

# Crystal Structure of LacI Member, PurR, Bound to DNA: Minor Groove Binding by $\alpha$ Helices

Maria A. Schumacher, Kang Yell Choi, Howard Zalkin, Richard G. Brennan\*

The three-dimensional structure of a ternary complex of the purine repressor, PurR, bound to both its corepressor, hypoxanthine, and the 16-base pair *purF* operator site has been solved at 2.7 Å resolution by x-ray crystallography. The bipartite structure of PurR consists of an amino-terminal DNA-binding domain and a larger carboxyl-terminal corepressor binding and dimerization domain that is similar to that of the bacterial periplasmic binding proteins. The DNA-binding domain contains a helix-turn-helix motif that makes base-specific contacts in the major groove of the DNA. Base contacts are also made by residues of symmetry-related  $\alpha$  helices, the "hinge" helices, which bind deeply in the minor groove. Critical to hinge helix–minor groove binding is the intercalation of the side chains of Leu<sup>54</sup> and its symmetry-related mate, Leu<sup>54'</sup>, into the central CpG–base pair step. These residues thereby act as "leucine levers" to pry open the minor groove and kink the *purF* operator by 45 degrees.

The purine repressor, PurR, is a 341-amino acid DNA-binding protein that functions as the master regulator of de novo purine biosynthesis and, to a lesser extent, de novo pyrimidine biosynthesis in *Escherichia coli* (1). Specifically, PurR represses transcription from ten polycistronic and monocistronic operons that encode the enzymes of de novo purine biosynthesis, and four genes encoding enzymes participating in de novo pyrimidine biosynthesis and salvage. In addition, PurR regulates the transcription of *glyA*, *gcv*, and *prs*, which encode enzymes for the synthesis of glycine, one-carbon units, and 5-phosphoribosyl-1-pyrophosphate (PRPP), respectively, all of which are necessary for de novo purine biosynthesis. Finally, PurR is autoregulated (2).

PurR belongs to the LacI (lactose repressor) family of transcription regulators of which there are more than 21 members (3). These proteins show strong sequence similarity indicative of a structural relationship. Sequence identity is greatest in the amino-terminal end and often exceeds 60 percent. Genetic and biochemical studies have shown that these proteins can be divided into two functional domains, an NH<sub>2</sub>-terminal DNA-binding domain, approximately the first 60 residues, and a larger COOH-terminal domain, approximately 280 residues, which imparts the functions of effec-

tor binding and oligomerization (3). Whereas the lactose, fructose, and raffinose repressors exist as tetramers (4), all other members of the LacI family appear to be dimeric (3). These proteins function as repressors by binding to operator sites, typically 16 to 18 base pairs (bp) long, that also display significant sequence identity (3). For most LacI family members, operator affinity is highest for the unliganded state of the protein. However, for PurR, binding to operator DNA is dependent on a corepressor. The corepressors for PurR are hypoxanthine and guanine (2), which bind cooperatively with an equilibrium dissociation constant ( $K_d$ ) of 9.3 and 1.5  $\mu$ M, respectively, (5).

Several studies have suggested that the structures of the effector binding domains of LacI family members are similar to the bacterial periplasmic binding proteins (PBPs) (6), and three-dimensional models of the effector binding domains of LacI (7) and the galactose repressor, GalR, (8) have been constructed on the basis of these studies. However, the only three-dimensional structural data available for the LacI proteins have been obtained from nuclear magnetic resonance (NMR) studies of the LacI headpiece, both free (9) and bound to DNA (10). From these studies, the presence of a helix-turn-helix (HTH) motif (11) was confirmed, and contacts between several side chains and bases, some of which had been implicated genetically (12), were determined. However, these studies left in question the structure of the effector binding and dimerization domain and the complete source of DNA sequence discrimination.

In addition, these studies could not address the question of how a signal in the form of a small effector molecule is transduced to effect gene regulation.

We now describe the crystal structure of PurR bound to both its corepressor, hypoxanthine, and a cognate DNA site containing the high affinity *purF* operator ( $K_d = 3.4 \times 10^{-9}$  M) (2). The structure of this LacI member confirms that the corepressor binding domain has a PBP-like fold and an NH<sub>2</sub>-terminal HTH DNA-binding motif. However, the key to DNA binding specificity resides not only in base specific contacts made by the HTH in the major groove but also in DNA deformability and contacts made to the minor groove by the "hinge" helix. Symmetry related residues Leu<sup>54</sup> and Leu<sup>54'</sup> from each hinge helix act as levers to pry open the minor groove thereby unwinding and kinking the DNA toward the major groove. The strong sequence similarity between the LacI members and their cognate DNA sites suggests that this is also the means by which other LacI members interact with their operators.

**PurR structure and corepressor binding.** The structure of the PurR-hypoxanthine-*purF* operator complex was solved by multiple isomorphous replacement (MIR) (Table 1). The asymmetric unit contains one PurR monomer-hypoxanthine-*purF* operator half-site, which requires the statistical disorder of 2 bp and the 5'-nucleoside overhangs because each *purF* operator half-site is not identical (Fig 1A). However, difference Fourier maps of an isomorphous complex, in which the *purF* operator was replaced by a perfect 16-bp palindrome (Fig. 1B), revealed no significant differences between these PurR-hypoxanthine-*purF* operator complexes (13). Our model, which includes residues 3 through 340 and 20 water molecules, has an R factor of 19.5 percent based on all data from 10.0 to 2.7 Å (Table 1). The stereochemistry of the model is very good and only two violations of  $\phi$ ,  $\psi$  space are found (14). A typical section of the current ( $2F_{obs} - F_{calc}$ ) electron density map and an "omit" map are shown in Fig. 2. A topology diagram of the PurR monomer and a stereo view of the biologically relevant PurR-hypoxanthine-*purF* operator are shown in Fig 3.

The NH<sub>2</sub>-terminal DNA-binding domain of PurR can be divided into two functionally important regions. The first contains the HTH motif that spans residues 4 to 23 (Fig 3). Following this structural element is a short loop, residues 24 to 29, helix 3, residues 30 to 43, and another short loop, residues 44 to 47. Helices 1 through 3 form a globular subdomain that is connected to the corepressor binding domain by the "hinge" or more properly, the hinge helix,

M. A. Schumacher and R. G. Brennan are in the Department of Biochemistry and Molecular Biology, Oregon Health Sciences University, Portland, OR 97201-3098, USA. K. Y. Choi and H. Zalkin are in the Department of Biochemistry, Purdue University, West Lafayette, IN 47907, USA.

\*To whom correspondence should be addressed.

residues 48 to 56, which is followed by four extended residues. The hinge helix, helix 4, constitutes the second DNA-binding element of PurR.

The corepressor binding domain (CBD), residues 61 to 340, has the shape of an oblate ellipsoid (axial ratio 2:1) and consists of two topologically similar subdomains, namely, the CBD NH<sub>2</sub>-subdomain, which is directly attached to the DNA-binding domain, and the CBD COOH-subdomain, which contains the COOH-terminus of PurR. The CBD NH<sub>2</sub>-subdomain is composed of a core of six parallel  $\beta$  strands, A to E and J, that are flanked on both sides by  $\alpha$  helices, I, II, III, and IX, and follows the topology,  $\beta_B\beta_A\beta_C\beta_D\beta_E\beta_J$  (Fig 3). The CBD COOH-subdomain is composed of a core of five parallel  $\beta$  strands, F to I and K, flanked by  $\alpha$  helices IV, V, VI, VII, and VIII. The  $\beta$ -sheet topology of this subdomain is  $\beta_F\beta_G\beta_H\beta_I\beta_K$ . Three crossover regions connect the two subdomains, the first two are from  $\beta$  strands to  $\alpha$  helices ( $\beta_E$  to  $\alpha_{IV}$  and  $\beta_I$  to  $\alpha_{IX}$ ) and the last from  $\beta$  strand to  $\beta$  strand ( $\beta_J$  to  $\beta_K$ ). These crossover regions, like those of the PBP, presumably act as a hinge to allow relative movements of the two subdomains upon ligand association and dissociation (15).

Several studies have predicted that the CBD is structurally similar to the PBPs (6), which are (i) monomeric, and (ii) constitute a large, structurally conserved family that participates in transport of metabolites across the bacterial membrane and, in some cases, chemotaxis (16). Comparisons of the CBD structure with several PBPs reveal that its highest similarity is to the ribose binding protein (RBP) (17). An overlay of the corresponding 144 alpha carbons (C $\alpha$ ) of the CBD and RBP results in a root mean squared deviation (rmsd) of 2.29 Å. A similar overlay with the glucose-galactose binding protein yields an rmsd of 2.48 Å. The

corresponding CBD and RBP NH<sub>2</sub>-subdomains are more similar than the CBD and RBP COOH-subdomains in that their C $\alpha$  overlays reveal the respective rmsd of 1.81 and 2.34 Å.

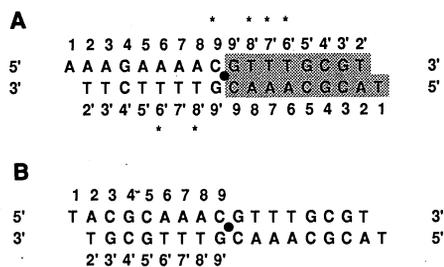
The CBD is solely responsible for binding corepressor (5), and the residues lining the interface of the CBD NH<sub>2</sub>- and COOH-subdomains contribute to high affinity ligand binding through a combination of polar, nonpolar, and aromatic interactions (Figs. 3 and 4). Six direct and water-mediated protein-hypoxanthine hydrogen bonds are provided by residues Asp<sup>275</sup>, Thr<sup>192</sup>, and Arg<sup>190</sup>. Asp<sup>275</sup> hydrogen bonds to N9 of hypoxanthine by way of its car-

boxylate O $\delta$ 1 atom (2.73 Å). This contact is anchored by Arg<sup>196</sup>, which donates hydrogen bonds from its NH2 and NH1 to the O $\delta$ 2 of Asp<sup>275</sup> (2.76 and 2.98 Å, respectively) (Fig 4). This interaction also neutralizes the charges of these side chains, which are buried in the ligand binding pocket. The importance of both residues in corepressor binding is underscored by the greatly diminished corepressor binding observed upon substitution of either residue with alanine (18). Thr<sup>192</sup>, which forms a side-chain hydrogen bond with the N7 of hypoxanthine (O $\gamma$ -N7, 2.70 Å), also participates in van der Waals contacts to the hypoxanthine by way of its C $\gamma$  methyl group (Fig 4). As

**Table 1.** Crystallographic analysis. Detailed conditions for crystallization of the PurR-hypoxanthine-*purF* operator complex have been described (34). Briefly, the crystals were grown by vapor diffusion from polyethylene glycol solutions and assume the orthorhombic space group C222<sub>1</sub> with cell dimensions,  $a = 175.85$  Å,  $b = 94.79$  Å, and  $c = 81.84$  Å. Intensity data were collected at room temperature with a San Diego Multiwire Systems (SDMS) Area Detector (35) and a Rigaku RU200-H rotating anode generator as the x-ray source set at 40 kV, 150 mA. The data were processed with software provided by SDMS. The structure was determined by multiple isomorphous replacement (MIR) and revealed one PurR monomer-hypoxanthine-*purF* operator half-site per asymmetric unit. Heavy atom parameters were refined and MIR phases were calculated with the program HEAVY (36) initially with data from 10.0 to 3.0 Å. The final figure of merit to 3.0 Å was 0.59. The initial electron density map used for tracing was generated after solvent flattening (37) and revealed clear density for most of the protein backbone, the hypoxanthine, and some of the DNA. FRODO (38) was used to build in a partial polyaniline model for the protein which was subsequently refined in real space with TNT (39). After ten cycles of real space refinement, the partial model was used as a source of phases which were combined with the MIR phases to give a new figure of merit of 0.78. The resulting map showed clear density for most of the side chains and the DNA. At that point PurR residues, 3 to 330, and the *purF* half-site were fit. After 20 cycles of positional refinement by means of TNT, the  $R$  factor dropped from 0.47 to 0.29. Several rounds of rebuilding were followed by more positional refinement, after which electron density for residues 331 to 340 was found and the data were extended to 2.7 Å. The  $R$  factor converged at 0.235 and tightly restrained refinement of  $B$  factor was begun. At the present stage of refinement, the agreement factor of the model to crystallographic data is 0.195, for all data from 10.0 to 2.7 Å.

| Item                                   | Native | Iodo-dC9 | Iodo-dU8*   | Iodo-dU7 | Iodo-dU6* | mHgCl <sub>2</sub> † | Thimerosal |
|--|--------|----------|-------------|----------|-----------|----------------------|------------|
| Resolution (Å)                         | 2.7    | 2.8      | 3.0         | 2.8      | 3.0       | 3.0                  | 3.0        |
| Unique reflections                     | 18,242 | 16,115   | 12,870      | 15,980   | 13,094    | 11,656               | 13,547     |
| Completeness (%)                       | 97     | 86       | 68          | 85       | 70        | 62                   | 72         |
| $I/\sigma(I)$ for data                 | 9.4    |          |             |          |           |                      |            |
| from 10.0 Å to 2.7 Å                   |        |          |             |          |           |                      |            |
| $I/\sigma(I)$ for data                 | 2.5    |          |             |          |           |                      |            |
| from 2.8 Å to 2.7 Å                    |        |          |             |          |           |                      |            |
| $R_{\text{symm}}$ (%)‡                 | 5.6    | 6.8      | 10.2        | 7.1      | 8.8       | 9.0                  | 8.1        |
| $R_{\text{iso}}$ (%)§                  |        | 12.2     | 14.9        | 11.0     | 15.1      | 19.2                 | 21.3       |
| Number of sites                        |        | 1        | 2           | 1        | 2         | 1                    | 1          |
| Phasing power                          |        | 1.27     | 0.89        | 1.25     | 1.18      | 1.02                 | 1.01       |
| $R_c$ ¶                                |        | 0.80     | 0.70        | 0.68     | 0.69      | 0.74                 | 0.71       |
| Mean overall figure of merit (to 3 Å)* |        | 0.59     |             |          |           |                      |            |
| Refinement statistics                  |        |          |             |          |           |                      |            |
| Resolution (Å)                         |        |          | 10.0 to 2.7 |          |           |                      |            |
| $R$ factor**                           |        |          | 0.195       |          |           |                      |            |
| Total number of atoms                  |        | 3007     |             |          |           |                      |            |
| Water molecules                        |        | 20       |             |          |           |                      |            |
| rms deviations                         |        |          |             |          |           |                      |            |
| Bond angles (°)                        |        | 2.35     |             |          |           |                      |            |
| Bond lengths (Å)                       |        | 0.012    |             |          |           |                      |            |

\*Indicates derivatives in which pseudo twofold related sites were both iodinated. †mHgCl<sub>2</sub> is an abbreviation for methyl-mercuric chloride. ‡ $R_{\text{symm}} = \sum |I_o - \langle I \rangle| / I_o$ , where  $I_o$  is the observed intensity,  $\langle I \rangle$  is the average intensity obtained from multiple observations of symmetry-related reflections. § $R_{\text{iso}} = \sum ||F_{\text{PH}}| - |F_{\text{P}}|| / \sum |F_{\text{P}}|$ , where  $|F_{\text{P}}|$  is the protein structure factor amplitude and  $|F_{\text{PH}}|$  is the heavy atom derivative structure factor amplitude. ||Phasing power is the rms ( $|F_{\text{H}}|/E$ ),  $|F_{\text{H}}|$  is the heavy atom structure factor amplitude, and  $E$  is the residual lack of closure. ¶ $R_c = \sum ||F_{\text{der}} \pm F_{\text{nat}}| - F_{\text{H(calcd)}}| / \sum |F_{\text{der}} - F_{\text{nat}}|$  for centric reflections, where  $F_{\text{H(calcd)}}$  is the calculated heavy atom structure factor. \*The figure of merit is  $\int P(\theta) \exp(i\theta) d\theta / \int P(\theta) d\theta$  where  $P$  is the probability distribution of  $\theta$ , the phase angle. \*\* $R$  factor =  $\sum |F_{\text{obs}} - F_{\text{calc}}| / \sum F_{\text{obs}}$ . The rms bond lengths and rms bond angles are the respective root-mean-square deviations from ideal values.



**Fig. 1.** (A) Sequence of the duplex deoxyoligonucleotide used in cocrystallization. The numbering scheme is used throughout the text and the locations of iodine-labeled bases in the structure determination are identified by an asterisk (\*). The half-site used in the design of the perfect palindrome site is shaded. (B) Sequence of the perfect palindromic *purF* operator site. The pseudo-dyad and dyad axes of each sequence is indicated by a dot (•).

observed in the structures of liganded PBPs (16, 17), side-chain ligand stacking interactions are also prevalent in PurR-hypoxanthine binding (Fig 4). Aromatic residues Tyr<sup>73</sup> and Phe<sup>221</sup> form a "sandwich" interaction with the corepressor in which Phe<sup>221</sup> is located centrally "above" the hypoxanthine base, and Tyr<sup>73</sup> interacts on the opposite face more with the purine imidazole ring. Phe<sup>74</sup> completes the stack by contacting N1 and C6 (Fig 4).

Whereas the above described interactions are important for purine binding affinity, Arg<sup>190</sup> is the key to corepressor specificity in that its side chain  $\epsilon$ -NH and NH<sub>2</sub> hydrogen bond to the O6 acceptor of hypoxanthine ( $\epsilon$ -NH-O6, 3.10 Å and NH<sub>2</sub>-O6,

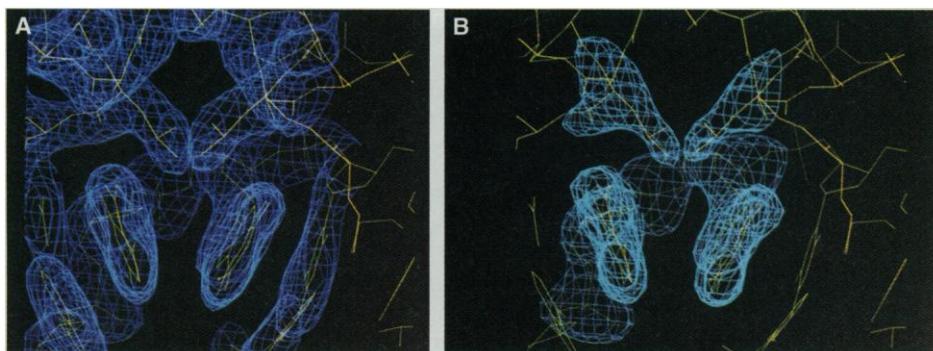
2.83 Å) (Fig 4). These interactions allow PurR to read the exocyclic atom at position 6 of a purine ring and, consequently, discriminate against the NH<sub>2</sub> donor group of adenine. The Arg<sup>190</sup> further contributes to hypoxanthine binding by making a bridging hydrogen bond from its NH<sub>2</sub> nitrogen to a water molecule (2.95 Å), which is, in turn, hydrogen-bonded to the N1 of the purine ring (3.03 Å).

Unlike the PBPs, PurR is a dimer (5), in which the dimerization interface of PurR excludes 2242 Å<sup>2</sup> of protein surface area from the solvent and is formed, for the most part, equally by both CBD subdomains (Fig 3). Subunit contacts between the two CBD NH<sub>2</sub>-subdomains are

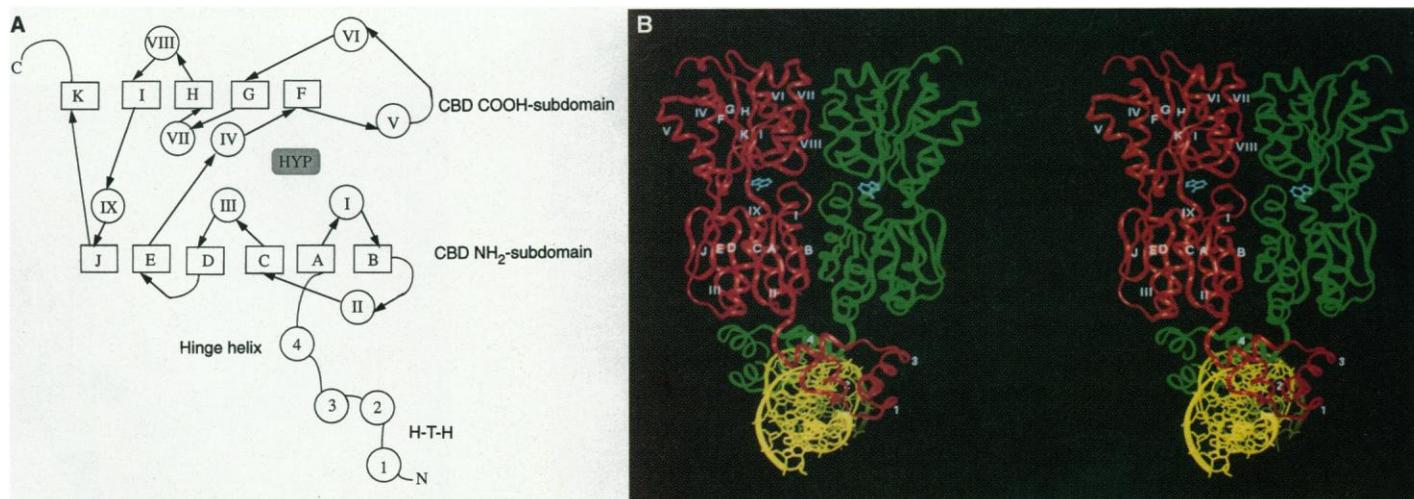
provided by residues 68 through 115 and include part of the loop before  $\alpha_1$ ,  $\alpha_1$ ,  $\beta_B$ ,  $\alpha_{II}$  and the turns in between (Fig 3). The dimerization region between the two CBD COOH-subdomains is noncontiguous and spans residues 223 to 229, 249 to 267, 278 to 285 and 328 to 329 (Fig 3). This region includes residues from  $\alpha_{VI}$ , the NH<sub>2</sub>-terminus of  $\alpha_{VII}$ ,  $\alpha_{VIII}$ , and  $\beta_K$ . Two parallel three-helix bundles formed between  $\alpha_{VI}$  and  $\alpha_{VII}$  of one monomer and  $\alpha_{VII}$  of the other monomer comprise most of the CBD COOH-subdomain interface. Contacts made between residues 328 to 329 and residues of  $\alpha_{VII}$  of the other subunit complete the dimerization interface.

Three cross-subunit contacts between the DNA-binding domain of one subunit and the CBD of the other are found between the main-chain carbonyl oxygen of Gln<sup>113</sup> and the backbone NH of Ala<sup>49</sup>, and the NH1 and NH2 of Arg<sup>115</sup> and side chain O $\gamma$  and main chain CO of Ser<sup>46</sup>. Linkage between the CBD NH<sub>2</sub>-subdomain of one subunit to the CBD COOH-subdomain of the other is provided by two salt bridges between Arg<sup>278</sup> and Glu<sup>70</sup> and Glu<sup>70</sup> and Arg<sup>278</sup>.

**DNA binding.** PurR makes extensive contacts to the bases and phosphate backbone in the major groove of the *purF* operator through its HTH, the loop that follows and helix 3. The HTH of PurR is unusual in that, unlike most structurally characterized HTH motifs, the invariant Gly of the turn is an Asn. However, the side chain of this branched amino acid residue does not alter



**Fig. 2.** Electron density showing how the crystallographically related leucine levers (Leu<sup>54</sup> and Leu<sup>54</sup>) intercalate into the minor groove and are wedged between the central Cyt<sup>9</sup>pGua<sup>9</sup> base pairs. The 45° roll of the central Cyt<sup>9</sup>pGua<sup>9</sup> base-pair step is evident. (A) The refined ( $2F_{obs} - F_{calc}$ ) electron density map contoured at 1.5  $\sigma$ . (B) Omit map ( $F_{obs} - F_{calc}$ ) with nucleotides Cyt<sup>9</sup>, and Gua<sup>9</sup> and residue Leu<sup>54</sup> omitted from the model refinement. The contour level is 3.7  $\sigma$ .



**Fig. 3.** (A) Topology diagram of the PurR monomer.  $\alpha$ -Helical segments are shown as circles and  $\beta$  strands as squares. The NH<sub>2</sub>-terminus and COOH-terminus of the protein are identified with N and C, respectively. The DNA-binding domain consists of (residues enclosed in parentheses) helix 1 (4 to 10) and helix 2 (15 to 23) (the HTH, labeled), helix 3 (30 to 43) and helix 4 (48 to 56) (the hinge helix, labeled). The corepressor binding domain or CBD consists of a CBD NH<sub>2</sub>-subdomain (labeled), which contains strand A (61 to 66), helix I (72 to 88), strand B (91 to 96), helix II (101 to 113), strand C (118 to 121), helix III (128 to 134), strand D (142 to 147), strand E (156 to 159), helix IX (294 to 310), strand J (318 to 320), a CBD COOH-subdomain

(labeled), helix IV (162 to 174), strand F (181 to 184), helix V (191 to 204), helix VI (223 to 234), strand G (241 to 245), helix VII (248 to 260), strand H (270 to 275), helix VIII (280 to 282), strand I (288 to 291), and strand K (324 to 326). The hypoxanthine corepressor is represented by a shaded rectangle. (B) Stereo view of the PurR-hypoxanthine-DNA complex. The DNA is shown as yellow stick bonds and the hypoxanthine as blue stick bonds. The PurR dimer is represented by a ribbon with one subunit colored green and the other red. The secondary structural elements are indicated in white for one monomer subunit. (This figure was generated with Biosym Insight II.)

the HTH structure and likely stabilizes the turn by making a hydrogen bond to the peptide backbone CO of residue 10 thereby capping helix 1. Superposition of the 20 C $\alpha$ s of the HTH of PurR, residues 4 to 23, on the corresponding C $\alpha$ s of the  $\lambda$  repressor results in an rmsd of 0.55 Å. A similar superposition of the PurR and LacI HTH (10) yields an rmsd of 0.59 Å (Fig. 5) that, when extended to include residues 24 through 43, that is, the loop and helix 3, yields an rmsd of 1.24 Å. Exclusion of the loop residues results in an rmsd of 0.88 Å.

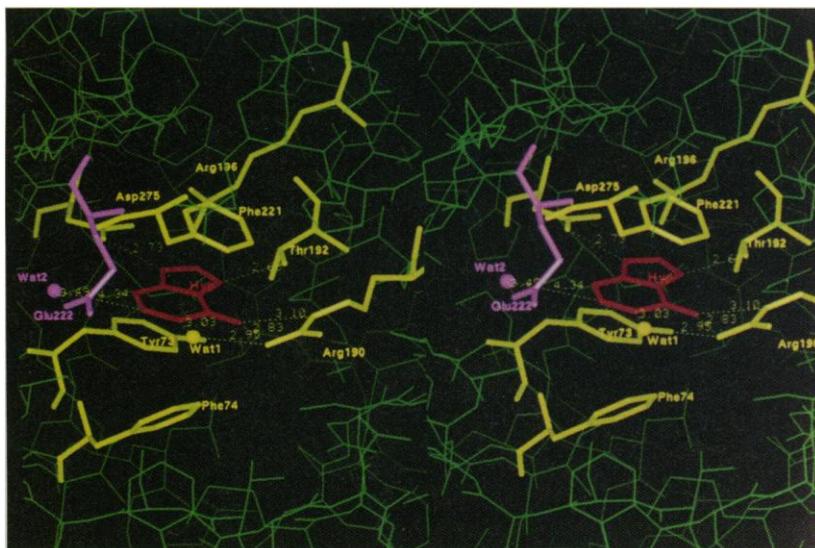
As observed in other prokaryotic HTH protein-DNA complexes, the recognition helix of PurR (helix 2) is positioned on the DNA such that the NH<sub>2</sub>-terminus points into the major groove (Fig. 6A). In this binding mode, helix 1 contributes one phosphate contact from the amide NH of Ile<sup>4</sup>, the first residue in helix 1, to phosphate 8'. The Ile<sup>4</sup> side chain also makes van der Waals contacts to the deoxyribose ring of Thy<sup>8'</sup>. Residues from the turn and helix 2 make six phosphate contacts. Ser<sup>14</sup> makes side-chain and main-chain NH hydrogen bonds to phosphate 3. The side chain of Thr<sup>17</sup> also hydrogen-bonds to phosphate 3, whereas the side chain NH<sub>2</sub> of Asn<sup>23</sup>, the last residue of the recognition helix, and the O $\gamma$  of Thr<sup>19</sup> form hydrogen bonds to phosphate 7' (Fig. 6, A and B). The Ile<sup>4</sup> and Asn<sup>23</sup>-phosphate contacts are analogous to the positioning contacts observed in other HTH protein-DNA complexes (20). Asn<sup>23</sup> also makes a key hydrogen bond via its side chain CO to the side chain NH<sub>2</sub> of hinge

helix residue Arg<sup>52</sup> thus providing a critical link between the globular subdomain, helices 1 through 3, and the hinge region. The loop that follows helix 2 provides two phosphate contacts, one between the amide NH of Ala<sup>29</sup> and phosphate 2, and a second, water-mediated contact from the main chain CO of Phe<sup>27</sup> to phosphate 2. Phe<sup>27</sup> also stacks against the deoxyribose ring of Ade<sup>2</sup>. From helix 3 and the short loop which follows, two phosphate contacts are made that involve the side chain of Thr<sup>32</sup> and phosphate 2, and the side chain of Tyr<sup>45</sup>, which hydrogen bonds to three oxygens of phosphate 8. Of the 13 protein-DNA backbone interactions only four phosphate groups are contacted via the major groove (Fig. 6, A and B).

Five residues from the three-helix globular subdomain participate in direct and water-mediated major groove base contacts (Fig. 6, A and B). Helix 1 contributes van der Waals contacts from the Cy methyl group of Ile<sup>4</sup> to the methyl groups of Thy<sup>7'</sup> and Thy<sup>8'</sup>. The first residue of the recognition helix, Thr<sup>15</sup>, participates in the only water-mediated base contact, in which its O $\gamma$  is hydrogen-bonded to a water (2.54 Å) that is hydrogen-bonded to the O4 atom of Thy<sup>7'</sup> (2.56 Å). The O $\gamma$  of Thr<sup>16</sup> bonds simultaneously to the N6 of Ade<sup>6</sup> and the O4 of Thy<sup>6'</sup> (Fig. 6, A and B). Our structure analysis of the complex of PurR-hypoxanthine and the perfect palindrome operator (Fig 1B) reveals that, when base pair 5 is C-G, the O $\gamma$  of Thr<sup>16</sup> makes an additional hydrogen bond to the N4 of Cyt<sup>5</sup>. The final

major groove contact is directed by a van der Waals interaction between the imidazole ring of His<sup>20</sup> and the side chain of Arg<sup>26</sup>, which is located in the loop that follows helix 2. This protein-protein contact positions the guanidino nitrogens to donate hydrogen bonds to the N7 (2.96 Å) and O6 (2.52 Å) acceptors of Gua<sup>4</sup> (Fig. 6, A and B). Accordingly, Gua<sup>4</sup> is the only guanine of the *purF* operator to display methylation interference sensitivity (20). Furthermore, this contact provides an explanation for the absolute conservation of guanine at position 4 in all known *pur* operators (1, 21).

**The hinge helix and minor groove binding.** The most striking and unanticipated feature of the PurR-hypoxanthine-*purF* operator complex is the binding of the DNA minor groove by helix 4, the hinge helix, and its dyadic partner (Fig. 6, C and D). The hinge helices interact with each other through a series of van der Waals contacts between the side chains of residues Val<sup>50</sup> and Val<sup>50'</sup>, and Leu<sup>54</sup> and Leu<sup>54'</sup> (Fig. 6D). The accommodation of the hinge helices in the minor groove requires its expansion. This is accomplished by the side chains of residues Leu<sup>54</sup> and Leu<sup>54'</sup>, which intercalate into the central Cyt<sup>9</sup>pGua<sup>9'</sup> base pair step and thereby act as "leucine levers" to pry open the minor groove (Figs. 2 and 6, C and D). As a result, two sets of van der Waals contacts are formed, one between the Cy methyl group of Leu<sup>54</sup> and O2 oxygen of Cyt<sup>9</sup> and the second between the Leu<sup>54</sup> C $\delta$ 1 and the deoxyribose O4' atom of Gua<sup>9'</sup> on the same strand (Fig. 6D). Operator site specificity is contributed to by hinge helix



**Fig. 4.** Stereo diagram of the corepressor binding pocket. The hypoxanthine molecule is shown in red. PurR residues participating in hypoxanthine binding and relevant hydrogen bonds are indicated in yellow. Also shown in yellow and represented as a sphere is a water molecule, Wat1, which participates in hypoxanthine binding. The architecture of the purine binding pocket provides insight into the ability of PurR to bind guanine with high affinity. Water molecule, Wat2 (in purple), is located within 4.34 Å of the hypoxanthine C2 and hydrogen bonds to the carboxylate side chain of Glu<sup>222</sup> (3.49 Å). A simple rotation of this side chain would provide a hydrogen bond to the guanine N2. (Figure generated with Biosym Insight II.)



**Fig. 5.** Superposition of the HTH motifs (helices 1 and 2) of PurR (magenta) and LacI (blue). Also shown is the side chain of PurR residue Asn<sup>12</sup>, which is found at the "invariant" glycine position of the turn. Although not included in the calculation, the C $\alpha$  atoms of the loop that follows helix 2, helix 3, and the residues that follow show a high degree of structural overlap. (Figure generated with Biosym Insight II.)

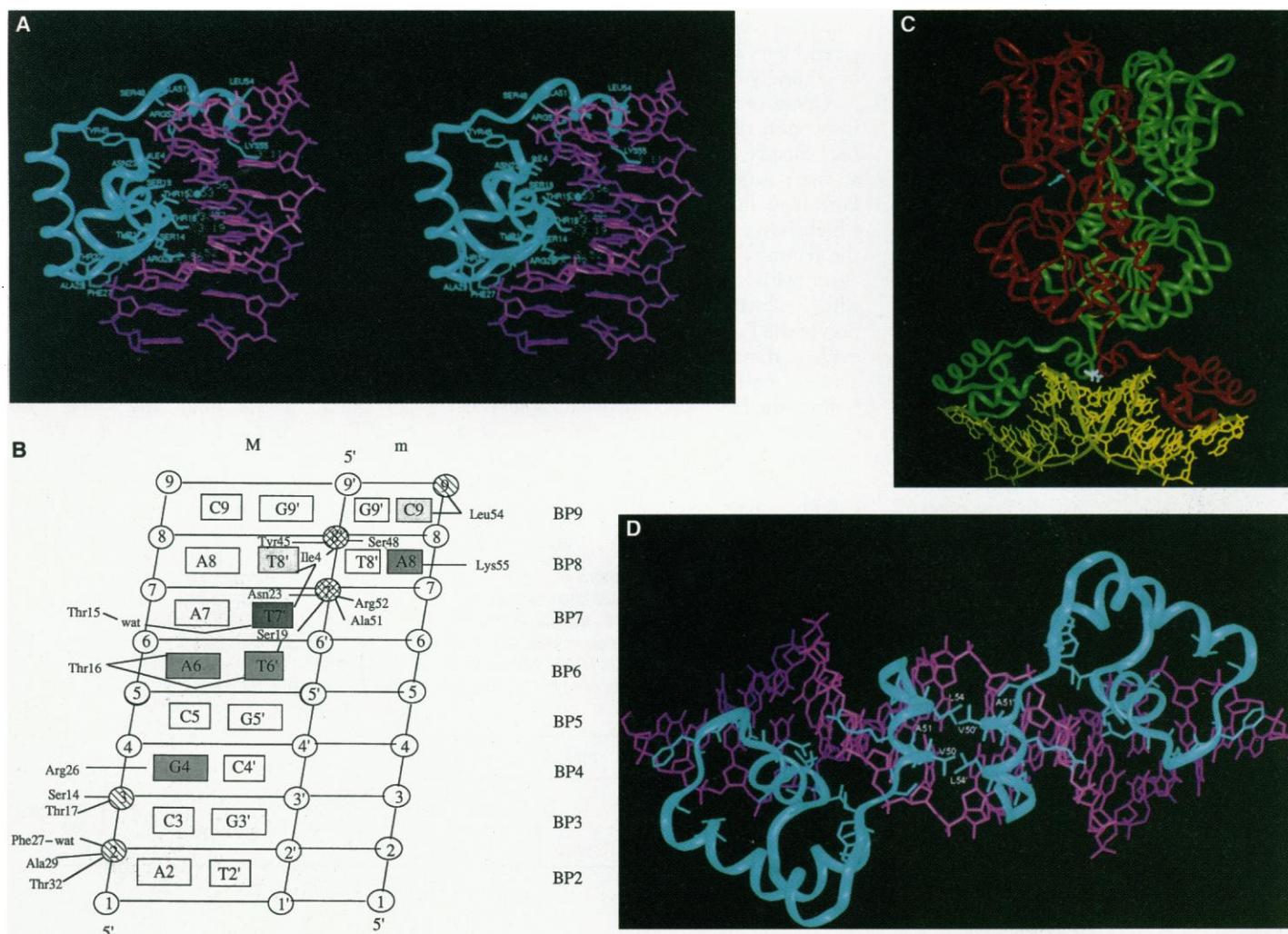
residue Lys<sup>55</sup> which makes a hydrogen bond from its  $\epsilon$ -NH<sub>3</sub> to the N3 of Ade<sup>8</sup> (3.11 Å) and van der Waals contacts to C2. Additional DNA backbone contacts are made by the amide group of Ser<sup>48</sup>, which contacts phosphate 8', the side chain of Arg<sup>52</sup>, which hydrogen-bonds to phosphate 7', and the C $\beta$  methyl group of Ala<sup>51</sup>, which makes van der Waals contacts to the deoxyribose ring of Ade<sup>7</sup>.

That hinge helix residues are critical for *purF* operator binding by PurR is corroborated by mutagenesis data showing that substitution of Leu<sup>54</sup> by lysine, serine, trypto-

phan, threonine, or arginine produces a PurR that cannot regulate *in vivo* transcription from a *purF-lacZ* fusion gene (22). However, replacement by methionine results in a near wild-type phenotype, an indication that the size and hydrophobic nature of this position are crucial to proper function. Substitution of hinge helix residues Arg<sup>52</sup> and Lys<sup>55</sup> with either alanine or glutamic acid results in similar deregulation of the *purF-lacZ* fusion gene with the glutamic acid substitutions being particularly severe (22).

**DNA structure and kinking.** The most

evident consequence of the hinge helices binding to the minor groove is a single 45° kink at the central Cyt<sup>9</sup>pGua<sup>9'</sup> base pair step, which bends the DNA toward the major groove and away from the protein (Figs. 2 and 6, C and D). The Cyt<sup>9</sup>pGua<sup>9'</sup> step displays an unusually large roll angle of 45° and helical rise of 6.4 Å and is clearly unstacked. This CpG step is also unwound with a twist angle of 27°. Helix insertion splay open the minor groove which can be quantitated by the Cyt<sup>9</sup>-Gua<sup>9'</sup> cross-groove O4' to O4' distance of 9.2 Å. The A·T base pair directly 5' of the



**Fig. 6.** PurR-DNA interactions. (A) Stereo figure showing DNA base contacts made by the DNA-binding domain of one PurR monomer. The DNA is shown in magenta stick bonds and PurR residues 3 to 56, which includes the helices 1 through 3 and the hinge helix, are represented by a blue ribbon. Residues contacting the DNA are depicted as blue sticks. Selected hydrogen bond distances are shown in blue. A blue sphere representing a water molecule which mediates the contact between Thr<sup>15</sup> and Thy<sup>7</sup> is also shown. (B) Schematic view of the PurR-palindromic *purF* operator interaction. The DNA is represented as a cylindrical projection. Contacts made to the major groove (M) are differentiated from contacts made to the minor groove (m). Bases involved in van der Waals interactions are shaded light gray, bases involved in hydrogen bonds are shaded medium gray, and bases involved in both are shaded dark gray. Deoxyribose and phosphate groups contacted from either the major or minor groove are filled in with

stripes while phosphate and sugar groups contacted from both the major and minor grooves are hatched. (C) Overall view of the PurR-hypoxanthine-DNA complex. The DNA is represented as yellow stick bonds, the hypoxanthine molecules as blue balls and sticks, and PurR as a ribbon with one monomer subunit green and the other red. Shown as white sticks are the side chains of Leu<sup>54</sup> and Leu<sup>54'</sup>. The expansion of the minor groove and severe kink caused by their insertion is evident. (D) View of the twofold related hinge helices and the minor groove of the *purF* operator. The crystallographic twofold axis is perpendicular to the plane of the figure and bisects the central CpG' base pair step. Shown as blue sticks are key hinge helix residues, Val<sup>50</sup> and Val<sup>50'</sup> (V<sup>50</sup> and V<sup>50'</sup>), Leu<sup>54</sup> and Leu<sup>54'</sup> (L<sup>54</sup> and L<sup>54'</sup>), and Ala<sup>51</sup> and Ala<sup>51'</sup> (A<sup>51</sup> and A<sup>51'</sup>). The two HTH motifs are located in the major grooves in the lower left and upper right. (Figure generated with Biosym Insight II.)

kink, which is contacted by a hinge helix residue Lys<sup>55</sup>, displays severe unwinding, a slide toward the minor groove and an A-DNA-like helical rise (Table 2). Analysis of the helical twist angles of the three central base-pair steps reveals that PurR locally unwinds the *purF* operator by 42°, a value similar to that reported for the unwinding of the *lac* operator by LacI (23).

The remainder of the *purF* operator half-site is also affected by PurR binding (Table 2). Measurement of the minor groove width reveals values that range from 7.8 to 10.3 Å (Table 2). Such expansion of the entire minor groove results directly from the insertion of the hinge helix and indirectly from the recognition helix compression of the major groove. Despite the unusually wide minor groove, the helical rise, twist, and slide of base pair steps 3 through 7 more closely resemble B-DNA (Table 2). The aberrant slide and twist angle values seen for Ade<sup>2</sup>-Thy<sup>2'</sup> could result from crystal packing effects as the PurR COOH-terminus, which is rich in aromatic and basic residues, contacts the 5'-nucleoside overhang. As a result, the DNA is not pseudo-continuous in the crystal. No bifurcated hydrogen bonds are observed between any base-pair steps.

Many DNA-binding proteins bend their DNA recognition sites (11, 24). However, DNA kinking has been observed directly in only three transcription factor-DNA complexes, the catabolite gene activator protein (CAP) (25), the TATA binding protein (TBP) (26, 27), and now PurR. CAP kinks its DNA binding site by ~40° at two dyad-related TpG steps (25). Stabilization of the kink is brought about by several protein-major groove and protein-phosphate backbone interactions emanating, in large part, from residues of the recognition helix. The plant *Arabidopsis thaliana* TBP kinks the TATA element at the 5'-TpA and ApG steps of a TATAAAAG binding site (26), whereas yeast TBP kinks the 5'-TpA and ApA steps of a TATATAAA binding site (27). Unlike CAP, TBP stabilizes its ~45° kinks through minor groove stacking and van der Waals interactions between phenylalanines, located in or near β strands, and the bases and deoxyribose rings of the TATA element. Somewhat of a hybrid, PurR (like CAP) uses an α helix to induce and stabilize its kink site, but like TBP, interacts in the minor groove by analogous van der Waals contacts to the base and deoxyribose ring. Nuclear magnetic resonance (NMR) studies on the sex determining factor SRY, bound to a high-affinity DNA site, suggest that this HMG protein interacts with the minor groove in a manner similar to PurR whereby an isoleucine side chain appears to intercalate into a TpT base pair step (28).

The energetic compensation for kinking the DNA and unstacking the most favorable base pair step (29) comes, in large part, from the extensive hinge helix-minor groove and recognition helix-major groove interactions and the entropy gain related to freeing waters of hydration from the DNA. Another contribution to the free energy of specific binding likely arises from the local folding of the hinge helix upon *purF* operator binding. The thermodynamics of DNA binding-induced protein folding have been described for several protein-DNA complexes and such folding contributes substantially to DNA binding affinity (30). Evidence that the hinge helix of PurR is unfolded in the absence of operator DNA is given by its susceptibility to rapid cleavage by a variety of proteases (5, 22).

#### Operator specificity of LacI members.

Inspection of a sequence alignment of 21 LacI members reveals that leucine is found at the position that corresponds to 54 in PurR in all but the cytidine repressor, CytR, which has a valine (3). Also conserved is the alanine corresponding to Ala<sup>51</sup> in PurR, again with the lone exception of CytR, which substitutes a glycine. From inspection of the PurR-*purF* operator complex it is evident that this position can be only an alanine or glycine because residue 51 directly abuts the DNA phosphate backbone (Fig. 6D). The nearly complete conservation of these key residues of the hinge helix, combined with the conservation of a central CpG base-pair step in all LacI family oper-

ators and biochemical studies demonstrating that several LacI family members bend their operators (31), suggests similar modes of minor groove binding by all LacI family members. However, modification of the operator site can alter binding affinity. A case in point is the wild-type *lac* operator, which contains an additional central C-G base pair that when deleted, binds LacI eight to ten times more tightly (32).

Operator discrimination by LacI family members relies on DNA deformability and base specific contacts. Sequence alignment of the recognition helices and position 55 of ten LacI proteins and their consensus operators (3) provides insight into the basis of DNA recognition by the LacI family (Fig. 7). In PurR, the second residue of the recognition helix, Thr<sup>16</sup>, hydrogen bonds to an A·T base pair at position 6, whereas in LacI the corresponding residue, Gln<sup>18</sup>, contacts a G·C base pair also at position 6 (10, 12) (Fig. 7). Of the eight remaining LacI members, only threonine, serine, and alanine are found and base pair 6 of their respective consensus operators is always an A·T, with the exception of the ribitol repressor, RbtR, which binds operators containing either C·G or A·T (Fig. 7). This correlation suggests that alanine, serine, and threonine strongly prefer an A·T pair at position 6 but that glutamine can bind only a G·C base pair. The identity of base pair 4 is determined directly or indirectly by the sixth residue of the recognition helix. In LacI, this residue is an arginine that inter-

**Table 2.** DNA helical parameters of the *purF* operator half-site. Inter-base pair and intra-base pair parameters of the *purF* operator site calculated with CURVES (40). Inter-base pair parameters refer to those between base pairs within the same strand and intra-base pair parameters refer to those within a base pair. Typical roll angles, rise values, and twist angles are 0°, 3.38 Å, and 34.3° for B-DNA and 0°, 2.56 Å, and 32.7° for A-DNA (41, 42). Minor groove width is defined as the shortest O4'-O4' distance minus 2.8 Å (the van der Waals radii of two oxygen atoms). The average minor groove width for B-DNA is 5.7 Å and for A-DNA is 11.7 Å (41).

| Base pair | Inter-base pair   |          |          | Slide (Å) | Intra-base pair     |            |        |
|-----------|-------------------|----------|----------|-----------|---------------------|------------|--------|
|           | Helical twist (°) | Roll (°) | Rise (Å) |           | Propeller twist (°) | Buckle (°) |        |
| 2         | A·T               | 5.69     | 14.36    | 3.60      | -1.02               | 7.27       | -19.20 |
| 3         | C·G               | 38.68    | -3.14    | 3.47      | -0.60               | -2.18      | -0.28  |
| 4         | G·C               | 49.69    | 2.87     | 3.48      | 0.35                | 1.01       | -0.49  |
| 5         | A·T               | 21.80    | -1.11    | 3.68      | -0.13               | -16.12     | 6.96   |
| 6         | A·T               | 41.84    | 0.71     | 3.52      | -0.01               | -1.74      | 3.59   |
| 7         | A·T               | 39.93    | 4.57     | 3.60      | -0.49               | -18.87     | 1.95   |
| 8         | A·T               | 16.93    | -1.65    | 2.84      | -1.17               | -22.93     | 7.17   |
| 9         | C·G               | 27.22    | 44.86    | 6.45      | 0.65                | -3.99      | 21.38  |
| 9'        | G·C               |          |          |           |                     | -3.99      | 21.38  |

| Minor groove width (Å) of <i>purF</i> operator site |       |       |       |       |       |       |       |       |       |
|---|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| T(1)  | A(2)  | C(3)  | G(4)  | A(5)  | A(6)  | A(7)  | A(8)  | C(9)  | G(9') |
|   | 9.26  | 10.31 | 9.46  | 8.50  | 7.80  | 8.98  | 9.75  | 9.20  | 9.20  |
|   | T(2') | G(3') | C(4') | T(5') | T(6') | T(7') | T(8') | G(9') | C(9)  |

acts with Gua<sup>4</sup> (10, 12). Other LacI family members that have an arginine at position 6, have either a G·C or C·G base pair at position 4 (Fig. 7). PurR recruits Arg<sup>26</sup> from the loop following the recognition helix to interact with Gua<sup>4</sup>, an interaction that is directed by His<sup>20</sup> located at position six of the recognition helix (Figs. 6A and 7). Hydrophobic residues at position six appear to select for a T·A base pair at operator position 4 (Fig. 7).

A third DNA discriminating contact is made by the hinge-helix residue that corresponds to Lys<sup>55</sup> in PurR. Lys55 contacts C2 and N3 of Ade8 (Fig. 6, A and D). Other LacI members that have a basic residue at this position also discriminate against a G·C

base pair at position 8, most likely the result of steric clash between the side chain and the guanine N2 exocyclic group. However, when this residue is small, as in GalR and LacI, there is no discrimination against G·C at position 8.

**Mapping LacI mutants.** Biochemical and genetic studies, which have resulted in the generation of more than 4000 LacI mutants (33), have provided insight into the structure and function correlates of this repressor in the absence of a complete three-dimensional structure. The phenotypic characterization of LacI mutants identifies (i) repressors defective in operator binding (I<sup>-</sup>), (ii) repressors defective in effector binding (I<sup>s</sup>), (iii) repressors that are tight

operator binders (I<sup>tb</sup>), and (iv) repressors defective in operator and effector binding.

Now that the structure of an intact LacI member is available we can begin to ascribe the structural basis for these mutant classes. When the LacI sequence is appropriately imposed on the PurR structure, it is evident that many of the mutations are clustered (Fig. 8). Most I<sup>-</sup> mutations are located in the DNA-binding domain with residues of the HTH and hinge helix being particularly sensitive to substitution (Fig. 8, blue). However, mutations leading to incorrect secondary, tertiary, or quaternary structures also lead to the I<sup>-</sup> phenotype. Many such substitutions are located in the effector binding domain and frequently map to  $\beta$  sheets that constitute the structural core of the protein or to the interior facing hydrophobic regions of helices.

The I<sup>s</sup> mutations are clustered mainly in or around the ligand binding cleft (Fig. 8, magenta), for example Ser<sup>193</sup>, which corresponds to a PurR residue Thr<sup>192</sup> (Fig. 4). Other mutants of this type are found in the dimerization interface in positions where substitutions likely perturb the opening and closing of the ligand binding pocket. The I<sup>tb</sup> mutants are confined to LacI residues Val<sup>24</sup>, Ser<sup>28</sup>, Val<sup>52</sup>, and Ser<sup>61</sup> and correspond to PurR residues Ile<sup>22</sup>, Arg<sup>26</sup>, Val<sup>50</sup>, and Thr<sup>59</sup> (Fig. 8, green). Of these residues all but Thr<sup>59</sup>, which is located in the short loop connecting the hinge helix to the CBD, are found in the DNA-binding domain of PurR. The LacI mutants, which display the I<sup>-</sup> and I<sup>s</sup> phenotypes, are tightly clustered, and most are located in the corepressor binding pocket, including those corresponding to PurR residues Ser<sup>191</sup>, Arg<sup>196</sup>, and Asp<sup>275</sup>, or in the dimerization interface, such as those corresponding to PurR residues Cys<sup>281</sup> and Tyr<sup>282</sup>, (Fig. 8, yellow).

In conclusion, the three-dimensional structure of the PurR-hypoxanthine-*purF* operator complex has revealed the atomic details of minor groove binding by  $\alpha$  helices and the mechanism by which the side chains of aliphatic amino acids can intercalate into a base pair step and kink its DNA binding site. Furthermore, the structure has broadened our understanding of the DNA binding specificities of the entire LacI family.

## REFERENCES AND NOTES

1. H. Zalkin and J. E. Dixon, *Prog. Nucleic Acid Res. Mol. Biol.* **42**, 259 (1992).
2. R. J. Rolles and H. Zalkin, *J. Bacteriol.* **172**, 5758 (1990); L. Meng, M. Kilstrup, P. Nygaard, *Eur. J. Biochem.* **187**, 373 (1990).
3. J. D. Gralla, in *Transcriptional Regulation 2*, S. L. McKnight and K. R. Yamamoto, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1992), p. 629; R. J. Rolles and H. Zalkin, *J. Biol. Chem.* **263**, 19653 (1988); M. J. Weichert and S. Adhya, *ibid.* **267**, 15869 (1992).

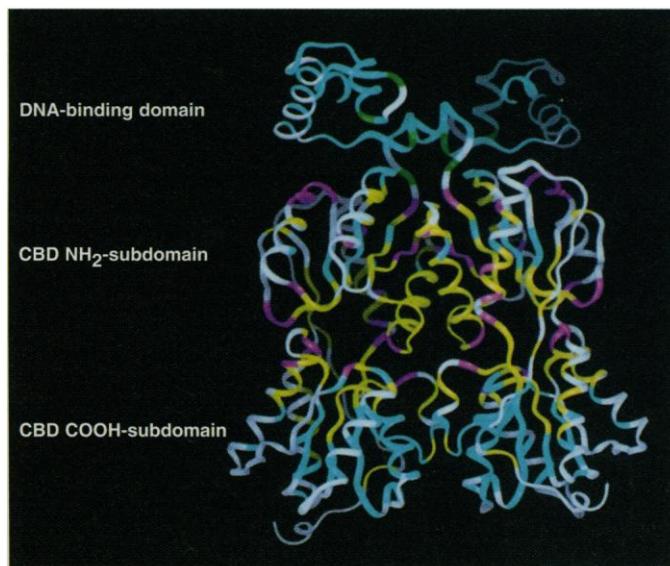
**Fig. 7.** LacI family member-operator site specificity. Residues of ten LacI family members and the base pairs they contact or are predicted to contact are shaded similarly. Specifically, residue 2 from helix 2, the recognition helix, contacts base pair 6; residue 6 from helix 2 contacts base pair 4; and LacI family member residues corresponding to PurR residue Lys<sup>55</sup> (indicated by asterisk) contact base pair 8. Upper, consensus operator half-sites from LacI family members, the operators of which have been well characterized. Q, any base; O, Ade or Cyt; R, Gua or Thy; S, Ade or Thy; Y, Cyt or Thy. The four-letter abbreviations for the LacI family members not provided in the text are: the galactose isorepressor (GalS), amy-lase repressor (CcpA), maltose repressor (Mall), rhamnose repressor, and sucrose repressor (ScrR). Lower, sequences of the putative or known HTH motifs of these LacI family members.

|      |    |   |   |   |   |   |   |
|------|----|---|---|---|---|---|---|
|      |    |   |   | 4 | 6 | 8 |   |
| PurR | 5' | A | C | G | C | A | A |
| GalR | 5' | G | T | G | R | A | A |
| GalS | 5' | G | T | G | R | A | A |
| LacI | 5' | T | T | G | T | G | A |
| CcpA | 5' | T | G | T |   | A | A |
| CytR | 5' | A | T | G | C | G | A |
| Mall | 5' | G | A | T | A | A | A |
| RafR | 5' | A | S | C | C | G | A |
| RbtR | 5' | G | C | T | Y | A | A |
| ScrR | 5' | C | T | A | A | A | C |

|      |   |             |         |       |                 |
|------|---|-------------|---------|-------|-----------------|
|      |   | Helix       | Turn    | Helix |                 |
|      |   |             |         | 2     | 6               |
| PurR | I | K D V A K R | A N V S | T     | T V S H V I N   |
| GalR | I | K D V A R L | A G V S | V     | T V S R V I N   |
| GalS | I | R D V A R Q | A G V S | V     | T V S R V L N   |
| LacI | L | Y D V A E Y | A G V S | Y     | T V S R V V N   |
| CcpA | I | Y D V R R E | A N V S | M     | T V S R A L N   |
| CytR | M | K D V A L K | A K V S | T     | A T V S R A L M |
| Mall | I | H D V A L A | A G V S | V     | T V S L V L S   |
| RafR | L | K A I A T T | L G I S | V     | T V S R A L G   |
| RbtR | I | Y D L A E L | S G V S | A     | A V S A I L N   |
| ScrR | I | K D I A E L | A G V S | K     | A T A S L V L V |

**Fig. 8.** LacI mutants mapped onto the three-dimensional structure of PurR [based on our previous sequence alignment (6)]. Labeled are the DNA-binding domain and the CBD NH<sub>2</sub>- and COOH-subdomains of PurR. The positions of I<sup>-</sup> mutants are blue, I<sup>s</sup> mutants are magenta, I<sup>tb</sup> are green, and mutations that lead to both the I<sup>-</sup> and I<sup>s</sup> phenotype are yellow.



4. R. Jaenicke, I. Muiznieks, C. Aslanidis, R. Schmitt, *FEBS Lett.* **260**, 233 (1990); S. Alberti, S. Oehler, B. von Wilcken-Bergmann, H. Kramer, B. Müller-Hill, *New Biol.* **3**, 57 (1991); A. Chakerian *et al.*, *J. Biol. Chem.* **266**, 1371 (1991); S. Alberti, S. Oehler, B. von Wilcken-Bergmann, B. Müller-Hill, *EMBO J.* **12**, 3227 (1993); W.-I. Chang, J. S. Olson, K. S. Matthews, *J. Biol. Chem.* **268**, 17613 (1993); J.-C. Cortay *et al.*, *ibid.* **269**, 14885 (1994).
5. K. Y. Choi and H. Zalkin, *J. Bacteriol.* **174**, 6207 (1992).
6. C. A. Mauzy and M. A. Hermodson, *Protein Sci.* **1**, 843 (1992); J. U. Bowie, R. Lüthy, D. Eisenberg, *Science* **253**, 164 (1991); N. B. Vartak *et al.*, *Res. Microbiol.* **142**, 951 (1991); M. A. Schumacher, J. R. Macdonald, J. Björkman, S. L. Mowbray, R. G. Brennan, *J. Biol. Chem.* **268**, 12282 (1993); C. S. Barbier and S. A. Short, *J. Bacteriol.* **174**, 2881 (1992); B. Müller-Hill, *Nature* **302**, 163 (1983).
7. J. C. Nichols, N. K. Vyas, F. A. Quijcho, K. S. Matthews, *J. Biol. Chem.* **268**, 17602 (1993).
8. M. Hsieh, P. Hensley, M. Brenowitz, J. S. Fetrow, *ibid.* **269**, 13825 (1994).
9. E. R. P. Zuiderweg, R. Kaptein, K. Wüthrich, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 5837 (1983); R. Kaptein, E. R. P. Zuiderweg, R. M. Scheek, R. Boelens, W. F. van Gunsteren, *J. Mol. Biol.* **182**, 179 (1985).
10. V. P. Chuprina *et al.*, *J. Mol. Biol.* **234**, 446 (1993); R. M. J. N. Lamerichs *et al.*, *Biochem.* **28**, 2985 (1989).
11. R. G. Brennan, *Curr. Opin. Struct. Biol.* **1**, 80 (1991); S. C. Harrison and A. K. Aggarwal, *Annu. Rev. Biochem.* **59**, 933 (1990); R. G. Brennan and B. W. Matthews, *J. Biol. Chem.* **264**, 1903 (1989); T. A. Steitz, *Q. Rev. Biophys.* **23**, 205 (1990); C. O. Pabo and R. T. Sauer, *Annu. Rev. Biochem.* **61**, 1053 (1992).
12. N. Lehming, J. Sartorius, S. Oehler, B. von Wilcken-Bergmann, B. Müller-Hill, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 7947 (1988).
13. M. A. Schumacher and R. G. Brennan, unpublished results.
14. R. A. Laskowski, M. W. MacArthur, D. S. Moss, J. M. Thornton, *J. Appl. Crystallogr.* **26**, 283 (1993).
15. A. J. Sharf, L. E. Rodseth, J. C. Spurlino, F. A. Quijcho, *Biochem.* **31**, 10657 (1992); G. A. Olah, S. Trahanov, J. Trehwella, F. A. Quijcho, *J. Biol. Chem.* **268**, 16241 (1993); B.-H. Oh *et al.*, *ibid.*, p. 11348.
16. F. A. Quijcho, *Curr. Opin. Struct. Biol.* **1**, 922 (1991); N. K. Vyas, *ibid.*, p. 732.
17. S. L. Mowbray and L. B. Cole, *J. Mol. Biol.* **225**, 155 (1992).
18. K. Y. Choi, F. Lu, H. Zalkin, *J. Biol. Chem.* **269**, 24066 (1994).
19. C. O. Pabo *et al.*, *Science* **247**, 1210 (1990).
20. F. Lu and H. Zalkin, unpublished results.
21. B. He, A. Shiau, K. Y. Choi, H. Zalkin, J. M. Smith, *J. Bacteriol.* **172**, 4555 (1990); B. He, J. M. Smith, H. Zalkin, *ibid.* **174**, 130 (1992); B. He and H. Zalkin, *ibid.*, p. 7121; B. He, K. Y. Choi, H. Zalkin, *ibid.* **175**, 3598 (1993); B. He and H. Zalkin, *ibid.* **176**, 1009 (1994).
22. K. Y. Choi and H. Zalkin, *ibid.*, p. 1767.
23. J. C. Wang, M. D. Barkley, S. Bourgeois, *Nature* **251**, 247 (1974).
24. A. A. Travers, *Curr. Opin. Struct. Biol.* **1**, 114 (1991); *ibid.* **2**, 71 (1992); P. J. Hagerman, *Biochim. Biophys. Acta* **1131**, 125 (1992).
25. S. C. Schultz, G. C. Shields, T. A. Steitz, *Science* **253**, 1001 (1991).
26. J. L. Kim, D. B. Nikolov, S. K. Burley, *Nature* **365**, 520 (1993); J. L. Kim and S. K. Burley, *Nature, Struct. Biol.* **1**, 638 (1994).
27. Y. Kim, J. H. Geiger, S. Hahn, P. B. Sigler, *Nature* **365**, 512 (1993).
28. C.-Y. King and M. A. Weiss, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 11990 (1993).
29. R. L. Ornstein, R. Rein, D. L. Breen, R. D. MacElroy, *Biopolymers* **17**, 2341 (1978).
30. R. S. Spolar and M. T. Record Jr., *Science* **263**, 777 (1994).
31. C. Zwieb, J. Kim, S. Adhya, *Genes Dev.* **3**, 606 (1989); J. Kim, C. Zwieb, C. Wu, S. Adhya, *Gene* **85**, 15 (1989); F. Lu and H. Zalkin, unpublished results.
32. J. R. Sadler, H. Sasmor, J. L. Betz, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 6785 (1983); A. Simons, D. Tils, B. von Wilcken-Bergmann, B. Müller-Hill, *ibid.* **81**, 1624 (1984).
33. R. O. Spotts, A. E. Chakerian, K. S. Matthews, *J. Biol. Chem.* **266**, 22998 (1991); W.-I. Chang, P. Barrera, K. S. Matthews, *Biochemistry* **33**, 3607 (1994); L. G. Kleina and J. H. Miller, *J. Mol. Biol.* **212**, 295 (1990); A. J. E. Gordon *et al.*, *ibid.* **200**, 239 (1988); A. E. Chakerian and K. S. Matthews, *J. Biol. Chem.* **266**, 22206 (1991); P. Markiewicz, L. G. Kleina, C. Cruz, S. Ehret, J. H. Miller, *J. Mol. Biol.* **240**, 421 (1994).
34. M. A. Schumacher, K. Y. Choi, H. Zalkin, R. G. Brennan, *J. Mol. Biol.* **242**, 302 (1994).
35. N. h. Xuong, C. Nielsen, R. Hamlin, D. Anderson, *J. Appl. Crystallogr.* **18**, 342 (1985).
36. T. C. Terwilliger and D. Eisenberg, *Acta Crystallogr.* **A39**, 813 (1983).
37. B.-C. Wang, *Methods Enzymol.* **115**, part B, 90 (1985).
38. T. A. Jones, *ibid.*, p. 157.
39. D. E. Tronrud, L. J. Ten Eyck, B. W. Matthews, *Acta Cryst.* **A43**, 489 (1987).
40. R. Lavery and H. Sklenar, *J. Biomol. Struct. Dynam.* **6**, 63 (1988).
41. G. Ravishanker *et al.*, *J. Biomol. Struct. Dynam.* **6**, 669 (1989).
42. W. Saenger, *Principles of Nucleic Acid Structure* (Springer-Verlag, New York, 1984).
43. We thank D. L. Beveridge and M. Young for help in the analysis of the DNA and P. A. Stange for technical assistance. Supported by March of Dimes grant 505 and by Public Health Service grant GM 49244 (R.G.B.), by a National Defense Science and Engineering Graduate Fellowship in Biosciences (M.A.S.), and by Public Health Service grant GM 24658 (H.Z.). The coordinates are being deposited in the Brookhaven Protein Data Bank and are available from R.G.B.

29 June 1994; accepted 12 September 1994