical to that for the native protein (31), except that all buffers contained 5 mM dithiothreitol to prevent Se oxidation. Wild-type PurR was crystallized at 20°C by hanging-drop vapor diffusion from a 1:1 mixture of protein solution [10 mg of PurR per ml, 10 mM HEPES (pH 8.0), 50 mM $(NH_4)_2SO_4$, 300 mM NaCl] and well solution (5% polyethylene glycol [PEG] 8000, 66.7 mM HEPES [pH 7.0], 167 mM Li_2SO_4). SeMet PurR was crystallized under similar conditions from a 1:1 mixture of protein solution [10 mg of PurR per ml, 10 mM HEPES (pH 8.0), 50 mM $(NH_4)_2SO_4$, 300 mM NaCl, 5 mM dithiothreitol] and well solution (5% PEG 8000, 83.3 mM HEPES [pH 7.0], 450 mM Li_2SO_4), which was equilibrated for 3 days and then streak seeded with crystals of wild-type PurR. Platelike crystals grew to an average size of 0.2 by 0.4 by 0.05 mm in approximately 1 month.

Data collection. The crystals were harvested in 200 μl of well solution and cryoprotected by five successive exchanges of 50 μl of harvesting solution with an equal volume of cryosolution (20% PEG 400, 10% PEG 8000, 200 mM NaCl, 500 mM Li_2SO_4, 51 mM HEPES [pH 7.0]) every 30 s. The crystals were immediately flash frozen in a gaseous N_2 stream at 100 K. Frozen crystals had a high mosaicity (~1.3°), which was reduced to ~0.55° by cryoannealing. The crystal was thawed by immersion in the final cryoprotectant solution for 2 min and flash frozen again in a gaseous N_2 stream at 100 K. All data used to solve the structure were collected from a single SeMet PurR crystal. Multiwavelength anomalous diffraction (MAD) data were recorded at beamline BM-14 at the European Synchrotron Radiation Facility (ESRF). X-ray fluorescence spectra from a SeMet crystal were recorded from 12.65 to 12.68 keV to select wavelengths for data collection. Due to beam time limitations, data sets for only two wavelengths at the Se K edge (249° at the inflection point, λ_1, 12.6645 keV, 0.9790 Å; and 237° at the peak, λ_2, 12.6671 keV, 0.9788 Å) were recorded at the ESRF on a Mar 345 imaging plate detector with 60-s exposures over 1° of crystal rotation per image. Several days later, data were collected from the same crystal (326.4° at the remote, λ_3, 8.0416 keV, 1.5418 Å) using CuKα radiation and an R-axis IIc imaging plate with 30-min exposures over 0.6° of crystal rotation per image. The useful data were limited to 294°, as after 8 days of data collection, diffraction quality deteriorated dramatically due to crystal decay and severe ice formation. Data were processed with the HKL package (24). Unit cell parameters in space group P1 are *a*, 65.1 Å; *b*, 72.2 Å; *c*, 83.0 Å; α, 84.8°; β, 84.0°; and γ, 67.5° with four copies (V_m = 2.8


A
B
β13
β4
β12
β5
α5
α4
3₁₀
β9
162
α8
β6
β8
β7
C
274
α6
α7
β11
168
β10
β2
β3
α3
α1
β1
α2
1
N

$Å^3$/Da, ~57% solvent) of the PurR polypeptide per asymmetric unit. The MAD data set used for phasing was produced by scaling the λ_2 and λ_3 data sets to the merged data for λ_1 by using SCALEIT from the CCP4 suite (7). Data quality is summarized in Table 1.

MAD structure determination. The PurR monomer has nine methionine residues, including the N-terminal methionine. Thus, up to 36 Se sites were expected in each asymmetric unit. Twenty-eight sites, arranged in four clusters, were located (18). Refinement of the Se partial structure and MAD phasing were performed by the pseudoisomorphous replacement approach using MLPHARE (23). Noncrystallographic symmetry operators were determined by superposition of four clusters of Se sites. Phases were refined to 2.7 Å by fourfold noncrystallographic symmetry averaging, histogram matching, and solvent flattening with a 32% solvent mask by using density modification (8) and the amplitudes from the λ_1 data set.

Model building and refinement. The model was built into a 2.7-Å electron density map by using the program O (14). The initial model was 83.5% complete and had an R_{work} value of 0.483 and an R_{free} value of 0.477 for all data between 20.0 and 2.4 Å. For two iterations of refinement, the model was refined against the 2.4-Å λ_1 data set with maximum-likelihood amplitude targets by using the Crystallography and NMR system (5). Subsequently, the model was refined with maximum-likelihood amplitude and phase probability targets in the Crystallography and NMR system against a wavelength-combined 2.2-Å data set ($\lambda_1+\lambda_2+\lambda_3$, Table 1). This data set, generated by merging data at λ_1 (20 to 2.2 Å), λ_2 (20 to 2.2 Å), and λ_3 (30 to 3.1 Å), was more complete at low resolution and more redundant overall. Numerous iterations of manual rebuilding and simulated annealing, using models with different omitted regions, were required to complete the model. The model of the protein core was restrained by fourfold noncrystallographic symmetry during all cycles of refinement except the last three. Atomic occupancies of the sulfate ions, HEPES molecules, and protein residues in dual positions were each refined once. Final R_{work} and R_{free} values for all data between 30.0 Å and 2.2 Å were 0.188 and 0.237, respectively. The final refined model consists of two PurR dimers, AB and CD, related by an approximate twofold screw axis, with 270 residues in monomers A and D and 269 residues in monomers B and C, eight sulfate ions, two HEPES molecules, six fragments of PEG, and 568 water molecules. Model quality is summarized in Table 2. This model has been deposited in the Protein Data Bank and is available with accession code 1p41.

Sequence and structure analysis. Homologs of PurR were identified in the sequence database using BLAST (2); multiple sequence alignment was done with CLUSTAL W (34); and structural similarity searches were done with DALI (13). Electrostatic surface potentials were calculated with GRASP (22). Structure superpositions were done with program O (14). Figure 1 was drawn with ALSCRIPT (3), Fig. 2, 3, 5, and 6 were drawn with MOLSCRIPT (19) and Raster 3D (21), and Fig. 4 was drawn with GRASP.

RESULTS AND DISCUSSION

Structure determination. The crystal structure of *B. subtilis* PurR was determined by SeMet MAD. PurR crystallized as a dimer (Fig. 2A), which is its quaternary state in solution (38). The final model, refined against complete 2.2-Å data, includes residues 2 to 162 and 168 to 274 for all four monomers in the asymmetric unit. In addition, monomers A and D include residues 275 to 276, and monomers B and C include the N-terminal methionine, which is involved in a crystal lattice contact. Residues 277 to 285 and the hexa-histidine tag at the C terminus of the polypeptide are disordered in all monomers. The model agrees well with the diffraction data (R_{work}, 0.188 and R_{free}, 0.237) and with stereochemical criteria ($RMSD_{bonds}$, 0.010). No residues are in forbidden regions of the Ramachandran plot. The estimated coordinate error of the final model is 0.24 Å (27).

Protein structure. The PurR polypeptide folds into two domains connected by a short three-residue linker (Fig. 2B). The domain interface buries 541 $Å^2$ of surface area in each monomer. The N-terminal domain comprises residues 1 to 74, and the C-terminal domain includes residues 77 to 285 (Fig. 3).

The N-terminal domain is a winged-helix domain, a subdivision of the HTH structural family (Fig. 3A). An HTH domain had not been predicted from the PurR sequence. The nearest structural neighbor is the winged-helix domain of the biotin operon repressor BirA (39), with an RMSD value of 1.5 Å over 60 C_α positions and 16% sequence identity (Fig. 3A). The PurR winged-helix domain consists of the canonical arrangement of secondary structures: α1-β1-α2-T-α3-β2-W-β3, where α2-T-α3 is the HTH motif, α3 is the recognition helix, and W is the wing (Fig. 1 and 2B). Most winged-helix proteins have a second wing following β3, but in PurR, a three-residue linker in this position connects to the C-terminal domain.

The C-terminal domain of *B. subtilis* PurR has a 13-residue sequence motif for PRPP binding, which is characteristic of the PRT family (Fig. 1) (32). As expected, this domain has the PRT fold consisting of a central parallel β sheet flanked by α helices (Fig. 2B). Secondary structures at the N and C termini of the PRT domain form a hood over the C-terminal edge of the central β sheet, as occurs in other PRT family members (Fig. 1 and 2B). Among the PRTs, *B. subtilis* PurR is most closely related to the adenine PRTases (APRTs). At the sequence level, the 21% identical wheat APRT is the closest homolog. The closest structural homolog among 38 other PRTs of known structure is the *Leishmania donvanii* APRT (25), with an RMSD of 1.7 Å for 151 C_α positions (Fig. 3B).

PurR crystallized as a symmetric dimer (Fig. 2A). Both domains contribute to a substantial dimer interface. The interface between the two N-terminal winged-helix domains consists of a well-packed core of 22 hydrophobic residues and three pairs of peripheral hydrogen bonds, burying 16% (747 $Å^2$) of the total domain surface area. In contrast, the interface between the two C-terminal PRT domains is less tightly packed and consists of a hydrophobic core of 20 residues burying 8% (788 $Å^2$) of the total domain surface area. The dimer found in the crystal clearly corresponds to the solution dimer because of the extensive hydrophobic interface between subunits. There are no significant differences between the AB and CD dimers in the crystallographic asymmetric unit. However, subunits A and D have similar crystal lattice environments, as do subunits B and C, and these subunits therefore have more similar atomic B-factor distributions, disordered residues, and atomic positions (Table 2).

Winged-helix domain and DNA binding. The HTH proteins are the most thoroughly studied group of prokaryotic DNA-

FIG. 2. (A) PurR dimer. The monomers are colored according to function (N termini, HTH motifs, and wings in blue; hoods in cyan; PP_i loops in gold; flexible loops in red; and PRPP motifs in purple). Other regions are green in one monomer and gray in the other and are semitransparent to aid in viewing the highlighted regions. In this view, the dimer dyad axis is vertical and in the plane of the page. Sulfate ions and a HEPES buffer molecule are rendered as bonds colored by atomic type (C, yellow; O, red; N, blue; and S, green). (B) PurR monomer. The stereo ribbon diagram is colored as the gray subunit in panel A. Secondary structures are labeled. Sulfate ions are shown as in panel A. Residues 163 to 167 and 275 to 285 are not displayed, as they are disordered.

TABLE 1. MAD data collection and phasing[a]

Wavelength	λ_1, 0.9790 Å	λ_2, 0.9788 Å	λ_3, 1.5418 Å	$\lambda_1 + \lambda_2 + \lambda_3$
Data				
Data range (Å)	20–2.2 (2.28–2.20)	20–2.2 (2.28–2.20)	30–2.7 (2.8–2.7)	30–2.2 (2.28–2.20)
>95% complete (Å)	3.6–2.2	3.6–2.2	30.0–2.7	30.0–2.2
Unique reflections	64,125	64,937	37,802	66,722
Avg. redundancy	2.5	2.6	3.0	4.3
% Completeness	94.5 (97.4)	94.4 (97.0)	97.2 (95.5)	97.6 (96.6)
% R_{sym}[b]	5.0 (18.0)	8.2 (18.7)	5.5 (18.7)	7.3 (19.4)
$\langle I/\sigma_I \rangle$	13.0 (3.6)	12.4 (3.6)	17.8 (4.2)	11.1 (3.6)
Phasing				
R_{cullis}[c]				
Dispersive		0.72	0.93	
Anomalous	0.85	0.78	0.77	
Phasing power[d]		1.47	0.54	

[a] Values in parentheses pertain to the outermost shell of data.
[b] $R_{sym} = \Sigma_{h,i}|I_{h,i} - \langle I_h \rangle|/\Sigma_{h,i} I_{h,i}$.
[c] R_{cullis} = root mean square (rms) lack of closure/rms isomorphous difference.
[d] Phasing power = $<|F_h|>$/rms lack of closure.

binding proteins. They exhibit a remarkable variety of schemes for sequence-specific DNA recognition and binding. While a DNA-PurR complex cannot be modeled from these other proteins, available biochemical, sequence, and structural information limits the possibilities for PurR.

DNA sequence recognition is generally achieved by interaction of the HTH recognition helix with the major groove of DNA (for example, reference 1). Winged-helix proteins generally follow this pattern, and in addition, one or both of the wings contact the DNA minor groove (for example, reference 6). However, this binding mode is not universal. There is considerable variability in the length, structure, and function of the wings in these proteins. For example, DNA sequence recognition by eukaryotic transcription factor RFX1 occurs by insertion of the wing into the major groove, while only one side chain of its recognition helix binds in the minor groove (12). In contrast, the wings of the catabolite activator protein do not interact with DNA (29).

Most HTH proteins are dimeric and bind to DNA inverted repeats with aligned protein and DNA symmetries. However, in contrast to canonical HTH or winged-helix proteins, two or more PurR dimers bind one inverted repeat, which encompasses considerably more DNA than in complexes with other HTH or winged-helix proteins. Precedent for two protein dimers binding one DNA inverted repeat is found in the multidrug-binding repressor QacR (30). Typical HTH proteins bind inverted repeats of about 20 base pairs, with recognition helices bound in adjacent major grooves on one side of the DNA duplex. The DNA inverted repeats recognized by PurR encompass more than 40 base pairs (14 base pairs per repeat and 16 to 17 intervening base pairs), and more than 70 base pairs are required for high-affinity protein binding (4).

Among winged-helix proteins, the extensive hydrophobic contact between domains is unique to PurR. The hydrophobic interface, which is continuous with the hydrophobic cores of the two domains, is formed by α1 and the recognition helix, α3. Close association of winged-helix domains is a common feature of PurR homologs because amino acid side chains in the interface (Leu9, Val10, Thr13, and Leu17 in α1 and Ile46, Ile47, Thr50, and Phe51 in α3) are conserved as hydrophobic in 21 eubacterial sequences. This seems to rule out control of DNA affinity by a classic allosteric switch in which relative positions of the reading heads change in response to PRPP binding in the PRT domain.

The most positively charged surface of HTH and winged-helix domains is consistently the DNA binding surface (11). This is striking in RFX1, where the wing rather than the recognition helix has a positive surface and binds DNA (12). The most positively charged surface of the PurR winged-helix domain (Fig. 4) is formed by the N-terminal tail (amino terminus, Lys2, and Arg4) and beginning of α1 (Arg5 and Arg8). These residues from both subunits form a positively charged swath

TABLE 2. Refinement statistics

Residue range	
Subunits A and D	2–162, 168–276
Subunits B and C	1–162, 168–274
HEPES molecules	1 each in monomers A and D
Sulfate ions	1 in monomers A and D, 3 in monomers B and C
Water molecules	568
Side chains with two conformers	33
Data range (Å)	20.0–2.2
R_{work} (%)[a]	18.8
R_{free} (%)[a]	23.7
Data cutoff	$I/\sigma_I > 0$
No. of reflections (% possible)	
Total	66,365 (94.4)
Working set	63,030 (89.7)
Test set	3,335 (4.7)
Average B values (Å^2)	
Main chain	48.0
Side chain	55.3
Sulfate	72.9
HEPES	105.9
PEG	87.2
Water	58.3
All atoms	57.7
RMS deviations from target values	
Bond lengths	0.010 Å
Bond angles	1.459°
Dihedral angles	23.5°
B factors between bonded atoms	3.04 Å^2
RMSD between monomers (C_α)	0.11–0.2 Å
Ramachandran outliers	None

[a] R factor = $\Sigma_h||F_{obs}| - |F_{calc}||/\Sigma_h|F_{obs}|$

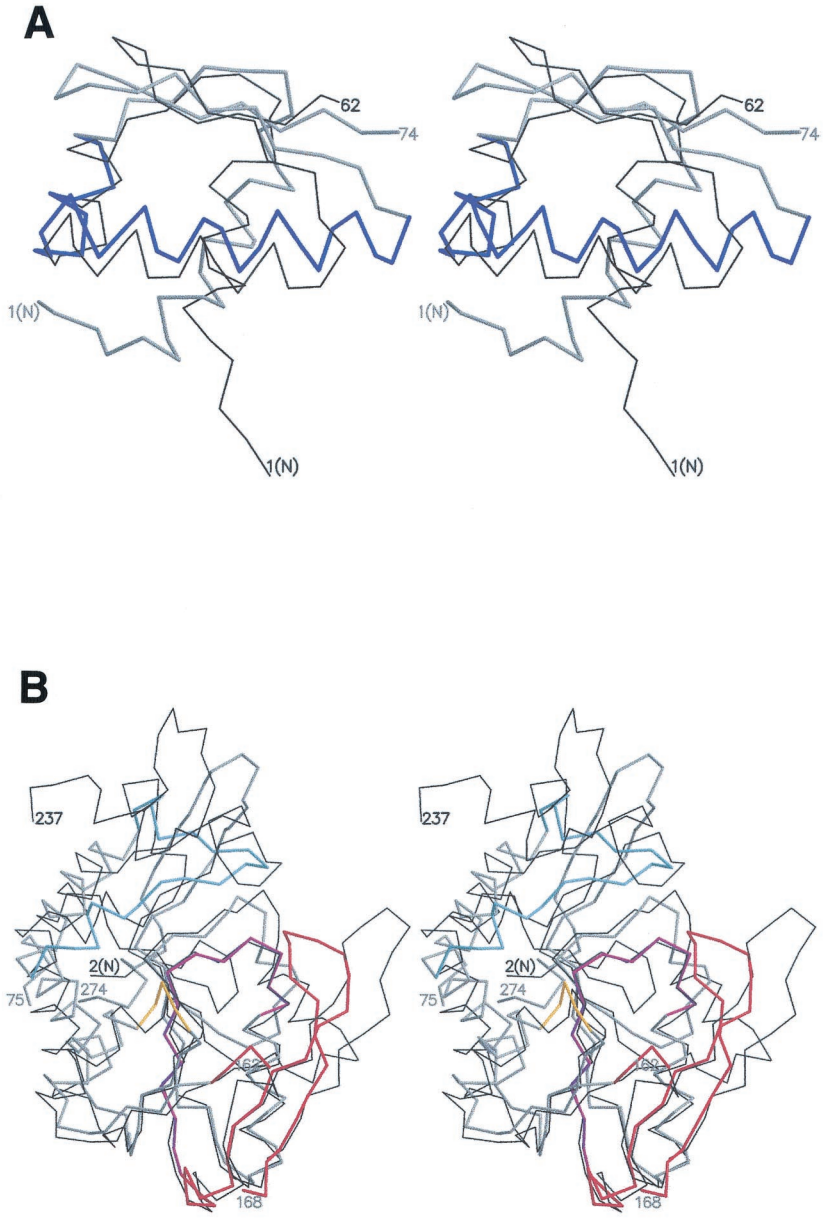


FIG. 3. (A) Winged-helix domain of PurR. C_α atoms of the winged-helix domains of PurR (thick, gray C_α trace with the HTH motif in blue) and *E. coli* BirA (37) (thin black trace) are superimposed in this stereo diagram. Conserved regions thought to be important to function (HTH motif, N terminus, and wing) are colored blue. The termini of each domain are labeled. (B) PRT domain of PurR. C_α atoms of the PRT domains of PurR (gray trace with functionally important regions colored as in Fig. 2) and *L. donovanii* APRTase (25) (thin black trace) The termini of each domain and gaps in the structure are labeled.

across the bottom surface of the PurR dimer (Fig. 4). The large positive surface is bounded by the carboxyl groups of Glu42 in the recognition helix, α3, separated by 24 Å across the dimer interface. A sulfate ion from the crystallization solution bound to Lys2, Arg8, Ser35, and the backbone of Phe3 demonstrates that this region is a suitable site for interaction with DNA phosphate. If DNA binds to the winged-helix domains of PurR, it surely binds to this positively charged surface.

The function of PurR homologs is unknown apart from the *L. lactis* transcription activator, also called PurR. However, a strong case is made for a common DNA-binding function by mapping conserved residues onto the structure (Fig. 5). All eubacterial homologs have the winged-helix domain and are thus presumptive DNA-binding proteins. The positively charged surface is conserved among the 21 eubacterial sequences. The length of the N-terminal tail varies (i.e., from 2 to 6 residues), but always includes one or two positively charged side chains in addition to the amino terminus (Fig. 1). Invariant residues cluster at the positively charged surface (Arg5 and Arg8), the N terminus of the recognition helix α3 (Lys37, Ser38, Ser40, Glu42, and Asp43), and the wing (Gly64, Gly67, and Gly68). These regions of the domain form a continuous

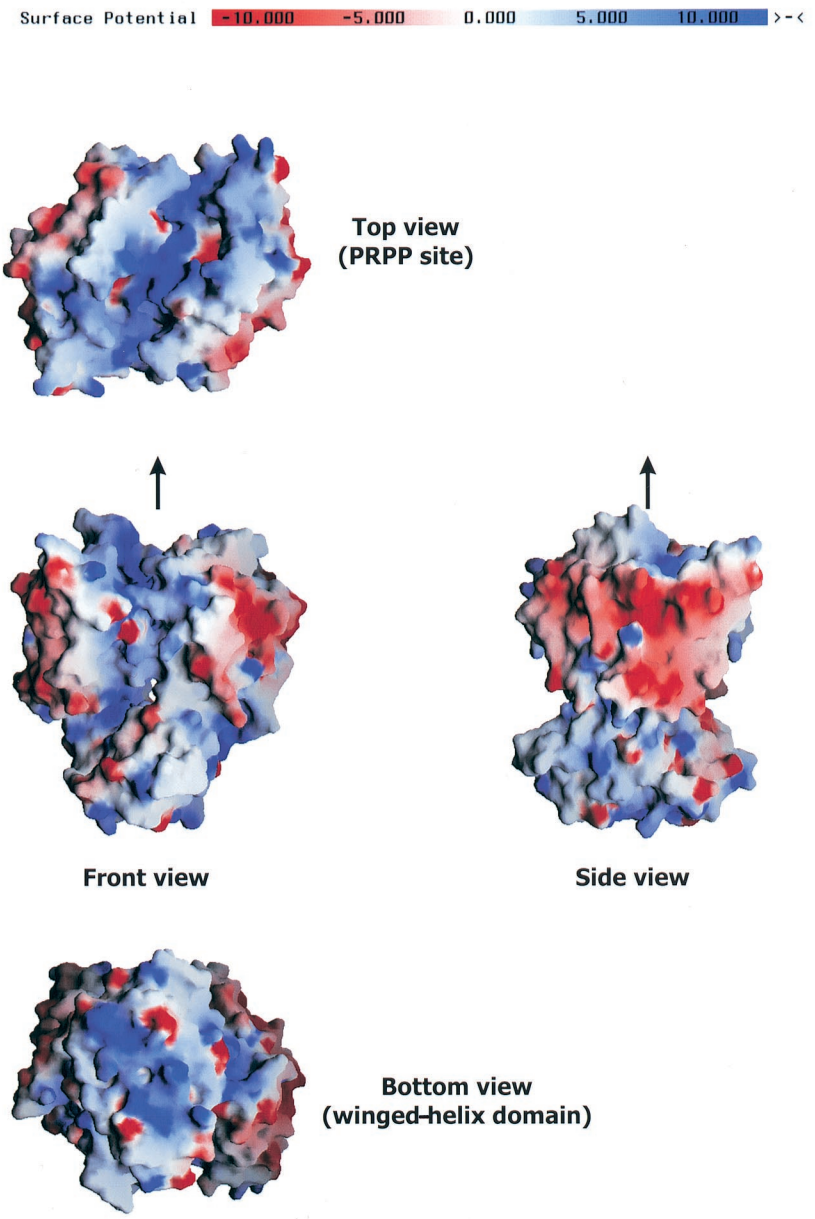


FIG. 4. Electrostatic surface potential of PurR. The electrostatic potential of the molecular surface is color coded from red (−10kT) to blue (+10kT). The four panels represent the front of the dimer as in Fig. 2A, the top of the dimer viewed into the PRPP site, the bottom of the dimer viewed onto the conserved surface of the winged helix domain, and the side of the dimer. The twofold axis of molecular symmetry is vertical in the front and side views and is indicated by an arrow. The dimer images appear asymmetric because of slight differences between subunits in the position of the flexible loop and in the identity of missing residues.

surface at the bottom of the PurR dimer, with the positive charges at the center and the wings in the outermost positions (Fig. 4).

PRT domain and PRPP binding. Most PRT family members are PRTases and catalyze the displacement of the α1-pyrophosphate of PRPP by a nitrogen-containing nucleophile to produce a β1-substituted ribose-5-phosphate and free pyrophosphate (PP_i). All PRTs have similar PRPP binding sites, comprising three structures at the C-terminal edge of the central β sheet (Fig. 6A). A PP_i loop (PurR residues 138 to 141) follows the first β strand of the central sheet and contains a characteristic nonproline *cis* peptide between Ala138 and Thr139 (Fig. 1). A long flexible loop (residues 160 to 188) between the second and third β strands is fully ordered in other PRTs only when closed over bound PRPP. A PRPP loop (residues 203 to 211) is within the PRPP-binding motif and follows the third β strand of the central sheet. The structure of the PRPP-binding motif (residues 199 to 211) is extremely well conserved (Fig. 3B), with pairwise RMS deviations of less than 1 Å for thirteen C_α atoms from PurR and various other PRTs. The PP_i and PRPP loops are adjacent and form the topological switch point of the PRT fold.

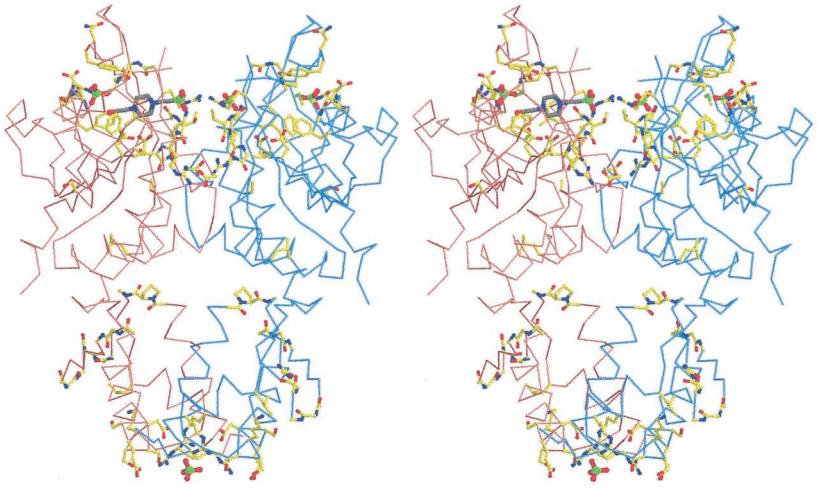


FIG. 5. Conservation among PurR homologs. Side chains of residues invariant among the 21 homologs in Fig. 1 are mapped onto the PurR dimer in this stereo diagram. Invariant residues cluster at opposite ends of the dimer on the surfaces of the winged-helix domain at the bottom and the PRT domain at the top. Sulfate ions and a buffer molecule are visible in these sites. Polypeptides are rendered as the C_α trace. View is identical to Fig. 2A.

Residues constituting these three structural regions are remarkably well conserved among PurR and its eubacterial homologs (Fig. 1), consistent with a common function. In addition to residues expected to contact PRPP directly, a second tier of conserved residues forms specific contacts to maintain precise positions for the three critical loops. For example, a hydrogen bond between the side chain of invariant Thr136 and the backbone carbonyl of conserved Ile202 anchors the PP_i loop to the PRPP loop; invariant Gly141 keeps the PP_i loop near invariant Asp203 and Asp204 in the PRPP loop; and invariant Gly210 and Gly214 permit the flexible loop to pack closely against the PRPP loop.

In PurR, as in many other PRT structures lacking PRPP, a sulfate ion from the crystallization solution mimics the binding of PRPP 5-phosphate (Fig. 6A) through hydrogen bonds with the amides of Met206, Lys207, Ala208, Gly209, invariant Gly210, and invariant Thr211 in the PRPP loop, with the side chain of Thr211, and with the amide of conserved Gly178 in the flexible loop. The sulfonate of a HEPES buffer molecule in subunits A and D and a sulfate ion in subunits B and C mimic the β-phosphate of PRPP pyrophosphate. This sulfate/sulfonate forms hydrogen bonds with side chains and backbone amides of invariant *cis*-Thr139 and Lys140 in the PP_i loop and an ionic contact with conserved Lys161 of the flexible loop. Two hydrogen bonds with the sulfate/sulfonate are formed by the side chain of invariant Arg160 from the flexible loop of the partner subunit in the dimer (Fig. 1 and 6A). All of these interactions with sulfate or sulfonate are characteristic of PRPP binding to other PRT proteins.

The PRT flexible loop of PurR is ordered through most of its length, forming an antiparallel β ribbon (β8 and β9 and an eight-residue connecting loop). The β ribbon partially covers the PRPP binding site through the hydrogen bond between Gly178 and the sulfate bound in the PRPP loop. A similar β ribbon conformation also occurs in the flexible loop of *L. donovanii* APRT (25), the nearest structural neighbor of PurR, although the loop is more open in APRT. Unlike other PRTs, the disordered region of the PurR flexible loop is not the tip of the loop but rather the connection between the PRT core and the first β strand of the loop (residues 163 to 167).

The presumed PRPP binding sites are clefts at the top of the PurR dimer separated by about 15 Å. Among the 21 eubacterial PurR homologs, invariant residues cluster around these sites consistent with a common PRPP-binding function for all of the proteins (Fig. 5). The invariant residues from the two subunits form a highly conserved top surface for the PurR dimer.

PRT family members have a hood above the central β sheet. In PurR, the hood (residues 92 to 110 and 252 to 264; Fig. 2B) does not resemble the hood of any other PRT protein, nor does it resemble any other protein motif. The hood of PRTases is responsible for binding the nitrogen-containing nucleophile cosubstrate (ammonia, adenine, guanine, hypoxanthine, xanthine, orotate, and uracil). PurR appears incapable of binding a nucleophile because conserved Tyr102 fills the space normally occupied by the nucleophile substrates in PRTases (Fig. 6B). Contacts of Tyr102 with invariant Phe205 in the PRPP loop fix the position of the tyrosine ring. The PurR hood is also stabilized by several interactions of conserved residues, including a bidentate salt bridge between invariant Arg96 and Asp107, a hydrogen bond between the carbonyl of invariant Gly100 and conserved Lys207 in the PRPP loop, and several hydrophobic interactions. Thus, the PurR hood seems designed to prevent nucleophile binding. The inability to bind a nucleophile explains both the observed absence of phosphoribosyltransferase activity in PurR and the inability of molecules with nucleobases to disrupt PurR-DNA binding (38).

DNA binding to PurR. How does the PurR dimer bind specifically to control site DNA? Using other winged-helix and HTH proteins as a precedent, we presume that the positively charged surface of a PurR winged helix dimer recognizes the conserved CGAA sequence in the center of a DNA PurBox.

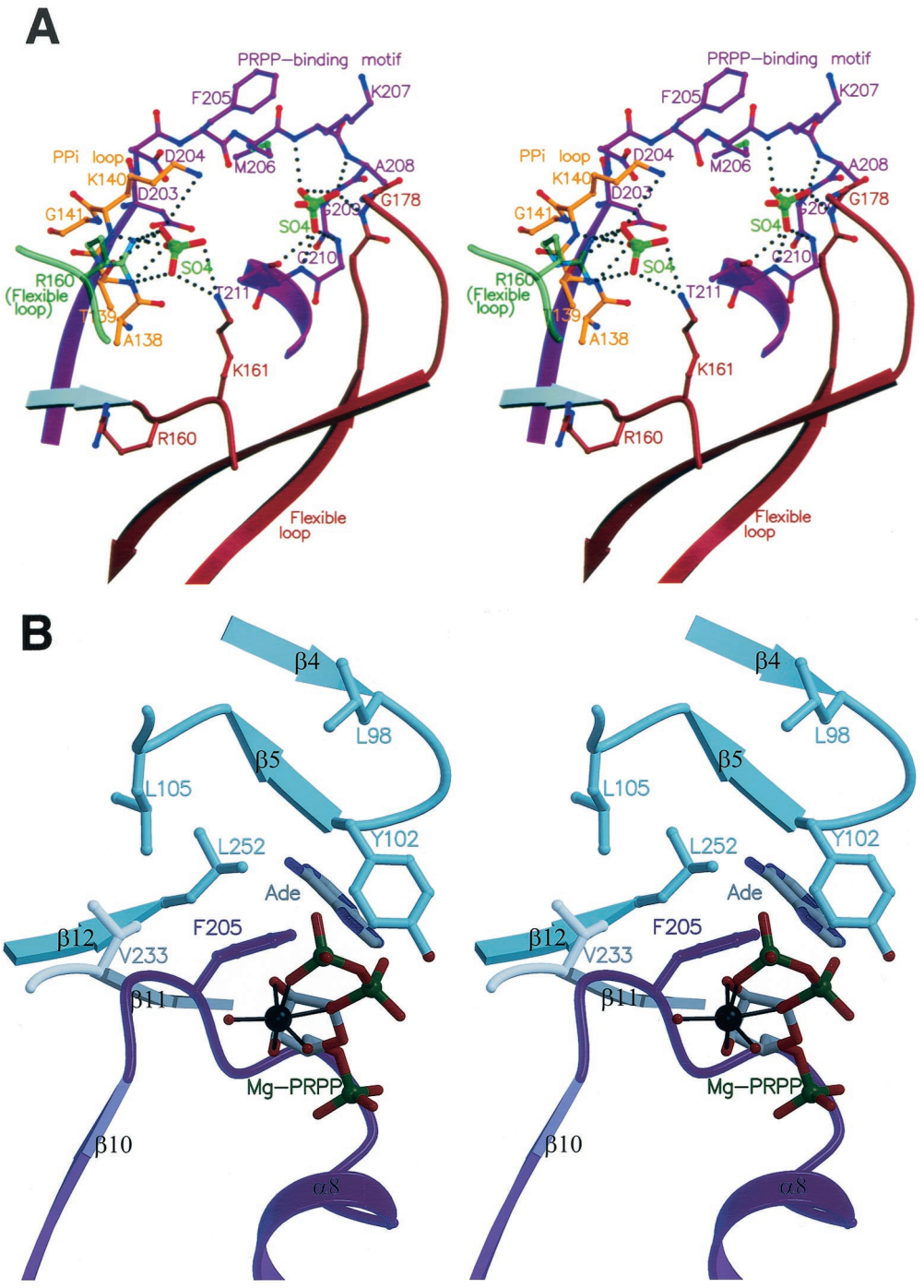


FIG. 6. (A) The PRPP-binding site of PurR. The stereo view is color coded, with the PRT motif in violet, the PP_i loop in gold, and the flexible loop in red. Arg160 from the flexible loop of the other subunit of the dimer is green. This site corresponds to the substrate-binding site for PRTases. Sulfates bound to this site are shown. Residues in the PRPP-binding site that interact with the sulfate are labeled. Hydrogen bonds are shown as dotted lines. (B) Occlusion of a potential nucleophile-binding site. The adenine is modeled based on superposition of the PRPP-binding motifs from PurR and an APRTase-adenine complex (PDBcode 1QB7) (25), while PRPP is modeled based on a similar superposition of a glutamine PRPP amidotransferase-PRPP analog complex (17). Protein elements are from PurR. Phe205 from the PurR PRPP loop and Tyr102 from the PurR hood occlude the space that would bind a nucleophile in the PRTases.

However, PurR protection of long stretches of control site DNA suggests that high-affinity DNA binding requires nonspecific interactions with additional regions of the winged-helix and/or PRT domains. Positively charged surfaces on the front, bottom (winged-helix domains) and top (presumed PRPP binding sites) of the PurR dimer are potential DNA binding sites (Fig. 4). Use of multiple binding surfaces also would explain both the requirement for more than 70 base pairs of

control site DNA and the alternating pattern of DNase I protected and hypersensitive sites. Similar modes of DNA binding have been proposed for winged-helix proteins such as OmpR (20) and ArgR (33).

The presumed PRPP binding site and flexible loop are candidates for direct DNA interactions with the PRT domain. These structural elements in the two subunits form a large, electropositive, conserved surface on the top of the PurR dimer—opposite the positive surface of the winged-helix domains (Fig. 4 and 5). The DNA backbone may bind directly in adjacent PRPP sites of the dimer, as these sites are tailored for a molecule with ribose and phosphate moieties. Local PRPP-induced changes in the top electropositive surface may reduce DNA affinity enough to relieve repression by PurR. For example, binding of electronegative PRPP would greatly reduce the basicity and DNA affinity of the top surface of the dimer. Conformational changes in the flexible loop, such as those induced by PRPP in other PRT proteins, would further alter the surface encountered by DNA. Additionally, PRPP may induce conformational changes at domain interfaces, thus modulating DNA binding to additional sites on the PurR dimer outside the winged-helix domains. Large PRPP- or DNA-induced changes at the dimer interface of the winged helix domains are unlikely because of the well-packed, compact, hydrophobic core between domains. However, the interfaces between the winged helix and PRT domains within a monomer and the two PRT domains within a dimer are less well packed and may be subject to PRPP- or DNA-induced changes.

Shin et al. proposed that DNA wraps around PurR based on their finding that PurR induces right-handed supercoils in DNA (31). This is consistent with the conserved and positively charged bottom and top surfaces of the PurR dimer (Fig. 4 and 5). The *B. subtilis* PurBox sequences (28) are remarkable in the high AT content bordering the central conserved CGAA motif. These sequences could facilitate DNA bending in the outer regions of the PurBoxes. The possibility of such a deformation of DNA is also consistent with the results of torsional constraint experiments (31).

Binding of two or more PurR dimers to palindromic PurBoxes is highly cooperative. Cooperative DNA binding could be achieved through DNA deformation, as in the DNA complexes of winged-helix proteins RFX1 (12) and QacR (30) or by protein-protein contacts such as those in the DNA complex of the winged-helix protein FadR (37, 40). The structure and binding properties of PurR are consistent with either mechanism of cooperativity.

ACKNOWLEDGMENTS

This work was supported by grants DK42303 and GM24658 from the U.S. Public Health Service.

REFERENCES

1. **Aggarwal, A. K., D. W. Rodgers, M. Drottar, M. Ptashne, and S. Harrison.** 1988. Recognition of a DNA operator by the repressor of phage 434: a view at high resolution. Science **242:**899–907.
2. **Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:**3389–3402.
3. **Barton, G. J.** 1993. ALSCRIPT: a tool to format multiple sequence alignments. Protein Eng. **6:**37–40.
4. **Bera, A. K., J. Zhu, H. Zalkin, and J. L. Smith.** 2003. Functional dissection of the *Bacillus subtilis pur* operator site. J. Bacteriol. **185:**4099–4109.
5. **Brünger, A. T., P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson, and G. L. Warren.** 1998. Crystallography & NMR system (CNS): a new software suite for macromolecular structure determination. Acta Crystallogr. D **54:**905–921.
6. **Clark, K. L., E. D. Halay, E. Lai, and S. K. Burley.** 1993. Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. Nature **364:**412–420.
7. **Collaborative Computational Project Number 4.** 1994. The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D **50:**760–763.
8. **Cowtan, K. D.** 1994. 'DM': an automated procedure for phase improvement by density modification. Joint CCP4 and ESF-EACBM Newsl. Protein Crystallogr. **31:**34–38.
9. **Ebbole, D. J., and H. Zalkin.** 1989. *Bacillus subtilis pur* operon expression and regulation. J. Bacteriol. **171:**2136–2141.
10. **Ebbole, D. J., and H. Zalkin.** 1989. Interaction of a putative repressor protein with an extended control region of the *Bacillus subtilis pur* operon. J. Biol. Chem. **264:**3553–3561.
11. **Gajiwala, K. S., and S. K. Burley.** 2000. Winged helix proteins. Curr. Opin. Struct. Biol. **10:**110–116.
12. **Gajiwala, K. S., H. Chen, F. Cornille, B. P. Roques, W. Reith, B. Mach, and S. K. Burley.** 2000. Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding. Nature **403:**916–921.
13. **Holm, L., and C. Sander.** 1993. Protein structure comparison by alignment of distance matrices. J. Mol. Biol. **233:**123–138.
14. **Jones, T. A., J. Y. Zou, S. W. Cowan, and M. Kjeldgaard.** 1991. Improved methods for building protein models in electron density maps and the location of error in these models. Acta Crystallogr. A **47:**110–119.
15. **Kilstrup, M., S. G. Jessing, S. B. Wichmand-Jørgensen, M. Madsen, and D. Nilsson.** 1998. Activation control of *pur* gene expression in *Lactococcus lactis*: proposal for a consensus activator binding sequence based on deletion analysis and site-directed mutagenesis of *purC* and *purD* promoter regions. J. Bacteriol. **180:**3900–3906.
16. **Kilstrup, M., and J. Martinussen.** 1998. A transcriptional activator, homologous to the *Bacillus subtilis* PurR repressor, is required for expression of purine biosynthetic genes in *Lactococcus lactis*. J. Bacteriol. **180:**3907–3916.
17. **Krahn, J. M., J. H. Kim, M. R. Burns, R. J. Parry, H. Zalkin, and J. L. Smith.** 1997. Coupled formation of an amidotransferase interdomain ammonia and a phosphoribosyltransferase active site. Biochemistry **36:**11061–11068.
18. **Krahn, J. M., S. Sinha, and J. L. Smith.** 1999. Successes and prospects for SeMet MAD and large structures. Trans. Am. Crystallogr. Assoc. **34:**27–38.
19. **Kraulis, P. J.** 1991. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. **24:**946–950.
20. **Martínez-Hackert, E., and A. M. Stock.** 1997. The DNA-binding of OmpR: crystal structure of a winged helix transcription factor. Structure **5:**109–124.
21. **Merritt, E., and D. J. Bacon.** 1997. Raster3D: photorealistic molecular graphics, p. 505–524. *In* C. W. J. Carter and R. M. Sweet (ed.), Macromolecular crystallography, part B, vol. 277. Academic Press, New York, N.Y.
22. **Nicholls, A., K. A. Sharp, and B. Honig.** 1991. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins **11:**281–296.
23. **Otwinowski, Z.** 1991. Maximum likelihood refinement of heavy atom particles, p. 80–85. *In* W. Wolf, P. R. Evans, and A. G. W. Leslie (ed.), Isomorphous replacement and anomalous scattering. Science and Engineering Research Council, Daresbury Laboratory, Warrington, United Kingdom.
24. **Otwinowski, Z., and W. Minor.** 1997. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. **276:**307–325.
25. **Phillips, C. L., B. Ullman, R. G. Brennan, and C. P. Hill.** 1999. Crystal structure of adenine phosphoribosyltransferase from *Leishmania donovani*. EMBO J. **18:**3533–3545.
26. **Rappu, P., B. S. Shin, H. Zalkin, and P. Mäntsälä.** 1999. A role for a highly conserved protein of unknown function in regulation of *Bacillus subtilis purA* by the purine repressor. J. Bacteriol. **181:**3810–3815.
27. **Read, R. J.** 1986. Improved fourier coefficients for maps using phases from partial structures with errors. Acta Crystallogr. A **42:**140–149.
28. **Saxild, H. H., K. Brunstedt, K. I. Nielsen, H. Jarmer, and P. Nygaard.** 2001. Definition of the *Bacillus subtilis* PurR operator using genetic and bioinformatic tools and expansion of the PurR regulon with *glyA*, *guaC*, *pbuG*, *xpt-pbuX*, *yqhZ-folD*, and *pbuO*. J. Bacteriol. **183:**6175–6183.
29. **Schultz, S. C., G. C. Shields, and T. A. Steitz.** 1991. Crystal structure of a CAP-DNA complex: the DNA is bent by 90°. Science **253:**1001–1007.
30. **Schumacher, M. A., M. C. Miller, S. Grkovic, M. H. Brown, R. A. Skurray, and R. G. Brennan.** 2002. Structural basis for cooperative DNA binding by two dimers of the multidrug-binding protein QacR. EMBO J. **21:**1210–1218.
31. **Shin, B. S., A. Stein, and H. Zalkin.** 1997. Interaction of *Bacillus subtilis* purine repressor with DNA. J. Bacteriol. **179:**7394–7402.
32. **Sinha, S. C., and J. L. Smith.** 2001. The PRT protein family. Curr. Opin. Struct. Biol. **11:**733–739.
33. **Sunnerhagen, M., M. Hilges, G. Otting, and J. Carey.** 1997. Solution structure of the DNA-binding domain and model for the complex of multifunctional hexameric arginine repressor with DNA. Nat. Struct. Biol. **4:**819–826.

34. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:**4673–4680.
35. **Tomchick, D. R., R. J. Turner, R. L. Switzer, and J. L. Smith.** 1998. Adaptation of an enzyme to regulatory function: structure of *Bacillus subtilis* PyrR, a *pyr* RNA-binding attenuation protein and uracil phosphoribosyltransferase. Structure **6:**337–350.
36. **Turner, R. J., E. R. Bonner, G. K. Grabner, and R. L. Switzer.** 1998. Purification and characterization of *Bacillus subtilis* PyrR, a bifunctional pyr mRNA-binding attenuation protein/uracil phosphoribosyltransferase. J. Biol. Chem. **273:**5932–5938.
37. **van Aalten, D. M., C. C. DiRusso, and J. Knudsen.** 2001. The structural basis of acyl coenzyme A-dependent regulation of the transcription factor FadR. EMBO J. **20:**2041–2050.
38. **Weng, M., P. L. Nagy, and H. Zalkin.** 1995. Identification of the *Bacillus subtilis pur* operon repressor. Proc. Natl. Acad. Sci. USA **92:**7455–7459.
39. **Wilson, K. P., L. M. Shewchuk, R. G. Brennan, A. J. Otsuka, and B. W. Matthews.** 1992. *Escherichia coli* biotin holoenzyme synthetase/*bio* repressor crystal structure delineates the biotin- and DNA-binding domains. Proc. Natl. Acad. Sci. USA **89:**9257–9261.
40. **Xu, Y., R. J. Heath, Z. Li, C. O. Rock, and S. W. White.** 2001. The FadR.DNA complex. Transcriptional control of fatty acid metabolism in *Escherichia coli*. J. Biol. Chem. **276:**17373–17379.