# Structural insights into the DNA recognition mechanism by the bacterial transcription factor PdxR

Ida Freda[1,†], Cécile Exertier[2,†], Anna Barile[2], Antonio Chaves-Sanjuan [3], Mirella Vivoli Vega[4], Michail N. Isupov[5], Nicholas J. Harmer [6], Elena Gugole[2], Paolo Swuec[7], Martino Bolognesi [3], Anita Scipioni[8], Carmelinda Savino[2], Martino Luigi Di Salvo[1], Roberto Contestabile [1,9], Beatrice Vallone[1,2], Angela Tramonti[2,*] and Linda Celeste Montemiglio [2,*]

[1]Department of Biochemical Sciences "A. Rossi Fanelli", Sapienza, University of Rome, Rome 00185, Italy, [2]Institute of Molecular Biology and Pathology, National Research Council, Rome 00185, Italy, [3]Department of Biosciences, Pediatric Clinical Research Center Romeo ed Enrica Invernizzi and NOLIMITS, University of Milano, Milano 20133, Italy, [4]School of Biochemistry, University of Bristol, University Walk, BS8 1TD Bristol, UK, [5]Geoffrey Pope Building, University of Exeter, Stocker Road, Exeter EX4 4QD, UK, [6]Living Systems Institute, University of Exeter, Stocker Road, Exeter EX4 4QD, UK, [7]Cryo-Electron Microscopy Core Facility, Human Technopole, Milano 20157, Italy, [8]Department of Chemistry, Sapienza, University of Rome, Rome 00185, Italy and [9]Istituto Pasteur-Fondazione Cenci Bolognetti, Sapienza, University of Rome, Rome 00185, Italy

## ABSTRACT

**Specificity in protein–DNA recognition arises from the synergy of several factors that stem from the structural and chemical signatures encoded within the targeted DNA molecule. Here, we deciphered the nature of the interactions driving DNA recognition and binding by the bacterial transcription factor PdxR, a member of the MocR family responsible for the regulation of pyridoxal 5′-phosphate (PLP) biosynthesis. Single particle cryo-EM performed on the PLP-PdxR bound to its target DNA enabled the isolation of three conformers of the complex, which may be considered as snapshots of the binding process. Moreover, the resolution of an apo-PdxR crystallographic structure provided a detailed description of the transition of the effector domain to the holo-PdxR form triggered by the binding of the PLP effector molecule. Binding analyses of mutated DNA sequences using both wild type and PdxR variants revealed a central role of electrostatic interactions and of the intrinsic asymmetric bending of the DNA in allosterically guiding the holo-PdxR–DNA recognition process, from the first encounter through the fully bound state. Our results detail the structure and dynamics of the PdxR–DNA complex, clarifying the mechanism governing the DNA-binding mode of the holo-PdxR and the regulation features of the MocR family of transcription factors.**

## GRAPHICAL ABSTRACT
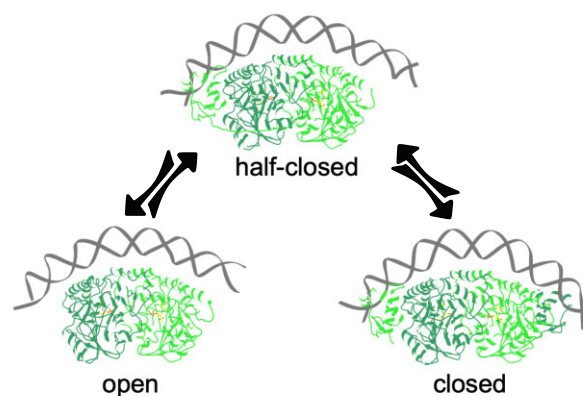


## INTRODUCTION

Gene expression is controlled by the activity of transcription factors (TFs). TFs are grouped in different families based on their domain composition. Usually, an effector domain senses specific environmental stimuli through the interaction with proteins or ligands, whereas a responsive

---

domain directly contacts the DNA through base-specific interactions (1).

The GntR family comprises a group of bacterial transcriptional regulators that share a common architecture. The N-terminal DNA-binding domain is composed of a winged type helix-turn-helix motif (wHTH), formed by the canonical three helical bundle and a β-strand hairpin (the wing) (2). The C-terminal effector binding and/or oligomerization domain varies and defines the classification of the GntR regulators into different subfamilies (3,4). Of these, the MocR subfamily is characterized by a large C-terminal domain that structurally resembles the fold type I pyridoxal 5′-phosphate (PLP)-dependent enzymes. As the archetype of these is aspartate aminotransferase (AAT) (5), this is called the AAT-like domain in the MocR subfamily. A long flexible linker connects the two domains of MocR (3,6). PLP is a highly versatile cofactor, contributing to around 4% of all known cellular enzymes, mostly reacting with amino acids (7). In MocRs, PLP, alone or in combination with other molecules that bind to it, can act as effector, inducing a conformational change in the protein that affects the TF DNA binding properties and triggers the transcription activation or repression. Some of the MocR family members are involved in the regulation of amino acid and vitamin B6 metabolism, whereas others regulate the catabolism of compounds used as substrates, like rhizophine and ectoine (8).

The most studied member of the MocR subfamily is GabR from *Bacillus subtilis*, a regulator of γ-aminobutyric acid (GABA) metabolism (9,10). To date, GabR is the only MocR TF whose crystal structure has been solved (11,12). GabR is a protein dimer, with a head-to-tail domain-swap arrangement in which each AAT-like domain interacts with the wHTH domain of the facing subunit (Supplementary Figure S1) (11). The structures of GabR in the apo form and in the presence of both PLP and GABA provided a starting point to unravel the mechanistic role of the effectors in the transcriptional activity of the GabR regulators (13,14). Functional and computational data suggest that large movements of the wHTH domain, with associated motions of the linker, accompany the apo-to-holo transition. These structural rearrangements regulate the DNA-binding and the transcriptional mechanism of GabR (11,15,16).

PdxR, a distinct MocR TF which shares ∼36% sequence identity with GabR, regulates the so-called DXP-independent pathway of the *de novo* biosynthesis of PLP in *Corynebacterium glutamicum* (17), *Streptococcus pneumoniae* (18), *Listeria monocytogenes* (19) and *Bacillus clausii* (20). This metabolic pathway is common to many other prokaryotes and involves the PLP synthase complex (21). When bound to PLP, PdxR acts as a repressor of its own *pdxR* gene and, in the absence of PLP, it acts as a direct activator of the *pdxST* operon, which encodes the PLP synthase subunits (17–20).

The biochemical and the DNA-binding properties of PdxR, as well as the organization of its target regulon, have been widely studied in the probiotic bacterium *B. clausii*, where the *pdxR* gene is adjacent to and transcribed divergently from the *pdxST* operon. In the *pdxR-pdxST* intergenic region, two direct (motif 1 and 2) and one inverted (motif 3) repeats represent the PdxR binding site (Figure 1). Functional data suggest that PdxR recognizes different combinations of these motifs on the target DNA depending on the presence of PLP, with the apo form showing comparable affinity for all the repeats whereas the holo form preferentially binds the inverted repeats (20). Therefore, PLP acts as an anti-activator by changing the DNA binding mode of PdxR and abolishing its activity as a transcriptional activator. Similarly to GabR, the switch in motif binding likely involves conformational changes of the wHTH domains through motion of the flexible linker (8,20). Although some aspects of the PdxR regulatory activity have already been clarified, a comprehensive understanding of the molecular mechanism underlying its function has been hindered by the absence of detailed structural information on the DNA-protein complex.

In this study, we integrated X-ray crystallography, single particle cryo-electron microscopy, DNA curvature prediction, and DNA binding studies to investigate the mechanism exploited by PdxR to bind and modulate the transcription of its target DNA. We observed that upon PLP-binding, the AAT-like domains of PdxR adopt a more extended conformation. Single particle analysis of the holo-PdxR–DNA cryo-EM data revealed the existence of multiple conformational states of the system, representing different snapshots along the dynamic path of the PdxR–DNA complex assembly. Binding assays performed on PdxR variants with modified DNA targets indicate that the specificity of the interaction is the result of an interplay between DNA shape recognition and direct DNA sequence readout, hinting at an allosteric role played by non-specific electrostatic contacts in modulating the asymmetric and sequential binding of the wHTH domains to the DNA.

## MATERIALS AND METHODS

### Construction of recombinant PdxR variants

The plasmid pET28*pdxR* (20) containing the coding sequence of the *pdxR* gene from *B. clausii* (strain Em Lm Pc rif Cf Cs Nv; Enterogermina N/R; Bacillus Genetic Stock Center) was used as a template in site-directed mutagenesis reactions performed by QuikChange methodology (Stratagene, La Jolla). Mutagenic primers, synthesized by Metabion International AG (Steinkirchen, Germany), are reported in Supplementary Table S1. All mutations were confirmed by sequence analysis (Microsynth AG, Balgach, Switzerland). Plasmids were purified using a NucleoSpin Plasmid kit from Macherey-Nagel.

### Expression and purification of apo- and holo-PdxR wild type and of the holo- variants

Transformed *Escherichia coli* BL21 (DE3) competent cells were grown at 37°C until the optical absorbance at 600 nm reached ∼0.5. PdxR overexpression was induced by the addition of 0.2 mM IPTG, then the temperature was lowered to 28°C for 18 h. Cells were harvested by centrifugation and the pellet was resuspended in lysis buffer containing 20 mM Tris·HCl, 0.5 M NaCl, 100 μM PLP, pH 8, and one tablet of cOmplete™ Protease Inhibitor Cocktail (Roche, Basel, Switzerland). Cells were lysed by sonication on ice, and

**Figure 1.** Schematic representation of the *B. clausii pdxR-pdxST* intergenic region, covering the *pdxR* and the *pdxS* promoters and the PdxR binding sites. The –10 and –35 regions of both promoters are highlighted in gray; the PdxR target direct repeats (motif 1 and motif 2) are in dark green and blue, respectively, and the inverted repeat (motif 3) is in light green. The arrows indicate the transcription orientation of the two genes.

lysate was centrifuged at 12 000–24 000 *g* for 30 min. The supernatant containing PdxR was recovered, and the proteins were precipitated using ammonium sulfate at 45% saturation, centrifuged for 30 min at 12000 *g* and resuspended in a solution containing 20 mM Tris·HCl, pH 8.0. The resuspended pellet was dialyzed overnight against 20 mM Tris·HCl, 0.5 M NaCl, pH 8, at 4°C, and then centrifuged for 30 min at 12000 *g*, filtered using 0.45 μm filters, and loaded onto a nickel-nitrilotriacic acid (Sigma-Aldrich, St. Louis, Missouri, USA) column pre-equilibrated with 20 mM Tris·HCl, 0.5 M NaCl, pH 8, using an ÄKTA prime FPLC system (Ge Healthcare, Chicago, Illinois, USA) at 1 ml/min. Protein was eluted with a linear gradient of 0–0.5 M imidazole in 20 mM Tris·HCl, 0.5 M NaCl, pH 8.0. Fractions containing PdxR were pooled, PLP was added (100 μM, final concentration), and the sample was dialyzed for 72 h against 50 mM HEPES sodium, pH 7.5, 0.5 M NaCl.

The concentration of PdxR (Mw~ 55 kDa, monomer) was determined by collecting the UV-visible spectrum using a theoretical extinction coefficient at 280 nm of 66 700 $M^{-1}$ $cm^{-1}$, estimated with the EXPASY PROTPARAM online tool (http://web.expasy.org/protparam). PLP content was calculated by adding 0.2 M NaOH and measuring the absorbance at 388 nm (22). Typically, 1 l of bacterial culture yielded ~10 mg of pure holo-PdxR. The same purification protocol was applied for all the holo-PdxR variant proteins.

To produce the apo form, pooled fractions eluted from the IMAC were incubated for 1 h at room temperature in the presence of 50 mM L-cysteine, and then dialyzed against 25 volumes of 20 mM Tris·HCl, 0.5 M NaCl, 5% glycerol, 20 mM L-cysteine, pH 8, overnight at 4°C with constant shaking. Sample was then concentrated and loaded onto a Sephadex 200 16/60 column at 1 ml/min and eluted in isocratic mode in 10 mM Na-Pipes, 250 mM NaCl, 50 mM trisodium citrate, pH 7. The apo formation was monitored following the disappearance of protonated PLP aldimine absorbing at 420 nm. Purified apo-PdxR was concentrated to at least 10 mg/ml for crystallization screening.

### Crystallization and structure determination

Crystals of apo-PdxR were obtained using the microbatch-under-oil method with protein at 5 or 10 mg/ml. The initial sparse matrix screen was set up with the Oryx6 crystallization robot (Douglas Instruments, Douglas House, East Garston, Hungerford, Berkshire, UK). The optimized crystallization solution contained 33% (vol/vol) PEG 400, 200 mM calcium acetate, and 200 mM imidazole, pH 7.5. Prior to flash freezing, crystals were soaked in cryoprotectant solution, containing 16% (w/vol) PEG 3350 and 30% (w/vol) PEG 400 in 10 mM Na-Pipes, pH 7.5, and 0.5 M NaCl.

X-ray diffraction data sets were collected at 100 K at the Diamond synchrotron beamline I03.

Data were indexed, integrated, and scaled using *Xia2* (23), *XDS* (24), and *Aimless* (25). The dataset was scaled in the space group *P*6122 to resolution of 2.8 Å. Phases were obtained by molecular replacement (MR) using the structure of GabR from *B. subtilis* (PDB code: 4N0B) (11) as a search model in *Molrep* (26), finding 10 copies in the asymmetric unit (four dimers plus two further dimers formed by crystallographic symmetry). Iterative cycles of model building and refinement of the structure were performed with *Coot* (27) and *Refmac5* v.8.0155 (28), 10-fold NCS averaging was performed by DM (29). The quality of the electron density allowed the reconstruction of 10 copies of the C-terminus of PdxR (residues 103–469) and of one of the 10 N-terminal domains (residues 3–83). No density for residues 84–102 was identified. A summary of statistics for data collection and refinement is reported in Supplementary Table S2.

### DNA fragments preparation

Various DNA fragments encompassing the wild type and modified PdxR target sequences were designed and purchased from Eurofins Genomics (Ebersberg, Germany) or Metabion International AG (Steinkirchen, Germany) as lyophilized forward and reverse single-stranded oligonucleotides (sequences are listed and schematically displayed in Supplementary Table S3). For the protein–DNA binding assay, the forward oligonucleotides were labeled at the 5′ end with the Cy3 fluorophore. Oligonucleotides were dissolved in sterile water to achieve a final concentration of 100 μM. Equimolar amounts of complementary oligonucleotides were mixed and denatured by heating to 95°C for 5 min, and then annealed by cooling to 25°C, reaching a final concentration of the double-stranded fragments of 50 μM.

### Holo-PdxR–DNA complex formation

Purified holo-PdxR in 50 mM HEPES, 0.2 M NaCl, pH 7.5, was incubated with a 48-bp synthetic DNA fragment containing the two direct and the inverted repeats of *pdxR-pdxST* intergenic region (wild type, Supplementary Table S3) in 1:2 ratio to generate the protein–DNA complex. The final sample was stored at 4°C for 48 h before cryo-electron microscopy (cryo-EM) experiments.

### Cryo-EM grid preparation and data collection

The concentrated solution of holo-PdxR–DNA complex was diluted with 50 mM HEPES, pH 7.5, to reach a

concentration of 0.5 mg/ml and to decrease the concentration of NaCl to 50 mM. Holey carbon grids (Quantifoil R1.2/1.3, Cu 300 Mesh, Quantifoil Micro Tools GmbH, SPT Labtech, China) were glow-discharged for 30 s at 30 mA using a GloQube system (Quorum Technologies, Lewes, UK). A 3.5 µl droplet of the sample containing the holo-PdxR–DNA complex was applied on a glow-discharged grid within the chamber of a mark IV Vitrobot (FEI, Thermo Fisher Scientific). The temperature was set at 4°C with 100% humidity. The sample was incubated 30 s, the excess solution was blotted for 4 s and the grid was rapidly plunged into liquid ethane.

Cryo-EM data were collected on a Talos Arctica electron microscope operated at 200 kV, using EPU 2.8 automated software (Thermo Fisher Scientific). Image acquisition was carried out at a nominal magnification of $120\,000\times$, corresponding to a calibrated pixel size of 0.889 Å/pixel at the specimen level, at nominal defocus values ranging between $-0.8$ and $-2.2$ µm. A total of 3284 movies were recorded using a Falcon 3EC direct electron detector (Thermo Fischer Scientific) operating in electron counting mode and total dose of 40 $e^-/A^2$, equally distributed over 40 movie fractions (1 $e^-/A^2$ per fraction).

**Cryo-EM data processing**

The main steps of data analysis are schematized in the workflow shown in Supplementary Figure S2. Collected movie frames were aligned and dose-weighted using *MotionCor2* (30) with 5 by 5 patches and B-factor of 150. Contrast transfer function (CTF) was estimated with *CTFFIND4.1* (31) on aligned and non-dose-weighted micrographs. Micrographs yielding resolution limit estimates $\leq 5$ Å were kept for further processing, discarding those containing relevant ice contamination or protein aggregates. The selected 3 197 micrographs were imported into *Relion 3.1* (32). After manual picking and the generation of 2D references, a total of 135 8491 particles were automatically picked and extracted. Several rounds of reference-free 2D classification were carried out, resulting in 267 988 of selected particles that were used to generate a *de novo* 3D model in *Relion*. The reconstructed model was low pass filtered to 60 Å and given as a reference for an initial round of 3D classification. The best and most populated class generated was chosen and used as a reference (filtered at 60 Å) for an additional round of 3D classification, run over the same set of particles. Inspection of the produced classes identified two distinct conformers of DNA-protein complex, displaying a DNA differently bound to the protein (Supplementary Figure S2): class 1 exhibited a compact shape with the DNA tightly and symmetrically bound to the protein which is anchored at both the ends of the double-strand filament (closed conformation); class 2 showed a single protein domain wrapping the DNA fragment and the second one missing, resulting in an asymmetric complex, with the DNA partially detached from the dimeric core of the protein (half-closed conformation). Particles belonging to these two classes (112 086 total particles) were selected and subjected to an additional round of 2D classification for further cleaning. The resulting 92 710 good particles were selected for further processing.

In parallel, autopicking was also performed with WARP (WARP 1.0.9) (33), using the deep convolutional neural network BoxNet2Mask_20180918 as a template. The resulting 321 382 particles were extracted and imported into *Relion 3.1*. Several rounds of 2D classifications were performed, yielding a set of 112 581 particles showing high resolution features, that were combined to the first refined particle dataset (see Supplementary Figure S2). To prevent invalidation of gold standard FSC by duplicates, particles that had been translated within 20 Å of their nearest neighbor were removed. The de-duplication procedure was performed twice. Individual stacks were subjected to a subsequent round of 3D classification, using the half-closed map as a reference model. Among the generated 3D classes, an additional conformer of the holo-PdxR–DNA complex emerged that differs from the others for the absence of densities at both the edges of the double strand filament (open conformation, Supplementary Figure S2). Particles belonging to different conformers (open, half-closed, and closed) were individually selected and independently processed.

After refinement with C1 symmetry and postprocessing, the resolution measured using a mask was estimated to be about 4.6 Å for the open map (56 199 particles), ~4.3 Å for the half-closed map (68 689 particles), and ~4.3 Å for the closed map (32 628 particles). Inspection of the reconstructed maps revealed an incorrect handedness that was reversed using the 'vop zflip' command of *UCSF Chimera* (34).

Bayesian particle polishing and additional masked refinement of the polished particles improved the resolution to 3.9 Å for the closed conformer, and 4.0 Å for the half-closed and the open conformers. Since the quality of the closed map does not allow a clear distinction between the nucleotide bases of the DNA, we took advantage of the pseudo-symmetry of this conformer to perform a parallel job of refinement imposing a C2 symmetry. This increased the resolution to about 3.7 Å.

CTF-refinement did not improve the resolution of the resulting reconstructions. The B-factors were automatically calculated by *Relion* in the post-processing of the reconstructed maps.

Features consistent with the estimated resolution, including clear helical pitch throughout the maps, and well-defined density for bulky side chains, were identified (Supplementary Figures S3–S6). The overall resolution of the maps obtained was estimated with the FSC = 0.143 criterion. The input maps were corrected for the modulation transfer function (MTF) of Falcon 3EC detector and sharpened using negative temperature B factors as estimated by *Relion*. Local resolution of the final maps was estimated in *Relion* (Supplementary Figures S3–S6).

Statistic information of the maps is reported in Supplementary Table S4. All reconstructed maps were sharpened using the Autosharpen tool in *Phenix* prior to model building. *UCSF Chimera* and *Coot* were utilized for graphical visualization of the cryo-EM density maps.

**Model building and refinement**

A basic requirement for the reconstruction of the holo-PdxR–DNA complex structures was the identification of

the correct orientation of the bound DNA fragment with respect to the protein. Indeed, the quality of the reconstructed maps was sufficient to distinguish DNA grooves and to locate the phosphate backbone, but it did not allow unambiguous assignment of the specific bases. As a guide for model building, we took advantage of the asymmetry of the half-closed map conformer, that we found to reflect the intrinsic asymmetric curvature of the PdxR DNA target resulting from the computational analysis of its intrinsic bendability (see below) (Supplementary Figure S7A). In fact, it must be considered that, apart from motifs 1 and 3, the absence of palindromic features in the sequence of the DNA fragment forces the DNA to assemble in a unique orientation with respect to the protein.

Hence, after creating an initial model of the DNA in a straight B conformation in *Coot*, we docked it into the half-closed map, following the asymmetry in the curvature suggested by the prediction. Thereafter, an extensive manual rebuilding and refinement was performed to obtain the right curvature and the best fit. As a further element in support of the correct assignment, we searched for good quality areas of the map corresponding to tracts of repeated bases (i.e. the A/T rich tracts, Figure 1). Supplementary Figure S7B shows a DNA tract where the density can only accommodate repetition of pyrimidine bases in one strand and purine bases on the other, but not *vice versa*.

Following this criterion, the DNA was fitted into the half-closed map with motif 1 bound to the visible wHTH domain, and motif 3 corresponding to the portion of DNA detached from the protein, where density accounting for the wHTH was missing.

Once a model for the DNA was built, a model for the structure of PdxR was obtained by using the coordinates of the crystal structure of apo-PdxR (PDB code: 7PQ9), comprising the wHTH domain and the N-terminal tract of the linker (residues 3–83) and the AAT-like domain and the C-terminal portion of the linker (from residue 103 up to residue 469). Both models were individually fitted as rigid bodies to the sharpened density map using *UCSF Chimera*, merged, and transferred to *Coot*, where the protein chain was joined to the DNA. The whole structure was rigid body refined against the half-closed map using the real space refinement tool in *PHENIX*. The model was inspected in *Coot* for manual model building and further refined using *phenix.real_space_refine* tool with all default settings and imposing secondary structure restraints.

Once the coordinate model for the half-closed conformer was built, it was used as a starting template to reconstruct both the closed (in C1 and C2 symmetry) and the open conformers obtained for the PdxR–DNA complex using *Coot*. For the closed conformer, a second wHTH domain was copied and docked to the corresponding region of the density, while in the case of the open conformer, the wHTH domain bound to DNA was removed from the coordinate model, given the absence of suitable density. Insufficient or lacking density was found at the N-and C-termini (residues 3–6 or 3–14 and 465–469) and at the linker region (from residues 80–87 to 100–104) of those maps, and the corresponding model coordinates were removed from the model. Also, the first two and the last five base pairs of the DNA fragment at the 5′ and 3′ ends respectively were removed

from the open structure due to weak density in those regions. Both the open and the closed (C1 and C2 symmetry) models were then refined using *phenix.real_space_refine* tool with all default settings and imposing secondary structure restraints. Since the quality of the map does not allow to clearly identify the DNA bases, for the structural analysis herein discussed the C2 symmetry map and the relative coordinate model was used as representative of the holo-PdxR–DNA closed conformer.

In all reconstructed maps, we observed extra density extending out from Lys309 to the PLP-cavity, clearly visible in both sites of the PdxR dimer. We assigned this density to a PLP molecule, whose model was imported in *Coot* and linked as a Schiff-base to Lys309 of both the AAT-like monomers of all maps, consistently with our spectroscopic analysis.

Final visual inspection was performed in *Coot* to adjust secondary structures, amino acid side chains, PLP and nucleotides into the density, and to correct Ramachandran outliers. The quality of the final model was validated using *MolProbity* (35) and the quality of side chain fit with *EMRinger* (36). Statistics for the final model are shown in Supplementary Table S4.

The web server STRIDE (37) was employed for secondary structure assignments (Supplementary Figure S8). PISA (38) was used to analyze the 3D structures, *UCSF Chimera* and Chimera X (39) were used to produce figures.

### Prediction of DNA topology from sequence and generation of DNA fragments with altered shape

The 3D structure of the 48-bp DNA target of PdxR was created through the Christoph Gohlke DNA Curvature Analysis service (C. Gohlke, http://www.lfd.uci.edu/~gohlke/dnacurve/) (40), using the AA wedge model (41). The same procedure was carried out for the modified DNA fragments (Supplementary Table S3).

The target DNA sequences with altered curvature were designed according to the DNA curvature model proposed by De Santis *et al.* (42). The curvature was calculated in modulus and phase adopting the roll, tilt, and twist angles theoretically evaluated for the 10 different dinucleotide steps by a conformational energy minimization process (42). The most curved fragments were obtained by mutating some nucleotides to increase the length of A and T tracts in phase with the double helix periodicity. On the contrary, breaking the A and T tracts by the insertion of C and G nucleotides along the wild type DNA sequence destroyed the 10-bp periodicity of the dinucleotide steps AA and TT and provided the straight fragment.

### Fluorescence DNA-binding assay

PdxR binding to target DNA fragments was analyzed by titrating PdxR against a fixed concentration of Cy3-labeled DNA in 20 mM Na-HEPES buffer, pH 8.0, containing 50 mM KCl, 5 mM $MgCl_2$, 2% (vol/vol) glycerol, 1 mM DTT. The Cy3-labelled DNA (5 nM) was mixed with the serially diluted PdxR protein (final concentration, 0–480 nM) in a final volume of 800 μl at 25°C. PLP (0.5 μM) was added to ensure complete saturation. Fluorescence emission spectra

were recorded from 550 nm to 650 nm upon excitation at 546 nm at 25°C using a FluoroMax-3 Jobin Yvon Horiba spectrofluorimeter (Kisshoin, Japan). Excitation and emission slits were set at 3 nm and 5 nm, respectively. Emission fluorescence values at 560 nm obtained from three different experiments were combined and analyzed using the Hill equation (Eq. 1) which gives the apparent dissociation constant ($K_{Dapp}$) values of the DNA-protein binding equilibria.

$$Fluorescence = \frac{F_{max} \times [PdxR]^n}{[PdxR]^n + K_{Dapp}^n} \quad (1)$$

where $F_{max}$ is the maximum fluorescence, and $n$ is the Hill factor.

## RESULTS

### Crystal structure of the apo-PdxR AAT-like domain

The structure of apo-PdxR, i.e. without PLP, was solved by X-ray crystallography at 2.8 Å resolution, using crystals of the full-length protein in the apo form. The quality of the electron density allowed the reconstruction of the C-terminal portion of PdxR (residues 103–469) that comprises the C-terminus of the interdomain linker (residues 103–106) and the entire AAT-like domain (Supplementary Figure S8). The asymmetric unit comprises ten PdxR protomers, with four complete PdxR dimers and two protomers that form dimers across crystallographic axes (Supplementary Figure S9). No density for the central region of the linker (residues 80–106) was identified. One isolated unit of the wHTH domain (residues 3–83) over 10 copies of PdxR dimer could be built in the asymmetric unit. This domain is trapped by crystallographic contacts between five protomers in the lattice. As no density was present for the linker, it was not clear which AAT-like domain protomer is the same chain as this wHTH domain. This precludes attributing functional significance to the interactions formed with the surrounding monomers within the crystal. Consequently, these contacts will not be discussed. As for GabR (43), in PdxR an extended loop (residues 109–120) directly connecting the linker to the AAT-like domain and exposed to the PLP cavity was identified. The AAT-like domain was found in the canonical aminotransferase homodimeric assembly, typical of all fold type I PLP-dependent enzymes (Supplementary Figure S10) (5). Each protomer is composed of a large (residues 107–387) and a small (residues 388–468) subdomain (Figure 2A). An extended dimer interface covers about 2 350 Å², involving 52 amino acids of the two monomers interacting through hydrogen bonds and salt bridges (38). A network of interactions involving aromatic residues located at the α4 helix of both monomers contributes to the stabilization of the dimer interface. This network starts from a stacking nucleus formed by a pair of facing Phe128 residues and extends over the N-terminus of the helix, involving Trp124, Phe121 and Phe119 (Figure 2B).

The PLP-binding cavity resides in the large subdomain in the proximity of the subunit interface. Within this cavity, we identified the quartet of residues typical of PLP-binding proteins: i) Lys309, the so-called 'active site lysine', conserved in aminotransferases (e.g. Lys258 in *E. coli* AAT)

as well as in MocR TFs (Lys312 in GabR) (11); ii) Ser308 and Arg316, the canonical residues that in the holo forms of AAT enzymes (Ser257 and Arg266 in *E. coli* AAT) and MocRs (Ser311 and Arg319 in GabR) contact the phosphate group of PLP (44); iii) the conserved Asp276 (Asp222 in *E. coli* AAT and Asp279 in GabR) interacting with the pyridinium nitrogen of PLP to maintain the effector in the protonated state (Figure 2C). In addition, in PdxR the PLP-binding pocket is surrounded by other residues, such as Tyr202 and Tyr278, that are also conserved in *B. subtilis* GabR (Tyr205 and Tyr281, respectively). However, the organization of these residues in the apo-PdxR markedly differs from the apo-GabR crystal structure, where a molecule of imidazole bound to the PLP cavity was detected (11). In PdxR, Tyr202 is rotated by approximately 75° with respect to the corresponding Tyr205 in GabR, and it forms a π-stacking interaction with Phe247. Tyr278 is rotated by ∼95° with respect to Tyr281 in GabR and it establishes hydrogen bonds with Asn142 of the other monomer and with Ser308, which assumes a different conformation with respect to the corresponding Ser311 in GabR (Supplementary Figure S11). In the absence of PLP and unintentionally bound ligands (i.e. imidazole in GabR), these residues in PdxR fill the cavity, with Tyr278 positioned at its center, supported by a network of π-stacking, H-bond, dipole-dipole, *van der Waals* interactions established with Tyr202, Ser308, Lys309, Arg316, Thr112, Val178, Asp279 and Asp276 of the same monomer and with His338 and Asn142 of the other monomer (Figure 2C).

### The cryo-EM analysis of *B. Clausii* holo-PdxR–DNA complex enables the identification of different conformational states

*Overall structures.* Our first attempts to obtain structural information about PdxR in the PLP-bound form were pursued through X-ray crystallography. However, crystallization screening carried out on the holo-PdxR, either in isolated form or bound to fragments of different sizes of its target DNA, proved to be unsuccessful. Instead, we performed cryo-EM experiments in which the holo- form of the PdxR dimer was analyzed in complex with a 48-bp double strand DNA fragment, covering the two direct and the inverted repeats predicted as PdxR binding sites (Figure 1 and Supplementary Table S3) (20). Single-particle analysis allowed us to reconstruct the maps of three forms of the holo-PdxR–DNA complex in which the DNA is differently bound, bent, and wrapped around the protein (Figure 3). Consistently with the high flexibility of the N-terminal regions of PdxR that most likely prevented particle alignment required for a clear 3D map reconstruction, the three conformers differ in the presence of density defining the linker and the wHTH domains. Depending on the extent of the contacts established and the consequent level of mutual wrapping of protein and DNA, we referred to the three states as the 'open', the 'half-closed' and the 'closed' conformers.

In the open conformer, the DNA-protein contact is only mediated by the ridge of the central AAT-like domain of PdxR, since both tails of the DNA fragment were found free from direct interactions with the protein. No density ascribable to the linkers and wHTH domains of both PdxR
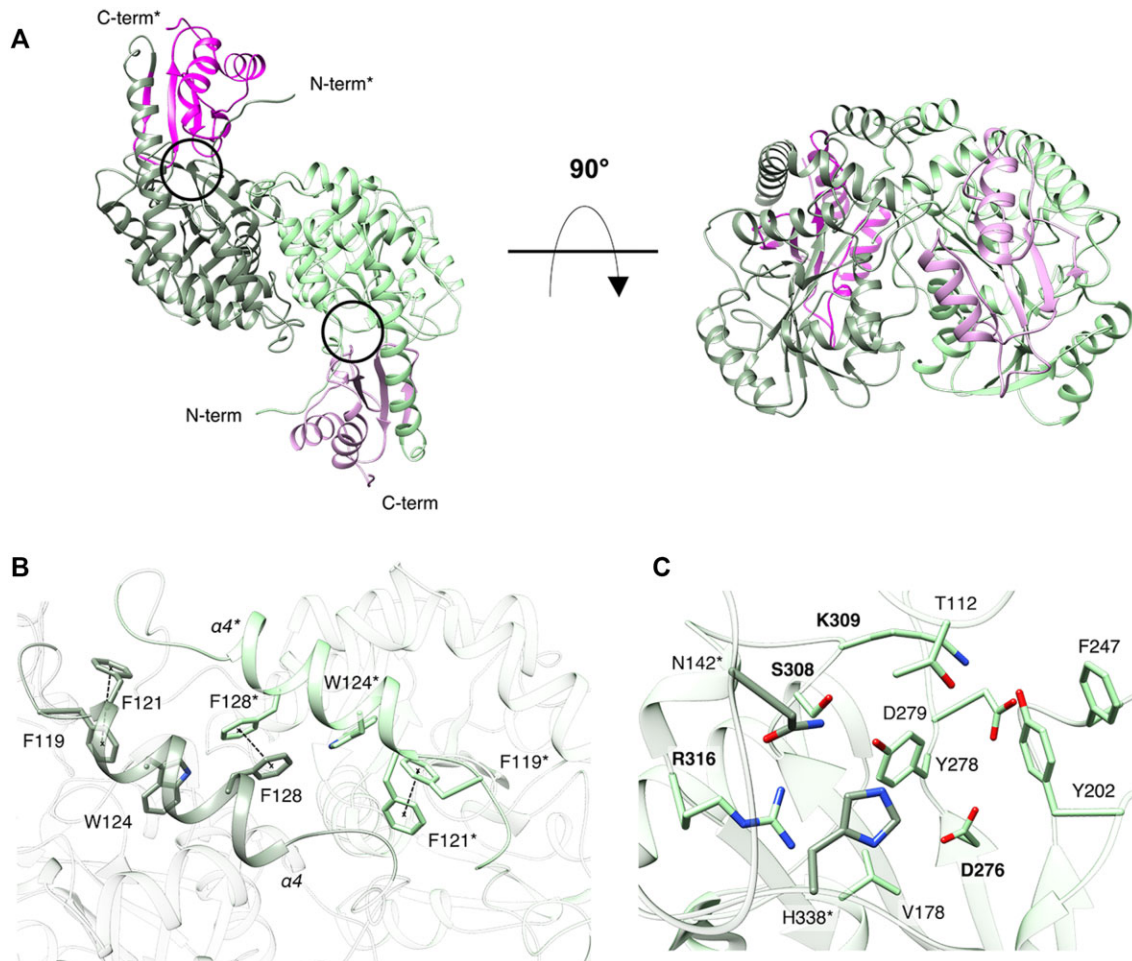
**Figure 2.** Structure of apo-PdxR AAT-like domain. (**A**) The homodimeric crystal structure of *B. clausii* apo-PdxR in ribbon representation (top and side view). Plum and light green colors were used to distinguish the small and the large subdomains, respectively, of one subunit; Magenta and dark green colors were used for the small and large subdomains of the other subunit. Black circles indicate the PLP cavity. (**B**) Interactions at the subunit interface. Pairs of Phe128-Phe128 and other aromatic residues at helix α4 involved in the intersubunit contacts are represented as sticks. Dashed lines show the stacking interactions. (**C**) PLP-binding site of apo-PdxR. Residues surrounding the empty pocket are represented as sticks and labelled. Bold residue labels indicate the quartet of residues typical of PLP-binding proteins. Asterisks refer to secondary structures and residues belonging to the flanking subunit.

monomers was detected in the map. In this form, the total angle of curvature of the DNA is about 130°.

In the half-closed complex, one of the two wHTH domains is anchored to the DNA tail at motif 1, while the other, missing in the density map, is most likely dissociated from the AAT-like domain of the adjacent monomer. Therefore, in this conformer the protein is not attached to the DNA tail at motif 3, resulting in an asymmetric assembly (Figure 3). From the open to the half-closed conformation, no change in the global curvature of the DNA was detected, which remains arched at about 130°.

In the closed conformation, the simultaneous binding of the two wHTH domains to DNA compacts the structure of the complex. In this arrangement, the DNA is strongly bent and wrapped around the protein, forming an arc of ~115° over PdxR and conferring an overall symmetry to the PdxR–DNA complex (Figure 3).

Since PLP binding to PdxR may trigger conformational changes including large reorganizations of the wHTH domain which moves jointly with the linker (20), the accumu-

lation of different conformational populations of the holo-PdxR–DNA complex might derive from diverse occupancies of PLP among the PdxR homodimers. Nevertheless, a close look at the PLP-binding site of all reconstructed maps of the complex revealed in all cases an extra density in both monomers of PdxR (Supplementary Figure S12), perfectly fitting with a molecule of PLP covalently bound to Lys309 (described below). Therefore, the structural differences observed among conformers do not result from a lack of or partial binding of the effector. Rather they indicate that holo-PdxR explores alternative conformational states upon binding to its target DNA. The co-existence of multiple conformers of the holo-PdxR–DNA complex revealed by single-particle cryo-EM suggests the existence of a dynamic equilibrium of these intermediates in solution.

*The apo-to-holo conversion of PdxR.* The superimposition of the crystallographic structure of apo-PdxR dimer to the open holo-PdxR–DNA complex obtained by cryo-EM (residues 107–464: overall rmsd mean on $C_\alpha$ carbons
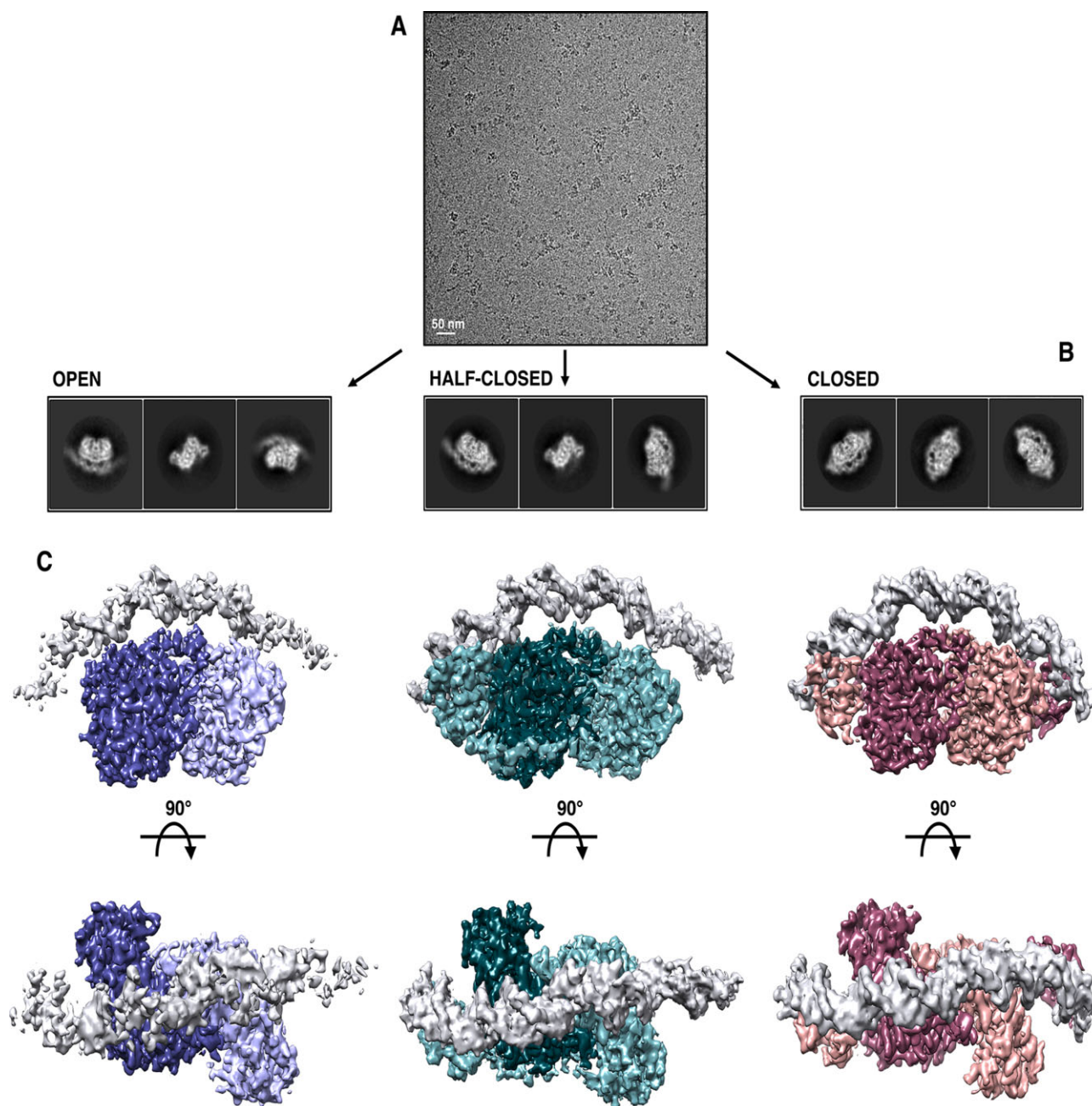
**Figure 3.** Cryo-EM structures of the PLP-bound PdxR in complex with its target DNA. (**A**) Representative cryo-EM micrograph for the holo-PdxR–DNA complex. (**B**) Representative 2D class averages of the open, the half-closed, and the closed conformer of the complex as calculated with *Relion*. (**C**) Side and top views of the cryo-EM 3D reconstructions of the holo-PdxR–DNA complex.

∼1.4 Å, with maximum displacement of about 7 Å), showed that the presence of PLP covalently bound to Lys309 at both the AAT-like domains alters the mutual positioning of the protein core subunits, inducing an overall structural movement that opens the dimer (Figure 4A, MovieS1). A closer inspection of the PLP binding pocket revealed that the effector induces pronounced local structural rearrangements that involve residues surrounding and lining the site (Figure 2C, 4B and Supplementary Figure S11). In more detail, upon PLP binding, the pyridine ring of the effector displaces Tyr278, which moves by about 10 Å, shifting Asp279

by around 2.5 Å (Figure 4C). The side chain of Tyr202 rotates by about 45°, directly interacting with the pyridine ring of PLP at the center of the cavity (Figure 4D). Following these movements, the entire AAT-like domain rearranges, with effects that extend up to the external surface (Figure 4E and F). Posed at the center of the cavity, the PLP ring contacts all the surrounding residues, most of which are conserved in GabR and in fold type I PLP-dependent enzymes (11,45). More specifically, i) the nitrogen atom of the PLP pyridine ring interacts with the side chain of the conserved Asp276 on β8 (11); ii) the hydroxyl group at the C3 of the
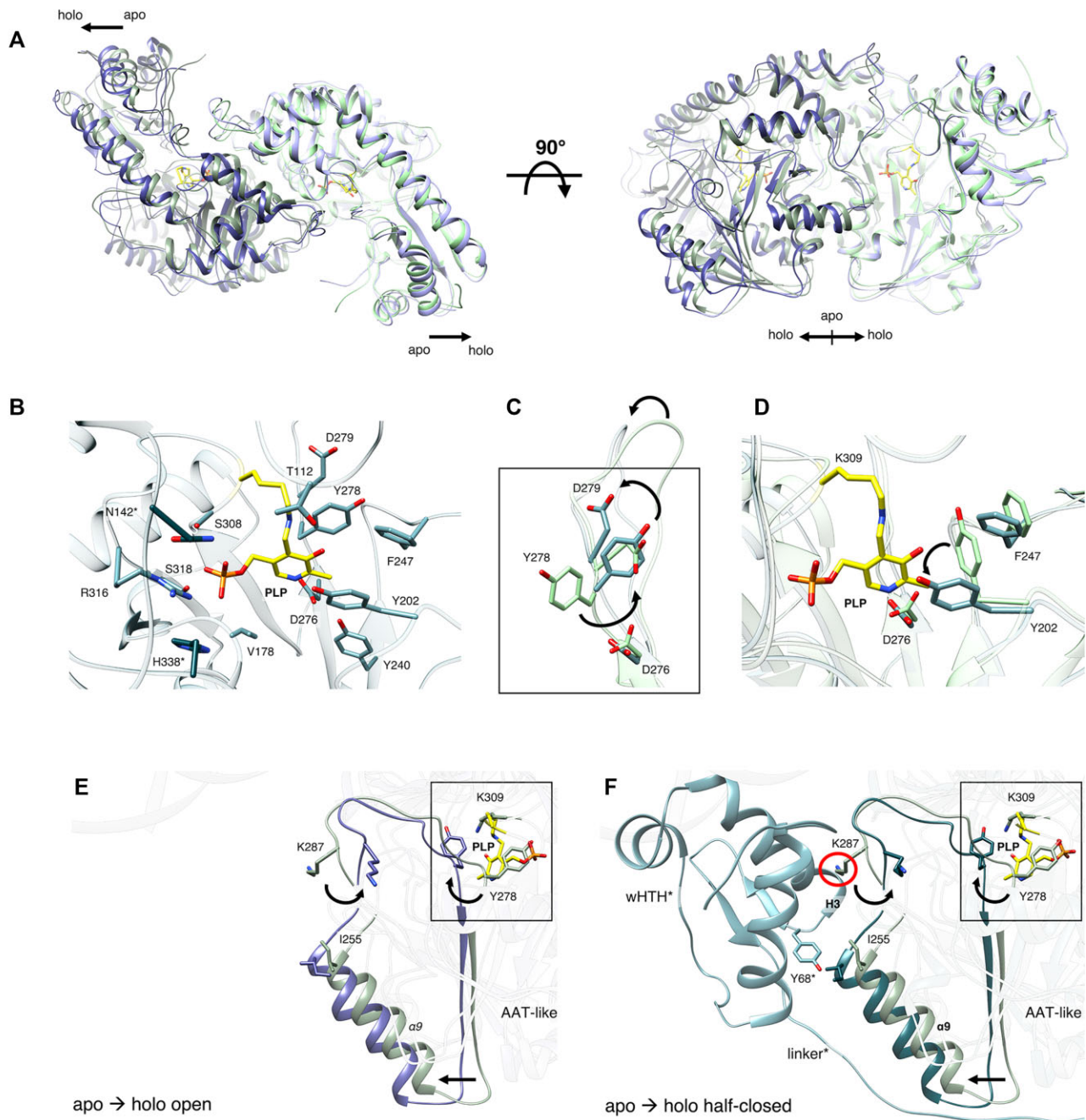
**Figure 4.** From apo- to holo-PdxR: structural changes at the PLP-binding site affect the overall structure of the AAT-like domain. (**A**) Superposition of the apo- (green) and holo-PdxR (open conformation, purple) shown as side and top views. Dark and light colors are used to distinguish monomers. Upon PLP binding, the AAT-like domain opens, arrows indicate the direction of the whole movement. PLP is shown as yellow sticks. (**B**) The structural organization of the PLP-binding site found in the holo-PdxR is reported. For the sake of simplicity, we used the half-closed holo-PdxR structure as representative since only minor variations were observed at the PLP-binding site comparing all the holo- structures. Residues interacting with PLP and located within ∼5 Å of the effector are represented as sticks and labelled. (**C, D**) Structural changes at the PLP-binding site: structural details of the transition explored by Tyr278, Asp279 and the 277–291 segment (panel C), and by Tyr202 once the molecule of PLP is bound to Lys309. (**E, F**) The structural effects of the PLP-binding transmitted to the external surface of the AAT-like domain are displayed. The rotation of Tyr278 induced by the presence of PLP within the cavity, zoomed in panel C, is highlighted by a black box. In panel E, the apo structure (green) is superimposed to the holo structure in the open conformation (purple); panel F shows the superposition of the apo (green) to the holo half-closed structures (dark and light blue color the AAT-like and the wHTH domain with the linker of the other monomer, respectively). As an example of helix α9-wHTH domain interaction, the Ile255-Tyr68 contact is shown. Lys287 and Ile255 on the AAT-like domain and Tyr68 located in the wHTH domain of the flanking monomer are represented as sticks and labelled. Red circle indicates the clash point between Lys287 and residues at α(H)3 of the wHTH domain that would occur if the structural rearrangement induced by PLP binding did not take place. Arrows indicate the direction of the apo-to-holo movements.

pyridine ring contacts the conserved Tyr202 on the β5-α7 turn, Phe247 on the β8-α9 turn and Tyr278 on β9-α10 turn, anchoring the PLP unit within an aromatic niche; iii) the phosphate group is involved in electrostatic interactions (*i.e.* hydrogen bonds and salt bridges) with the conserved Ser308 on the β10-η3 turn and Arg316, and with Ser318 on β11; iv) the imine nitrogen forms a hydrogen bond with Thr112 on the extended loop (Figure 4B). Further arrangements of the Asn108-Ile114 and Cys133-His145 segments and of the tract defined by Ser342-Ser335 of the other monomer (not shown) contribute to expanding the PLP-cavity volume, providing the space required for the positioning of PLP.

Notably, the structural arrangement of the PLP-binding site found in the holo-PdxR closely resembles the arrangement observed in the crystal structure of the apo-GabR (Supplementary Figure S11). Here, a molecule of imidazole derived from purification or crystallization buffers was detected bound to the site (11). This observation led us to conclude that the presence of imidazole in the PLP-binding cavity of GabR stabilizes a holo-like conformation. In addition, we noted that the holo-PdxR conformation could be accommodated in the crystal lattice (Supplementary Figure S13), suggesting that the crystallographic structure of apo-PdxR is the preferred apo conformation and represents the first *bona fide* apo-structure of the AAT-like domain of a member of the MocR family.

Only minor structural variations were observed at the PLP-site when superposing the AAT-like domain of the open, the half-closed, and the closed structures of the holo-PdxR dimer (residues 107–464: overall rmsd on $C_\alpha$ carbons ∼0.6 Å, with maximum displacement of about 3.5 Å). Therefore, once the effector is bound, the structural organization of the site is only negligibly affected by the overall transition of the protein occurring upon the DNA-complex formation.

The analysis of the structural changes involving the AAT-like domain shows that the ensemble of the rearrangements induced by PLP entry extends over the entire AAT-like domain, leading to a global opening of the dimer. It is reasonable to suppose that this movement could contribute to the modulation of the DNA binding process since it may affect the structure of the linker and the position of wHTH domains of PdxR. Indeed, a crucial structural response that has been proposed to be directly induced by the effector binding to MocRs is a change in the position of the wHTH domains. This is associated with a rearrangement of the linker region (15,20). Being endowed with a high level of flexibility, this region is generally hard to describe from a structural point of view. To date, the only information available on the conformation of the linker comes from the crystallographic structures of GabR (11,12). In our data (apo and holo) the PdxR linker is described by a good quality density map for the amino acids placed at the terminal regions (residues 80–83/87 and 101/103–106, depending on the conformer and the monomer). The density was partially defined or absent for the central portion (residues 83–100) in most structures, preventing a clear-cut identification of the spatial position adopted by this segment upon the apo-to-holo conversion. However, a defined but not continuous density outlining the linker path was identified in one of

the monomers of the half-closed holo-PdxR. This density, starting from Gln80, running along the AAT-like domain of the flanking monomer, and extending towards Arg106, was used to build the entire tract of the linker (Figure 5A and Supplementary Figure S14A). According to the reconstruction, the segment including residues 100–106 protrudes into the most external region of the PLP-binding cavity, with the side chain of Lys101 extending in the direction of the phosphate group of the effector. Interactions established between the ε-amino group of the lysine side chain and i) Asn108 and His113 on the extended loop, ii) Tyr423 on the β15, iii) the backbone carbonyl of Asn142 (of the flanking monomer), sustain Lys101 position and orientation (Figure 5B). The spatial organization of the reconstructed linker in PdxR suggests that it moves close to the PLP-cavity upon PLP binding (Supplementary Figure S14B), as previously proposed (15,20,43,46).

In contrast with the linker described in GabR, which forms an α-helix at the C-terminal region upon PLP binding (11,12), our data did not reveal the presence of secondary structure elements within the linker.

One unclear aspect of the regulatory mechanism of the PdxR transcriptional activity is how the effector binding triggers PdxR conformational changes that affect the DNA binding mode, and so modulate PdxR function. We note that the displacement of Tyr278 upon PLP binding at the core of the AAT-like domain is likely to be crucial to the apo-to-holo structural change. This displacement directly affects the arrangement of the structural elements that compose the external surface at the side of the large subdomain (Figure 4E and F). This area constitutes the docking site of the wHTH of the other monomer observed in both the half-closed and the closed conformations. Indeed, the entire tract 255–288 undergoes a discrete structural repositioning (Figure 4E). Particularly, the flip of the loop 283–288 results in a rotation of about 30° in Lys287. In the apo structure the side chain of this residue is exposed to the solvent, while in the holo state it is oriented towards the AAT-like domain. The superposition of the apo onto the holo half-closed structure reveals that the flipping of the loop 283–288 generates a more favorable interacting surface for the wHTH domain, that otherwise would collide with Lys287 and other elements of the 283–288 loop (Figure 4F). Likewise, the concomitant movement of helix α9 towards the solvent supports the interaction with the wHTH domain and the linker, since it favors the approaching of residues that will constitute the binding site for both the structural elements (Figure 4F and Supplementary Figure S14A).

Together, these observations indicate that structural variations of the AAT-like domain upon PLP binding may guide a different positioning of the wHTH domain and the linker to which it is connected, suggesting a possible mechanism underlying the apo-to-holo switch in DNA binding mode. Indeed, the structural changes induced by PLP binding to the AAT-like domain could alternatively hamper (apo) or favor (holo) the approach of the mobile domains to the PdxR core, thus leading to a selective stabilization of alternative conformations of PdxR in the apo and holo states that have different abilities to bind the DNA.

Notably, these rearrangements are observed in all the holo structures, even in those where the wHTH domain and
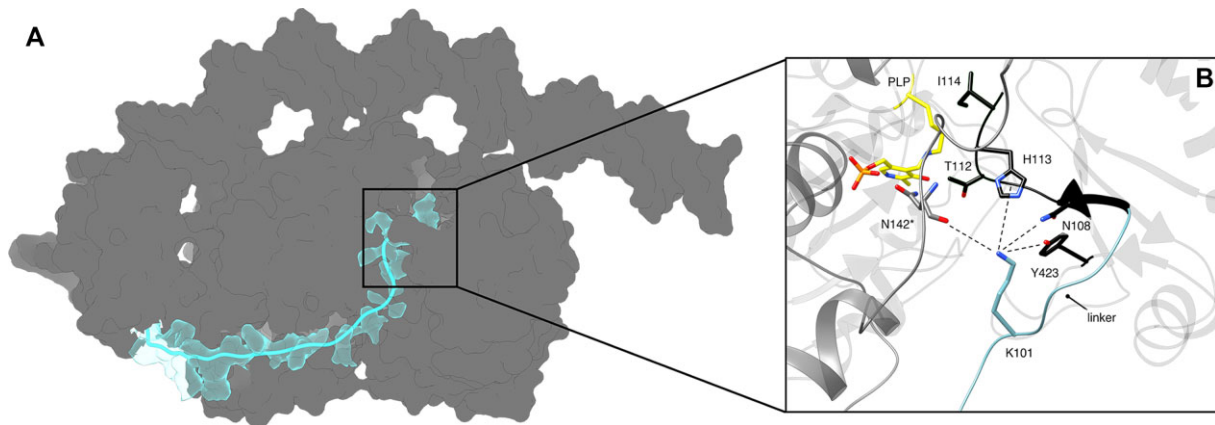
**Figure 5.** From apo- to holo-PdxR: the structural organization of the linker region. (**A**) The linker path in the half-closed PdxR structure. The structure is represented as a dark gray surface, while the linker is shown in cyan ribbon representation, surrounded by the reconstructed cryo-EM density map (solid representation). (**B**) Close-up view of the external side of the PLP-binding site, where the C-terminus of the linker and the extended loop (residues 100–114) approaches. Lys101 and the interacting residues located within ∼5 Å, Thr112 and Ile114 are represented as sticks. The PdxR subunits are shown as gray and black ribbons. Hydrogen bonds and dipole-dipole interactions are represented using dashed lines. The asterisk indicates residues belonging to the flanking subunit.

the linker have sufficient conformational flexibility that they are not visible in the cryo-EM maps (i.e. the open and the half-closed structures). This indicates that these movements are a direct consequence of PLP binding and that they occur before the structural changes that bring the linker and the wHTH domains close to the core domain of PdxR. Binding to DNA may further induce changes in the structural organization of PdxR that lead to the conformations observed by cryo-EM in the final complexes. Indeed, the linker and wHTH domains may be sampling a discrete range of conformations, and DNA binding captures them in one of these poses.

*Holo-PdxR–DNA interaction.* The structural transitions induced in PdxR upon PLP binding to the AAT-like domain have a strong impact on how PdxR interacts with its target DNA. The cryo-EM structures reveal a multi-step structural reorganization of the complex. Such variations involve concerted movements of both molecular partners, with the accumulation of at least three intermediates. Holo-PdxR and DNA must establish a large network of specific and non-specific interactions to achieve the highest level of binding, DNA bending, and wrapping, typical of the closed structure.

The recognition of specific DNA sequences is mediated by the wHTH domain of PdxR, which is structurally conserved among the GntR family (Figure 6A-B). The observed DNA-binding mode closely resembles the one found in FadR-DNA structure (47). The helices and the wing are positioned in a conserved orientation with respect to the nucleic acid (Supplementary Figure S15): the helices α(H)2 and the so-called 'recognition helix' α(H)3 occupy the DNA major groove, while the wing contacts the minor groove. In the PdxR–DNA half-closed conformation, the binding to DNA is achieved through a single wHTH domain that contacts the first direct repeat (motif 1, Figure 1). This interaction stabilizes the initial partial bending of the DNA observed in the open structure: its double strand is arranged on the longitudinal plan of PdxR with an asymmetric shape.

In its closed conformation, the holo-protein binds DNA with the two wHTH domains anchoring both the first direct (motif 1) and its palindromic inverted repeat (motif 3) on the opposite sides of the DNA filament (Figure 1 and Figure 3C). In this state, the DNA molecule is largely bent to allow the simultaneous contact between the two wHTH domains of PdxR and the two DNA-binding motifs, longitudinally spanning over the protein surface.

Three residues of the wHTH domain of PdxR were proposed to be involved in DNA interaction, namely Arg43 and Ser52 located on the HTH motif, and Lys75 on the wing (20). The cryo-EM structures of both the half-closed and closed holo-PdxR–DNA conformers validate the involvement of Arg43 and Lys75 in the interactions with the DNA binding motifs, while Ser52 contacts neighboring nucleotides. In addition, other interactions formed by flanking residues, i.e. Lys42 and Lys44 on the α(H)2, Gln53, Ans54, Thr55, and Glu57 on the α(H)3, and Arg74 and Phe77 on the wing, were found to contribute to the sequence readout interaction with the DNA motifs 1 and 3 (Figure 6A, B). Other residues in the wHTH domain further stabilize the DNA binding at the cognate sites by means of electrostatic interactions established with the phosphate backbone at a minor groove, namely Arg10, Pro15, Leu16, and Tyr17 on the α(H)1, and Ser41 on the α(H)1- α(H)2 loop.

Besides the specific contacts formed with the canonical DNA binding domain, a very relevant element of PdxR–DNA interaction is represented by the non-specific contacts established between basic residues distributed at the ridge of the AAT-like domain and the DNA phosphate backbone of the bridging sequence between motifs 1 and 3 (Figure 6C). The electrostatic forces lining at the top of the dimeric core of the protein mediate the DNA binding even in the absence of interactions with the wHTH domains, as observed in the open structure. Several residues contribute to the overall positive charge of the AAT-like ridge surface. Among those, we identified: i) Lys126 and Lys129 on α4 located at the interface between the two large subdomains of the AAT-like domain; ii) Lys360, Arg364, Lys367, Arg370, Arg371 at
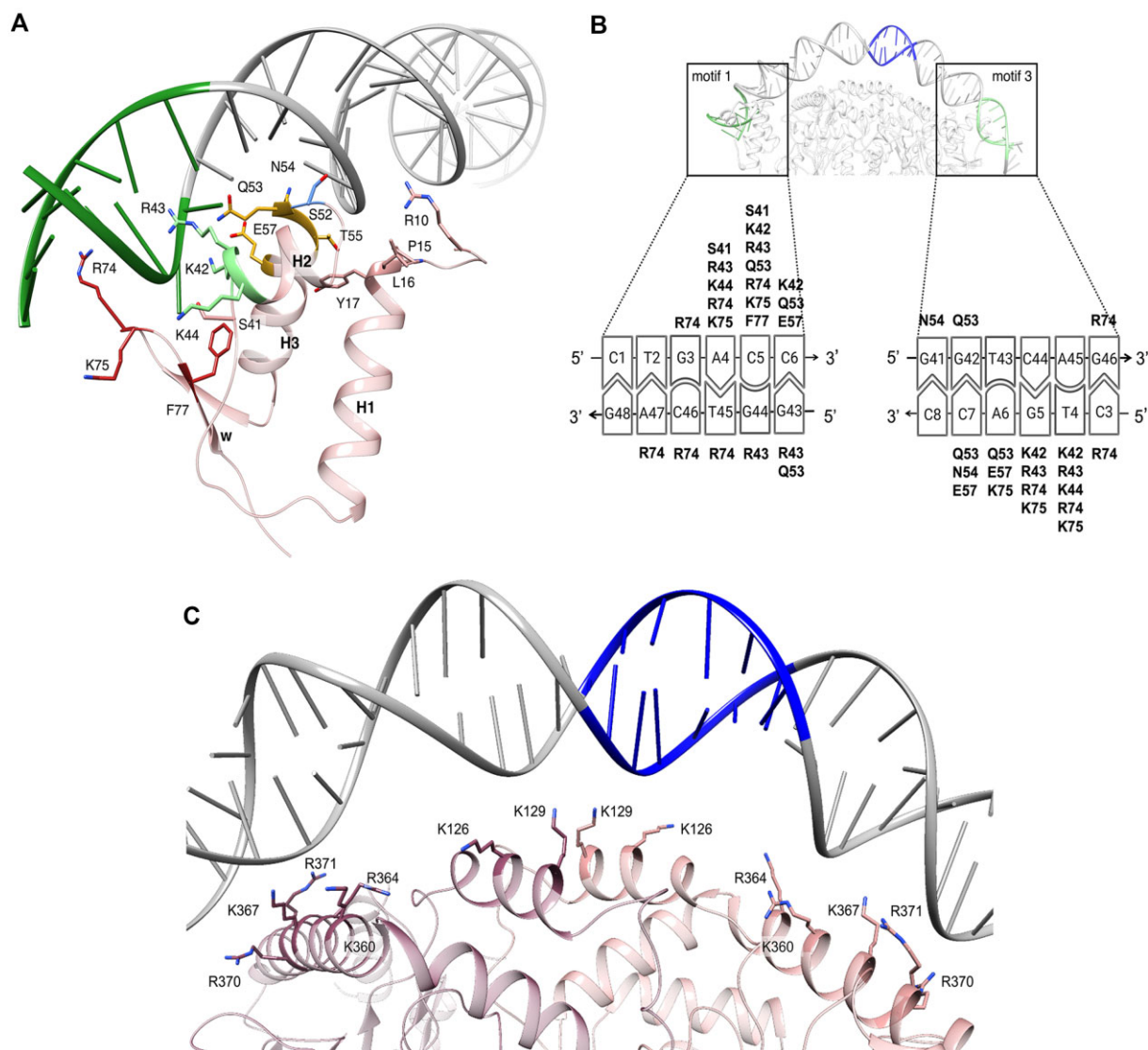
**Figure 6.** Sequence-specific and non-specific interactions guide the holo-PdxR–DNA recognition. (**A**) The wHTH domain forms specific interaction with the target DNA. Representative ribbon depiction of the wHTH domain of the holo-PdxR–DNA in the closed conformation (light pink). Residues interacting with the DNA are represented as sticks, labelled, and differently colored according to the secondary structure element they belong to (α(H)2, green; α(H)3, orange; w = wing, red). DNA is shown as gray ribbon. (**B**) Schematic summary of the holo-PdxR–DNA main contacts formed through the wHTH domain. The different distribution of the identified contacts on motifs 1 and 3 is likely due to coordinate uncertainties associated with the local resolution of the EM maps. (**C**) Non-specific contacts formed with the AAT-like domain. Basic residues of the PdxR AAT-like domain interacting with the DNA backbone are highlighted and represented in dark and light pink sticks, depending on the PdxR monomer. In all panels motifs 1, 2, and 3 are colored in dark green, blue and light green respectively.

the α13 (Figure 6C). These residues are longitudinally distributed over the protein–DNA central interface. A similar positive charge distribution involved in DNA binding was also observed in GabR (48). Of relevance in our structure is the Lys129 couple: the positively charged, long, and flexible side chains of these residues experience several conformational changes along the apo- to holo-PdxR–DNA transition towards their final position in the closed conformation (Supplementary Figure S16). Only in the closed state the side chains of both Lys129 are directed towards the narrow minor groove located at the site of maximum curvature of the DNA, where the A/T-rich tract and motif 2 converge. The pronounced bending of the DNA upon the full PdxR binding produces a deformation at that site that enhances

its local negative electrostatic potential, complementing the positively charged tips of both lysines 129 and conferring some extent of specificity to these electrostatic interactions, as described in (49).

*The dynamics of the holo-PdxR–DNA complex formation.*
Formation of the holo-PdxR–DNA complex requires numerous, multidirectional, and sequential protein–DNA contacts to be established. All the structured protein domains of PdxR, including the AAT-like domain, are involved and cooperate in DNA binding, which is sustained by both longitudinal (non-specific) and lateral (specific) interactions. Hence, DNA recognition requires not merely the canonical DNA binding domains; rather, the whole protein

contributes to the process. The large extent of interconnection and mutual wrapping achieved in the closed state complex require conformational rearrangements that involve both the protein and the DNA and are correlated to their individual structural flexibility. Undoubtedly, in PdxR—as in other MocR TFs—most of the conformational plasticity is provided by the dynamics of the linker region. Nevertheless, the movements of the AAT-like domain upon PLP binding further contribute to the structural transition of PdxR. Given the correlation between DNA sequences and their propensity to bending (50), we investigated the intrinsic bendability of the 48-bp DNA fragment computationally using a DNA curvature prediction tool (www.lfd.uci.edu/~gohlke/dnacurve/) (40). The global structure is calculated starting from the nucleotide sequence. The resulting 3D-model of DNA revealed a natural propensity to bend in one direction of the investigated fragment, with a total angle of curvature of about 165°. The overall bending brings the two cognate sites (motifs 1 and 3) closer to each other (Figure 7A and Supplementary Figure S17A). Unsurprisingly, the region of the sequence where the DNA bends corresponds to an adenine-thymine rich tract (5′-TTTTCTTAAAAA-3′), known as a curvature-inducing feature (51), which is located in the first half of the 48-bp fragment (Supplementary Table S3). Alternative 3D DNA structures were also generated, following different algorithms, but all converge to similar results (Supplementary Figure S17A).

In the light of these observations, we hypothesize that when the holo-PdxR binds the DNA, the electrostatic interactions established with the ridge of the AAT-like domain induce an initial conformational change of the nucleic acid. Following its intrinsic bendability, the DNA fragment asymmetrically bends around the dimeric core of PdxR in the sole direction allowed by its sequence (as seen in the open conformation). Such movement induces an initial increase of the DNA curvature of about 35° with respect to the unbound fragment (Figure 7A, B). This structural transition favors the establishment of the first specific contact with the wHTH domain of PdxR at motif 1, most likely supported by movements of the linker region (half-closed conformation). This interaction, by altering the DNA structure, facilitates the anchoring of the second wHTH domain at motif 3, further compacting and closing the protein–DNA complex (closed conformation). Notably, given the head-to-tail arrangement of the PdxR monomers, the simultaneous binding of the two wHTH domains, symmetrically located at the opposite ends of PdxR, can only occur on opposite DNA strands, as observed in the closed structure (Figure 3B). In this state, the DNA must further bend an additional ~15° to allow the binding of the two wHTHs to the inverted repeats (motifs 1 and 3), thus adopting a symmetric 'S' shape (Figure 7B).

While the trajectory of the structural changes followed by the DNA molecule is clearly defined in the cryo-EM structures (Figure 7), that of PdxR can be only inferred given the absence of density in the regions that mostly rearrange upon DNA binding, i.e. the linker and the wHTH domain. We cannot exclude that the DNA may explore different conformational states in the unbound form, and that the confor-

mation that we observed in our open structure is the result of the interaction with the protein selecting the conformation showing higher affinity.

### The PdxR mode of DNA recognition: DNA binding experiments

The structural data here reported indicate that the shape features of the target DNA, its bending propensity, and specific recognition sequences all contribute to the process of the holo-PdxR–DNA recognition. To investigate the relevance of both DNA sequence and conformation in the mechanism of holo-PdxR–DNA binding, we produced altered forms of the target DNA fragment to test how these variations affect the binding capability of the holo-PdxR. The wild type DNA sequence was modified with the aim to independently perturb: i) the base readout, specifically modifying the sequence of the two inverted repeats recognized by holo-PdxR, and ii) the shape readout, changing or disrupting the intrinsic bendability of the 48-bp DNA fragment.

Three modified DNA fragments containing modified sequences at motifs 1 (M1), 2 (M2), 3 (M3) and simultaneously at motifs 1 and 3 (M13) were produced to impair binding to the direct and inverted repeats (Supplementary Figure S18A). To exclude that such substitutions introduced alterations in the overall DNA shape, the 3D structure of the four modified fragments was calculated (www.lfd.uci.edu/~gohlke/dnacurve/) (40). A minor change in the curvature was observed only for the M2 mutant (Supplementary Figure S17B). Two additional DNA sequences were designed to alter the intrinsic bendability of the original DNA fragment based on the nearest-neighbor curvature model (42). In the first mutant, the nucleotide sequence between motifs 1 and 3 was modified to achieve the maximum possible curvature of the 48-bp segment, conserving the same direction of bending of the wild type. This produced a bent DNA. In the second one, the introduced nucleotide substitutions aimed at totally abolishing the DNA curvature. The latter two DNA fragments were called 'bent' and 'straight', respectively (Supplementary Figure S18B). All DNA fragments (sequences reported in Supplementary Table S3) were labelled at the 5′ terminus with the Cy3 fluorophore. The affinity of the holo-PdxR for the modified DNA fragments was analyzed by measuring the change in the emitted fluorescence of the labelled DNA fragments upon titration with PdxR. These data showed that the introduced mutations at motifs 2 and 3 do not impair holo-PdxR–DNA binding. A 4-times reduced affinity was observed when motif 1 was modified, whereas modification of both the cognate sequences of motifs 1 and 3 causes a 20-fold decrease in affinity. Minor effects are observed when the DNA curvature is more ('bent' fragment) or less ('straight' fragment) pronounced than in the wild type (Table 1, Supplementary Figure S19).

An interesting finding from the structural data is the possible role of the electrostatic forces between protein and DNA in mediating the initial stages of the complex formation. Following the first encounter between the two molecules, these contacts may also favor conformational
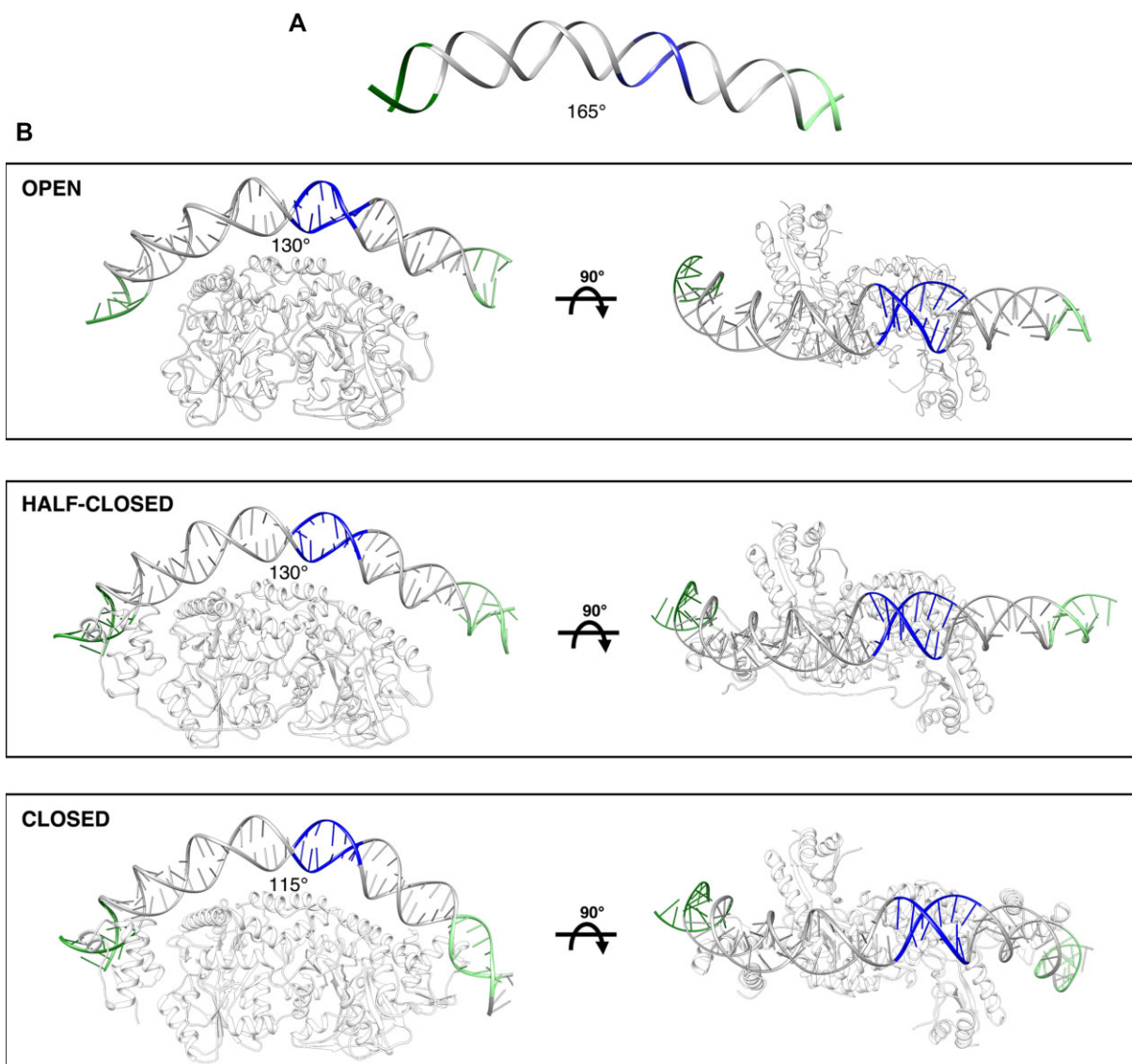
**Figure 7.** DNA structural dynamics upon holo-PdxR binding. (**A**) The predicted curvature of the 48-bp DNA containing PdxR target sequences generated by (C. Gohlke, www.lfd.uci.edu/~gohlke/dnacurve/) is shown. (**B**) Side and top views of the holo-PdxR–DNA complex in the open (top), half-closed (middle), and closed (bottom) conformers. In both panels the DNA molecule is depicted in gray ribbon representation. PdxR cognate binding sites are colored in dark green (motif 1) and light green (motif 3); motif 2 is colored in blue. The protein is represented as white transparent ribbons.

changes of the nucleic acid, promoting the preliminary rearrangements observed in the open to the half-closed transition and the late structural rearrangements leading to the closed conformer. To evaluate the contribution of these interactions to the protein–DNA recognition process, we produced three PdxR variants, namely K126Q/K129Q, K360Q/R364Q, and R370Q/R371Q, that were analyzed for their DNA binding capability (Supplementary Figure S19). These mutations, removing positive charges from PdxR, impair the protein DNA binding compared to the wild type (Table 1). As already seen for the wild type protein, strong effects are observed when residues binding either motif 1 or both motifs 1 and 3 are altered. In addition, a noticeable inability to bind the DNA is observed for K126Q/K129Q with the bent fragment (Table 1).

## DISCUSSION

The structural analyses of PdxR reported in this work, integrated with DNA binding affinity assays, provide clues to decipher the mechanism by which the holo form of the protein (with PLP bound to it) binds to its target DNA.

The crystal structure of apo-PdxR revealed the organization of the AAT-like domain in the absence of PLP, providing the first structural description of the empty effector binding domain of a MocR TF (Figure 2). Comparison with the holo-PdxR–DNA open structure revealed that, once PLP binds to the AAT-like domain, PdxR undergoes structural rearrangements that originate from the PLP-binding site and extend over the whole AAT-like domain (Figure 4). We propose that these structural changes alter the external surface of the large subdomain of the

**Table 1.** DNA binding experiments and mutational analysis. The $K_{Dapp}$ values are the mean $\pm$ SEM of at least three independent fluorescence binding assays. For those fragments whose binding could not be detected, $K_D$ was not determined (nd). The experiments were performed using the holo form of PdxR (PdxR) variants

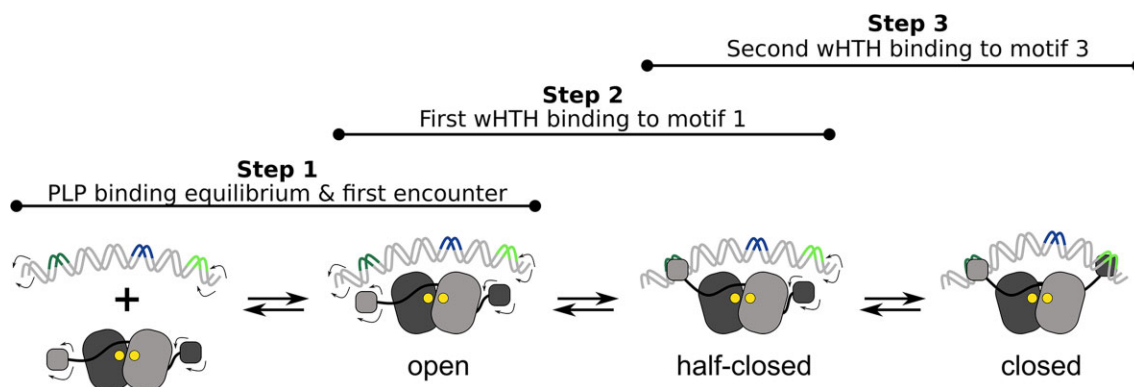| DNA fragments (48 bp) | $K_{Dapp}$ (nM) | | | |
| --- | --- | --- | --- | --- |
| | PdxR wt | K126/K129Q | K360/R364Q | R370Q/R371Q |
| **Wild type** | 16.5 $\pm$ 0.7 | 36.6 $\pm$ 1.7 | 33.8 $\pm$ 2.6 | 18.8 $\pm$ 0.8 |
| **M1** | 64.3 $\pm$ 5.7 | 80.4 $\pm$ 3.3 | nd | 83.0 $\pm$ 5.9 |
| **M2** | 14.8 $\pm$ 1.2 | 41.8 $\pm$ 3.4 | 36.9 $\pm$ 2.3 | 50.9 $\pm$ 3.9 |
| **M3** | 18.7 $\pm$ 2.7 | 42.2 $\pm$ 2.7 | 36.9 $\pm$ 2.6 | 41.2 $\pm$ 2.3 |
| **M13** | 356 $\pm$ 111 | nd | nd | nd |
| **Bent** | 25.8 $\pm$ 7.0 | nd | 38.8 $\pm$ 3.2 | 46.5 $\pm$ 3.3 |
| **Straight** | 22.8 $\pm$ 2.7 | 61.3 $\pm$ 3.0 | 54.8 $\pm$ 5.4 | 47.5 $\pm$ 1.8 |



**Figure 8.** The multi-step mechanism of binding of holo-PdxR to DNA. Scheme of the binding mechanism of holo-PdxR–DNA. The process involves conformational changes of both binding partners, which are allosterically driven by electrostatic interactions. The two subunits of the PdxR dimer are shown in dark and light gray, the linker is represented as a black line. PdxR binding sites are colored in dark green (motif 1) and light green (motif 3); motif 2 is colored in blue.

AAT-like domain most likely affecting the position of the linker and wHTHs. The consequent change of the interface between the wHTH and AAT-like domains may determine the change in affinity for the cognate sites of the target DNA.

The cryo-EM structures of holo-PdxR in complex with its target DNA revealed that the structural changes occurring upon effector binding further extend when the first contact with DNA is established (Figure 3). The open conformer of the holo-PdxR–DNA complex indicates a role of electrostatic forces in the initial encounter with the nucleic acid. Non-specific electrostatic interactions are in many cases the driving forces securing the TF longitudinal motion along the DNA that facilitates the rapid identification of the target sequence. In this process, the natural geometric properties of DNA play a central role (52–54). Indeed, DNA shape features, such as minor groove narrowing, induce a variation of the local electrostatic potential. In this case, the enhanced negative electrostatic environment attracts positively charged amino acids of the binding protein through the phenomenon of the electrostatic focusing (55), thus generating a specific element of recognition (56,57). Similarly, in the case of PdxR, the intrinsic properties of the unbound target DNA sequence, which is bendable in one direction, are responsible for a curvature prone region that forms a narrow minor groove. This favors the establishment of shape readout interactions with the basic residues at the ridge of the AAT-like domains. The presence of an A/T-rich sequence located at one end of the DNA fragment

confers asymmetric properties to the nucleic acid, guiding the binding direction of PdxR, as previously described for histones in nucleosomes (58). Hence, following the natural asymmetric curvature of the DNA, those initial forces may induce an additional bending of the DNA, as observed in the open and in the half-closed structures. This would promote the anchoring of one of the two wHTH domains to the first DNA binding site, namely motif 1, switching the non-specific protein–DNA recognition event to a specific one. Since our data lack some information about the linker and the wHTH domain positions, we can only assume that from the half-closed to the closed conformers of the complex other changes in the protein structure occur, leading the second wHTH domain to interact with motif 3 (closed conformer). Such interaction is coupled to a pronounced curving of the nucleic acid, which further bends over the protein surface, forming a 'S'-shaped symmetric protein–DNA complex (Figure 7). We propose that the binding of one wHTH domain can alter the structure and flexibility of either or both of the DNA and the protein, allosterically facilitating the interaction between motif 3 with the second wHTH domain. The closure of this conformational state is instrumental to the gene regulation activity of the PLP-bound PdxR.

The relative contribution of base recognition and shape readout to the DNA binding process is highlighted well by the results of DNA binding assays obtained with mutated DNA fragments and protein variants (Table 1). The replacement of positively charged residues in PdxR nonspecifically

decreases the affinity of all DNA fragments. On the other hand, DNA mutations at the binding motifs have a specific effect on affinity. More specifically, we observed that M1 mutation reduces the binding affinity while mutations of motifs 2 and 3 do not show any significant effect. This finding, in conjunction with the cryo-EM structure of the PdxR–DNA complex in the half-closed state, supports the hypothesis that motif 1 interacts first with the protein in the binding process. When M1 and M3 mutations are combined, a drastic effect is observed with all protein forms. This result indicates that the simultaneous mutation of both motifs not only perturb the first sequence-specific binding step but, most remarkably, impairs the subsequent events that lead to the formation of the final closed conformation of the complex.

Importantly, the impairing effect on the DNA binding of mutation K126Q/K129Q is strong when it is coupled to the 'bent' DNA. This observation, combined with the cryo-EM structures of protein–DNA complex intermediates, shows that residues K126 and K129, which contact the most deformed region of the DNA, where motif 2 is located (Figure 6), play a major contribution in the DNA-shape recognition.

Together, our data highlight a possible allosteric role of the positively charged residues at the AAT-like domain ridge in modulating the asymmetric and sequential binding of the wHTH domains to the DNA; indeed, the recognition between protein and DNA is often allosterically modulated (59–62). As already mentioned, in PdxR, the direction of the binding process is defined by the asymmetry of the natural DNA fragment curvature. Recently, a similar phenomenon has been described using the model of the electrostatics-mechanical coupling. More precisely, exploring the DNA binding properties of the TF PU.1, this concept has been introduced to quantitatively explain how non-specific protein–DNA interactions may induce electrostatics-driven DNA shape deformation that modulates the overall protein–DNA binding mechanism (54).

Based on our structural and functional data, we conclude that a plausible mechanism describing the binding reaction of holo-PdxR to its target DNA would involve multiple steps implying allosterically driven conformational changes experienced by both the binding partners, as formalized in Figure 8. Step 1. The initial step is governed by shape readout interactions. Electrostatic forces mediate the first contact with PdxR and induce and stabilize the asymmetric DNA curvature. Step 2. The DNA curvature enables motif 1 to come closer to one of the two wHTH domains favoring its docking. The second step is therefore guided by a base readout recognition. We cannot assess yet if this process is only mediated by DNA structural changes or if the protein may also explore different conformational states before adopting the one that allows the correct interaction with motif 1, observed in the structure. Step 3. In the third step, the structural changes induced on either the DNA or the protein by the binding of the first wHTH domain cooperatively favor the binding of the second wHTH domain, which then becomes sequence specific, forming a closed symmetric complex.

In conclusion, with the caveat that our *in vitro* observations are limited to a 48-bp long DNA fragment, these results represent a fundamental step to clarify the molecular mechanism at the basis of the PdxR transcriptional regulation induced by PLP binding and by the formation of a protein–DNA complex. Given the sequence and the predicted structural homology of the different members of the MocR family and the common positive charge distribution over the surface of the ridge of the AAT-like domains (51,63), we believe that the model here proposed for PdxR could be extended to GabR and to other MocR TFs. Our findings may also have a more general relevance in the context of protein–DNA binding mechanism since they highlight a critical role of the non-specific electrostatic interaction in modulating the recognition of DNA sequences. In addition, the identification of a structural plasticity in the PdxR–DNA system reinforces the general view that shows that wrapped protein–DNA complexes are not static but dynamic entities that explore distinct conformational states, in compliance with the dynamism required by their specific biological function. This has also been observed in histone-DNA assembly in nucleosomes (58) and other DNA regulatory systems (62,64). The presence of an A/T-rich sequence on DNA in proximity to the binding site that confers asymmetric flexibility to a specific region of the nucleic acid represents a common feature of all those systems.

## DATA AVAILABILITY

Atomic coordinates and structure factors for the reported apo-PdxR crystal structure have been deposited with the RCSB Protein Data Bank (PDB) under accession number 7PQ9. The cryo-EM maps of the holo-PdxR–DNA complex in the open, half-closed, and closed (C1 and C2 symmetry) conformation and the relative coordinates generated and analysed in the current study have been deposited in the Electron Microscopy Data Bank (EMDB) and in the PDB under accession code EMD-14960 (PDB 7ZTH), EMD-14778 (PDB 7ZLA), EMD-14852 (PDB 7ZPA) and EMD-14801 (PDB 7ZN5), respectively.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

## FUNDING

## REFERENCES

1. Balleza,E., López-Bojorquez,L.N., Martínez-Antonio,A., Resendis-Antonio,O., Lozada-Chávez,I., Balderas-Martínez,Y.I., Encarnación,S. and Collado-Vides,J. (2009) Regulation by transcription factors in bacteria: beyond description. *FEMS Microbiol. Rev.*, **33**, 133–151.
2. Aravind,L., Anantharaman,V., Balaji,S., Babu,M.M. and Iyer,L.M. (2005) The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS Microbiol. Rev.*, **29**, 231–262.
3. Rigali,S., Derouaux,A., Giannotta,F. and Dusart,J. (2002) Subdivision of the helix-turn-helix GntR family of bacterial regulators in the FadR, HutC, MocR, and YtrA subfamilies. *J. Biol. Chem.*, **277**, 12507–12515.
4. Suvorova,I.A., Korostelev,Y.D. and Gelfand,M.S. (2015) GntR Family of bacterial transcription factors and their DNA binding motifs: structure, positioning and Co-evolution. *PLoS One*, **10**, e0132618.
5. Schneider,G., Käck,H. and Lindqvist,Y. (2000) The manifold of vitamin B6 dependent enzymes. *Structure*, **8**, R1–R6.
6. Bramucci,E., Milano,T. and Pascarella,S. (2011) Genomic distribution and heterogeneity of MocR-like transcriptional factors containing a domain belonging to the superfamily of the pyridoxal-5'-phosphate dependent enzymes of fold type I. *Biochem. Biophys. Res. Commun.*, **415**, 88–93.
7. Percudani,R. and Peracchi,A. (2003) A genomic overview of pyridoxal-phosphate-dependent enzymes. *EMBO Rep.*, **4**, 850–854.
8. Tramonti,A., Nardella,C., di Salvo,M.L., Pascarella,S. and Contestabile,R. (2018) The MocR-like transcription factors: pyridoxal 5'-phosphate-dependent regulators of bacterial metabolism. *FEBS J.*, **285**, 3925–3944.
9. Belitsky,B.R. and Sonenshein,A.L. (2002) GabR, a member of a novel protein family, regulates the utilization of γ-aminobutyrate in Bacillus subtilis. *Mol. Microbiol.*, **45**, 569–583.
10. Belitsky,B.R. (2004) Bacillus subtilis GabR, a protein with DNA-binding and aminotransferase domains, is a PLP-dependent transcriptional regulator. *J. Mol. Biol.*, **340**, 655–664.
11. Edayathumangalam,R., Wu,R., Garcia,R., Wang,Y., Wang,W., Kreinbring,C.A., Bach,A., Liao,J., Stone,T.A., Terwilliger,T.C. *et al.* (2013) Crystal structure of Bacillus subtilis GabR, an autorepressor and transcriptional activator of gabT. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 17820–17825.
12. Okuda,K., Ito,T., Goto,M., Takenaka,T., Hemmi,H. and Yoshimura,T. (2015) Domain characterization of Bacillus subtilis GabR, a pyridoxal 50-phosphate-dependent transcriptional regulator. *J. Biochem. (Tokyo)*, **158**, 225–234.
13. Park,S.A., Park,Y.S. and Lee,K.S. (2017) Crystal structure of the C-terminal domain of Bacillus subtilis GabR reveals a closed conformation by γ-aminobutyric acid binding, inducing transcriptional activation. *Biochem. Biophys. Res. Commun.*, **487**, 287–291.
14. Wu,R., Sanishvili,R., Belitsky,B.R., Juncosa,J.I., Le,H.V., Lehrer,H.J.S., Farley,M., Silverman,R.B., Petsko,G.A., Ringe,D. *et al.* (2017) PLP and GABA trigger GabR-mediated transcription regulation in Bacillus subtilis via external aldimine formation. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, 3891–3896.
15. Milano,T., Gulzar,A., Narzi,D., Guidoni,L. and Pascarella,S. (2017) Molecular dynamics simulation unveils the conformational flexibility of the interdomain linker in the bacterial transcriptional regulator GabR from Bacillus subtilis bound to pyridoxal 5'-phosphate. *PLoS One*, **12**, e0189270.
16. Frezzini,M., Guidoni,L. and Pascarella,S. (2019) Conformational transitions induced by γ-amino butyrate binding in GabR, a bacterial transcriptional regulator. *Sci. Rep.*, **9**, 19319.
17. Jochmann,N., Götker,S. and Tauch,A. (2011) Positive transcriptional control of the pyridoxal phosphate biosynthesis genes pdxST by the MocR-type regulator PdxR of Corynebacterium glutamicum ATCC 13032. *Microbiology*, **157**, 77–88.
18. El Qaidi,S., Yang,J., Zhang,J.-R., Metzger,D.W. and Bai,G. (2013) The vitamin B6 biosynthesis pathway in Streptococcus pneumoniae is controlled by pyridoxal 5'-phosphate and the transcription factor PdxR and has an impact on ear infection. *J. Bacteriol.*, **195**, 2187–2196.
19. Belitsky,B.R. (2014) Role of pdxr in the activation of vitamin B6 biosynthesis in Listeria monocytogenes. *Mol. Microbiol.*, **92**, 1113–1128.
20. Tramonti,A., Fiascarelli,A., Milano,T., di Salvo,M.L., Nogués,I., Pascarella,S. and Contestabile,R. (2015) Molecular mechanism of PdxR - A transcriptional activator involved in the regulation of vitamin B6 biosynthesis in the probiotic bacterium Bacillus clausii. *FEBS J.*, **282**, 2966–2984.
21. Fitzpatrick,T.B., Amrhein,N., Kappes,B., Macheroux,P., Tews,I. and Raschle,T. (2007) Two independent routes of de novo vitamin B6 biosynthesis: not that different after all. *Biochem. J.*, **407**, 1–13.
22. Peterson,E.A. and Sober,H.A. (1954) Preparation of crystalline phosphorylated derivatives of vitamin B6. *J. Am. Chem. Soc.*, **76**, 169–175.
23. Winter,G. (2010) Xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.*, **43**, 186–190.
24. Kabsch,W. (2010) *XDS. Acta Crystallogr. D Biol. Crystallogr.*, **66**, 125–132.
25. Evans,P.R. and Murshudov,G.N. (2013) How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.*, **69**, 1204–1214.
26. Vagin,A. and Teplyakov,A. (2010) Molecular replacement with MOLREP. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 22–25.
27. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 486–501.
28. Murshudov,G.N., Skubák,P., Lebedev,A.A., Pannu,N.S., Steiner,R.A., Nicholls,R.A., Winn,M.D., Long,F. and Vagin,A.A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.*, **67**, 355–367.
29. Liebschner,D., Afonine,P.V., Baker,M.L., Bunkoczi,G., Chen,V.B., Croll,T.I., Hintze,B., Hung,L.W., Jain,S., McCoy,A.J. *et al.* (2019) Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. Sect. Struct. Biol.*, **75**, 861–877.
30. Zheng,S.Q., Palovcak,E., Armache,J.-P., Verba,K.A., Cheng,Y. and Agard,D.A. (2017) MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods*, **14**, 331–332.
31. Rohou,A. and Grigorieff,N. (2015) CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.*, **192**, 216–221.
32. Scheres,S.H.W. (2012) RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.*, **180**, 519–530.
33. Tegunov,D. and Cramer,P. (2019) Real-time cryo-electron microscopy data preprocessing with Warp. *Nat. Methods*, **16**, 1146–1152.
34. Pettersen,E.F., Goddard,T.D., Huang,C.C., Couch,G.S., Greenblatt,D.M., Meng,E.C. and Ferrin,T.E. (2004) UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.
35. Williams,C.J., Headd,J.J., Moriarty,N.W., Prisant,M.G., Videau,L.L., Deis,L.N., Verma,V., Keedy,D.A., Hintze,B.J., Chen,V.B. *et al.* (2018) MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.*, **27**, 293–315.
36. Barad,B.A., Echols,N., Wang,R.Y.R., Cheng,Y., Dimaio,F., Adams,P.D. and Fraser,J.S. (2015) EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy. *Nat. Methods*, **12**, 943–946.
37. Heinig,M. and Frishman,D. (2004) STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res.*, **32**, W500–W502.
38. Krissinel,E. and Henrick,K. (2007) Inference of macromolecular assemblies from crystalline State. *J. Mol. Biol.*, **372**, 774–797.

39. Goddard,T.D., Huang,C.C., Meng,E.C., Pettersen,E.F., Couch,G.S., Morris,J.H. and Ferrin,T.E. (2018) UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci.*, **27**, 14–25.

40. Gohlke,C. (2023) *DNA Curvature Analysis service*. https://www.cgohlke.com/dnacurve/.

41. Bolshoy,A., Mcnamarat,P., Harringtont,R.E. and Trifonov,E.N. (1991) Curved DNA without A-A: experimental estimation of all 16 DNA wedge angles (DNA curvature/anomalous migration/Eulerian angles). *Proc. Natl. Acad. Sci. U.S.A.*, **88**, 2312–2316.

42. De Santis,P., Palleschi,A., Savino,M. and Scipioni,A. (1990) Validity of the nearest-neighbor approximation in the evaluation of the electrophoretic manifestations of DNA curvature. *Biochemistry*, **29**, 9269–9273.

43. Frezzini,M., Narzi,D., Sciolari,A.M., Guidoni,L. and Pascarella,S. (2020) Molecular dynamics of an asymmetric form of GabR, a bacterial transcriptional regulator. *Biophys. Chem.*, **262**, 106380.

44. Milano,T., Contestabile,R., Presti,A.L., Ciccozzi,M. and Pascarella,S. (2015) The aspartate aminotransferase-like domain of Firmicutes MocR transcriptional regulators. *Comput. Biol. Chem.*, **58**, 55–61.

45. Catazaro,J., Caprez,A., Guru,A., Swanson,D. and Powers,R. (2014) Functional evolution of PLP-dependent enzymes based on active-site structural similarities. *Proteins Struct. Funct. Bioinform.*, **82**, 2597–2608.

46. Milano,T., Angelaccio,S., Tramonti,A., di Salvo,M.L., Contestabile,R. and Pascarella,S. (2016) Structural properties of the linkers connecting the N- and C- terminal domains in the MocR bacterial transcriptional regulators. *Biochim. Open*, **3**, 8–18.

47. van Aalten,D.M.F., DiRusso,C.C., Knudsen,J. and Wierenga,R.K. (2000) Crystal structure of FadR, a fatty acid-responsive transcription factor with a novel acyl coenzyme A-binding fold. *EMBO J.*, **19**, 5167–5177.

48. Amidani,D., Tramonti,A., Canosa,A.V., Campanini,B., Maggi,S., Milano,T., di Salvo,M.L., Pascarella,S., Contestabile,R., Bettati,S. *et al.* (2017) Study of DNA binding and bending by Bacillus subtilis GabR, a PLP-dependent transcription factor. *Biochim. Biophys. Acta - Gen. Subj.*, **1861**, 3474–3489.

49. Rohs,R., West,S.M., Sosinsky,A., Liu,P., Mann,R.S. and Honig,B. (2009) The role of DNA shape in protein–DNA recognition. *Nature*, **461**, 1248–1253.

50. Munteanu,M. (1998) Rod models of DNA: sequence-dependent anisotropic elastic modelling of local bending phenomena. *Trends Biochem. Sci.*, **23**, 341–347.

51. Hizver,J., Rozenberg,H., Frolow,F., Rabinovich,D. and Shakked,Z. (2001) DNA bending by an adenine-thymine tract and its role in gene regulation. *Proc. Natl. Acad. Sci. U.S.A.*, **17**, 8490–8495.

52. Marcovitz,A. and Levy,Y. (2011) Frustration in protein–DNA binding influences conformational switching and target search kinetics. *Biophys. Comput. Biol.*, **108**, 17957–17962.

53. Barbi,M. and Paillusson,F. (2013) Protein–DNA electrostatics: toward a new paradigm for protein sliding. *Adv. Protein Chem. Struct. Biol.*, **92**, 253–297.

54. Chen,X., Tsai,M.Y. and Wolynes,P.G. (2022) The role of charge density coupled DNA bending in transcription factor sequence binding specificity: a generic mechanism for indirect readout. *J. Am. Chem. Soc.*, **144**, 1835–1845.

55. Honig,B. and Nicholls,A. (1995) Classical electrostatics in biology and chemistry. *Science*, **268**, 1144–1149.

56. Chiu,T.P., Rao,S., Mann,R.S., Honig,B. and Rohs,R. (2017) Genome-wide prediction of minor-groove electrostatic potential enables biophysical modeling of protein–DNA binding. *Nucleic. Acids. Res.*, **45**, 12565–12576.

57. Joshi,R., Passner,J.M., Rohs,R., Jain,R., Sosinsky,A., Crickmore,M.A., Jacob,V., Aggarwal,A.K., Honig,B. and Mann,R.S. (2007) Functional specificity of a hox protein mediated by the recognition of Minor groove structure. *Cell*, **131**, 530–543.

58. Winogradoff,D. and Aksimentiev,A. (2019) Molecular mechanism of spontaneous nucleosome unraveling. *J. Mol. Biol.*, **431**, 323–335.

59. Chaires,J.B. (2008) Allostery: DNA does it, too. *ACS Chem. Biol.*, **3**, 207–209.

60. Moretti,R., Donato,L.J., Brezinski,M.L., Stafford,R.L., Hoff,H., Thorson,J.S., Dervan,P.B. and Ansari,A.Z. (2008) Targeted chemical wedges reveal the role of allosteric DNA modulation in protein - DNA assembly. *ACS Chem. Biol.*, **3**, 220–229.

61. Rosenblum,G., Elad,N., Rozenberg,H., Wiggers,F., Jungwirth,J. and Hofmann,H. (2021) Allostery through DNA drives phenotype switching. *Nat. Commun.*, **12**, 2967.

62. Yoshua,S.B., Watson,G.D., Howard,J.A.L., Velasco-Berrelleza,V., Leake,M.C. and Noy,A. (2021) Integration host factor bends and bridges DNA in a multiplicity of binding modes with varying specificity. *Nucleic Acids Res.*, **49**, 8684–8698.

63. Al-Zyoud,W.A., Hynson,R.M.G., Ganuelas,L.A., Coster,A.C.F., Duff,A.P., Baker,M.A.B., Stewart,A.G., Giannoulatou,E., Ho,J.W.K., Gaus,K. *et al.* (2016) Binding of transcription factor GabR to DNA requires recognition of DNA shape at a location distinct from its cognate binding site. *Nucleic Acids Res.*, **44**, 1411–1420.

64. Etheve,L., Martin,J. and Lavery,R. (2016) Dynamics and recognition within a protein–DNA complex: a molecular dynamics study of the SKN-1/DNA interaction. *Nucleic Acids Res.*, **44**, 1440–1448.