2008

Structural Basis of the Transcriptional Regulation of the Proline Utilization Regulon by Multifunctional PutA

Yuzhen Zhou
University of Nebraska - Lincoln

John D. Larson
University of Missouri-Columbia, Columbia, MO

Christopher A. Bottoms
University of Missouri-Columbia, Columbia, MO

Emilia C. Arturo
University of Missouri-Columbia, Columbia, MO

Michael T. Henzl
University of Missouri-Columbia, Columbia, MO

See next page for additional authors

doi:10.1016/j.jmb.2008.05.084 J. Mol. Biol. (2008) 381, 174–188

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Structural Basis of the Transcriptional Regulation of the Proline Utilization Regulon by Multifunctional PutA

Yuzhen Zhou[1]†, John D. Larson[2]†, Christopher A. Bottoms[3], Emilia C. Arturo[2], Michael T. Henzl[4], Jermaine L. Jenkins[4], Jay C. Nix[5], Donald F. Becker[1]* and John J. Tanner[2,4]*

[1]*Department of Biochemistry, University of Nebraska-Lincoln, Lincoln, NE 68588, USA*

[2]*Department of Chemistry, University of Missouri-Columbia, Columbia, MO 65211, USA*

[3]*Department of Computer Science, University of Missouri-Columbia, Columbia, MO 65211, USA*

[4]*Department of Biochemistry, University of Missouri-Columbia, Columbia, MO 65211, USA*

[5]*Molecular Biology Consortium, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*

Received 1 April 2008; received in revised form 22 May 2008; accepted 31 May 2008
Available online 7 June 2008

Edited by J. E. Ladbury

The multifunctional *Escherichia coli* proline utilization A (PutA) flavoprotein functions both as a membrane-associated proline catabolic enzyme and as a transcriptional repressor of the proline utilization genes *putA* and *putP*. To better understand the mechanism of transcriptional regulation by PutA, we have mapped the *put*-regulatory region, determined a crystal structure of the PutA ribbon–helix–helix domain (PutA52, a polypeptide corresponding to residues 1–52 of *E. coli* PutA) complexed with DNA, and examined the thermodynamics of DNA binding to PutA52. Five operator sites, each containing the sequence motif 5′-GTTGCA-3′, were identified using gel-shift analysis. Three of the sites are shown to be critical for repression of *putA*, whereas the two other sites are important for repression of *putP*. The 2.25-Å-resolution crystal structure of PutA52 bound to one of the operators (operator 2; 21 bp) shows that the protein contacts a 9-bp fragment corresponding to the GTTGCA consensus motif plus three flanking base pairs. Since the operator sequences differ in flanking bases, the structure implies that PutA may have different affinities for the five operators. This hypothesis was explored using isothermal titration calorimetry. The binding of PutA52 to operator 2 is exothermic, with an enthalpy of −1.8 kcal/mol and a dissociation constant of 210 nM. Substitution of the flanking bases of operator 4 into operator 2 results in an unfavorable enthalpy of 0.2 kcal/mol and a 15-fold-lower affinity, showing that base pairs outside of the consensus motif impact binding. Structural and thermodynamic data suggest that hydrogen bonds between Lys9 and bases adjacent to the GTTGCA motif contribute to transcriptional regulation by fine-tuning the affinity of PutA for *put* control operators.

Keywords: proline utilization A; X-ray crystallography; isothermal titration calorimetry; ribbon–helix–helix; proline catabolism

Introduction

Proline is used as a source of carbon, nitrogen, and energy through two oxidative steps catalyzed by proline dehydrogenase (PRODH) and Δ^1-pyrroline-5-carboxylate dehydrogenase (P5CDH).[1–7] In enteric bacteria such as *Escherichia coli*, proline utilization requires two genes, *putP* and *putA*. The former encodes the PutP high-affinity Na^+-proline transporter, and the latter encodes the multifunctional flavoprotein proline utilization A (PutA).[8,9] PutA is unique in that it functions both as a transcriptional repressor of the *put* genes and as a membrane-associated bifunctional proline catabolic enzyme.[2,10–12] The enzymatic and transport functions

*Corresponding authors. E-mail addresses: dbecker3@unl.edu; tannerjj@missouri.edu.

† Y.Z. and J.D.L. contributed equally to this research.

Abbreviations used: PutA, proline utilization A; PutA52, polypeptide corresponding to residues 1–52 of *Escherichia coli* PutA; PRODH, proline dehydrogenase; P5CDH, Δ^1-pyrroline-5-carboxylate dehydrogenase; RHH, ribbon–helix–helix; ITC, isothermal titration calorimetry; WT, wild type; PDB, Protein Data Bank; EDTA, ethylenediaminetetraacetic acid; PutA47, polypeptide corresponding to residues 1–47 of *Escherichia coli* PutA; dsDNA, double-stranded DNA.

of the *putA* and *putP* genes, respectively, are conserved among different Gram-negative bacteria, whereas the genetic organization and regulatory mechanisms that control the expression of these genes are highly divergent.[7,10–18] The focus of this work is to provide a molecular and structural understanding of the regulation of *put* genes in *E. coli* by PutA.

PutA from *E. coli* combines PRODH, P5CDH, and transcriptional regulatory activities into a single polypeptide of 1320 amino acids.[2,19] Insights into the organization of the functional domains in PutA have been gained from molecular dissection and characterization of truncated PutA proteins. The PRODH and P5CDH active sites are located within residues 261–612 and 650–1130, respectively, with the PRODH active site utilizing a FAD cofactor and with P5CDH activity requiring NAD^+. Structural studies have shown that the PRODH domain forms a unique $(\beta\alpha)_8$ barrel,[20,21] and that reduction by dithionite causes dramatic conformational changes in the FAD ribityl chain.[22] Molecular dissection studies showed that the DNA-binding domain is contained in residues 1–47.[23] Subsequently, the crystal structure of PutA52 (a polypeptide corresponding to residues 1–52 of *E. coli* PutA) was solved, showing that PutA is a member of the ribbon–helix–helix (RHH) family of transcriptional regulators.[23,24]

While knowledge of PutA structure and function continues to build, a considerable gap in our understanding of critical PutA–DNA interactions in the *put* control DNA region remains. To further understand the regulation of proline metabolism in *E. coli*, we have identified the PutA binding sites in the *put*-regulatory region, elucidated the roles of these operators in repressing the expression of *putA* and *putP*, determined the crystal structure of PutA52 bound to one of the identified operators, and investigated the thermodynamics of DNA binding to PutA52 using isothermal titration calorimetry (ITC).

Results

Identification of PutA binding sites

Initial localization of PutA binding sites in the *put* control DNA region was performed by gel mobility shift assays using different fragments of the 419-bp *put* control DNA. Systematic evaluation of different regions of the *put* control DNA indicated that PutA does not bind to the 1- to 170-bp region immediately downstream of *putP* (Fig. 1a, lanes 3 and 4). However, PutA was observed by gel mobility shift assays to bind to regions 183–231 and 342–412 of the *put* control DNA (data not shown). Additional assays indicated that PutA binds to oligonucleotides 183–210, 342–365, and 388–412 (Fig. 1a, lanes 5–10). Previously, we showed by gel mobility shift assays that PutA also binds to oligonucleotides 211–231, with an apparent binding stoichiometry of one DNA duplex per PutA dimer.[26]

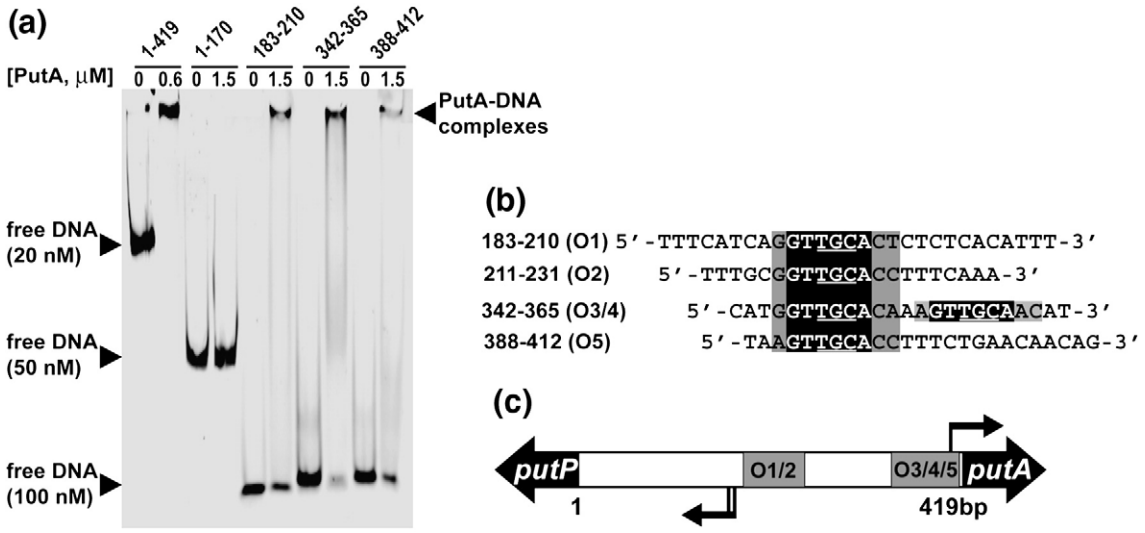

Fig. 1. Localization of PutA binding sites in the *put* control DNA region. (a) Gel mobility shift assays of PutA with different regions of *put* control DNA. Separate binding mixtures of PutA (0–1.5 μM) with full-length *put* control DNA 1–419 bp (20 nM), *put* control DNA region 1–170 bp (50 nM), and oligonucleotides (100 nM each) 183–210 bp, 342–365 bp, and 388–412 bp were incubated at 20 °C for 20 min in 50 mM Tris buffer (pH 7.5; 100 mM NaCl). The PutA–DNA complexes were separated by native polyacrylamide gel (4%) electrophoresis at 4 °C. The positions of the uncomplexed DNA and protein–DNA complexes are indicated. (b) Sequences of the five PutA binding sites. The consensus sequence GTTGCA is highlighted in black. Flanking nucleotides predicted to contact PutA based on the crystal structure are shaded in gray. The TGC sequence that was changed to CAT for mutational analysis is underlined. (c) Schematic representation of the intergenic *put* control region in *E. coli* and transcriptional directions of *putP* and *putA*. The black arrows show the transcription start sites determined previously.[25] The five PutA binding sites are labeled as O1, O2, O3, O4, and O5.

Sequence alignment of the four oligonucleotides that bound to PutA (oligonucleotides 183–210, 211–231, 342–365, and 388–412) revealed a GTTGCA consensus sequence. This motif is present in each of the four oligonucleotides (Fig. 1b), and it appears five times in the 183- to 412-bp region of the *put* control DNA (Fig. 1b). Thus, five potential operator sites denoted O1–O5 were proposed, as shown in Fig. 1c.

The proposed binding sites were further examined by changing each one from GTTGCA to GTCATA by site-directed mutagenesis of the *put* control DNA. Gel mobility shift assays show that simultaneously mutating all five sites disrupts PutA binding to the *put* control DNA (Fig. 2a, Δ12345), confirming that PutA specifically recognizes only the five binding sites in the *put* control region. Gel mobility shift assays were then used to test PutA binding to the five sites incrementally using PutA52 to resolve the different complexes. As shown in Fig. 2b, decreasing the number of binding sites in the *put* control DNA reduces the observed mobility shift of the protein–DNA complex. This further confirms that the *put* control DNA contains five PutA binding sites and suggests that PutA52 is able to bind all five sites simultaneously.

Autorepression of *putA*

Cell-based reporter gene assays were performed to test the role of each PutA binding site in repressing the expression of *putA*. For these assays, *E. coli* strain JT31 *putA*$^-$ *lacZ*$^-$ was cotransformed with PutA-pUC18 and P_{putA}:*lacZ*-pACYC184 constructs [wild-type (WT) and single or multiple operator site mutations in the *put* control DNA]. Western blot analysis confirmed the expression of PutA. Consistent with previous results, PutA repressed expression of the *lacZ* reporter gene by over 75% relative to control cells (pUC18 alone and WT P_{putA}:*lacZ* construct) (Fig. 3a, WT).[22] Mutations of O1 (Δ1) and O2 (ΔO2), singly or in combination (ΔO1–O2), did not increase β-galactosidase activity. Because PutA repression of the *lacZ* reporter gene (~73%) was not diminished by mutating operator sites O1 and O2, PutA binding to these sites is not necessary for repressing transcription of *putA*. Mutating O3 (ΔO3) greatly reduced *lacZ* reporter gene expression in the control cells (data not shown) to ~10% of WT *put* control DNA. Because of the intrinsically low reporter gene expression of the ΔO3 mutant construct, we were not able to directly assess the impact of site O3 on PutA repression of *putA*. O3 is located in the −35 region of the *putA* promoter (Fig. 1c); thus, the mutation at site O3 most likely decreases the binding of the σ subunit of *E. coli* RNA polymerase to the −35 element. We thus consider O3 to be an important operator for autorepression of *putA*, despite the fact that we could not test its role using the reporter gene assay. Mutating site O4 (ΔO4) or O5 (ΔO5) increased β-galactosidase activity and lowered repression of the *lacZ* reporter gene to about 50% relative to the control cells (Fig. 3a). Simultaneously mutating O4 and O5 (ΔO4–O5) generated an additive effect with a 3-fold increase in β-galactosidase activity relative to WT, resulting in only 20%

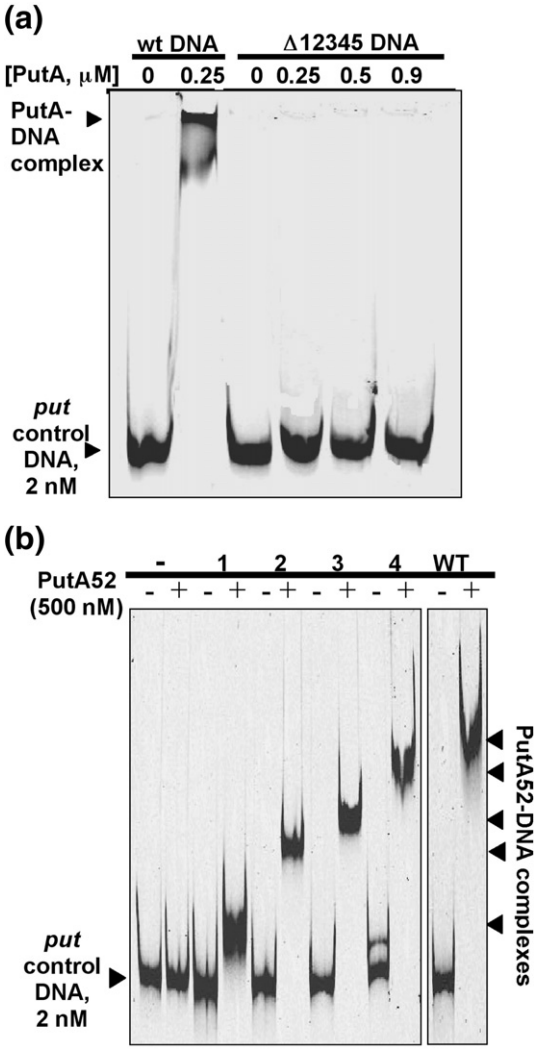

Fig. 2. Gel mobility shift assays of full-length PutA and PutA52 with WT *put* control DNA and *put* control DNA with an increasing number of mutated binding sites. (a) Separate binding mixtures of full-length PutA (0–0.25 μM) with WT *put* control DNA, and of full-length PutA (0–0.9 μM) with mutant *put* control DNA (Δ12345) were incubated at 20 °C for 20 min in 50 mM Tris buffer (pH 7.5; 250 mM NaCl) containing calf thymus DNA (100 μg/ml). The PutA–DNA complexes were separated using a native polyacrylamide gel (4%) at 4 °C. (b) PutA52 (500 nM) was incubated in separate binding mixtures with WT *put* control DNA (WT) and *put* control DNA with: (–) no available binding sites, Δ1–5; (1) one available binding site, Δ1–4; (2) two available sites, Δ3–5; (3) three available sites, Δ12; and (4) four available sites, Δ5. Binding reactions were carried out in 50 mM Tris buffer (pH 7.5; 50 mM NaCl) at 20 °C in the presence of calf thymus competitor DNA (100 μg/ml). The complexes were separated by native polyacrylamide gel (15%) electrophoresis at 4 °C. The positions of the uncomplexed DNA and protein–DNA complexes are indicated. TRdyed-700-labeled *put* control DNA was used for all of the above assays.

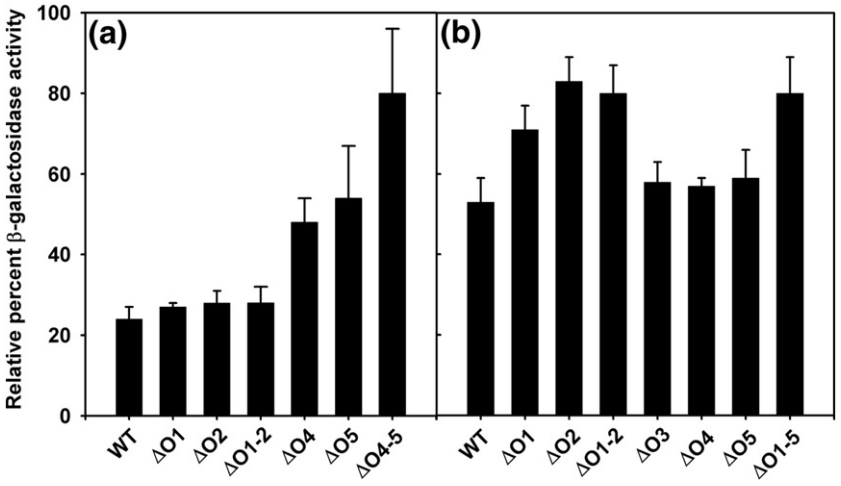

Fig. 3. β-Galactosidase activity from *lacZ* reporter constructs containing various mutations in the *put* control DNA. (a) Relative percent β-galactosidase activity in *E. coli* strain JT31 containing PutA-pUC18, WT P_{putA}:*lacZ*, and various P_{putA}:*lacZ* mutant constructs. (b) Relative percent β-galactosidase activity in *E. coli* strain JT31 containing PutA-pUC18, WT P_{putP}:*lacZ*, and various P_{putP}:*lacZ* mutant constructs. Percent β-galactosidase activity is the β-galactosidase activity from JT31 cells expressing PutA and the various *lacZ* reporter constructs relative to β-galactosidase activity from JT31 control cells containing only the matching *lacZ* reporter construct (WT or mutant) and pUC18 alone (no PutA). Percent activity data are reported as mean±standard error of the mean of four independent experiments. For the P_{putA}:*lacZ* reporter construct, β-galactosidase activity ranged from 700±80 mU for WT P_{putA}:*lacZ* to 2100±420 mU for the ΔO4–O5 P_{putA}:*lacZ* mutant construct. For P_{putP}:*lacZ*, the range of β-galactosidase activity was from 1220±130 mU to 1900±140 mU for WT P_{putP}:*lacZ* and P_{putP}:*lacZ* mutant ΔO2 construct, respectively.

repression of the *lacZ* reporter gene. Thus, O3, O4, and O5 are the most critical sites for PutA autorepression of *putA*.

Regulation of *putP* by PutA

The binding sites that are critical for regulating *putP* were also identified. In these assays, *E. coli* strain JT31 *putA*$^-$ *lacZ*$^-$ was cotransformed with the PutA-pUC18 construct and the P_{putP}:*lacZ*-pACYC184 construct (WT and single or multiple operator site mutations in the *put* control DNA). These results are shown in Fig. 3b. PutA repressed *lacZ* reporter gene expression by about 47% relative to control cells (Fig. 3b, WT). Apparently, PutA represses *putP* promoter activity less than the *putA* promoter, consistent with previous results suggesting that PutA is a stronger regulator of *putA* than *putP*.[25,27] Mutation of O1 (ΔO1) increased β-galactosidase activity, thereby decreasing the repression of the *lacZ* reporter gene to about 30% relative to the control cells. Mutating O2 singly (ΔO2) or in combination with O1 (ΔO1–O2) resulted in about 20% repression relative to the control cells. In contrast to the *putA* promoter, mutation of O3, O4, and O5 individually (ΔO3, ΔO4, and ΔO5) (Fig. 3b) or in combination (ΔO3–O5) (data not shown) did not significantly increase β-galactosidase activity or alter the repression of the *lacZ* reporter gene. Mutating all five binding sites (ΔO1–O5) resulted in the same repression of *lacZ* expression (20%) as ΔO1–O2 *put* control DNA (Fig. 3b). Thus, PutA binding to O1 and O2 is responsible for repressing the *putP* promoter.

Overall structure of PutA52 bound to O2

The crystal structure of PutA52 bound to O2 was solved in order to understand the three-dimensional structural basis of DNA recognition by PutA. This structure is the first one of a PutA RHH domain bound to DNA, and it is currently the highest-resolution structure of a RHH–DNA complex. The asymmetric unit contains one PutA52 dimer bound to one O2 duplex (Fig. 4).

Each PutA52 chain adopts the RHH fold, which consists of a β-strand (β1), followed by two α-helices (αA and αB). The two protein chains assemble into a dimer featuring an intermolecular two-stranded antiparallel β-sheet (Fig. 4).

The bound DNA ligand adopts the B conformation, based on analysis of projected phosphorus positions (z_P) using 3DNA.[28] Values of z_P<0.5 are diagnostic of B-form DNA, whereas values of z_P>1.5 Å indicate A-form DNA.[28,29] All but three of the 17-bp steps of O2 have z_P<0.5 Å. The three exceptions have z_P=0.52–0.58 Å. Thus, binding of PutA52 to O2 does not cause significant distortion of the DNA from the expected B conformation. Also, the double helix displays no discernable curvature (Fig. 4).

The β-sheet of PutA52 inserts into the DNA major groove (Fig. 4). Residues of the sheet contact DNA bases, while residues near the N-terminus of αB interact with the DNA backbone. This general mode of binding is typical of RHH proteins.[30]

Although the five operators that we identified each contain the 6-bp consensus sequence of GTTGCA (Fig. 1b), the structure shows that PutA52 contacts a larger fragment of DNA. A plot of the surface area buried by nucleotides in the protein–DNA interface is shown in Fig. 5a. The bimodal shape of the plot reflects the 2-fold symmetries of the protein dimer and the DNA double helix. The surface area calculations, along with detailed inspection of the protein–DNA interface, show that the footprint of PutA52 encompasses the 9-bp fragment from G6:C16 to C14:G8 (see boxed base pairs in Fig. 4). Note that this fragment contains the GTTGCA motif. Interactions

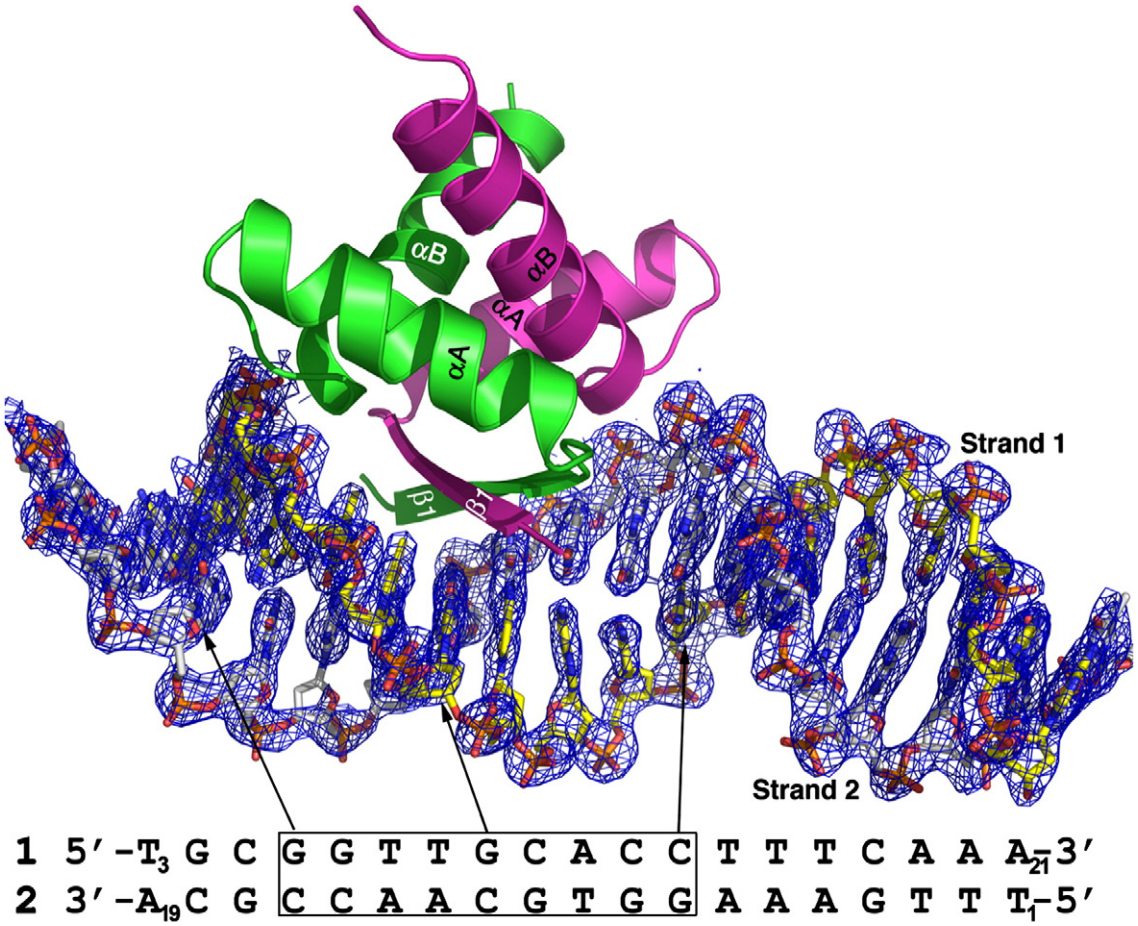

Fig. 4. Overall structure of PutA52 bound to O2. PutA52 chains A and B are in green and magenta, respectively. DNA is represented with sticks, with strand 1 in yellow and strand 2 in white. The electron density map is a $2F_o - F_c$ map with F_c and phases calculated from the final model. The contour level is 1σ. The box encloses the sequence of the 9-bp fragment contacted by PutA52.

with the 9-bp fragment are summarized schematically in Fig. 5b and shown in detail in Fig. 6.

Interactions with DNA bases

Structures of RHH domains bound to DNA show that, typically, two polar residues and one Arg/Lys from each β-strand form hydrogen bonds to DNA bases. In PutA, this critical triad corresponds to Thr5, Gly7, and Lys9, and all three residues interact with DNA bases. We note that these residues are identically conserved among PutAs.[24]

Lys9 binds to the pair of guanine bases located at the 5′ ends of each strand (Fig. 5b). Lys9 of chain A interacts with the guanine bases of strand 2, while Lys9 of the B chain interacts with the guanine bases of strand 1. The two sets of interactions are nearly identical (Fig. 6); this is expected, since they involve the palindromic ends of the DNA fragment. Each Lys9 forms four hydrogen bonds—two with each base of the guanine pair. These interactions are shown for Lys9(B) in Fig. 7a. The hydrogen bond distances are 2.5–3.1 Å for the inner base (G7 of strand 1; G9 of strand 2) and 3.2–3.5 Å for the outer base (G6 of strand 1; G8 of strand 2). We note that only guanine has two appropriately placed hydrogen bond acceptors for interaction with Lys9, so these interactions appear to enforce a preference for binding a 9-bp fragment containing GG at the 5′ ends of both strands.

Thr5 forms hydrogen bonds with three different base pairs and both DNA strands. In chain A, the hydroxyl of Thr5 donates a hydrogen bond to T8 of strand 1 (Fig. 7a), while the backbone carbonyl accepts a hydrogen bond from C12 of strand 2 (Fig. 7b). Since the hydrogen bond with T8 involves the palindromic GGT end of the DNA, one might expect Thr5(B) to form an analogous interaction with T10 of strand 2. Interestingly, Thr5(B) accepts a hydrogen bond from C11 of strand 1 (Fig. 7b), rather than hydrogen bonding with T10. The expected 2-fold symmetry is broken by a conformational change of Thr5(B). The χ_1 angle of Thr5(B) is +60°, whereas this angle is −60° for Thr5(A). We note that Thr5 has $\chi_1 = -60°$ in all chains of ligand-free PutA52 structures [Protein Data Bank (PDB) codes 2AY0 and 2GPE]. Thus, binding to DNA induced a conformational change in Thr5(B), which introduces asymmetry in PutA52.

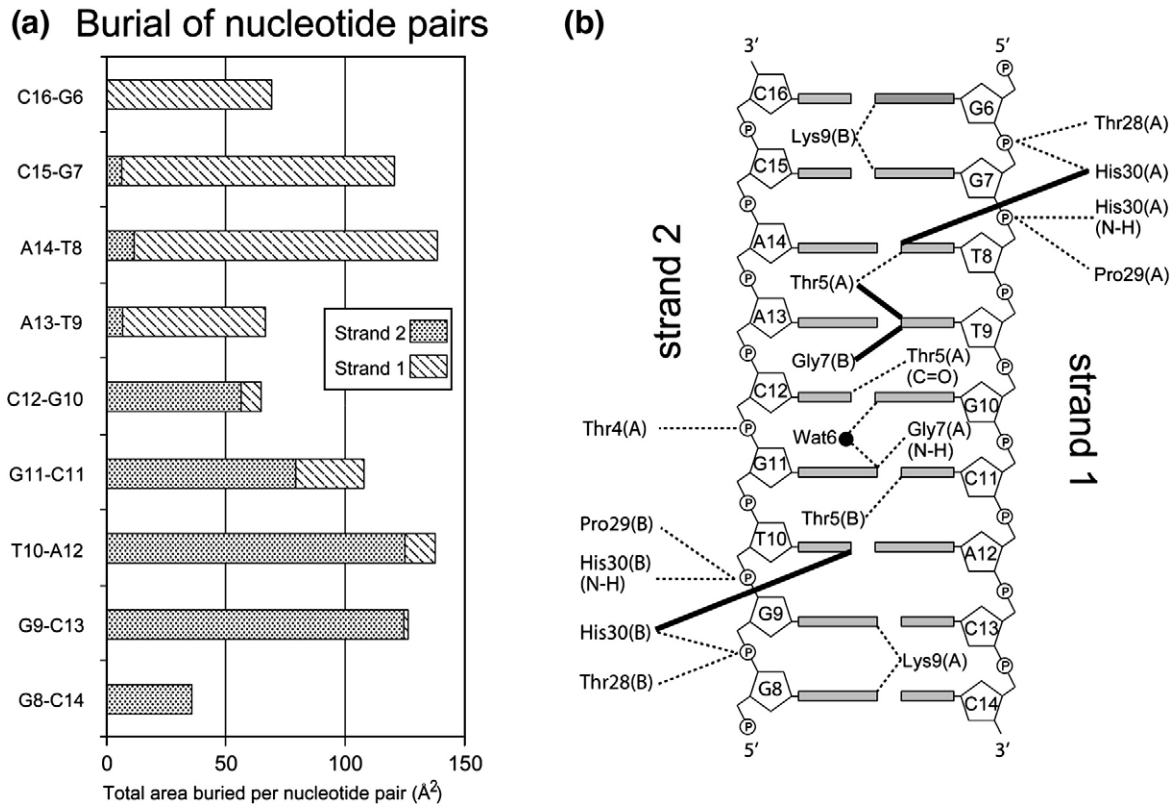

Fig. 5. Footprint of PutA52 on O2 derived from the crystal structure. (a) Surface area contributed by DNA nucleotides to the protein–DNA interface. (b) Schematic diagram of protein–DNA interactions. Dotted lines indicate electrostatic interactions. Thick solid lines denote steric interactions.

Gly7 helps confer sequence specificity, despite lacking a side chain. In chain A, Gly7 donates a hydrogen bond (2.9 Å) to the N7 atom of G11 (Fig. 7b). In chain B, Gly7 forms van der Waals interactions with the C5 methyl of T9 (Fig. 7c). Note also the close contacts between DNA bases and Thr5(A) in this

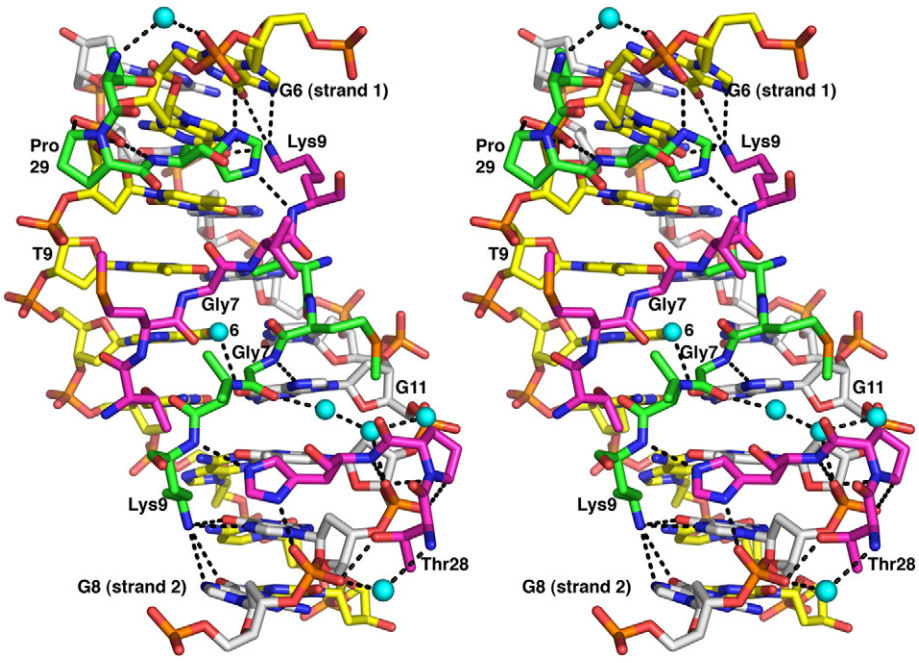

Fig. 6. Detailed stereographic view of the protein–DNA interface. PutA52 chains A and B are in green and magenta, respectively. DNA strands 1 and 2 are in yellow and white, respectively. Cyan spheres represent water molecules. Dotted lines indicate electrostatic interactions.

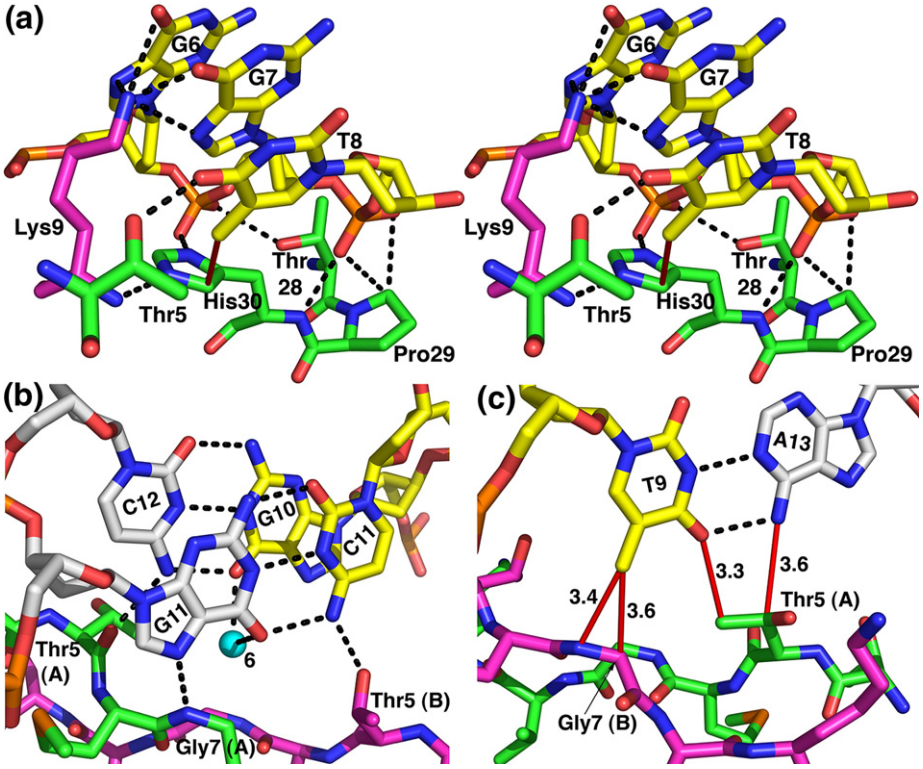

Fig. 7. Close-up views of selected protein–DNA interactions. In all three panels, PutA52 chains A and B are in green and magenta, respectively; DNA strands 1 and 2 are in yellow and white, respectively; and black dotted lines indicate electrostatic interactions. (a) Stereographic view of interactions formed by the GGT ends of the 9-bp DNA fragment. The red dotted line denotes steric contact between His30 and T8. (b) Close-up view of interactions involving the two base pairs, G10:C:12 and C11:G11. The cyan sphere represents a water molecule. (c) Close-up view of van der Waals interactions formed between the protein and the bases of T9 and A13. The red lines denote these interactions.

region of the structure (Fig. 7c). The tight packing of the T9:A13 base pair against Gly7 and Thr5 could contribute to sequence specificity.

Finally, there are no water molecules bridging the protein with DNA bases. There is, however, one water molecule (Wat6) that is strategically located at the protein–DNA interface on the pseudo 2-fold axis that relates the two chains (Fig. 6). It is equidistant from the two Gly7 residues of the β-sheet, and forms hydrogen bonds with G10 of DNA strand 1 and G11 of strand 2 (Fig. 7b). Wat6 appears to fill the void created by the lack of a side chain at residue 7. Indeed, mutation of Gly7 *in silico* to any other residue causes steric clash with this water molecule, as well as with DNA bases.

Interactions with the DNA backbone

Thr28, Pro29, and His30 bind the DNA backbone. Thr28 is the N_{cap} of αB, while Pro29 and His30 are the first two residues of αB. The interactions display nearly perfect 2-fold symmetry (Fig. 6), so just one set of interactions will be described. The side chains of Thr28 and His30 form electrostatic interactions with the phosphate group connecting the two G nucleotides at the 5′ end of the 9-bp fragment (Fig. 7a). In addition, the backbone of His30 donates a hydrogen bond to the phosphate group of the T nucleotide at the 5′ end of the 9-bp fragment (T8 of strand 1; T10 of strand 2; see Fig. 7a). Finally, the C^{δ} atom of Pro29 forms close contacts (3.4 Å) with oxygen atoms of the phosphate backbone (Fig. 7a).

ITC

The binding of O2 to PutA52 at pH 8.0 was studied using ITC to gain insights into the thermodynamic basis of DNA recognition. In Tris buffer, the association reaction was evidently endothermic (Fig. 8a), whereas in phosphate buffer at the same pH, the reaction was weakly exothermic (Fig. 8b). Since the enthalpy of ionization of Tris (11 kcal/mol) differs substantially from that of dihydrogen phosphate (1 kcal/mol), these results suggest that the DNA-binding event is coupled to the ionization reaction of the buffer at pH 8.0. Moreover, the fact that the titration in Tris yielded the more endothermic result implies proton uptake by the protein–DNA complex during association.

The data from the four titrations with O2 were fitted simultaneously, as described in Materials and Methods, to estimate the intrinsic binding enthalpy, association equilibrium constant, and number of protons transferred (Fig. 8c). This analysis shows that the binding of O2 to PutA52 is intrinsically exothermic, with $\Delta H = -1.8$ kcal/mol (Table 1) and $K = 4.8 \times 10^6\ M^{-1}$, which corresponds to $K_d = 210$ nM (Table 1). The latter value agrees favorably with the

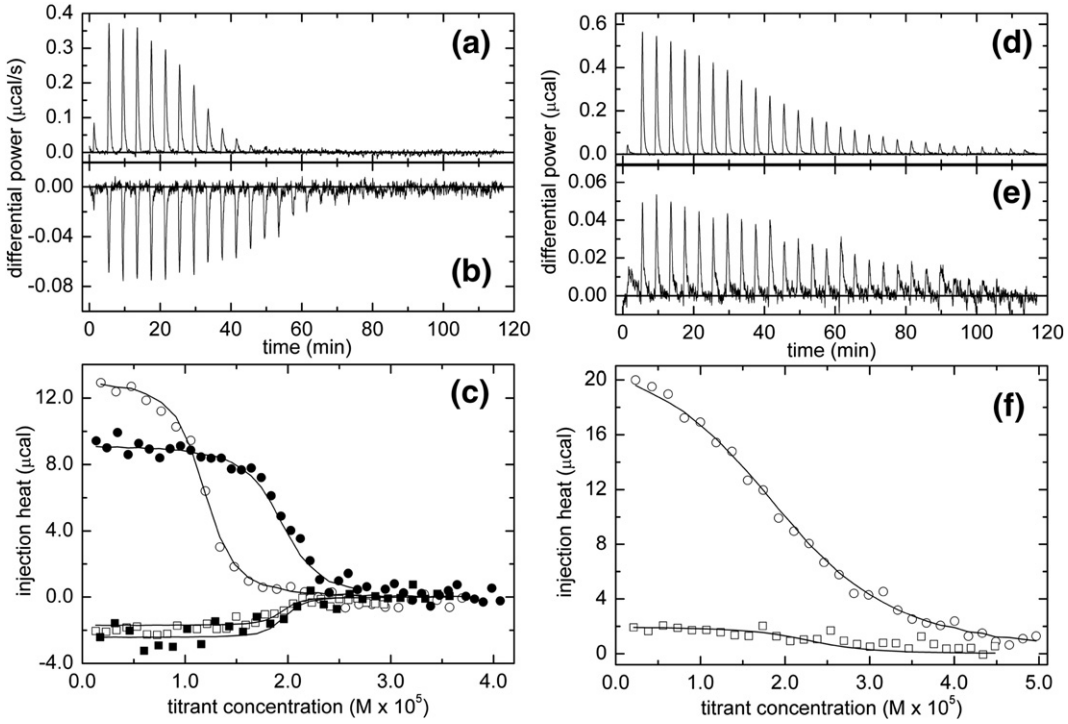

Fig. 8. ITC-based analysis of the interaction between PutA52 and the O2 and O2bf4 duplexes. (a) Raw data for the titration of 12 μM PutA52 with 0.15 mM O2 (10-μl additions) in NaCl and Tris (pH 8.0). (b) Raw data for the titration of 21 μM PutA52 with 0.14 mM O2 (10-μl additions) in NaCl and phosphate (pH 8.0). (c) Integrated data for titrations with the O2 duplex: 0.15 mM O2 *versus* 12 μM PutA52 in NaCl and Tris (pH 8.0; open circles); 0.15 mM O2 *versus* 21 μM PutA52 in NaCl and Tris (pH 8.0; filled circles); 0.14 mM O2 *versus* 21 μM PutA52 in NaCl and phosphate (pH 8.0; two trials, open and filled squares). (d) Raw data for the titration of 22 μM PutA52 with 0.19 mM O2fb4 (10-μl additions) in NaCl and Tris (pH 8.0). (e) Raw data for the titration of 25 μM PutA52 with 0.17 mM O2fb4 (10-μl additions) in NaCl and phosphate (pH 8.0). (f) Integrated data for titrations with the O2bf4 duplex: 0.19 mM O2bf4 *versus* 22 μM PutA52 in NaCl and Tris (pH 8.0; circles); 0.17 mM O2bf4 *versus* 25 μM PutA52 in NaCl and phosphate (pH 8.0; squares).

estimate from the gel-shift analysis of K_d<200 nM for O2 binding to full-length PutA.[26] The estimated number of protons transferred to the protein–DNA complex is 0.7.

A second set of titrations was performed using oligonucleotide O2fb4, which is identical with O2 except that the bases flanking the GTTGCA motif are those of O4. These measurements were performed to assess the impact of bases outside of the consensus motif on affinity. As with O2, the apparent enthalpy of binding of O2fb4 to PutA52 at pH 8.0 is dependent on buffer choice. In Tris buffer, the association appears to be strongly endothermic (Fig. 8d), but in phosphate buffer, the reaction is nearly isenthalpic (Fig. 8e).

Global analysis of the data from the two O2fb4 titrations (Fig. 8f) shows that binding of this oligonucleotide to PutA52 is marginally endothermic, with an intrinsic enthalpy change of only 0.18 kcal/mol (Table 1). The association constant from global fitting is $K=3.2\times10^5$ M^{-1}, which corresponds to $K_d=3100$ nM. As with the O2 titrations, there is an uptake of 0.7 proton during binding, suggesting that the binding mechanisms of the two ligands are qualitatively similar. Notice, however, that the association constant for O2 is 15 times higher than that for O2fb4. These results show that bases outside of the consensus motif impact the affinity of PutA52, and presumably PutA, for *put* control sites.

The binding of PutA52 is entropy-driven for both ligands. This result, combined with the observation that the protein–DNA interface is nearly devoid of bound water molecules, suggests that desolvation of macromolecular surfaces is important for DNA binding. We note that the free energy of the RHH protein MetJ binding to a metbox operator also includes a substantial favorable entropic component at 25 °C, particularly in the absence of the corepressor *S*-adenosylmethionine.[31]

Table 1. ITC data for DNA binding to PutA52 at 298 K

Ligand	K (M^{-1})	K_d (nM)	ΔH (kcal/mol)	$-T\Delta S$ (kcal/mol)	n[a]
O2	$4.8\pm0.5\times10^6$	210±20	−1.79±0.04	−7.3±0.2	1.0
O2fb4	$3.2\pm0.2\times10^5$	3100±200	0.18±0.05	−7.7±0.2	1.0

[a] Ligand to protein stoichiometry (dsDNA to PutA52 dimer).

Discussion

Transcriptional regulation of the *put* regulon

Based on the arrangement of the five PutA–DNA binding sites, PutA most likely represses the *put*

genes by hindering the σ^{70}-dependent binding of *E. coli* RNA polymerase to the *putA* and *putP* promoter regions.[32] We did not find additional PutA consensus binding sites in the coding regions of *putA* and *putP*, indicating that PutA binds only to the *put* intergenic region. Previous reports suggested that proline, via PutA, regulates the expression of *putA* more tightly than *putP*.[25,27] Here we have shown that PutA is a stronger repressor of the *putA* promoter than the *putP* promoter. Therefore, *putP* expression appears to be regulated relatively weakly by PutA, which would allow proline uptake under a variety of environmental conditions, leading to subsequent activation of *putA* expression. In addition to proline-specific regulation by PutA, the *put* genes are also responsive to global regulators. The cAMP receptor protein has been proposed to function as an activator by increasing *putA* and *putP* promoter activity in nutrient-poor environments.[25]

We evaluated *put* control DNA sequences from other bacteria in which PutA contains the conserved RHH domain and is predicted to function as an autogenous transcriptional repressor. The GTTGCA sequence was found in every *putA* promoter region of the 39 genome sequences analyzed.

We also examined *put* control DNA sequences in bacteria that share the same genetic organization of *putA* and *putP* as found in *E. coli*. Supplementary Figure S1 (Supplemental Material) shows an alignment of *put* control DNA sequences from *E. coli*, *Shigella boydii*, *Salmonella typhimurium*, *Klebsiella aerogenes*, *Yersinia pestis*, and *Pseudomonas putida*. The length of the intergenic region ranges from 361 bp in *P. putida* to 577 bp in *Y. pestis*. Each sequence has at least three exact repeats of the GTTGCA motif. Operators O1, O3, and O4 are present in all six sequences. O2 is present in all six organisms, except in *P. putida.* O5 is the least conserved site. *Y. pestis* and *P. putida* have additional exact repeats of the motif that do not align with the *E. coli* sites. These results suggest that the GTTGCA motif is the fundamental transcriptional control element of the PutA autogenous repression system. Therefore, the biochemical and structural results reported here for *E. coli* are likely applicable to other organisms in which PutA serves as a transcriptional repressor.

Importance of the GTTGCA motif for PutA operator binding

Mutation of the consensus GTTGCA motif to GTcatA was found to severely impact binding to PutA, based on gel-shift analysis. In fact, mutation of all five operators eliminated binding. Moreover, this mutation was found to affect gene transcription, as monitored by cell-based reporter assays. These results show that recognition of the middle of the consensus motif is essential for PutA binding and proper transcriptional control.

The basis for these results is evident from the crystal structure. The mutated triplet corresponds to base pairs T9:A13, G10:C12, and C11:G11 of the structure. As shown in Fig. 5b, the protein directly contacts T9, C12, C11, and G11. Mutation of T9 to C eliminates van der Waals interactions between the thymine C5 methyl and Gly7(B) (Fig. 7c). Replacement of C12 with T eliminates the hydrogen bond with the backbone carbonyl of Thr5(A) (Fig. 7b). In fact, this mutation would position two hydrogen bond acceptors—carbonyl of Thr5(A) and O4 carbonyl of thymine—close to each other, which is unfavorable. Moreover, the C5 methyl of the thymine would clash with Thr4(A). Interestingly, changing C11:G11 into T:A is predicted to have little effect on the protein–DNA interaction surface. Mutation of G11 to A would preserve the hydrogen bond donated by Gly7(A) (Fig. 7b), since both adenine and guanine have the hydrogen bond donor N7. Likewise, one can imagine the thymine O4 carbonyl engaging Thr5(B) in a hydrogen bond analogous to the one formed with the C11 in Fig. 7b. Thus, the observed deleterious effect on binding due to the GTcatA triple mutation appears to be primarily due to the change of TG into CA.

Whereas *E. coli* has five exact repeats of the GTTGCA motif, *Y. pestis* has single nucleotide substitutions in O1 (GTTaCA) and O2 (GTTGtA) (Supplementary Fig. S1). As described in the preceding paragraph, the G-to-A variation in O1 is predicted to disrupt the protein–DNA interface. On the other hand, the GTTGtA variation is likely to be accommodated by the protein without significant structural penalty. Hence, the *Y. pestis* O1 site may have a limited role in transcriptional control. We note that *Y. pestis* has an additional GTTGCA sequence motif upstream of O1, and that this site could substitute for a nonfunctional O1.

Importance of bases flanking the GTTGCA motif

Given the conservation of the GTTGCA motif in *put* intergenic regions, it is not surprising that interactions with these bases are essential for binding. Interestingly, the structure reveals that the footprint of PutA52 extends beyond the GTTGCA motif, indicating that bases flanking the consensus sequence may be important for operator recognition.

The structure shows that PutA52 contacts a 9-bp fragment, which we denote X*GTTGCA*YZ. Lys9 interacts with XG and the complementary bases of Y and Z. Based on the structure, it appears that G is preferred for X and CC is preferred for YZ because this sequence maximizes the number of hydrogen bonds formed by Lys9 (eight in total, four with each DNA strand).

Operators 1–3 of *E. coli* have G for base X, while operators 4 and 5 have A (Fig. 1b). Substitution of A in place of G at position X can be simulated by imagining adenine in place of G6 in Fig. 7a. This change would eliminate one of the hydrogen bonds formed by Lys9. All five operators have C for position Y, except O4, which has A. The effect of this variation can be seen by changing G7 in Fig. 7a into thymine. Elimination of one hydrogen bond is again predicted. Operators 1 and 3 have T and A,

respectively, at position Z, requiring Lys9 to interact with A and T, respectively. Both variations would eliminate one hydrogen bond with Lys9, relative to the optimal case of O2. This simple hydrogen bond inventory analysis implies that PutA exhibits different affinities for the five operator sites, with O2 predicted to have the highest affinity (eight hydrogen bonds to Lys9) and with O4 having the lowest affinity (six hydrogen bonds).

This hypothesis was tested with ITC experiments that compared the binding of PutA52 to two ligands differing only in the bases flanking the consensus motif: O2 and O2fb4. The former ligand is the one used in crystal structure determination. The latter is identical with O2, except that it has the flanking bases of O4 (X, Y = A). These two ligands thus mimic the two extremes of the predicted affinity spectrum of the PutA operator sites.

The ITC analysis showed that PutA52 binds to O2 with 15-times-higher affinity than O2fb4 (Table 1), in agreement with our structure-based predictions. The binding enthalpy accounts almost entirely for the difference in affinity (Table 1); ΔH for O2 is more exothermic than that for O2fb4 by about 2 kcal/mol. That the difference in affinity is enthalpic in origin is consistent with our prediction that Lys9 forms more hydrogen bonds with the flanking bases of O2 than with the flanking bases of O4. These results suggest that the five operator sites are nonequivalent in terms of binding affinity, which is potentially significant because differential binding could be important for proper transcriptional regulation.

In light of these results, it is interesting to note that PutA is a weaker repressor of the *putP* promoter than the *putA* promoter, yet the highest-affinity operator (O2) is involved in repression of *putP*, whereas the lowest-affinity operator (O4) is involved in repression of *putA*. We suggest that PutA–DNA binding affinity is only one of several factors that should be considered when assessing the potential impact of the operators on transcriptional repression. Other key factors include the number of transcription start sites and the location of the operators relative to the promoter regions. The *putP* gene was previously shown to have multiple transcription start sites and three functional promoters.[25] The three *putP* promoters are positioned 14, 26, and 58 bp upstream of the O1 site. Thus, PutA binding at O1 and O2 is not predicted to directly interfere with RNA polymerase at each of the *putP* promoters, resulting in weaker repression of the *putP* gene by PutA. On the other hand, the *putA* gene has only one promoter and a single transcriptional start site, with O3 and O4 located within the *putA* promoter region. Thus, these sites are positioned optimally for PutA to interfere with RNA polymerase, which could explain why PutA is a stronger repressor of *putA* expression.

Parenthetically, this ITC analysis underscores the value of examining binding reactions in buffers having distinct ionization enthalpies. Besides revealing the involvement of protonation in protein–ligand associations, inclusion of the buffer ionization enthalpy can, in select cases, significantly improve the quality of the titration data. For example, the intrinsic binding enthalpy for the interaction between PutA52 and the O2fb4 oligo (0.18 kcal/mol) is effectively zero. Absent the contribution of phosphate or Tris ionization, this binding reaction would be invisible by ITC. Although the intrinsic enthalpy for the PutA52/O2 association is somewhat larger (−1.79 kcal/mol), the interaction would nonetheless be difficult to characterize in phosphate buffer alone because the heat of buffer ionization reduces the observed enthalpy to just −0.8 kcal/mol. In striking contrast, the highly endothermic Tris ionization event renders the reaction much more amenable to analysis and facilitates, via a global-fitting strategy, treatment of the data collected in phosphate buffer.

Possible roles of PutA-specific residues

PutA is a unique member of the RHH family of transcription factors. With a polypeptide chain in excess of 1300 residues, PutA is the largest protein known to contain an RHH domain. Furthermore, to our knowledge, PutA is the only protein to have a flavin redox regulatory domain coupled to an RHH domain.

PutA is also distinguished from other RHH proteins at the primary sequence level.[24] In particular, Gly7 and Pro29 are absolutely conserved among PutAs, yet rarely found in other RHH domains.

We suggest three possible roles for Pro29. First, proline may facilitate initiation of αB and thus help position Thr28 and His30 for interaction with the DNA backbone. Second, Pro29 may provide a steric "backstop" for the DNA backbone and thereby contribute to the recognition of a nine-nucleotide fragment of B-form DNA. A third possibility is that Pro29 may be donating C–H···O hydrogen bonds to the DNA backbone. This suggestion is based on the observation that Pro29 C^{δ} forms close contacts with oxygen atoms of the phosphate backbone (Fig. 7a).

There is a precedent for proline C^{δ} donating hydrogen bonds. For example, when proline is located in the middle or C-terminus of an α-helix, C^{δ} donates hydrogen bonds to the backbone carbonyl 3–5 residues preceding the proline.[33] These unconventional hydrogen bonds enable proline to appear in α-helices despite lacking a free N–H group for the classic i to $i+4$ hydrogen bonding pattern. We note that the CH···O distance of 3.4 Å observed in the PutA52/O2 complex is identical with the average distance for proline intrahelix C–H···O hydrogen bonds.[33]

The unique location of Pro29 at the beginning of αB, juxtaposed to the DNA backbone, also supports a hydrogen bonding role. Analysis of RHH–DNA structures shows that there are two conserved hydrogen bonds donated by the backbone of residues at the N-terminus of αB to the DNA backbone.[30] We observe only one of these conserved hydrogen bonds, and it involves the backbone N–H of His30 (Fig. 7a). Since proline does not have a free N–H group, the second conserved hydrogen bond is

missing. We suggest that the unconventional C–H···O hydrogen bonds substitute for the missing conserved hydrogen bond.

Gly7 represents an interesting sequence variation for the RHH family. Typically, this position of the β-sheet is occupied by a polar residue such as Thr or Asn, which forms hydrogen bonds with DNA bases. The PutA52/O2 structure shows that Gly7 participates in base recognition, despite lacking a side chain. It donates a hydrogen bond to a guanine base and forms van der Waals contacts with the C5 methyl group of thymine.

Gly7 may underlie a more global structural aspect of DNA recognition by PutA: deep penetration into the major groove. Absence of a side chain at this position allows the β-sheet to penetrate further into the major groove, compared to other RHH proteins. As a measure of the depth of penetration, we calculated the distance of closest approach between the DNA axis and each of the C^{α} atoms of the three canonical β-sheet residues responsible for base recognition (Thr5, Gly7, and Lys9 in PutA). The values for PutA52/O2 are 5.0 Å, 5.1 Å, and 8.2 Å for chain A, and 7.3 Å, 6.0 Å, and 7.4 Å for chain B. The corresponding values for Arc,[34] MetJ,[35] CopG,[36] NikR,[37] omega,[38] and FitA[39] bound to DNA span the ranges 8.0–11.9 Å, 5.8–8.1 Å, and 8.2–11.7 Å, respectively, for these three residue positions. Thus, PutA52 penetrates deeper into the major groove than the other RHH proteins. The role of this close encounter in the transcriptional regulation of the *put* regulon is unknown, but the universal conservation of Gly7 in PutAs and its absence in the rest of the RHH family suggest functional significance.

Materials and Methods

Materials

Chemicals and buffers were purchased from Fisher Scientific and Sigma-Aldrich, Inc., unless otherwise stated. Restriction endonucleases and T4 DNA ligase were purchased from Fermentas and Invitrogen, respectively. BCA reagents used for protein quantitation were obtained from Pierce. Goat anti-rabbit secondary antibody was purchased from Amersham, Inc. *E. coli* strains XL-blue and BL21 DE3 pLysS were purchased from Stratagene. *E. coli* strain JT31 *putA*$^{-}$ *lacZ*$^{-}$ was a generous gift from J. Wood (University of Guelph, Guelph, ON, Canada). Synthetic oligonucleotides for site-directed mutagenesis, cloning, DNA-binding assays, and cocrystallization were purchased from Integrated DNA Technologies. All experiments used NANOPure water. LB medium and Terrific broth were used for general culture growth and protein production, respectively, while M9 minimal medium was used for cell-based transcription assays.

Gel-shift assays

Full-length PutA and PutA52 were expressed as C-terminally His-tagged proteins from vector pET23b (Novagen) and purified as described previously.[23,24,40] The C-terminal His tags were retained after purification. Purified full-length PutA was dialyzed into 50 mM Tris (pH 7.5) containing 10% glycerol and stored at −70 °C. PutA52 was dialyzed into 50 mM phosphate buffer (pH 7.3) containing 200 mM NaCl and stored at −70 °C. The concentrations of the PutA proteins were determined using the BCA method (Pierce) with bovine serum albumin as standard, and spectrophotometrically using molar extinction coefficients of 12,700 M^{-1} cm^{-1} at 451 nm for PutA and 6970 M^{-1} cm^{-1} at 280 nm for PutA52.[24,41]

Nondenaturing gel electrophoretic mobility shift assays were used to test the binding of the PutA proteins to the *put* control intergenic DNA as previously described.[23,41] Different regions of the *put* control DNA (*putC*) were PCR-amplified (nonlabeled) using synthetic primers and purified. The purified products were incubated with full-length PutA (0, 0.6, and 1.5 μM) for 20 min at 20 °C in 50 mM Tris buffer (pH 7.5; 100 mM NaCl). The protein–DNA complexes were separated using a native polyacrylamide gel (4%) at 4 °C. The gel was then stained with ethidium bromide and visualized by Bio-Rad Quantity One. Binding assays with synthetic oligonucleotides corresponding to base pairs 183–210 (O1), 342–365 (O3/4), and 388–412 (O5) of *putC* were performed similarly. The concentration of oligonucleotide for these assays was 100 nM. The duplex DNA of each oligonucleotide was prepared by annealing the complementary oligonucleotides in buffer [10 mM Tris pH 8.0, 50 mM NaCl, and 1 mM ethylenediaminetetraacetic acid (EDTA)] by first heating at 95 °C for 5 min and then gradually cooling down the sample to room temperature.

Gel-shift assays utilizing fluorescently labeled *put* intergenic DNA were also performed. The synthetic oligonucleotide (M13 forward primer) was 5′-end-labeled with IRdye-700 (LI-COR, Inc.) and used as one of the primers in PCR to amplify WT *put* intergenic DNA or mutant *put* intergenic DNA containing different combinations of PutA binding site mutations. The resulting IRdye-700-labeled *put* intergenic DNA was purified and quantitated by measuring the nucleic acid concentration at 260 nm and the absorbance of the IRdye-700 at 685 nm using an extinction coefficient of 170 mM^{-1} cm^{-1} in accordance with the recommendations of the manufacturer. PutA (0–900 nM) or PutA52 (0–200 nM) was incubated with 2 nM *put* intergenic DNA in a total volume of 25 μl in 50 mM Tris and 50–250 mM NaCl (pH 7.5) containing 10% glycerol for 20 min (20 °C) before electrophoresis. Calf thymus competitor DNA (100 μg/ml) was also added to the binding mixtures to prevent nonspecific protein–DNA interactions. The PutA–DNA and PutA52–DNA complexes were separated using native polyacrylamide gel electrophoresis at 4 °C. The gels were visualized using a LI-COR Odyssey Imager.

LacZ reporter assays

E. coli strain JT31 *putA*$^{-}$ *lacZ*$^{-}$ was cotransformed with PutA-pUC18 and the reporter construct P_{putA}:*lacZ*-pACYC184 or P_{putP}:*lacZ*-pACYC184. Details about the cloning procedures and the primers used for generating the above constructs are provided in Supplementary Table S1 (Supplemental Material). To test which PutA binding sites are critical for repressing *put* gene expression, *E. coli* strain JT31 *putA*$^{-}$ *lacZ*$^{-}$ containing different combinations of the PutA-pUC18 construct and the P_{putA}:*lacZ* or P_{putP}:*lacZ* reporter constructs was grown at 37 °C in M9 minimal medium supplemented with ampicillin (50 μg/ml), kanamycin (40 μg/ml), and chloramphenicol (34 μg/ml) to an OD_{600} of ~1.0. PutA expression from the

lac promoter on pUC18 was not induced, as no isopropyl-β-D-thiogalactopyranoside was added to the culture medium. Cells from the various cultures were pelleted, resuspended in Tris–HCl buffer (20 mM, pH 7.5), and broken using the B-PER II bacterial protein extraction reagent from Pierce (20 mM Tris–HCl, pH 7.5).[23] β-Galactosidase activity assays were performed in 1 ml of 100 mM sodium phosphate (pH 7.3) containing 1 mM $MgCl_2$, 50 mM β-mercaptoethanol, and 2 mM *o*-nitrophenyl-β-D-galactopyranoside. The initial velocity was determined by measuring the increase in absorbance at 420 nm. The reported β-galactosidase activities are values averaged from four independent experiments.

The expression of PutA was confirmed by Western blot analysis using an antibody directed against PutA47 (a polypeptide corresponding to residues 1–47 of *E. coli* PutA). PutA47 was purified without a 6× His tag as described previously.[23] Antiserum directed against purified PutA47 was prepared by Proteintech, Inc. For Western blot analysis, cell pellets from 5 ml of culture grown in minimal medium (OD ~1.0) were resuspended in 100 μl of SDS sample buffer and boiled for 10 min. After SDS denaturing electrophoresis, the protein bands were transferred onto a sequi-blot polyvinylidene difluoride membrane (0.2 μm pore size; Bio-Rad) using an EBU-4000 semidry electrophoretic blotting system. Immunoreactive bands were detected using enhanced chemiluminescence Western blot analysis reagents (Amersham, Inc.).

Crystallization

Exhaustive crystallization experiments were conducted using two different RHH domain constructs paired with several different DNA fragments, as described in Supplemental Material. Successful cocrystallization required the use of an RHH domain construct consisting of *E. coli* PutA residues 1–52 (PutA52) fused to a cleavable N-terminal histidine tag. Expression and purification protocols for this protein have been described.[24] The N-terminal tag was removed by proteolysis prior to crystallization trials as described.[24] The purified tag-free protein was concentrated using an Amicon Ultra centrifugal filtration device (MWCO = 5000) up to 10.8 mg/ml in a buffer consisting of 20 mM Tris (pH 8.0), 500 mM NaCl, and 20 mM imidazole. The protein concentration was estimated with the BCA method.

The oligonucleotide used for cocrystallization with PutA52 corresponds to nucleotides 211–231 of the *put* control region, which contains the second of five operator sites for PutA (denoted O2 in Fig. 1):

5′-TTTGCGGTTGCACCTTTCAAA-3′ (strand 1)
3′-AAACGCCAACGTGGAAAGTTT-5′ (strand 2).

Each DNA strand was dissolved in 10 mM Tris, 50 mM NaCl, and 1 mM EDTA (pH 8.0) to a concentration of 6 mM. The two strands were annealed as follows. Equal volumes of the oligonucleotide solutions were combined, and the mixture was placed in a water bath at room temperature. The temperature of the bath was then set to 94 °C. Once the target temperature had been reached, the power to the bath was turned off to slowly cool the sample back to room temperature.

The PutA52 and double-stranded DNA (dsDNA) stock solutions were mixed so that the molar ratio of PutA52 dimer to dsDNA was approximately 1:3. The mixture was injected onto a Sephacryl S-100 HiPrep 16/60 gel-filtration column equilibrated with 10 mM Tris, 50 mM NaCl, 0.5 mM EDTA, and 0.5 mM DTT (pH 8.0). Fractions were pooled and concentrated to 11.6 mg/ml (BCA assay) using an Amicon Ultra centrifugal filtration device (MWCO = 10,000).

The PutA52–DNA mixture was inputted to several crystal screens to identify initial crystallization conditions. The crystals used for data collection were obtained directly from Index Screen reagent 54, which consists of 30% polyethylene glycol (PEG) monomethyl ester 550, 50 mM $CaCl_2$, and 100 mM bis-Tris (pH 6.5).

The crystals were prepared for low-temperature data collection by soaking them in a solution of 32% PEG monomethyl ether 550, 50 mM $CaCl_2$, 100 mM bis-Tris (pH 6.5), and 15% PEG 200. The crystals were then picked up with Hampton loops and plunged into liquid nitrogen.

X-ray diffraction data collection, phasing, and refinement

Several crystals were analyzed at Advanced Light Source beamline 4.2.2 using a NOIR-1 charge-coupled device detector. The data were integrated with MOSFLM[42] and scaled with SCALA.[43] The crystals belong to space group *C*2, with unit cell parameters of $a = 90.9$ Å, $b = 44.1$ Å, $c = 55.2$ Å, and $\beta = 101.5°$. The asymmetric unit contains one PutA52 dimer and one DNA duplex, corresponding to 50% solvent and $V_M = 2.3$ Å^3/Da.[44,45] The best data set had a high-resolution limit of 2.25 Å and consisted of 180 frames collected with an oscillation angle of 1° per frame, an exposure time of 7 s per frame, and a

Table 2. Data collection and refinement statistics

Wavelength (Å)	1.24
Diffraction resolution (Å)	54–2.25
Number of observations	36,754
Number of unique reflections	10,293
Redundancy	3.6 (3.6)
Completeness (%)	99 (99)
R_{merge} (I)	0.059 (0.434)
R_{pim} (I)	0.037 (0.262)
Average I/σ	16.1 (2.9)
Wilson B-factor (Å^2)	52
PDB accession code	2RBF
Number of protein chains	2
Number of protein residues	89
Number of protein atoms	709
Number of DNA strands	2
Number of DNA nucleotides	37
Number of DNA atoms	757
Number of water molecules	27
R_{cryst}	0.206 (0.218)
R_{free}[a]	0.246 (0.277)
RMSD[b]	
Bond length (Å)	0.010
Bond angle (°)	1.6
Ramachandran plot[c] (%)	
Favored	100.0
Allowed	0.0
Outliers	0.0
Average B-factors (Å^2)	
Protein	47
DNA	49
Water	45

Values for the outer-resolution shell are given in parenthesis. The outer shell is 2.37–2.25 Å for processing and 2.31–2.25 Å for refinement.
[a] Five percent random R_{free} test set.
[b] Compared to the parameters of Engh and Huber.[46]
[c] The Ramachandran plot was generated with RAMPAGE.[47]

detector distance of 120 mm. Data collection and processing statistics are listed in Table 2.

The structure was solved using molecular replacement, with a PutA52 dimer serving as the search model. CNS was used for molecular replacement calculations.[48] The fast-direct method was used for cross-rotation function calculations. Prior to translation function calculations, the orientations from the cross-rotation function calculation were optimized with Patterson correlation refinement using two groups corresponding to the two protein chains of the dimer. The top solution from the translation function calculation had a correlation coefficient of 0.257. Rigid-body refinement resulted in a model with $R_{cryst}=0.488$ and $R_{free}=0.493$ for data at 3.0 Å resolution. Simulated annealing refinement in CNS lowered the *R*-factors to $R_{cryst}=0.450$ and $R_{free}=0.467$ for all reflections.

The model from simulated annealing was used as the starting point for iterative cycles of model building in Coot[49] and refinement with TLS in REFMAC5.[50] After a few cycles, the *R*-factors for a protein-only model were $R_{cryst}=0.416$ and $R_{free}=0.427$. At this point, electron density representing DNA base pairs was evident. The DNA part of the model was gradually built up over about two dozen cycles of model building and refinement. Solvent was added during the latter stages of model building.

The final model consists of 1 PutA52 dimer, 1 dsDNA molecule, and 27 water molecules. The modeled protein chains include residues 3–46 for chain A and residues 4–48 for chain B. The DNA strands include nucleotides 4–21 for strand 1 and nucleotides 1–19 for strand 2. Electron density near the end of the DNA duplex containing nucleotides 1–3 of strand 1 and nucleotides 19–21 for strand 2 was rather weak, and indicated fraying of the base pairs and possibly more than one conformation. But the density was not of sufficient quality to allow reliable modeling of this end of the DNA ligand. We note that this end of the oligonucleotide is far from the protein–DNA interface. Refinement statistics are listed in Table 2.

Structure analysis

Structures were analyzed graphically using Coot and PyMOL.[51] CNS was used to calculate buried surface area.[48] DNA conformation was analyzed with 3DNA.[28] The depth of penetration of the β-sheet into the DNA major groove was analyzed for RHH proteins. An operational definition of penetration depth—the shortest distance between selected C^{α} atoms of the β-sheet and the DNA helical axis—was adopted for this purpose. DNA helical axes were calculated using 3DNA.[28] The 9–10 bp corresponding to the region contacted by the protein were used for the axis calculation. Distances between C^{α} atoms and DNA axes were calculated using a program written by D. Coventry, which implements the theory by P. Bourke.

ITC

ITC experiments were conducted at 25 °C with a VP-ITC calorimeter (MicroCal, LLC). Prior to analysis, the protein and oligonucleotide were dialyzed extensively against the appropriate reaction buffer, which was either 50 mM Tris, 100 mM NaCl, and 1 mM EDTA (pH 8.0), or 50 mM sodium phosphate, 100 mM NaCl, and 1 mM EDTA (pH 8.0). The dimeric quaternary structure of PutA52 was confirmed by equilibrium analytical ultracentrifugation under conditions similar to those used in ITC experiments (20 μM PutA52 in 50 mM Tris and 50 mM NaCl, pH 8.0). The data could be fitted very well to a single-species model with an apparent molecular mass corresponding to that of the homodimer (~13.5 kDa). No evidence of a monomeric species was present at 20 μM. Sample and titrant were degassed under vacuum immediately before being loaded into the sample cell and buret, respectively. Following thermal equilibration, aliquots (7 or 10 μl) of titrant were added to the 1.41-ml sample at 240-s intervals. A 2-μl preinjection was included at the start of each titration. The heat associated with this addition—invariably inaccurate due to diffusion of titrant from the buret during the equilibration period—was neglected during the fitting process.

Samples of PutA52 were titrated with two oligonucleotides, designated O2 and O2fb4 (O2 with flanking bases of O4). O2 is the oligonucleotide used for cocrystallization. O2fb4 is identical with O2, except that the base pairs flanking the consensus motif are those of operator 4 (Fig. 1b):

5′-TTTGC*A*GTTGCA*A*CTTTCAAA-3′ (strand 1)
3′-AAACG*T*CAACGT*T*GAAAGTTT-5′ (strand 2).

Experiments were conducted in both phosphate and Tris—two buffers with distinct ionization enthalpies. The raw data were integrated with software supplied with the instrument. Blank titrations (injection of titrant into buffer) were performed for each oligonucleotide–buffer combination. The average injection heats associated with these experiments were used to correct the corresponding protein titrations for the nonspecific heat of mixing/dilution.

The apparent protein–DNA binding enthalpies differed profoundly in Tris and phosphate, indicating that the PutA52–DNA interaction is accompanied by protonation. Accordingly, the data from the two buffer systems were subjected to simultaneous least squares analysis, employing a model that explicitly includes the heat of buffer ionization. The following equation describes the cumulative heat after the *i*th titrant addition:

$$Q_i = V\left[\mathrm{M}\right]_t(\Delta H + n\Delta H_{buf})\left[\frac{K_{ap}[\mathrm{DNA}]}{1+K_{ap}[\mathrm{DNA}]}\right] \quad (1)$$

where *V* is the sample cell volume, $[\mathrm{M}]_t$ is the total protein concentration, ΔH is the intrinsic binding enthalpy, ΔH_{buf} is the heat of buffer ionization, *n* is the number of protons taken up by the protein–DNA complex during the binding reaction, K_{ap} is the apparent association constant, and [DNA] is the concentration of free DNA.

Because protein–DNA binding is linked to protonation, the apparent free-energy change for the reaction (ΔG_{ap}) includes a contribution from buffer ionization (ΔG_{buf}):

$$\Delta G_{ap} = \Delta G + \Delta G_{buf} \quad (2)$$

Thus,

$$-RT\ln K_{ap} = -RT\ln K - nRT\ln K_{buf} \quad (3)$$

or

$$K_{ap} = K(K_{buf})^n \quad (4)$$

where *K* is the intrinsic binding constant for the protein–DNA association and K_{buf} is equal to:

$$K_{buf} = \frac{[\mathrm{B}]}{[\mathrm{BH}^+]} = \exp[2.303(\mathrm{pH} - \mathrm{p}K_a)] \quad (5)$$

In Eq. (5), $[\mathrm{BH}^+]$ and [B] represent the concentrations of the conjugate acid and base forms of the reaction buffer,

respectively, and pK_a is the appropriate value for the particular buffer under consideration. K, ΔH, and n were global-fitting parameters; pH was a fixed global parameter; and ΔH_{buf} and pK_a were fixed titration-specific parameters obtained from the literature.[52]

The ith injection heat (q_i) was modeled as the difference in the cumulative heats associated with the ith and $(i+1)$th additions:

$$q_i = (Q_i - Q_{i-1}) + \frac{dV_i}{V}\left(\frac{Q_i + Q_{i-1}}{2}\right) \quad (6)$$

The second term in Eq. (6) is a correction for the heat associated with the volume of solution displaced from the sample cell by the ith titrant addition, where dV_i is the volume of the ith injection. Fitting was performed in Origin (version 7.5; OriginLab), employing a LabTalk script generated in-house.

PDB accession code

Atomic coordinates and structure factor amplitudes have been deposited in the PDB[53] as entry 2RBF.

Acknowledgements

This research was supported by National Institutes of Health grants GM065546 (J.J.T.) and GM061068 (D.F.B.), and National Science Foundation grant MCB0091664 (D.F.B.). C.A.B. was supported by a postdoctoral fellowship from the National Library of Medicine (2-T15-LM07089-14). We thank Damian Coventry and Paul Bourke for providing the computer code used in penetration depth calculations. Part of this research was performed at the Advanced Light Source, which is supported by the Director, Office of Science, Office of Basic Energy Sciences, Materials Sciences Division, US Department of Energy, under contract no. DE-AC03-76SF00098, at Lawrence Berkeley National Laboratory. This work is a contribution of the University of Nebraska Agricultural Research Division, supported in part by funds provided through the Hatch Act. This publication was also made possible by National Institutes of Health grant P20 RR-017675-02 from the National Center for Research Resources.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2008.05.084

References

1. Abrahamson, J. L. A., Baker, L. G., Stephenson, J. T. & Wood, J. M. (1983). Proline dehydrogenase from *Escherichia* K12, properties of the membrane-associated enzyme. *Eur. J. Biochem.* **134**, 77–82.
2. Brown, E. & Wood, J. M. (1992). Redesigned purification yields a fully functional PutA protein dimer from *Escherichia coli*. *J. Biol. Chem.* **267**, 13086–13092.
3. Scarpulla, R. C. & Soffer, R. L. (1978). Membrane-bound proline dehydrogenase from *Escherichia coli*. *J. Biol. Chem.* **253**, 5997–6001.
4. Straub, P. F., Reynolds, P. H. S., Althomsons, S., Mett, V., Zhu, Y., Shearer, G. & Kohl, D. H. (1996). Isolation, DNA sequence analysis, and mutagenesis of a proline dehydrogenase gene (*putA*) from *Bradyrhizobium japonicum*. *Appl. Environ. Microbiol.* **62**, 221–229.
5. Menzel, R. & Roth, J. (1981). Purification of the putA gene product. *J. Biol. Chem.* **256**, 9755–9761.
6. Krishnan, N. & Becker, D. F. (2005). Characterization of a bifunctional PutA homologue from *Bradyrhizobium japonicum* and identification of an active site residue that modulates proline reduction of the flavin adenine dinucleotide cofactor. *Biochemistry*, **44**, 9130–9139.
7. Krishnan, N. & Becker, D. F. (2006). Oxygen reactivity of PutA from *Helicobacter* species and proline-linked oxidative stress. *J. Bacteriol.* **188**, 1227–1235.
8. Chen, C.-C. & Wilson, T. H. (1986). Solubilization and functional reconstitution of the proline transport system of *Escherichia coli*. *J. Biol. Chem.* **261**, 2599–2604.
9. Jung, H. (2002). The sodium/substrate symporter family: structural and functional features. *FEBS Lett.* **529**, 73–77.
10. Menzel, R. & Roth, J. (1981). Regulation of genes for proline utilization in *Salmonella typhimurium*: autogenous repression by the putA gene product. *J. Mol. Biol.* **148**, 21–44.
11. Ostrovsky De Spicer, P., O'Brian, K. & Maloy, S. (1991). Regulation of proline utilization in *Salmonella typhimurium*: a membrane-associated dehydrogenase binds DNA *in vitro*. *J. Bacteriol.* **173**, 211–219.
12. Vílchez, S., Molina, L., Ramos, C. & Ramos, J. L. (2000). Proline catabolism by *Pseudomonas putida*: cloning, characterization, and expression of the *put* genes in the presence of root exudates. *J. Bacteriol.* **182**, 91–99.
13. Chen, L.-M. & Maloy, S. (1991). Regulation of proline utilization in enteric bacteria: cloning and characterization of the *Klebsiella put* control region. *J. Bacteriol.* **173**, 783–790.
14. Vílchez, S., Manzanera, M. & Ramos, J. (2000). Control of expression of divergent *Pseudomonas putida put* promoters for proline catabolism. *Appl. Environ. Microbiol.* **66**, 5221–5225.
15. Jafri, S., Evoy, S., Cho, K., Craighead, H. G. & Winans, S. C. (1999). An Lrp-type transcriptional regulator from *Agrobacterium tumefaciens* condenses more than 100 nucleotides of DNA into globular nucleoprotein complexes. *J. Mol. Biol.* **288**, 811–824.
16. Keuntje, B., Masepohl, B. & Klipp, W. (1995). Expression of the *putA* gene encoding proline dehydrogenase from *Rhodobacter capsulatus* is independent of NtrC regulation but requires an Lrp-like activator protein. *J. Bacteriol.* **177**, 6432–6439.
17. Nakada, Y., Nishijyo, T. & Itoh, Y. (2002). Divergent structure and regulatory mechanism of proline catabolic systems: characterization of the putAP proline catabolic operon of *Pseudomonas aeruginosa* PAO1 and its regulation by PruR, an AraC/XylS family protein. *J. Bacteriol.* **184**, 5633–5640.
18. Lee, J. H. & Choi, S. H. (2006). Coactivation of *Vibrio vulnificus* putAP operon by cAMP receptor protein and PutR through cooperative binding to overlapping sites. *Mol. Microbiol.* **60**, 513–524.

19. Ling, M., Allen, S. W. & Wood, J. M. (1994). Sequence analysis identifies the proline dehydrogenase and pyrroline-5-carboxylate dehydrogenase domains of the multifunctional *Escherichia coli* PutA protein. *J. Mol. Biol.* **245**, 950–956.
20. Zhang, M., White, T. A., Schuermann, J. P., Baban, B. A., Becker, D. F. & Tanner, J. J. (2004). Structures of the *Escherichia coli* PutA proline dehydrogenase domain in complex with competitive inhibitors. *Biochemistry*, **43**, 12539–12548.
21. Lee, Y. H., Nadaraia, S., Gu, D., Becker, D. F. & Tanner, J. J. (2003). Structure of the proline dehydrogenase domain of the multifunctional PutA flavoprotein. *Nat. Struct. Biol.* **10**, 109–114.
22. Zhang, W., Zhang, M., Zhu, W., Zhou, Y., Wanduragala, S., Rewinkel, D. *et al.* (2007). Redox-induced changes in flavin structure and roles of flavin N(5) and the ribityl 2′-OH group in regulating PutA–membrane binding. *Biochemistry*, **46**, 483–491.
23. Gu, D., Zhou, Y., Kallhoff, V., Baban, B., Tanner, J. J. & Becker, D. F. (2004). Identification and characterization of the DNA-binding domain of the multifunctional PutA flavoenzyme. *J. Biol. Chem.* **279**, 31171–31176.
24. Larson, J. D., Jenkins, J. L., Schuermann, J. P., Zhou, Y., Becker, D. F. & Tanner, J. J. (2006). Crystal structures of the DNA-binding domain of *Escherichia coli* proline utilization A flavoprotein and analysis of the role of Lys9 in DNA recognition. *Protein Sci.* **15**, 2630–2641.
25. Nakao, T., Yamato, I. & Anraku, Y. (1988). Mapping of the multiple regulatory sites for *putP* and *putA* expression in the *putC* region of *Escherichia coli*. *Mol. Gen. Genet.* **214**, 379–388.
26. Zhang, W., Zhou, Y. & Becker, D. F. (2004). Regulation of PutA–membrane associations by flavin adenine dinucleotide reduction. *Biochemistry*, **43**, 13165–13174.
27. Wood, J. M. & Zadworny, D. (1979). Characterization of an inducible porter required for L-proline catabolism by *Escherichia coli* K12. *Can. J. Biochem.* **57**, 1191–1199.
28. Lu, X. J. & Olson, W. K. (2003). 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* **31**, 5108–5121.
29. Lu, X. J., Shakked, Z. & Olson, W. K. (2000). A-form conformational motifs in ligand-bound DNA structures. *J. Mol. Biol.* **300**, 819–840.
30. Schreiter, E. R. & Drennan, C. L. (2007). Ribbon–helix–helix transcription factors: variations on a theme. *Nat. Rev. Microbiol.* **5**, 710–720.
31. Hyre, D. E. & Spicer, L. D. (1995). Thermodynamic evaluation of binding interactions in the methionine repressor system of *Escherichia coli* using isothermal titration calorimetry. *Biochemistry*, **34**, 3212–3221.
32. Browning, D. F. & Busby, S. J. (2004). The regulation of bacterial transcription initiation. *Nat. Rev. Microbiol.* **2**, 57–65.
33. Chakrabarti, P. & Chakrabarti, S. (1998). C–H˙O hydrogen bond involving proline residues in alpha-helices. *J. Mol. Biol.* **284**, 867–873.
34. Raumann, B. E., Rould, M. A., Pabo, C. O. & Sauer, R. T. (1994). DNA recognition by β-sheets in the Arc repressor–operator crystal structure. *Nature*, **367**, 754–757.
35. Somers, W. S. & Phillips, S. E. V. (1992). Crystal structure of the met repressor–operator complex at 2.8 A resolution reveals DNA recognition by β-strands. *Nature*, **359**, 387–393.
36. Gomis-Ruth, F., Sola, M., Acebo, P., Parraga, A., Guasch, A., Eritja, R. *et al.* (1998). The structure of plasmid-encoded transcriptional repressor CopG unliganded and bound to its operator. *EMBO J.* **17**, 7404–7415.
37. Schreiter, E. R., Wang, S. C., Zamble, D. B. & Drennan, C. L. (2006). NikR–operator complex structure and the mechanism of repressor activation by metal ions. *Proc. Natl Acad. Sci. U. S. A.* **103**, 13676–13681.
38. Weihofen, W. A., Cicek, A., Pratto, F., Alonso, J. C. & Saenger, W. (2006). Structures of omega repressors bound to direct and inverted DNA repeats explain modulation of transcription. *Nucleic Acids Res.* **34**, 1450–1458.
39. Mattison, K., Wilbur, J. S., So, M. & Brennan, R. G. (2006). Structure of FitAB from *Neisseria gonorrhoeae* bound to DNA reveals a tetramer of toxin–antitoxin heterodimers containing pin domains and ribbon–helix–helix motifs. *J. Biol. Chem.* **281**, 37942–37951.
40. Zhu, W. & Becker, D. F. (2003). Flavin redox state triggers conformational changes in the PutA protein from *Escherichia coli*. *Biochemistry*, **42**, 5469–5477.
41. Becker, D. F. & Thomas, E. A. (2001). Redox properties of the PutA protein from *Escherichia coli* and the influence of the flavin redox state on PutA–DNA interactions. *Biochemistry*, **40**, 4714–4722.
42. Leslie, A. G. (2006). The integration of macromolecular diffraction data. *Acta Crystallogr. Sect. D*, **62**, 48–57.
43. Evans, P. (2006). Scaling and assessment of data quality. *Acta Crystallogr. Sect. D*, **62**, 72–82.
44. Matthews, B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491–497.
45. Kantardjieff, K. A. & Rupp, B. (2003). Matthews coefficient probabilities: improved estimates for unit cell contents of proteins, DNA, and protein–nucleic acid complex crystals. *Protein Sci.* **12**, 1865–1871.
46. Engh, R. A. & Huber, R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallogr. Sect. A*, **47**, 392–400.
47. Lovell, S. C., Davis, I. W., Arendall, W. B., III, de Bakker, P. I., Word, J. M., Prisant, M. G. *et al.* (2003). Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins*, **50**, 437–450.
48. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. Sect. D*, **54**, 905–921.
49. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. Sect. D*, **60**, 2126–2132.
50. Winn, M. D., Isupov, M. N. & Murshudov, G. N. (2001). Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. Sect. D*, **57**, 122–133.
51. DeLano, W. L. (2002). *The PyMOL Molecular Graphics System.*; http://www.pymol.org.
52. Goldberg, R. N., Kishore, N. & Lennen, R. M. (2002). Thermodynamic quantities for the ionization reactions of buffers. *J. Phys. Ref. Chem. Data*, **31**, 231–370.
53. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H. *et al.* (2000). The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242.