SAPIENZA
UNIVERSITÀ DI ROMA

PhD Course in Biochemistry

Cycle XXXIV (2018-2021)

Cryo-EM structure of PdxR from *Bacillus clausii* in complex with its target DNA

PhD Candidate

Ida Freda

Tutors

Prof. Beatrice Vallone

Dr. Carmelinda Savino

PhD Coordinator

Prof. Stefano Gianni

December 2021

Table of Contents

1. Introduction

1.1 The regulation of transcription in bacteria

In all living organisms, cells express specific subsets of genes to respond and to adapt to changes in environmental conditions. The regulation of gene expression often involves rearrangements in genetic material, DNA-binding protein recruitment or modulation of protein production by post-transcriptional modifications.

Bacteria have evolved sophisticated mechanisms to produce an adequate amount of proteins and metabolites at the right time to maintain homeostasis and to respond to different stimuli and nutrient availability.

In prokaryotes, the regulation of gene expression mainly occurs at the transcriptional level, where transcription factors (TFs) play a crucial role.

The transcription unit of bacterial chromosomes comprises a regulatory region, transcription start and termination sites. The regulatory region contains the promoter and the TF binding sites, where TFs bind to regulate the RNA polymerase (RNA pol) access to target sequences.

TFs are sequence-specific DNA-binding regulatory proteins coupling environmental stimuli and gene transcription, and their expression/activity also need to be modulated. For example, the DNA-binding affinity of TFs can be modified by the binding of effector ligands, whose concentration varies according to nutrients/metabolites availability. The concentration of TFs in

cells can be tuned though other mechanisms, involving covalent modifications, proteolysis or sequestration (Browning & Busby, 2004).

When TFs bind to specific DNA sites, they can activate or repress transcription with several mechanisms. An example of negative regulation is the impairment of transcription through steric hindrance; in fact, repressors can interact directly with the promoter region or with a downstream site, interfering with RNA pol binding or with its progression along DNA (**Figure 1.1A)**. Contrarily, some positive regulators bind to a promoter's upstream site, leaving space for the recruitment of RNA pol. Other activators can induce a conformational change in DNA, facilitating a better alignment of the promoter region for the polymerase binding (Bervoets and Charlier, 2019). A TF can exclusively act as an activator/repressor or switch its activity according to the target promoter.

Several families of bacterial TFs have been identified and classified on the basis of different domains that modulate their regulatory activity. TFs are generally composed of at least two domains, one functioning as signal sensor through the binding with metabolites and other allosteric modulators, and the other one working as responsive domain, which directly interacts with DNA (Balleza et al., 2009).

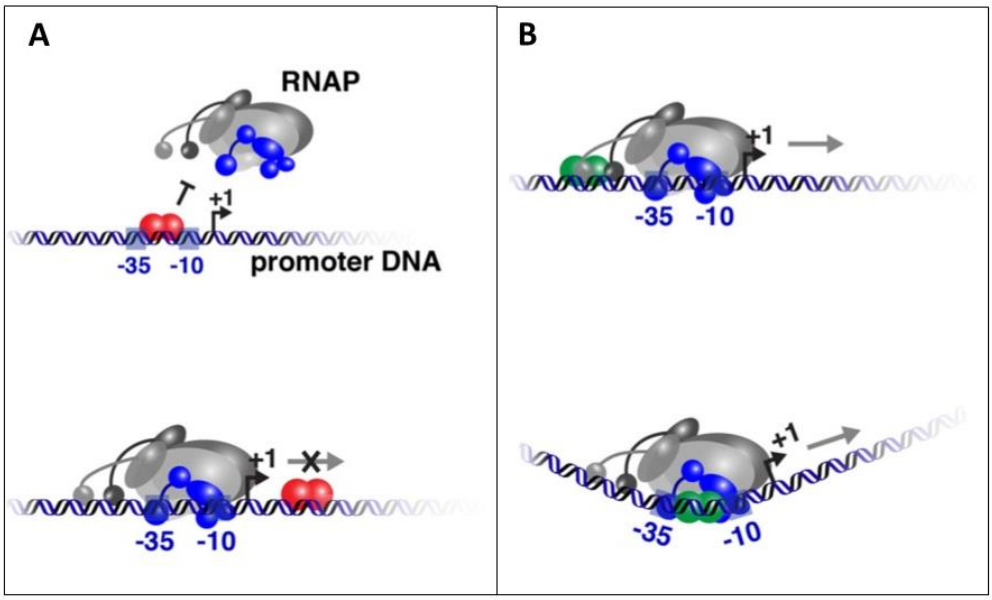


Figure 1.1 Examples of regulation mechanism of transcription factors. A: Negative regulation of repressors (red coloured symbols) by steric hindrance. The TF occupies the RNA pol binding site (upper panel); the TF can also bind a downstream site, interfering with the progression of RNA pol (lower panel). B: Activation of transcription by positive regulators (green symbols). The TF can bind upstream of the promoter, allowing recruitment of RNA pol (upper panel). Another mechanism of activation relies on the remodelling of the promoter region to favour the binding of RNA pol (lower panel; adapted from Bervoets and Charlier, 2019).

Structural, functional, phylogenetic studies, and sequence comparisons allowed the classification of DNA-binding proteins, and specifically TFs, into several groups on the basis of similar structural motifs and of their mechanism of DNA-recognition. The most common and studied binding motifs found in gene regulatory proteins are the helix-turn-helix (HTH), homeodomain, zinc

finger, leucine zipper, helix-loop-helix and β-sheet DNA-binding proteins (Pabo and Sauer, 1992).

1.2 The Helix-turn-helix DNA-binding proteins and the GntR family of transcriptional regulators

The HTH structure was the first DNA-binding motif discovered and investigated in detail (Henikoff et al., 1988; Brennan and Matthews, 1992; Gallegos et al., 1993; Nguyen and Saier, 1995). In HTH proteins, the DNA-recognition unit is extremely conserved, showing strong similarities among the different members of this family. The basic structural elements that compose the HTH domain are three helices and one turn located between the 2nd and the 3rd helix (Aravind et al., 2005, **Figure 1.2A**).

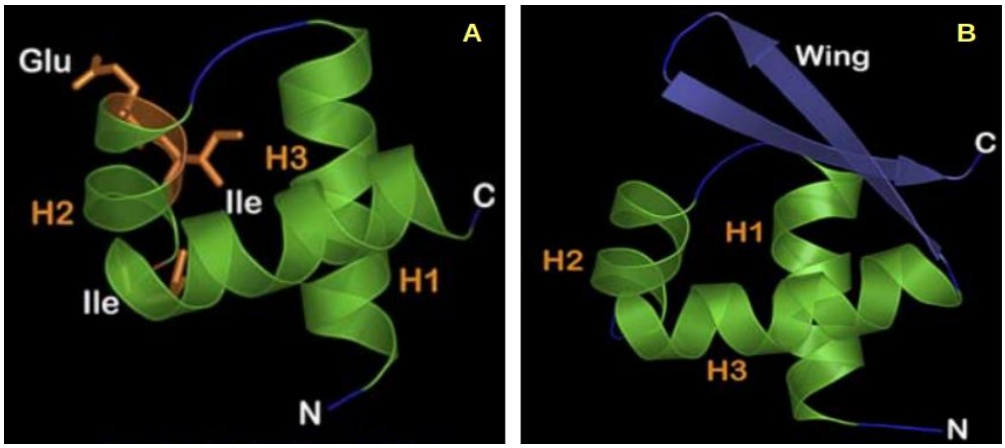


Figure 1.2 Structural features of the HTH domain. A: Structure of a canonical tri-helical HTH domain (pdb code: 1K78). B: Structure of a winged HTH domain (pdb code: 1SMT) Helices are shown in green ribbons, labelled with orange "H", turns and strand are in blue. Conserved residues are represented in orange sticks. (Adapted from Aravind et al., 2005).

Among the HTH proteins, the GntR family represents a large group of TFs distributed among prokaryotes and regulating diverse biological processes. Their name originates by the identification of the negative regulator of the gluconate operon GntR in *Bacillus subtilis* (Fujita and Fujita,1987; Haydon and Guest 1991).

The TFs belonging to the GntR family present a very similar molecular architecture, consisting of a N-terminal HTH DNA-binding domain linked to a C-terminal signalling domain through a linker of variable length (**Figure 1.3**)

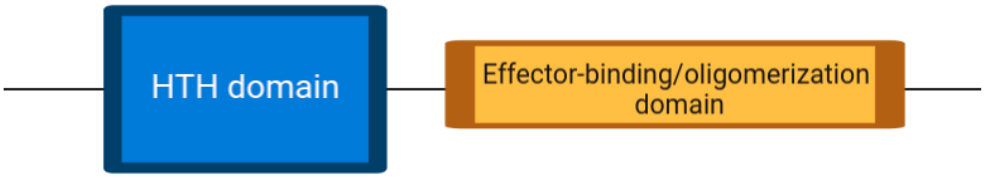


Figure 1.3 Schematic representation of a GntR transcription factor, showing the N-terminal HTH DNA-binding domain and the C-terminal effector-binding and/or oligomerization domain (Created with BioRender.com).

The N-terminal domain structure is highly conserved among the different members of the family. The HTH motif is constituted by helices H_2 and H_3 and a tight turn, where helix H_3 is the "recognition helix" that invades the DNA major groove. Of notice, the DNA-binding motif of GntR-like proteins is classified as winged HTH (wHTH) for the presence of a β-strand hairpin unit (the "wing", **Figure 1.2B**), which provides an additional DNA-binding interface, by interacting with target bases of the minor groove (Rigali et al., 2002; Aravind et al., 2005).

The C-terminal domain is involved in oligomerization and/or in effector binding, and it imposes steric constraints to the HTH domain, influencing the DNA-binding properties through conformational changes occurring upon ligand binding (Rigali et al., 2004). It presents several differences among GntRs conferring heterogeneous functions and specificity for effector ligands to each member of the group (Rigali et al., 2002).

Based on sequence comparisons and structural prediction studies performed on their C-terminal domain, the GntR TFs are further divided into subfamilies named FadR, HutC, MocR, YtrA, PlmA and AraR.

The FadR subfamily is the most represented, covering the 40% of the GntR regulators. The name originates from the regulator of fatty acid metabolism FadR, whose crystal structure was solved both in the absence and in the presence of DNA (van Aalten et al., 2000; Xu et al., 2001). The α-helical C-terminal domain consists of 160-170 amino acids. Members of FadR subfamily bind molecular effectors, such as carboxylic acids, and regulate various metabolic pathways involving substrates such as aspartate (AnsR), pyruvate (PdhR), glycolate (GlcC), galactonate (DgoR), lactate (LldR), malonate (MatR), or gluconate (GntR; Suvorova et al., 2015).

The HutC proteins, covering about 30% of GntR regulators, form the second most represented subfamily. The C-terminal domain consists of α-helices and β-sheets and it is capable of binding effectors such as histidine (HutC), fatty acids (FarR), sugars (TreR) and alkylphosphonates (PhnF; Aravind & Anantharaman, 2003).

The third subfamily, called MocR, presents the longest C-terminal domain among GntRs, which is formed by 350 amino acids and is homologous to the fold type I pyridoxal 5'-phosphate (PLP)-dependent enzymes.

The YtrA group, constituting the 6% of GntRs, contains the shortest α-helical C-terminal domain of the family, which is composed of only 50 amino acids (Rigali et al., 2002).

The PlmA subfamily represents a small group and comprises TFs only found in cyanobacterial.

Similarly, the AraR is a minor subfamily. Its C-terminal domain is homologous to the one present in the GalR/LacI family. AraR proteins modulate the expression of genes encoding transporters for the uptake and utilization of L-arabinose (Suvorova et al., 2015).

The composition in secondary structure and the comparison of the C-terminal domains of the different GntR groups are showed in **Figure 1.4**.

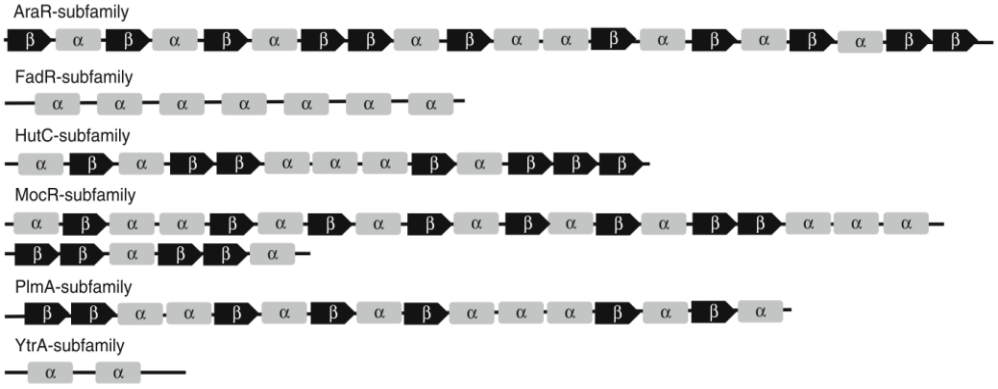


Figure 1.4 Comparison between the C-terminal domains of the different groups of GntR family. The composition in α-helices and β-sheets is shown in a schematic representation (Adapted from Hoskisson and Rigali, 2009).

The object of this research work is a TF belonging to MocR subfamily, Therefore, in the following sections we will give more detailed information on this group of GntR regulators.

1.3 MocR transcription factors

The subfamily of TFs MocR is constituted by proteins which are structurally composed by a small N-terminal DNA-binding domain connected by a flexible linker to a large C-terminal region. As mentioned above, the latter domain resembles the fold type I PLP-dependent aminotransferases, enzymes that catalyse the transfer of an amino group from an amino acid to an acceptor α-ketoacid, using PLP as a cofactor (Schneider et al., 2000).

The name MocR derives from the GntR regulator identified within the *moc* locus in *Rhizobium meliloti,* responsible of the degradation of rhizopine (Rossbach et al., 1994). Many other MocR proteins were then recognized, and some were also experimentally characterized, such as GabR from *Bacillus subtilis* (Belitsky, 2004), which regulates the utilization of γ-amminobutirric acid (GABA), that will be more extensively discussed later.

Studies focused on the genomic distribution of MocRs highlighted that they are widespread among *Eubacteria*, but their presence in bacterial phyla is not homogeneous, since some species do not possess any MocR TF, while others are characterized by the presence of multiple MocR TFs (Bramucci et al., 2011).

MocRs regulate several metabolic pathways of amino acids or vitamin B_6, as well as of compounds not connected with the classical substrates of PLP-dependent enzymes, like ectoine or rhizopine (Tramonti et al., 2018). Experimental evidence demonstrated that PLP plays a crucial role in the

control of gene expression carried out by MocRs, as described in sections **1.5.4** and **1.7.3**.

In several cases, the operons controlled by MocR TFs have been characterized. For example, TauR activates the expression of taurine utilization genes in *Rhodobacter capsulatus* (Wiethaus et al., 2008); GabR activates the expression of *gabT* and *gabD* encoding GABA aminotransferase and succinic semialdehyde dehydrogenase, respectively, forming a pathway for glutamate generation through GABA (Belitsky and Sonenshein, 2002); PdxR is involved in the regulation of the PLP biosynthesis in several bacteria and it acts on *pdxST* genes (Jochmann et al., 2011; El Qiadi et al., 2013; Belitsky, 2014, Liao et al., 2015, Tramonti et al., 2015).

The mechanism of recognition between DNA-binding domains of TFs and their target sites on nucleic acid is usually conserved in each DNA-binding protein class (Morozov and Siggia, 2007). However, even if the sequence conservation between the HTH domains within the same family is high, there can be a certain variability in consensus DNA sequences among different subfamilies (Badis et al., 2009), as in the case of GntRs.

For MocR subfamily, no conserved palindromic sequences were experimentally identified or computationally predicted (Rigali et al., 2002). For example, direct repeats in the target DNA sequences recognized by GabR (Belitsky, 2004), TauR (Wiethaus et al., J, 2008) or PdxR (Jochmann et al., N 2011) do not present any evident similarity.

MocR regulators represent a very interesting object of study, given their involvement in PLP and amino compounds metabolism and the presence of the protein scaffold of PLP-dependent enzymes. In fact, although the MocR members share the evolutionary route followed by GntR TFs (Rigali et al., 2002), the high structural similarity of the C-terminal domain with PLP-

dependent enzymes opens questions on the possibility that it might confer a catalytic activity to MocRs, an issue which is still under debate.

1.4 PLP-dependent enzymes and the structural similarity with the C-terminal domain of MocRs

PLP virtually represents the most versatile cofactor, being utilized by a large number of enzymes diffused in all living organisms, for reactions that mainly involve amino acids (Jansonius et al., 1998; Schneider et al., 2000; Christen and Mehta, 2001). Examples of reactions catalysed by PLP-dependent enzymes include the transfer of amino groups, decarboxylation, conversion of L- and D-amino acids, α-, β- or γ-eliminations.

In the resting state, PLP is bound to a lysine residue through a Schiff base linkage, forming an internal aldimine. Lysine is exchanged with a substrate upon its binding to the active site, and an external aldimine is formed between the substrate and PLP, triggering catalysis (Eliot and Kirsch, 2004).

The first half of transamination reaction consists in the transfer of an amino group from an amino acid to an α-ketoacid or an aldehyde, in which PLP acts as a transitory acceptor of the amino group from the donor substrate and is converted to pyridoxamine 5'-phosphate (PMP; **Figure 1.5**). The amino group is then transferred to the acceptor substrate and PLP is concomitantly regenerated.

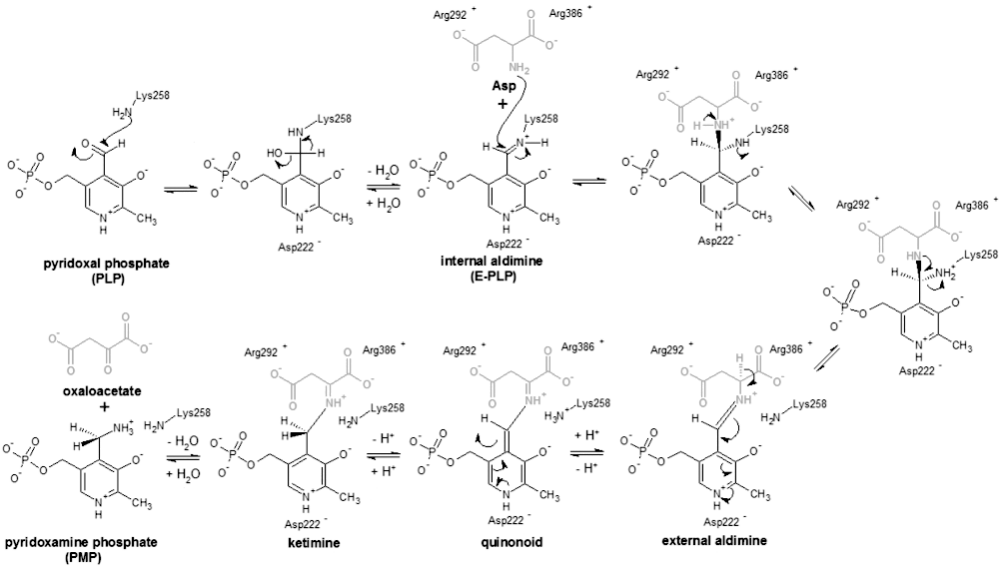


Figure 1.5 First half reaction catalysed by aspartate aminotransferase from *E. coli*. An internal aldimine is generated through a Schiff base linkage between the aldehyde carbon of PLP and Lys258. The amino group of aspartate displaces the ε-amino group of Lys258, forming an external aldimine, through a nucleophilic attack. The abstraction of the hydrogen attached to the C_α of the aspartate forms a quinonoid intermediate, which is then reprotonated at the aldehyde carbon, generating a ketimine intermediate. Finally, ketimine is hydrolyzed to PMP and oxaloacetate (Adapted from Kirsch et al., 1984).

PLP-dependent enzymes have been classified in five different fold types, according to their primary, secondary, and tertiary structures (Grishin et al., 1995)

Fold type I transferases are represented by aspartate aminotransferase (AAT) and this group contains the best structurally characterized vitamin B_6-dependent enzymes. The members of this family are generally dimeric, but they can potentially arrange in larger quaternary structures. Each monomer is itself composed by a large domain containing a seven-stranded β-sheet and a

C-terminal smaller domain of a three- or four-stranded β-sheet covered by helices on one side **(Figure 1.6A**).

In AAT, the lysine involved in the interaction with PLP is located in the large subdomain of each monomer on a short helix following a β-strand. The phosphate group contacts the N-terminal part of an adjacent helix while the PLP aromatic ring faces the neighbouring β-strands. The active site lies in a pocket at the interface of the two monomers, and it is formed by residues from both subunits that interact with the cofactor (Schneider et al., 2000). During its catalytic activity, AAT undergoes an open-to-closed transition upon interaction with substrate, involving both the domains. The conformational change buries the substrate in the active site, placing it close to PLP in the correct orientation and position for catalysis (McPhalen et al., 1992). As previously mentioned, the members of MocR subfamily show an architecture that closely resembles the fold type I PLP-dependent enzymes such as the one found in AAT. Given the structural homology between AAT and the C-terminal domain of MocRs, the structural rearrangement detected for AAT upon substrate binding is expected also to occur in these TFs upon DNA binding, which is likely at the basis of their mechanism of transcription regulation.

1.5 Structural characterization and DNA-binding properties of MocRs: the case of GabR from *Bacillus subtilis*

The best characterized member of the MocR subfamily of TFs is GabR from *Bacillus subtilis*. GabR is a γ-aminobutyric acid (GABA)- and PLP-dependent activator of *gabTD* operon encoding the enzymes involved in the degradation of extracellular GABA, which represents an important nitrogen source for the

bacterium. GabR also behaves as a GABA-independent negative autoregulator, acting on *gabR* gene (Belitsky and Sonenshein, 2002; Belitsky, 2004).

To date, GabR is the only member of the MocR subfamily whose structure has been solved, representing a model for the other MocRs, that likely share the same architecture.

Experimental data obtained in solution and *in crystallo* showed that GabR is characterized by a dimeric assembly, with a head-to-tail domain-swap arrangement. Each monomer is composed of an N-terminal winged HTH domain (wHTH), a C-terminal AAT-like domain, and a long linker (29 amino acids) connecting the two domains (**Figure 1.6**). The PLP binding cleft is located in each monomer at the large interface between the two AAT-like domains, as expected given the structural similarity with aminotransferases. A second interface is formed between the wHTH domain of one monomer and the AAT-like domain of the other monomer, and the relatively low free energy involved in the formation of this interface suggests some extent of conformational dynamics of the quaternary structure in solution and the possibility of large-scale movements of the DNA-binding domains (Edayathumangalam et al., 2013).

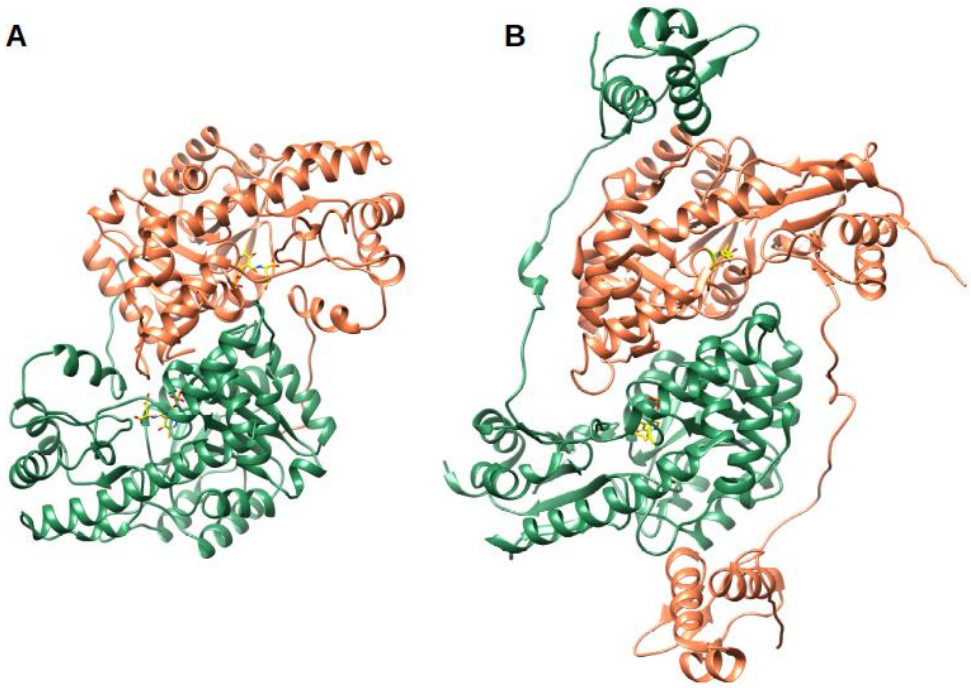


Figure 1.6 Comparison of the structures of aspartate aminotransferase and of GabR in their PLP-bound conformations. A: Crystal structure of dimeric *E. coli* AAT. (pdb code: 1ARG) B: Crystal structure of dimeric *B. subtilis* GabR (pdb code: 4N0B). The central core of GabR dimer shows a high structural homology with AAT. In each monomer, the effector-binding the AAT-like domain is connected to the wHTH domain through a linker of 29 residues. In both panels, the two subunits are represented in orange and green cartoons. PLP molecules are depicted in yellow sticks (Figures made using Chimera; Pettersen et al., 2004).

1.5.1 The wHTH DNA-binding domain

For the majority of the structurally characterized GntR regulators, the structure in complex with their target DNA is missing. Two GntRs whose DNA-bound structures are available are FadR and AraR (Xu et al., 2001; Jain and Nair, 2013).

Therefore, all the information currently available for GntRs about residues in the wHTH domain that are involved in the interaction with DNA only arises from the study of these two systems.

Among the residues identified in FadR as crucial for recognition and binding of DNA, Arg35 (helix H_2), Thr44 (turn), Thr46, Thr47 (helix H_3) and Lys67 (wing) have emerged since these are conserved among the GntR family members. In GabR they correspond to Arg43, Ser52, Asn54 and Ser55 on the HTH and Lys75 on the wing, and they constitute the major basic surface of the wHTH.

The superimposition of the structure of the GabR wHTH domain of GabR with the structure of FadR in complex with its 19-bp target DNA (**Figure 1.7**), revealed that, similarly to FadR, the wHTH of GabR could invade the major groove of DNA, while the wing might contact the minor groove. The model allowed to predict also the positioning of the GabR above-mentioned residues with respect to DNA, and their mode of interaction (Edayathumangalam et al., 2013).

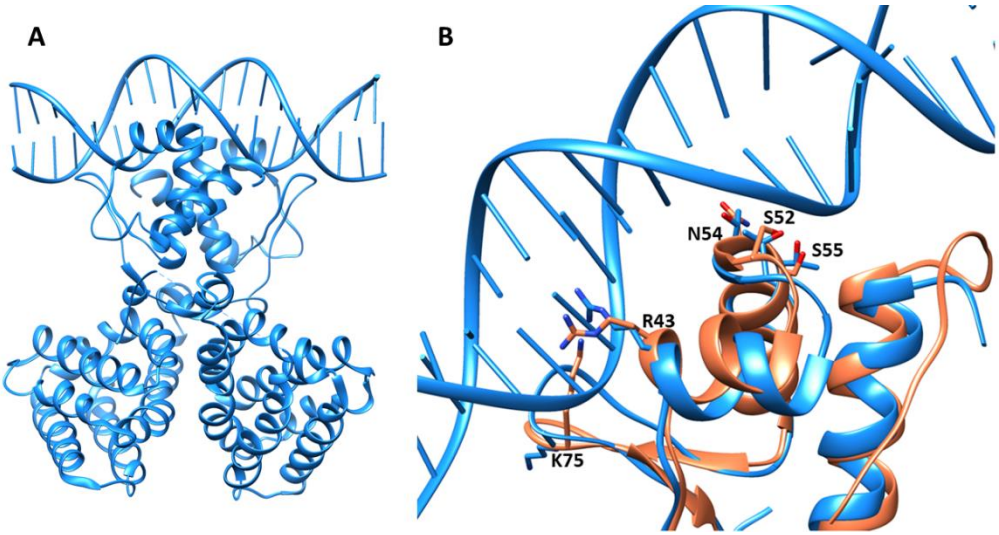


Figure 1.7 Putative DNA-binding residues in GabR on the basis of the homology with the FadR wHTH domain. A: Overall crystal structure of *E. coli* FadR in complex with 19-bp DNA (pdb code: 1H9T), represented in blue cartoons. B: GabR wHTH domain superimposed with the wHTH domain of FadR-DNA complex structure. Structures of FadR and GabR (pdb code: 4N0B) are represented in blue and orange cartoons, respectively. FadR DNA binding residues are highlighted and depicted in blue sticks; GabR putative binding residues are in orange sticks and labelled (Figures made using Chimera; Pettersen et al., 2004).

1.5.2 The C-terminal: an AAT-like domain

The crystal structure of *B. subtilis* GabR revealed that, similarly to fold type I aminotransferases (Eliot and Kirsch, 2004), the C-terminal domain is divided into two subdomains, the larger one containing the PLP-binding site. A conserved lysine residue forming a Schiff base with PLP, typical of aminotransferases, in GabR monomer corresponds to Lys312.

Other residues involved in PLP binding in GabR are conserved among aminotransferases, such as Arg319 and Ser311 (Arg266 and Ser257 in *E. coli*

AAT) interacting with the PLP phosphate group and Asp279 (Asp222 in *E. coli* AAT), that maintains PLP in the protonated state through its interaction with the pyridinium nitrogen (Edayathumangalam et al., 2013). These observations emphasize the similarity between the regulatory domain of GabR, and more in general of MocRs, and AATs.

1.5.3 The polypeptide linker and its putative functional role

The N- and C-terminal domains are connected by a flexible polypeptide linker that in MocRs can be of variable length, ranging from 10 to 150 residues with an average size of 32 residues (Milano et al., 2016). The length and the composition in amino acids might reflect the functional properties of the linker (Mattison et al., 2002; Gokhale and Khosla, 2000) and influence the specific mechanism of DNA-recognition of the different members of the subfamily.
GabR crystal structure showed that the 29-residue long linker does not establish significant interactions with any of the two domains, suggesting this structural element to be endowed with high flexibility in solution (Edayathumangalam et al., 2013).

1.5.4 Conformational rearrangements in GabR structure

GabR is a transcriptional activator of *gabTD* genes, and it acts in the presence of both GABA and PLP (Belitsky, 2004). It has been proposed that GabR undergoes a conformational change upon GABA binding that can be necessary for transcription activation. This structural rearrangement has been described as an open-to-close transition that occurs at the level of the AAT-like domain, similar to what observed in fold type I aminotransferases (McPhalen et al.,

1992; Edayathumangalam et al., 2013; Okuda et al., 2015a; Okuda et al., 2015b).

The structure of GabR C-terminal AAT-like domains, obtained crystallizing the protein in presence of PLP and GABA, captured the protein in an asymmetric dimer, one in a closed form, with GABA forming an external aldimine with PLP, and one in an open form, with only PLP bound as an internal aldimine. The superposition of the two subunits revealed that binding of GABA favours the closure of the AAT-domain (**Figure 1.8**).

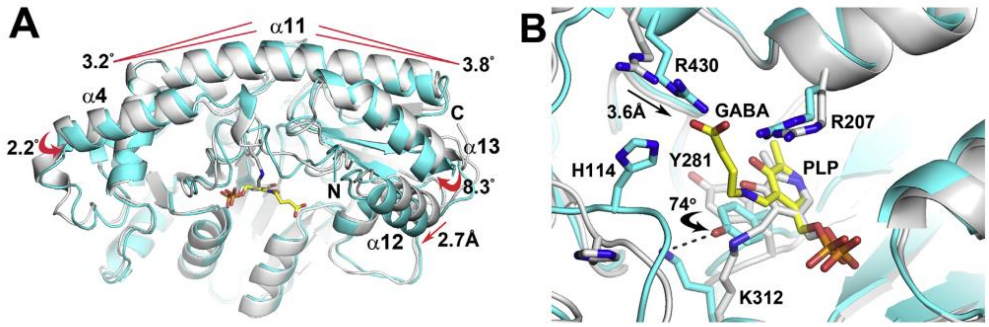


Figure 1.8 Structural superimposition of the two monomers in *B. subtilis* GabR AAT-like dimer (pdb code: 5X03). A: Cartoon representation of GabR AAT-like dimer bound to the PLP-GABA adduct (closed form, light blue) and of the open GABA-free form (light grey). B: View of the binding pocket of superimposed PLP-GABA-bound form (light blue) and of GABA-free form (light grey). Residues involved in PLP or GABA binding are represented in sticks and labelled. PLP and PLP-GABA adduct are in light grey and yellow, respectively (Adapted from Park et al., 2017).

The comparison between the structures of the full-length GabR dimer and of the AAT-like domains suggested that the GABA-induced open-to-close transition could weaken the interactions between AAT-like and wHTH

domains, through the loss of some hydrogen bonds (Park et al., 2017, **Figure 1.9**).

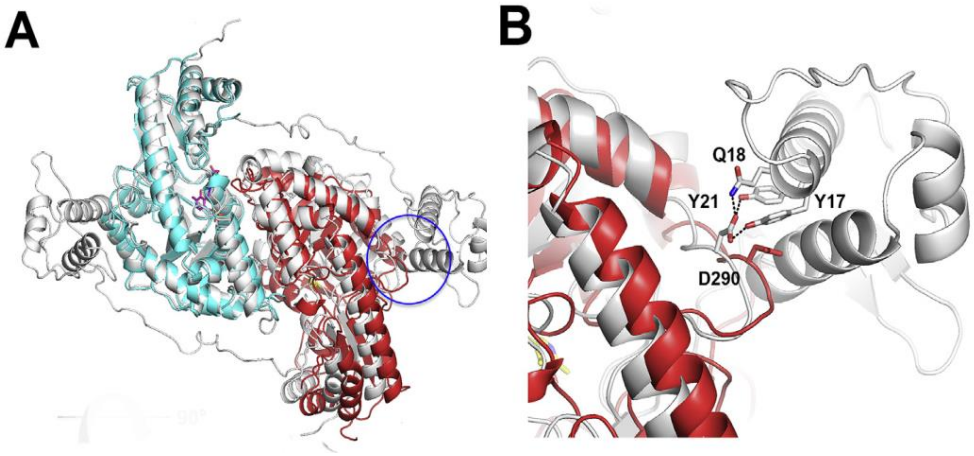


Figure 1.9 GABA-induced conformational transition in *B. subtilis* GabR. A: superimposition of the crystal structure of full-length GabR dimer (light grey; pdb code: 4MGR) and of the AAT-like dimer (pdb code: 5X03). The GABA-bound and GABA-free monomers are in blue and red, respectively. B: Close-up view of the interface between N-terminal and C-terminal domain. Residues involved in the interaction between the two domains are shown as sticks and labelled. Hydrogen bonds are indicated by black dotted lines (Adapted from Park et al., 2017)

Moreover, molecular dynamics simulations of PLP-bound GabR showed that the wHTH domain and the linker represent the most flexible portions of the structures. Therefore, comparisons of molecular dynamics calculations for the apo- and PLP-bound forms of GabR suggested that, in the presence of PLP, the interaction between the wHTH domain and the AAT-like domain is destabilized, and the linker likely facilitates a conformational transition through which the wHTH domain detaches from the protein core. The flexibility conferred to the wHTH domain by the unstructured linker might be

at the basis of the mechanism of DNA binding regulation since it influences the interaction with the DNA, thus tuning the transcriptional activity of GabR (Milano et al., 2017).

1.5.5 The interaction of GabR with DNA and the proposed mechanism of transcriptional regulation

GabR binds to the target DNA as a dimer, regardless of the presence or absence of PLP and GABA (Al-Zyoud et al., 2016, Amidani et al., 2017). The *gabR-gabTD* intergenic region is 49 bp long and contains two direct repeats (D1 and D2), separated by 34 bp (Belitsky, 2004). Another conserved sequence has an inverted orientation with respect to the first direct repeat (I3; **Figure 1.10**). It is controversial whether GabR recognizes the two direct repeats of the *gabR-gabTD* intergenic region, or one direct and the inverted repeat, since the interaction with two direct repeats is not compatible with the symmetrical and inverted orientation of the wHTH domains in the structure of GabR (Edayathumangalam et al., 2013).

Figure 1.10 Organization of the *B. subtilis gabR-gabTD* intergenic region. The sequences highlighted in red correspond to the repeats proposed as GabR binding sites. Red arrows indicate the orientation of the two direct repeats (D1 and D2) and the inverted repeat (I3). The box comprises the -35 sequence of g*abTD* promoter (Adapted from Nardella et al., 2020).

It has been proposed that in absence of GABA, GabR binds the inverted repeat I3 impairing the access of RNA pol. Upon GABA binding, a conformational change may facilitate the interaction of GabR with the two direct repeats D1 and D2, the recruitment of RNA pol and the transcription of the g*abTD* operon (Nardella et al., 2020).

Of note, the DNA sequence containing the sites recognized by GabR is characterized by an intrinsic propensity to bend, presenting adenine/thymine-rich tracts. Small angle X-ray scattering (SAXS), atomic force microscopy (AMF) and electrophoretic mobility shift assays (EMSA) data demonstrated that the proper shape, phasing and bendability of DNA, as well as the base readout, are essential for GabR to bind the cognate sites with high affinity. In fact, mutations introduced to alter DNA phase or bendability were found to negatively affect the stability of GabR-DNA complexes. In addition to the natural bendability, the binding of the two wHTH domains induced a more pronounced curvature to DNA (~ 80°) (Al-Zyoud etal., 2016; Amidani et al., 2017).

GabR is the most studies MocR transcriptional regulator, and many details have been elucidated that could be at the basis of its interaction with target DNA sites and of its mechanism of transcriptional regulation.

Nevertheless, many aspects of the molecular mechanism of the MocRs are not yet clarified, due to the lack of structural information on these TFs in complex with DNA.

Another MocR TF widely characterized at the functional level is PdxR, a regulator involved in vitamin B_6 metabolism predicted to be structurally homologous to *B. subtilis* GabR that could underlie similarities in their mechanism of action. Therefore, structural studies on this regulator might give

insight into the modulation of vitamin B_6 synthesis and also provide a model to understand the mechanism of the other members of the MocR subfamily.

1.6 Vitamin B_6/PLP metabolism and its regulation

PLP, the effector molecule of MocR TFs, is the active form of vitamin B_6. B_6 vitamers form a group of six compounds containing a pyridine ring, differing in a variable group at the 4' position of the pyridine ring, which can be an aldehyde (pyridoxal, PL), an alcohol (pyridoxine, PN) or an amine (pyridoxamine, PM).

Only plants and microorganisms are able to synthetize vitamin B_6 *de novo*, while all other organisms have to assume it through their diet, although it is involved in a multitude of biological functions.

The *de novo* biosynthesis of PLP follows two independent and mutually exclusive pathways (**Figure 1.11**), the deoxyxylulose 5-phosphate (DXP)-dependent pathway and the DXP-independent one.

Largely studied in *Escherichia coli* and in other bacteria, the DXP-dependent pathway involves six different enzymes. In the last steps, PNP synthase forms PNP through the condensation of 3-hydroxy-1-amino acetone phosphate and deoxy-D-xylulose-5-phosphate and PNP oxidase converts PNP in PLP (Fitzpatrick et al., 2007).

The DXP-independent pathway, discovered in fungi, plants and some bacterial strains, relies on a 3-carbon sugar, a pentose phosphate and glutamine for PLP synthesis, that is operated by PLP synthase complex, formed by PdxS and PdxT subunits, also named Pdx1 and Pdx2 (Strohmeier et al., 2006, Raschle et al., 2007, Fitzpatrick et al., 2007).

Furthermore, a third biosynthetic route is present in most organisms, that differs from the two *de novo* pathway described above, known as "salvage pathway". It recycles B_6 vitamers from nutrients and protein turnover and interconvert them using pyridoxal kinase (PLK) and PNP oxidase (PNPOx; Di Salvo et al., 2011, Parra et al., 2018).

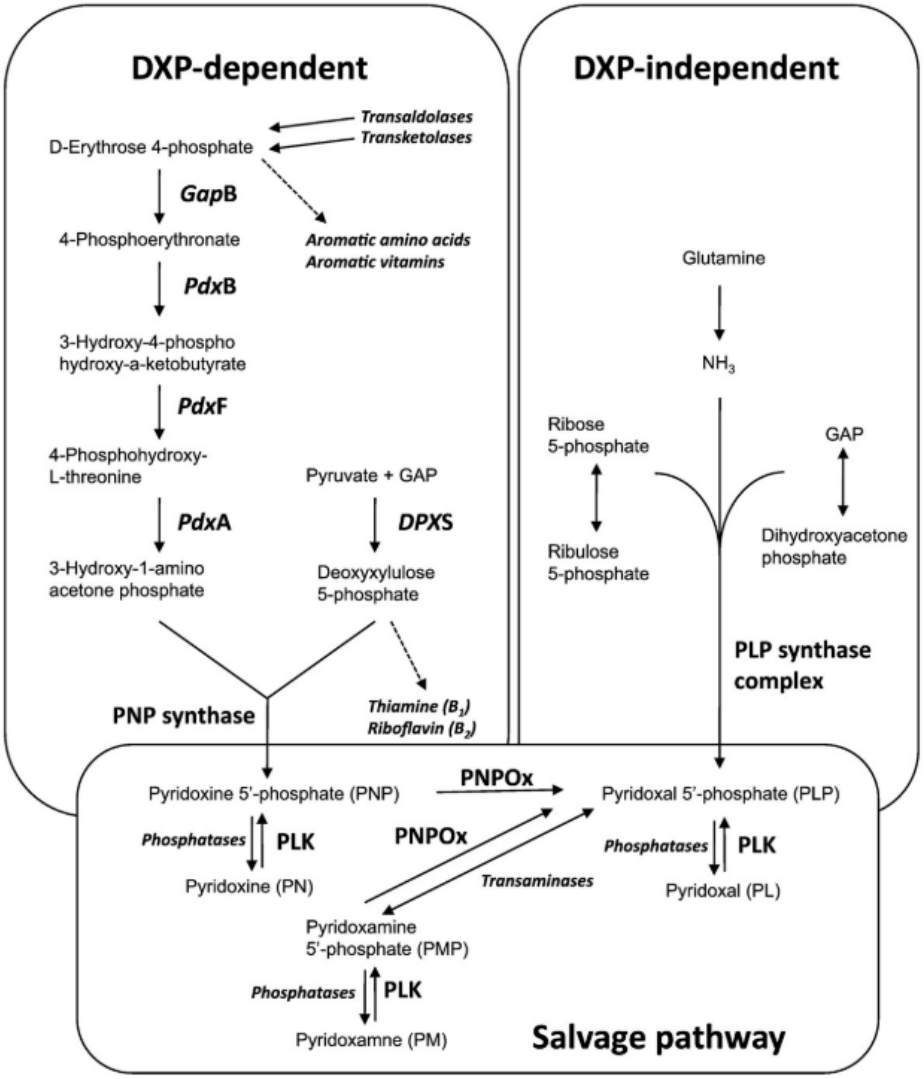


Figure 1.11 Schematic representation of the two PLP biosynthetic pathways and of the salvage pathway (Di Salvo et al., 2011).

PLP is required as an enzyme cofactor in a multitude of biochemical reactions, so its availability is essential for all living organisms. Nevertheless, its concentration in cells must be finely tuned, given the high reactivity of the aldehyde moiety that could generate harmful toxic effects.

In several prokaryotes in which the PLP biosynthesis relies on DXP-independent route, the expression of PLP synthase subunits is found to be regulated by PdxR, a transcriptional regulator belonging to MocR TFs (Jochmann et al., 2011; El Qaidi et al., 2013; Belitsky, 2014; Tramonti et al., 2015).

1.7 The transcriptional regulator PdxR

PdxR is a member of MocR subfamily of transcriptional regulators involved in the control of vitamin B_6 metabolism, since it regulates the expression of PLP synthase complex, that catalyses *de novo* biosynthesis of PLP in many bacteria.

The role of PdxR in expression regulation has been investigated in several microorganisms, such as *Corynebacterium glutamicum* (Jochmann et al., 2011), *Streptococcus pneumoniae* (El Qaidi et al., 2013), *Listeria monocytogenes* (Belitsky, 2014) and *Bacillus clausii* (Tramonti et al., 2015). It has been elucidated that, in these organisms, PdxR activates the expression of the *pdxST* operon, encoding for PdxS and PdxT subunits of PLP synthase complex, responsible of the synthesis of PLP. In presence of PLP, PdxR acts as a repressor of its own *pdxR* gene. PLP is then an anti-activator, and it seems to reduce PdxR regulatory activity (**Figure 1.12**).

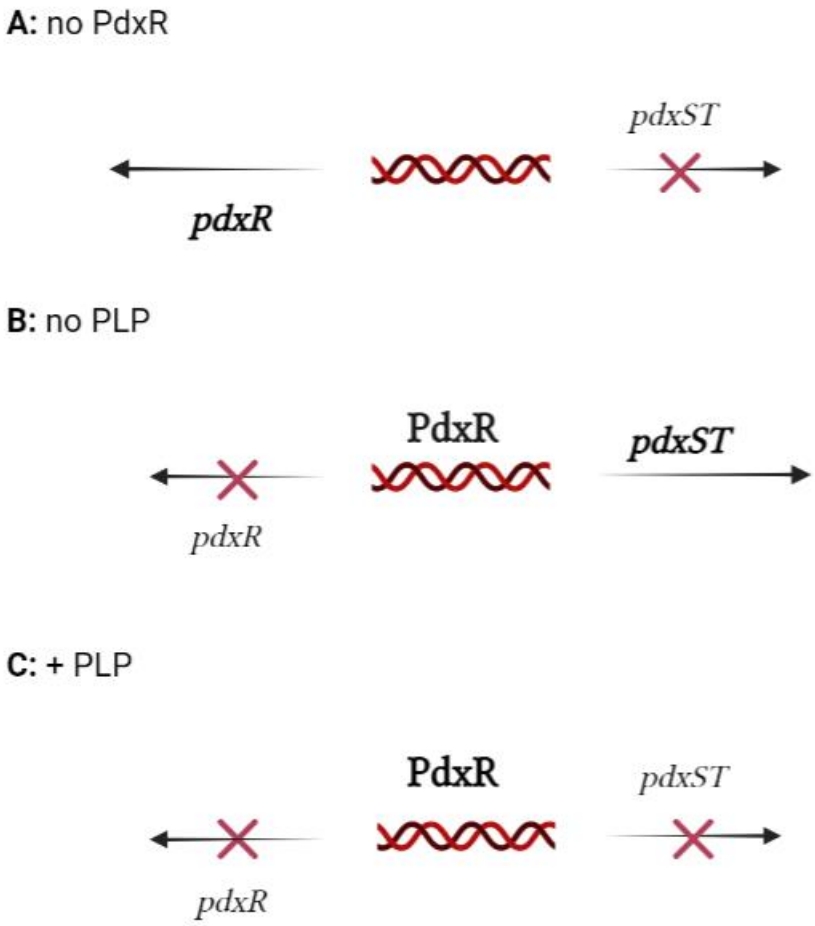


Figure 1.12 Schematic representation of the regulatory activity of PdxR. In absence of PdxR, the transcription of *pdxR* gene is activated. When PdxR is not bound to PLP, it activates the transcription of *pdxST* genes, encoding for the subunits of PLP synthase. In its PLP-bound form, PdxR acts as an autorepressor and does not activate the transcription of PLP synthase genes.

In contrast with *B. subtilis* GabR, that necessitates the binding of both PLP and GABA to exert its function, only PLP is necessary for PdxR activity. Moreover,

it has been demonstrated that PLP is the only B_6 vitamer that is able to bind PdxR (Belitsky, 2014, Tramonti et al., 2015).

1.7.1 PdxR target DNA: the *pdxR-pdxST* intergenic region

The organization of the *pdxR-pdxST* operon in *Bacillus clausii* was studied in Tramonti et al., 2015. Computational analyses allowed to predict the position of the putative -35 and -10 regions of both *pdxST* and *pdxR*. These predictions revealed the presence of a CTGACC direct repeat and of its palindromic inverted repeat (named motifs 1 and 3) interspaced by 34 bp, and of a further CTGACA direct repeat (motif 2) at 18 bp from motif 1.
Other bacteria, such as *L. monocytogenes* (Belitsky, 2014) and *C. glutamicum* (Jochmann et al., 2011), show a similar organization of the promoter region in two direct and one inverted repeat (**Figure 1.13**).

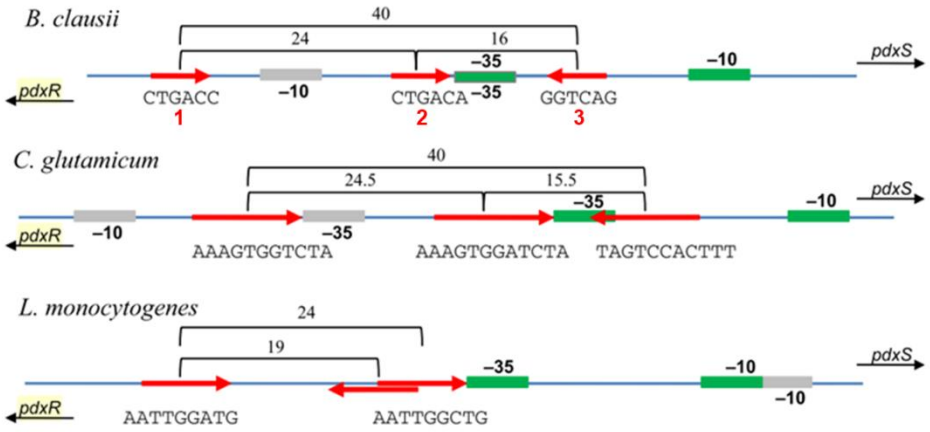


Figure 1.13 Organization of the *pdxR-pdxST* intergenic regions in *B. clausii*, *C. glutamicum* and *L. monocytogenes*. DNA repeats and their orientation are represented by red arrows. Regions at -35 and -10 are highlighted in green for *pdxST* and in grey for *pdxR* (Adapted from Tramonti et al., 2015).

1.7.2 Structural modelling of PdxR from *Bacillus clausii*

Most of the structural and biochemical information currently available on PdxR comes from studies performed on the isoform expressed in *Bacillus clausii* (Tramonti et al., 2015).

The primary structure of PdxR is conserved among different MocRs belonging to *Firmicutes* phylum (Bramucci et al., 2011), and it shares a 36% of sequence identity with *B. subtilis* GabR, the best characterized MocR so far. The sequence conservation allowed to construct a 3D homology model of PdxR, using the crystal structure of GabR as a template (**Figure 1.14**).

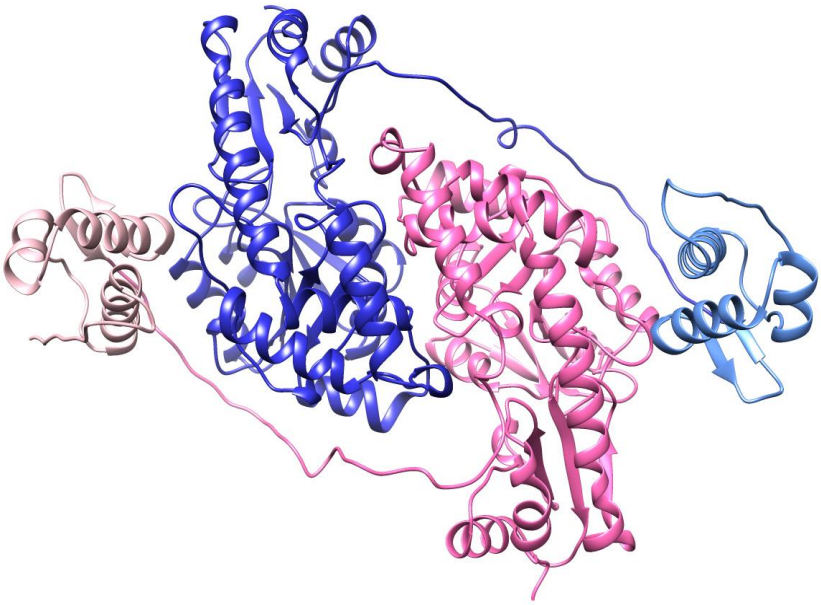


Figure 1.14 Three-dimensional model of bcPdxR. Cartoon representation of bcPdxR homology model, predicted using the crystal structure of bsGabR as a template (pdb code: 4N0B). The HTH domains are shown in light pink and blue; the AAT-like domains are shown in dark pink and blue (Figure made using Chimera; Pettersen et al., 2004).

PdxR was predicted to be formed by an wHTH N-terminal domain and by an AAT-like C-terminal domain, containing the PLP binding site. Alignment with GabR highlighted the presence of three conserved residues on the wHTH domain, namely Arg43, Ser52 and Lys75, that in GabR are predicted to play a crucial role in DNA binding (Edayathumangalam et al., 2013).

The putative PLP-interacting residues in *B. clausii* PdxR are Asp276 and Lys309. Absorption spectra of PLP-bound PdxR (holo-PdxR) present two bands at 335 nm and 425 nm, that may be attributed to the enolimine and the ketoenamine form, respectively, of the Schiff base (Tramonti et al., 2015). A

mutation of Lys309 generated a protein unable to bind PLP, confirming its role in cofactor binding. Moreover, a mutant lacking the wHTH domain and the long linker showed a very low affinity for PLP, indicating that these two regions are essential for the interaction of PLP with the AAT-like domain, even if they are not directly involved in its binding.

It has also been shown that PdxR is prevalently dimeric in solution in both the holo- and the apo- form, and that it assumes different conformations. Given the overall similarity with GabR, it was hypothesized that upon binding to PLP, PdxR undergoes a conformational transition driven by movements of the long linker connecting the wHTH and the AAT domains, that results in a detachment of the wHTH binding domain from the AAT core (Tramonti et al., 2015; Edayathumangalam et al., 2013), as better described in the following paragraph.

1.7.3 PdxR interaction with DNA

The entire *pdxR-pdxST* regulon covers a region of ~120 bp containing the three 6-bp motifs recognized by PdxR (two direct and one inverted repeat).

EMSA experiments revealed that the apo- form of PdxR binds with comparable affinity to all the combinations of motifs, reaching the highest affinity with a segment covering all the repeats (motifs 1, 2 and 3). When PLP is bound to PdxR, the interaction with the first direct repeat and the inverted repeat (motifs 1 and 3) becomes predominant, altering the specificity of PdxR-DNA binding. A structural prediction of PdxR bound to DNA, obtained by the superimposition of the 3D model of PdxR with FadR-DNA crystal structure, provided further evidence that a GabR-like domain swap structure can bind only inverted repeats (**Figure 1.15**). In fact, the predicted distance between the

two wHTHs (~96 Å) is highly compatible with the 34-bp distance separating the first direct repeat and the inverted one (~100 Å).

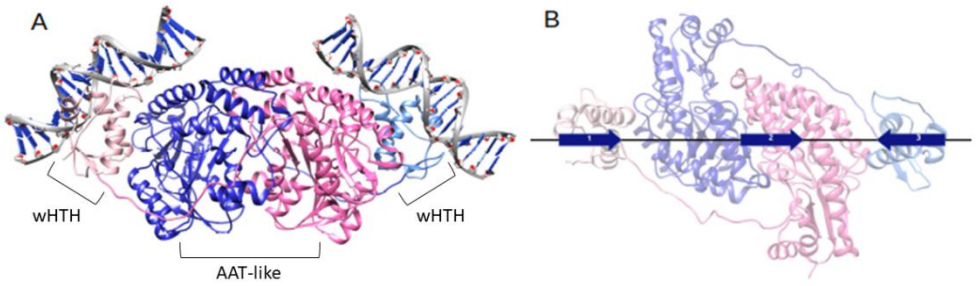


Figure 1.15 Scheme of the putative interaction between bcPdxR and its cognate DNA. A: Ribbon representation of the homology model of bcPdxR-DNA, predicted using bsGabR and FadR-DNA structures as templates. The wHTH domains are shown in light pink and light blue; the AAT-like domains are shown in dark pink and blue; DNA chains are in grey. B: Binding of bcPdxR to the direct and inverted repeats (motifs 1 and 3). Blue arrows indicate the three hypothetical binding sites in *pdxR-pdxST* intergenic region (Figures made using Chimera; Pettersen et al., 2004).

The change in the recognition of DNA binding motifs experienced by PdxR seems to be mediated by a large structural rearrangement of the wHTH domains that occurs through the flexible linker. As suggested for GabR (Park et al., 2017, Milano et al., 2017), this conformational change may involve the detachment of the wHTH DNA-binding domain from the central protein core. The electrostatic potential of PdxR model shows that a longitudinal region of the AAT-like dimer is covered by positively charged residues. Therefore, the interaction of the wHTH domains with DNA is also supported by basic residues distributed on the surface of the central protein core (**Figure 1.16**).

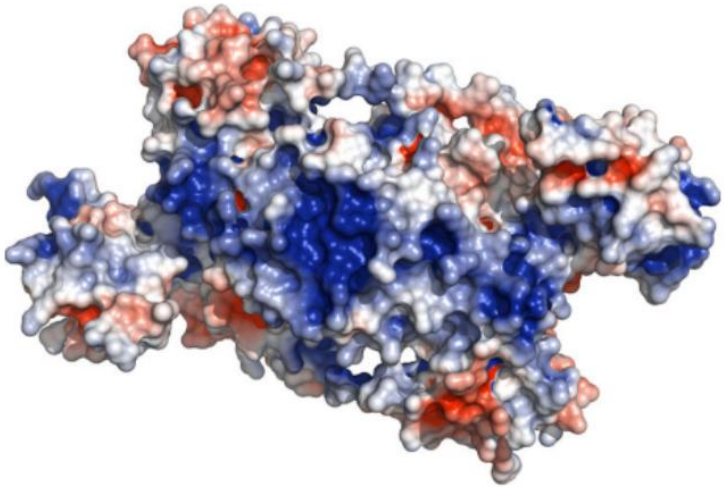


Figure 1.16 Qualitative electrostatic potential of bcPdxR structure. Surface representation of the homology model of bcPdxR. Red and blue regions correspond to negatively and positively charged areas, respectively (Adapted from Tramonti et al., 2015).

A molecular mechanism for the regulation of PdxR activity has been proposed. According to this model, when PLP is not bound, the apo-form of PdxR can adopt different conformations, driven by the unstructured and mobile linker that may undergo large movements, also involving the wHTH domain.

Apo-PdxR can assume both open and closed conformations. In the open form, it is capable of binding all combination of repeats (1+2, 2+3, 1+3) of the *pdxR-pdxST* intergenic region, while in the closed one it binds motifs 1+3. The open conformation of the apo-form would allow the recruitment of RNA pol on the -35 and -10 regions of *pdxST* promoter, ultimately promoting the transcription of PLP synthase genes, while the interaction of the pol with -35 and 10 regions of *pdxR* gene would be hindered.

PLP binding to the AAT-like domains shifts the equilibrium towards the closed conformation, that induces the wHTHs anchoring to the direct and the inverted repeat (1+3). This conformation inhibits the transcription of *pdxR* gene and at the same time does not activate the expression of *pdxST* operon (**Figure 1.17**; Tramonti et al., 2015).

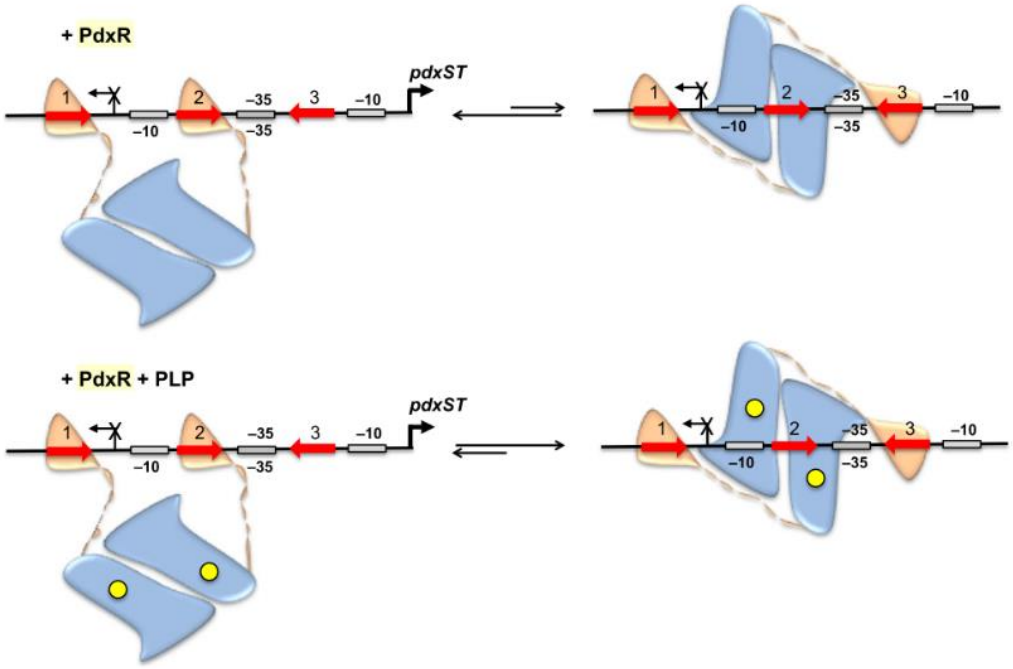


Figure 1.17 Proposed molecular mechanism of bcPdxR transcription regulation. Apo-PdxR is in equilibrium between an open and a closed conformation. The wHTH domains interact with the direct repeats 1 and 2, or with the direct repeat (1) and the inverted repeat (3). In presence of PLP, the closed conformation is favoured, and PdxR binds one direct and one inverted repeat (Adapted from Tramonti et al., 2015). The putative mechanism and the shift between the two conformations are described in the main text.

Overall, the studies conducted on PdxR and on the homologous GabR helped to clarify many aspects of their regulatory activity, such as the target regulons and the specific DNA binding sites, as well as the interaction with cofactors and conformational transitions influencing the DNA-binding mode.

Functional studies and computer modelling allowed to hypothesize the molecular mechanism of different MocR regulators. Nevertheless, structural and further functional studies are necessary to provide additional information on PdxR, necessary to define the molecular processes at the basis of its activity.

2. Aim of the work

MocR is a subfamily of transcriptional regulators widespread in bacteria and implicated in a plethora of biological processes. The recognition of their specific DNA target sequences is regulated by binding to pyridoxal-5'-phosphate (PLP) at the C-terminal domain. Indeed, this regulatory unit contains a binding site for the effector that strongly resembles the one found in PLP-dependent enzymes, such as aspartate aminotransferases (AAT). Changing the affinity of the transcription factor (TF) for DNA, the binding to PLP represents the molecular event at the basis of the mechanism of gene expression regulation followed by MocRs.

To date, the most characterized MocR is GabR from *Bacillus subtilis*, a transcription factor involved in the regulation of GABA metabolism. Functional analyses aiming at investigating its mode of interaction with DNA have clarified several aspects of its functions. GabR is the only MocR that has been structurally characterized in the presence of its effectors, GABA and PLP. However, the structure of GabR in complex with its target DNA is still missing. Given the high sequence similarity, one would expect that the members of MocR subfamily share a common structure and a similar mechanism of transcription regulation. Nevertheless, the current information available does not allow conclusions to be drawn since most of MocRs have not been fully

experimentally characterized yet and only theoretical models of their mechanism of action have been postulated on the basis of functional analyses and computer simulations.

At present, a crucial missing piece of information is the structure of MocRs bound to their target DNA. In fact, the lack of structural information on individual members of MocRs both alone and in the complex with their target DNA hampers the possibility to define common or uncommon tracts of the molecular mechanism of gene expression regulation followed by these TFs.

This PhD thesis is focused on the structural and functional characterization of the MocR transcriptional regulator PdxR, an activator of the genes encoding PLP synthase and an autorepressor. For this study, we used PdxR from *Bacillus clausii*, a probiotic organism widely used for pharmacological preparations, which has been previously characterized at a functional level.

The study of PdxR structure have been for long time pursued by mean of X-ray crystallography. Nevertheless, the crystallization of PdxR in the apo- and holo- forms, bound or unbound to the target oligonucleotide, was unsuccessful. The only diffracting crystals were obtained for a truncated form of PdxR comprising the C-terminal AAT-like domain dimer, in the absence of PLP (data not published). However, this 2.8 Å resolution structure confirms the structural similarity of the C-terminal domain of PdxR with the one of GabR and the AAT enzymes, but it does not provide information on the wHTH domains, on the overall conformation of the protein, and on the binding mode adopted by PdxR when it is in complex with the target DNA.

In the following chapters of this thesis, I describe the experimental procedures and the results obtained in the structural and functional characterization of PdxR bound to its target DNA by means of single particle cryo-electron

microscopy (cryo-EM) combined to computational analyses and electrophoresis mobility shift assays (EMSA).

In the last five years, cryo-EM has achieved a "resolution revolution" thanks to the technological advancements that allow to observe biological samples at a resolution comparable with the one achievable by crystallography. The technique is appropriate for the study of large macromolecules, complexes and heterogeneous systems that are difficult to pack in a crystalline form. Moreover, since in a cryo-EM experiment the sample is measured in conditions closer to the native state, it potentially allows to obtain information on more than one conformer of the system from the analysis of a single dataset. For all these reasons, it represents the technique of choice for the structural study of PdxR-DNA complex and to explore its conformational dynamics.

In this study the full-length PdxR in its PLP-bound form was used in complex with a 48-bp oligonucleotide containing the target DNA repeats of the *pdxST* operon. We took advantage of a 200 kV microscope (FEI Talos Arctica) equipped with a Falcon III direct electron detector and of a service provided by CINECA to perform the single particle analysis on a GPU-equipped computer cluster.

The cryo-EM structure of holo-PdxR bound to DNA has provided information about the DNA-binding mode adopted by this TF, pinpointing the amino acid residues and nucleotides directly involved in the interaction, and allowed to detect different conformations of the complex.

Moreover, since the recognition between TFs and target DNA relies on both DNA sequence and shape readout, a second purpose of this work was to predict the intrinsic curvature of PdxR target DNA and to understand how this affects PdxR binding, probing the role of both the sequence and the DNA shape readout. To this purpose, modified forms of the PdxR target DNA that alter its

natural bending and the sequences of the cognate binding sites were produced, and their binding capability to PdxR was analysed by EMSA.

The structural study, integrated with these methodologies, helps to unveil the rules governing the interaction with the DNA and so the regulation of gene expression of PdxR, providing a more general model for other similar TFs belonging to MocR subfamily.

3. Materials and Methods

3.1 Cloning, expression, purification

The genomic *Bacillus clausii* DNA was purified in accordance with the procedure indicated for *Bacillus subtilis* (http://2012.igem.org/Team:LMU-Munich/Lab_Notebook/Protocols). 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 amplified by PCR using the following primers: BCPdxR_for: 5'–ggcCATGGAGCTGCTATGGTGCG–3'; BcPdxR_rev: 5'–ggctcgaggctgccgcgcggcaccagAAACCCCCATGCATTCAACAC–3'. The amplified ~1400-bp was inserted into a pET28(+) expression vector using NcoI and XhoI restriction enzymes.

The construct was transformed in *E. coli* BL21 (DE3) competent cells. A single colony was picked, suspended in 20 mL LB containing kanamycin at a concentration of 40 mg/L, and the pre-culture was incubated overnight at 37 °C and 180 rpm.

The overnight pre-culture was used to inoculate 2 L LB + kanamycin (40 mg/L) and bacterial cells were allowed to grow approximately for 5 hours at 37°C, 180 rpm until OD_{600} reached ~0.5, then 0.2 mM IPTG was added to induce PdxR overexpression and temperature was lowered to 28°C.

Bacterial suspension was centrifuged after 18 hours and the cell-containing pellet was suspended in a lysis buffer containing 20 mM Tris·HCl, pH 8.0, 0.5 M NaCl, 100 μM pyridoxal 5'-phosphate (PLP) and one tablet of cOmplete™ Protease Inhibitor Cocktail (Roche).

Cells were lysed by sonication on ice for 5 minutes, 70 % amplitude, in short 20 seconds pulses with 20 seconds intervals.

Lysate was centrifuged 12000 *g* for 30 minutes, pellet was discarded, and supernatant was recovered for subsequent purification steps.

Supernatant was precipitated using ammonium sulphate at 60% saturation, centrifuged 30 minutes at 12000 *g* and resuspended in 100 mL of 20 mM Tris·HCl, pH 8.0.

The resuspended pellet was dialysed overnight against 1 L of 20 mM Tris·HCl, pH 8.0, 0.5 M NaCl at 4°C.

The sample was centrifuged 30 minutes at 12000 *g* to remove precipitated protein, filtered using 0.4 μm filters, and loaded onto a nickel-nitrilotriacetic acid (Sigma-Aldrich) column pre-equilibrated with 20 mM Tris·HCl, pH 8.0, 0.5 M NaCl using ÄKTA prime FPLC system (GE Healthcare) at 1 mL/min.

The column was washed with 50 mL of the same buffer, 50 mL of the same buffer containing 20 mM imidazole and 30 mL of the same buffer containing 100 mM imidazole. Protein was eluted with 40 mL of 20 mM Tris·HCl, pH 8.0, 0.5 M NaCl containing 250 mM imidazole. Fractions of 2 mL were collected at every step and analysed by SDS-PAGE to assess the presence of the molecular weight band at about 55 kDa corresponding to PdxR.

Fractions containing PdxR were pooled, PLP (100 μM, final concentration) was added, and the sample was dialysed 72 hours against 1 L of 50 mM Hepes, pH 7.5, 0.5 M NaCl.

Spectroscopic measurements allowed to calculate protein concentration using a theoretical extinction coefficient at 280 nm of 66700 $M^{-1}cm^{-1}$, calculated with EXPASY PROTPARAM online tool (http://web.expasy.org/protparam). PLP content was calculated adding 0.2 M NaOH and measuring the absorbance at 388 nm (Peterson and Sober, 1954).
The final yield from 2 L of bacterial culture was ~ 20 mg of pure holo-PdxR (from now on, called simply PdxR).

3.2 PdxR-DNA complex formation

Purified PdxR was incubated with a 48-bp synthetic DNA fragment containing the two direct and the inverted repeats (in bold) of *pdxR-pdxST* intergenic region:
(5'**CTGACC**TCATCATTTTCTTAAAAA**CTGACA**CTTACAATGT**GGT CAG**TT-3') in 1:2 ratio, to generate the protein-DNA complex.
The final sample was stored at 4°C for 48 hours before cryo-electron microscopy (cryo-EM) experiments.

3.3 Sample preparation for cryo-electron microscopy experiments

The most conventional and fast procedure to assess sample quality, homogeneity prior the cryo-EM experiment is to visualize it by negative staining EM (NS-EM) at room temperature. This experiment is very quick, easy to perform and inexpensive. For a NS experiment, few microliters of the sample are applied on a grid and coated with a stain containing heavy atoms (e.g., uranyl acetate), and subsequently dried. This procedure allows to image

biological specimen since the electron-dense stain increases the contrast (the signal-to-noise ratio) of the image and data collection does not need to be carried out at low temperature.

For PdxR-DNA complex, NS-EM was not suitable for preliminary screening, since uranyl ions contained in the stain tend to precipitate when reacting with phosphate groups of the backbone of nucleic acids. For this reason, PdxR-DNA samples were imaged directly by using a cryo-electron microscope.

For the cryo-EM experiment, the stock solution of 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 down to 50 mM. Preliminary screening showed that higher salt concentration results in a lower image contrast. This is due to the electron scattering from ions which is stronger than the scattering of electrons from water causing a background noise that covers the signal of the macromolecular sample.

Sample preparation for cryo-EM requires the application of the specimen to be imaged on a grid, reducing the solution to a thin layer. We used holey carbon perforated grids (Quantifoil™ R1.2/1.3, 300-mesh) mounted on copper grid support in which particles are distributed in 1.2 μm circular holes with an interspace of 1.3 μm of a foil of ~12 nm thick of carbon. The regular distribution of the holes over the grid later facilitates the procedure of automatic data collection. Since the support surface is highly hydrophobic, grids are usually treated and hydrophilized with low-energy plasma (glow discharger), to allow a homogeneous spreading of aqueous solutions and retaining particles on the grid surface.

The general procedure of grid preparation and vitrification follows the approach developed by Jacques Dubochet and colleagues (Dubochet,1988), called plunge-freezing, which is now performed by means of a robotic semi-

automatic cryo-plunger containing a chamber with controlled temperature and humidity. We used a Vitrobot Mk IV (Thermo Fischer Scientific) operating at 4 °C and 100% humidity.

A droplet of 3.5 μL of the sample containing the PdxR-DNA complex was applied on a holey carbon grid previously glow-discharged for 45 s at 40 mA using a GloQube system (Quorum Technologies). After its application on the grid, the sample is incubated for a certain time (waiting time) and the excess liquid is wicked away blotting the grid with filter paper before plunging it in liquid ethane cooled at liquid nitrogen temperature for vitrification. The waiting and blotting time can appreciably influence the number and the distribution of particles on the grid surface, so we tested different vitrification conditions, varying these parameters. Vitrification was carried out by rapidly plunging the grid into liquid ethane cooled at liquid nitrogen temperature, around -180 °C.

Plunge-frozen grids were tested on a FEI Talos Arctica (Thermo Fischer Scientific) 200 kV microscope at the cryo-EM facility of the University of Milan (Italy), CEMIL, initially collecting a low number of images to evaluate sample quality, particle concentration and distribution in the grid holes and the ice thickness, prior a more extensive data collection. The initial screening of the prepared grids allowed the selection of the best one, that contains the most adequate number of well separated particles, that was then used for data collection. The grid selected was prepared by incubating the sample for 30 s (waiting time), then blotted for 4 s.

3.4 Data collection

Cryo-EM data were acquired continuously for seven days at CEMIL using a FEI Talos Arctica (Thermo Fischer Scientific), a 200 kV field emission electron microscope, using EPU automated data collection software (Thermo Fischer Scientific), with the specimen constantly maintained at cryogenic temperature.

Image acquisition was carried out with a nominal magnification of 120,000×, corresponding to a calibrated pixel size of 0.889 Å on the object scale, at nominal defocus values ranging between -0.8 and -2.2 μm.

3284 movies were recorded using a Falcon 3EC direct electron detector (Thermo Fischer Scientific) operating in electron counting mode with 1 s exposure time and total dose of 40 e^-/A^2, distributed over 40 movie fractions (1 e^-/A^2 per fraction).

3.5 Image processing and Single Particle Analysis

Collected data were imported in RELION 3.1 (Scheres, 2012) for all the subsequent analysis steps. Movie frames were aligned and dose-weighted using MotionCor2 to correct beam-induced overall movement across the image (Zheng et al., 2017). Contrast transfer function (CTF) was estimated with CTFFIND4.1 (Rohou and Grigorieff, 2015) on aligned and non-dose-weighted micrographs. Micrographs showing evident ice contamination were removed and those reporting resolution estimate of 5 Å or better were selected for further analysis.

To generate particle templates, 1121 particles were manually picked from several micrographs with different defocus values. Particles were then

extracted and subjected to 2D classification. Automatic particle picking was performed using the best 2D classes as a reference, resulting in 1,358,491 picked particles, with a subsequent extraction using a particle box size of 256 pixels, rescaled to 128 pixels to speed up the calculations.

Three rounds of reference-free 2D classification were carried out. In this process, particles are rotated, translated in all possible orientations, and matched to each other, then averaged together to form a certain number of classes containing different views of the specimen. Particles not representing any view are removed at this step, cleaning up the dataset.

A total of 267,988 particles selected from several 2D classes was used to generate an *ab initio* 3D model reconstructed from the different orientations of the particles, that was provided as a reference for an initial round of 3D classification. The most populated class among the four generated was chosen as a reference for the first round of 3D classification task, using multiple classes (K=4). The model was low-pass filtered to 40 Å. Among the four classes produced, it was possible to distinguish two classes differing for the conformation adopted by the DNA oligomer in the complex and for the absence (or the presence) of one of the wHTH domains in the PdxR dimer. We called the two conformers "open" and "closed", that represented ~23% (61,940 particles) and ~19% (50,146 particles) of the total particles, respectively. The two resulting 3D maps were then analysed separately.

Each map was subjected to an additional 2D and then a 3D classification round (K=2) with the aim to discard noisy classes and to keep only well-defined and aligned ones that could contribute to high-resolution reconstructions.

A final selection of 24,430 particles corresponding to the closed conformer, and 33,371 particles for the open conformer was obtained. The selected classes were independently refined using RELION 3D auto-refinement tool and post-

processing, followed by Bayesian polishing to correct beam-induced motion at the particle level. CTF refinement (per-particle defocus, beam-tilt estimation, per-micrograph astigmatism) and subsequent 3D refinement of both sets of particles showed no significant improvement.

Final refinement of polished particles was performed applying extended and soft-edged masks (to exclude solvent contribution) to a 40 Å low-pass filtered reference maps.

Masked 3D refinement of the closed conformation was done either imposing no symmetry (C1) or forcing map calculation with a C2 symmetry. Indeed, both the protein and the DNA show an intrinsic symmetry and the overall C1 map of the complex reconstructed in the closed state displays a C2 symmetry. Being the open conformer map clearly asymmetric, no symmetry was imposed during refinement.

For all the 3D maps, post-processing using the same masks was repeated for map sharpening and resolution estimation.

3.6 Cryo-EM maps resolution estimation and sharpening

The overall resolution of the three maps of PdxR-DNA complex (open and closed with C1 and C2 symmetry) was estimated from Fourier Shell Correlations (FSC) = 0.143 criterion, based on the "gold-standard" protocol (Rosenthal and Henderson, 2003; Scheres and Chen, 2012), using a mask around the complex density. A final resolution of 3.9 Å was estimated for the closed conformer refined with C1 symmetry, 3.7 Å for the map refined imposing a C2 symmetry, and 3.9 Å for the open map.

The input maps were corrected for the modulation transfer function (MTF) of Falcon 3EC direct detector and sharpened using negative temperature B factors

as estimated by RELION (**Table 3.1).** Since global resolution estimated with a post-processing job does not consider variations in resolution throughout the reconstructed map, Local Resolution tool in RELION 3.1 was used to estimate the resolution of the different regions composing the 3D maps, calculating the contribution of each part using a post-processing-like procedure with a soft spherical mask that is moved around the entire map. Averaged half-maps and B-factor estimated with post-processing were used as inputs for Local Resolution task. Output maps were visualized in UCSF Chimera X (Pettersen et al., 2021) and coloured according to the different resolution values (**Figures 4.13, 4.14, 4.15**).

Since the closed map calculated with the C2 symmetry achieved a higher resolution for the protein-DNA interacting portions, this map together with the map of the complex in the open state were used to build the atomic models. The two electron density maps were sharpened using the Autosharpen tool in Phenix (Liebschner et al., 2019) prior to model buiding. UCSF Chimera (Pettersen et al., 2004) and Coot (Emsley and Cowtan, 2004) were utilized for graphical visualization of the electron density maps.

3.7 Model building of PdxR-DNA complex

Following the density of the closed and open maps reconstructed by single particle analysis, the initial atomic model for PdxR-DNA complex was assembled as follows.

Coordinates of the crystal structure of PdxR aspartate aminotransferase (AAT)-like domain dimer, comprising also part of the linker (residues 103-464, data not published), were used as a starting model to build the central core of PdxR-DNA complex for both open and closed conformers.

The two winged helix-turn-helix (wHTH) domains and the remaining residues of the linker (residues 1-102) were extracted by the PdxR structure previously obtained by homology modelling (Tramonti et al. 2015) and kindly provided by Prof. Stefano Pascarella, Sapienza University of Rome.
The 48-bp double stranded B-DNA was built in Coot in a straight conformation.
PLP molecule was imported in Coot and linked as a Schiff-base to Lys309 of both the AAT-like monomers of the open and closed maps.

3.7.1 Open conformation

The model of open PdxR-DNA complex was built first, since the map appeared better reconstructed than the closed one, and more details were detectable, although the resolution was slightly lower.
The crystal structure of the AAT-like domain dimer was rigid-body fitted into the central core of the sharpened electron density map at 3.9 Å resolution using UCSF Chimera (Pettersen et al., 2004). Model was inspected in Coot and the two monomers of the AAT-like domain were adjusted using the "jiggle fit" option (Brown et al. 2015).
The same procedure was adopted to fit one of the two wHTH domains (monomer A) in the map, since no density was reconstructed for the wHTH of the monomer B. Moreover, the density corresponding to the long flexible loop that connects each wHTH with the AAT-like domain was also not clearly visible, so residues 84-103 were removed from the model, as well as the N-terminal residues 1-9.
The straight 48-bp B-DNA was partially fitted as rigid body into the density. However, the pronounced curvature adopted in the complex, clearly visible in

the density map, only allowed a small portion of the DNA to be correctly fitted. The DNA molecule was then subjected to a long work of manual building and refinement in different regions imposing strong geometry restraints, until an adequate curvature and fit was achieved. At the resolution obtained, the density showed major and minor grooves, and even bumps for the phosphates in the backbone; this facilitated and allowed the reconstruction of a reliable atomic model. 46 of the 48 di-nucleotides were built, since the last two were not defined by density (n. 47 and 48 of 5'→3' strand; n. 1 and 2 of 3'→5' strand). All the chains corresponding to the AAT-like and wHTH domains and to the DNA were merged, and the whole structure was rigid body refined using real space refinement tool in Phenix, against the overall map at 3.9 Å, imposing secondary structure and Ramachandran restraints for protein.

Final visual inspection was performed in Coot to manually adjust secondary structures, amino acid side chains, PLP molecules and nucleotides into the density, and to correct Ramachandran outliers. The final model was validated using MolProbity (Williams et al., 2018) and EMRinger (Barad et al., 2015). Statistics for the final model are shown in **Table 3.1**.

3.7.2 Closed conformation

The map of the closed conformer chosen for model building was the one reconstructed imposing a C2 symmetry, showing a global resolution of 3.7 Å. Similarly to the open PdxR-DNA, the atomic structure of the closed conformer was built starting from the merged model comprising the AAT-like dimer crystal structure, the wHTH domain of both A and B monomers from the above-mentioned homology model and the 48-bp B-DNA that was taken by the open model (3.7.1).

The overall structure was initially rigid body positioned in the final sharpened open map using USCF Chimera, followed by model building and refinement procedure manually performed using Coot. PLP was imported and linked to Lys309 of both monomers.

The DNA was adjusted following the approach used for the open conformer. The density in correspondence to the last base pair T-A (n. 48 of the 5'→3' strand; n. 1 of the 3'→5' strand) was absent, and it was removed from the model, as well as protein residues 1-9, and the linker (residues 84-103).

The final model was validated using MolProbity and EMRinger (**Table 3.1**).

Table3.1 Cryo-EM data collection, refinement and validation statistics.

*model quality is still not optimal and requires further improvements.

	PdxR-DNA open complex	**PdxR-DNA closed complex with C1 sym**	**PdxR-DNA closed complex C2 sym**
Data collection and processing			
Magnification	120K	120K	120K
Voltage (kV)	200	200	200
Electron exposure (e–/Å^2)	40	40	40
Pixel size (Å)	0.889	0.889	0.889
Symmetry imposed	C1	C1	C2
Initial particle images (no.)	1358491	1358491	1358491
Final particle images (no.)	33371	24430	24430
Map resolution (Å)	3.86	3.86	3.73
FSC threshold	0.143	0.143	0.143
Map resolution range (Å)	3.7-7.0	3.7-7.0	3.6-6.5
Refinement			
Initial model used	PdxR AAT domain obtained by X-ray (*unpublished data*), wHTH domain obtained by homology modelling (Tramonti et al., 2015), DNA generated by COOT (Emsley et al., 2010)	PdxR-DNA open	PdxR-DNA open
Model resolution (Å)	3.86	3.86	3.73
FSC threshold	0.143	0.143	0.143
Map sharpening *B* factor (Å^2)	-117.9	-126.3	-111.1
Model composition			
Non-hydrogen atoms	8415	-	9035
Protein residues	797	-	869
Nucleotides	92	-	94
B factors (Å^2)			
Proteins (min, max)	25.3 (9.5, 81.5)	-	90.7 (30.0, 163.8)
Nucleotides (min, max)	6.1 (0.0, 20.0)		4.8 (0.0, 20.0)
R.m.s. deviations			
Bond lengths (Å)	0.013	-	0.01
Bond angles (°)	2.37	-	1.802
Validation			
MolProbity score	2.97	-	3.57
Clashscore	33.05	-	58.7
Poor rotamers (%)	4.68	-	10.57
EMRinger score	1.06	-	1.39
Ramachandran plot			
Favored (%)	92.67	-	89.94
Allowed (%)	7.33	-	10.06
Outliers (%)	0.00	-	0

3.8 Prediction of DNA topology from sequence

The 3D structure of the PdxR target DNA was determined with the Christoph Gohlke DNA Curvature Analysis service (http://www.lfd.uci.edu/~gohlke/dnacurve/) using the AA wedge model (Young and Beveridge, 1998). Same procedure was carried out for the mutants described in Section **3.9**. A 48-bp fragment was used in all cases.

3.9 Electrophoretic Mobility Shift Assays (EMSA)

DNA mobility shift assays were performed using purified PdxR (from 25 nM to 800 nM) and different DNA fragments (wild type and modified forms) at a fixed concentration of 10 nM.

The modifications were introduced to alter the sequence of the PdxR target repeats (known as motifs 1 and 3) or the bendability of the oligonucleotide (sequences are listed in **Table 3.2**).

The synthetic double-stranded DNA fragments were generated by annealing of complementary forward (FWD) and reverse (REV) oligonucleotides.

The resulting fragments were incubated at 22 °C for 20 min in 10 μL of binding buffer [20 mM HEPES pH 8.0, 50 mM KCl, 5 mM $MgCl_2$, 1 mM DTT, 0.05% (v/v) NP-40, 30 μg/mL BSA and 5% (v/v) glycerol] with increasing concentrations of purified PdxR (25 nM, 50 nM, 100 nM, 200 nM, 400 nM, 800 nM).

A 10-fold excess PLP with respect to protein was added to ensure complete saturation.

Dna fragment (48bp)	Sequence
WT	CTGACCTCATCATTTTCTTAAAAACTGACACTTACAATGTGGTCAGTT
M13	CTTCTTTCATCATTTTCTTAAAAACTGACACTTACAATGTAAGAAGTT
BENT	CTGACCAAAACAAAAAATCTCCAAAAAACTCGAAAAAAGTGGTCAGTT
STRAIGHT	CTGACCTCATCATCTCCCTACACACTGACACTTACAATGTGGTCAGTT

Table 3.2 Oligonucleotides used for EMSA. Underlined letters indicate nucleotides of the direct and inverted repeats (motifs 1 and 3). Red letters indicate mutated nucleotides.

Samples were loaded onto 5% non-denaturing polyacrylamide gels in 0.5× TBE buffer (45 mM Tris-borate, 1 mM EDTA). Gels were run at room temperature in 0.5× TBE buffer, then incubated with SyBR Green (Sigma-Aldrich) for 10 minutes and visualized on UV transilluminator.

The addition of the protein to DNA fragments determined the formation of slowly migrating bands, whose density increases as protein concentration is increased. These retarded bands correspond to protein-DNA complexes. Densitometric measurements of the slower bands (DNA-protein complex) were transformed into percentage of DNA-protein complex with respect to total DNA and yielded dependence of binding on protein concentration. Least square fitting of these data to the Hill equation (**Eq. 1**) allowed the estimation of the apparent dissociation constants (K_{Dapp}) reported in the main text, which are expressed as the average ± standard deviation of values obtained in three or more independent EMSA experiments. These do not correspond to actual dissociation constant values, rather they coincide with the protein concentration required to shift 50% of free DNA.

Eq. 1 $$\%free\ DNA = 100 - \left(100 \times \frac{[PdxR]^n}{[PdxR + K_d^n}\right) + const$$

4. Results and Discussion

4.1 Cryo-EM data collection of *Bacillus clausii* holo-PdxR-DNA complex

The main task of this work was to unveil the structure of the complex formed by the MocR transcriptional regulator PdxR in its pyridoxal 5'-phosphate (PLP)-bound form (holo-PdxR, that, for the sake of simplicity, we shall hereafter call PdxR) and a 48-bp DNA containing the direct and inverted repeats representing its binding sites.

Cryo-Electron Microscopy (cryo-EM) was employed to acquire images for the reconstruction of high-resolution electron density maps by single particle analysis and to build the structure of PdxR from *Bacillus clausii* bound to its target DNA.

Single particle analysis follows a trial-and-error approach which is necessary to select only the best particles among the thousands or millions collected, that allow the high-resolution reconstruction of the object of interest. The analysis of the PdxR-DNA system has required a long and complex work because, as illustrated below, the sample was conformationally heterogeneous. For the sake of brevity and simplicity, here only the main and most significant steps of the analysis performed will be described.

4.1.1 Sample preparation and cryo-EM screening

Prior to collect cryo-EM data, a preliminary step of optimization of sample concentration, grid preparation and freezing conditions is usually carried out. We performed screening and data collection at the cryo-EM facility of the University of Milan (CEMIL), Italy.

Freshly purified PdxR was incubated for 48 hours with its target 48-bp DNA to allow complex formation, as described in sections **3.1** and **3.2**. Few microliters of sample were applied on glow-discharged holey carbon grids (Quantifoil™ R1.2/1.3, 300-mesh). An initial screening was performed to identify the best conditions of protein concentration and buffer. We found that a sample concentration of 0.5 mg/mL allowed an adequate number of particles to be deposited within the grid holes, and that a concentration of 50 mM of NaCl in the final buffer was optimal to preserve the stability of the complex and, at the same time, to minimize the background signal generated by ion scattering. Six different vitrification conditions were then tested, this time only varying the time of incubation of the sample on the grid and the blotting time (as described in section **3.3**), using a Vitrobot system. Screening was performed at the cryo-electron microscope using a FEI Talos Arctica 200 kV (Thermo Fischer Scientific).

Ice thickness and particles distribution are the main aspects which are considered for the evaluation of the best vitrified grid, suitable for data collection. The grid blotted for 4 seconds after a sample incubation of 30 seconds showed the best particle distribution and adequate ice thickness and it was used for an extended data collection.

4.1.2 Data collection and micrograph inspection

A full dataset of 3284 movies (or micrographs) was collected using a Falcon 3EC direct electron detector (Thermo Fischer Scientific). Micrographs were recorded in a low-dose mode to prevent radiation damage to the sample (1 e^-/A^2 per movie frame). A representative collected micrograph is showed in **Figure 4.1A**. It displays a random and homogeneous distribution of particles in the vitreous ice layer, and the absence of ice contamination or particle aggregation reflects the good quality of the sample and of the vitrification procedure adopted.

To evaluate the good contrast of the collected images and eventually present aberrations, we also inspected the power spectra of the micrographs.

The aberrations of the electron microscope are mathematically described by the Contrast Transfer Function (CTF), an oscillating function that modulates the image contrast in the Fourier space and must be calculated to restore the actual contrast of the cryo-EM data. The effect of CTF can be seen in the alternating light and dark rings (Thon rings) composing the power spectra of the micrographs, which show the relation between image contrast and spatial frequency (Orlova and Saibil, 2011; Carroni and Saibil, 2016).

The circular and symmetric Thon rings in the power spectrum density of the micrographs indicate that the effects of drifting, astigmatism, or other optical aberrations were negligible (**Figure 4.1B)**

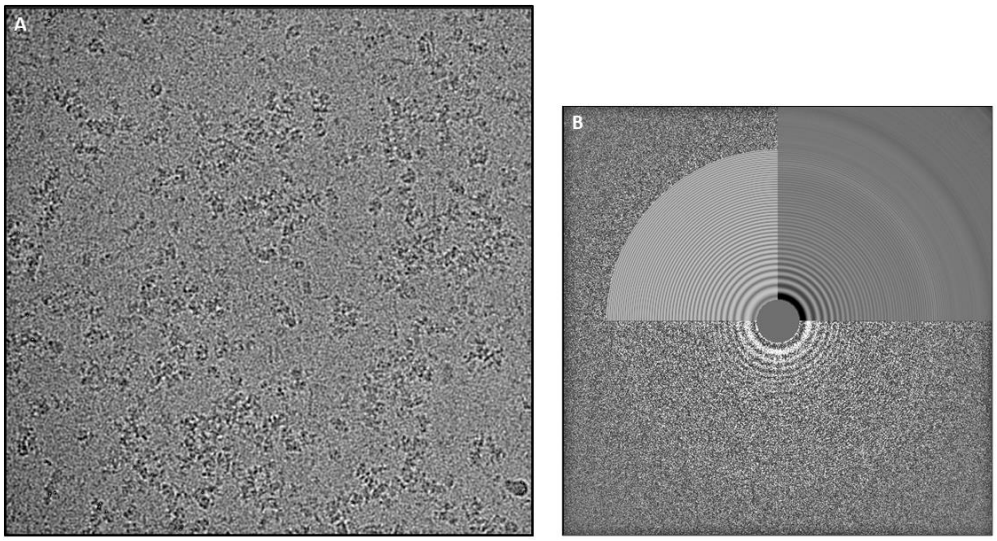


Figure 4.1 Inspection of cryo-EM micrograph of holo-PdxR-DNA complex. **A**: A representative micrograph from the data collection, showing the well-distributed particles of PdxR-DNA complex. **B**: Power spectrum of the same micrograph. Thons rings are visible as concentric circles radiating from the centre.

4.2 Single Particle Analysis

One of the biggest issues in Cryo-EM data analysis is the large computational resources required for single particle image processing and data storage, that cannot be sustained by a common personal computer. A typical cryo-EM data set is constituted by hundreds or thousands of images, each one being a movie of 2 gigabytes. Therefore, a single dataset consists of about 2 up to 5 TB of data that require at least 190 CPU-hours for each run of image processing. To speed-up the calculations, all software performing single particle analysis were implemented to use graphics processors (GPUs) to address the most

computationally intensive steps of cryo-EM structure determination workflow. Therefore, the usage of GPU-based architecture is fundamental due to the highly demanding computational cost of a typical run of image processing.

In order to analyze the cryo-EM data collected we took advantage of the computing resources provided by the high-performance computing facility CINECA (https://www.hpc.cineca.it/).

All data were imported and stored on a cloud infrastructure called MARCONI 100, provided by ISCRA CINECA, in which we remotely performed cryo-EM data processing using RELION (REgularized LIkelihood OptimizatioN) 3.1 (Scheres, 2012), a widely use program suite for single-particle analysis based on a Bayesian approach.

4.2.1 Image processing

Prior to work with single particles, images containing the 2D projections of the specimen of interest must be processed.

The 3284 movies collected were first aligned for beam-induced motion correction and drift with MotionCor2 (Zheng et al., 2017); then, CTFwas calculated using CTFFIND4.1 (Rohou and Grigorieff, 2015). After an inspection of the whole data set, images containing relevant ice contamination or protein aggregation were removed and micrographs showing CTF resolution limits better than 5 Å were selected for further analysis (3197 total images selected).

4.2.2 Particle picking and extraction

Cryo-EM data analysis relies on a machine-learning approach, where the human eye plays a crucial role in choosing the first set of particles from the collected micrographs to train the software in picking all the possible 2D projections resembling the specimen of interest, if manual selection is carried out.

For PdxR-DNA data set, 1121 particles were manually picked from a selected group of micrographs with different defocus values and used as references for the automated particle picking procedure on the entire set of images (an example of manual and automated picking is shown in **Figure 4.2**)

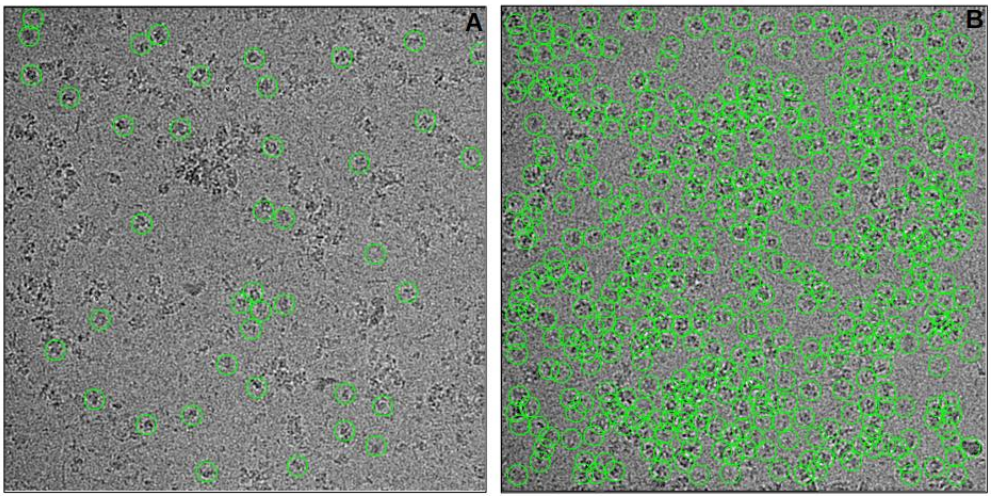


Figure 4.2 Example of micrographs showing manually and automatically picked particles (in green circles). **A:** Manual picking: representative micrograph where a low number of particles were selected by eye to train the program. **B:** Autopicking: representative micrograph displaying all the particles selected after software training.

A total of 1,358,491 particles were collected and extracted from the original micrographs. From this point forward, all subsequent tasks were performed on these selected single particles.

4.2.3 2D classification

After extraction of the autopicked particles, one needs to identify those particles that are suitable for high-resolution structure determination. One of the most effective ways of selecting suitable particles is 2D classification (Scheres, 2016). Therefore, extracted particles were subjected to a reference-free 2D classification, a procedure that classifies particles according to their shape and spatial orientation, generating a certain number of 2D class averages containing views of differently orientated particles (Orlova and Saibil, 2011). The aim of the 2D classification step is to remove junk particles, contained in smeared or noisy classes that do not include the particles of interest, thus reducing their negative impact on the further steps of 3D reconstruction.

The first round of 2D classification iteratively generated 100 class averages (K=100) from which we selected all the particles resembling a protein-DNA complex (**Figure 4.3 A**). We proceeded with two additional rounds of 2D classification, setting the same number of classes (K=100) and selecting the best and highly populated ones. Selection has been made on the basis of signal-to-noise ratio, that corresponds to those classes showing strong, white density for the macromolecular complex and a black and featureless background (Scheres, 2016), and of their distribution, meaning how many particles were averaged in each class, and particle shape. We were mostly guided by the presence of the DNA, being clearly detectable in several classes. Some

examples of the resulting “good classes” are shown in **Figure 4.3 B**. A total of 267,988 particles were selected for further analysis.

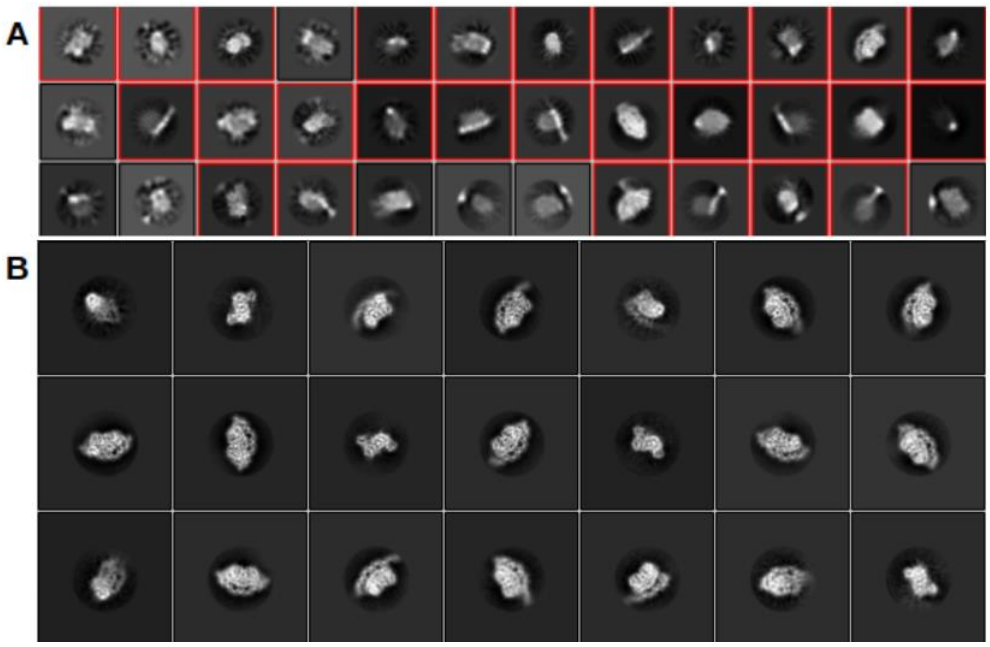


Figure 4.3 2D class averages resulted by different rounds of 2D classification on PdxR-DNA complex. **A:** 2D classes generated by the first round of 2D classification sorted by their distribution. 36 of 100 classes are showed. Classes selected for subsequent rounds of classifications are in red squared. **B:** Examples of best 2D class averages obtained with the last round of 2D classification.

At this step, class averages populating multiple orientations should appear and should be selected. Therefore, 2D class averages containing multi-angle views of the particles were selected and used to generate an initial low resolution 3D model of PdxR-DNA complex.

4.2.4 Generation of a 3D initial model and 3D classification

A preliminary low resolution three-dimensional map of PdxR-DNA complex was generated using the 3D Initial Model tool in RELION. The resulting map is shown in **Figure 4.4**.

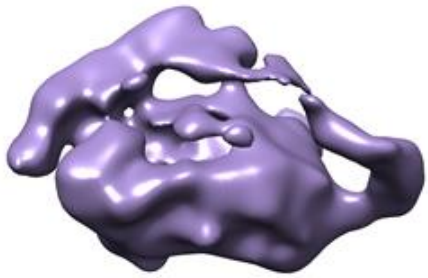


Figure 4.4 Low resolution PdxR-DNA initial model generated *de novo* from the selected 2D classes.

This model provided a first clue on the architecture of the complex, even if was not possible to detect any detailed structural information, being the map at low resolution.

This model was then used as a reference map for the first round of 3D classification, using K=4 and without imposing any symmetry. The reference map was low pass filtered at 60 Å to prevent model bias.

Among the 4 classes generated (**Figure 4.5A**), the most populated class that showed strong protein-like features was selected, and it was further used as a reference (filtered at 60 Å) for an additional round of 3D classification with K=4 run over the same set of particles. The idea was to provide a better-defined starting 3D reference to facilitate the alignment procedure performed

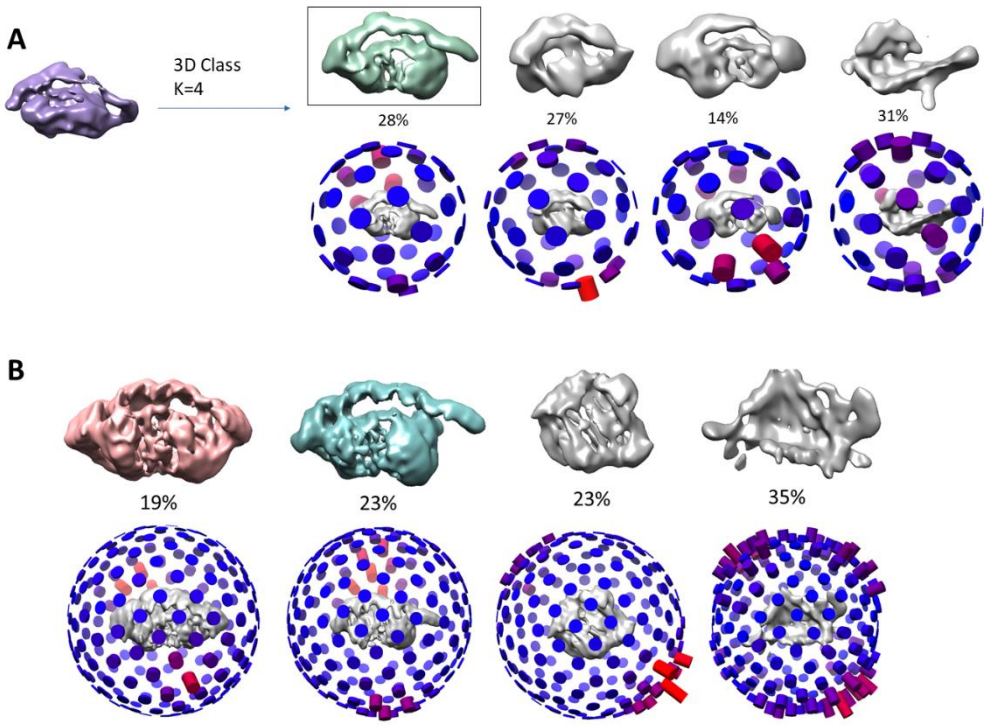
by RELION. The difference in classification efficiency reflects the dependence upon a starting model of the reference-based 3D classification procedure. Therefore, the better is the starting model the better is the result of the 3D classification.

The maps generated with the second round of 3D classification revealed the presence of two possible conformations of the PdxR-DNA complex, showing a density ascribable to DNA differently bound to a density most likely belonging to the protein (**Figure 4.5B**).

Overall, both reconstructed maps resemble the three-dimensional structure predicted by homology modelling for both GabR and PdxR in the DNA-bound form (Al Zyoud et al., 2016; Tramonti et al., 2015), as shown in **Figure 1.13**, section 1.7.3), with the domain-swap homodimer and the double strand of DNA bound to wHTH domains.

Class 1 (red coloured map) exhibits a more compact shape in which the DNA molecule looks tightly and symmetrically bound to both monomers of PdxR (closed conformation). On the contrary, in class 2 (sea green coloured map) the DNA looks partially detached, with one of the winged helix-turn-helix (wHTH) domain missing in the density (open conformation).

Class 1 and 2 were selected and kept separate in all subsequent steps of the single-particle analysis.



.

Figure 4.5 Schematic representation of two rounds of 3D classification. **A**: The four classes generated with the first round of 3D classification using the 3D initial model as a reference (violet map) are represented. The first class (in green) was selected for the second round of 3D. **B**: The four classes generated by the second round of 3D classification are shown. The first two classes (red and sea green map, respectively) represent two different conformations of PdxR-DNA complex and were selected for subsequent analysis. The angular distribution of particle projections is shown in the same orientation and at the bottom of the respective map. The height and colour (from blue to red) of the cylinders are proportional to the number of particles that populate that view.

4.2.5 Refinement of the closed conformer

The map corresponding to the closed conformation of PdxR-DNA complex (red coloured map in **Figure 4.5 B**) was subjected to a further round of 2D classification in order to perform an additional "cleaning up" procedure to remove noisy averages that could impair a high-resolution reconstruction. Best classes were selected and subjected to a 3D classification with K=2. This step produced a most populated class, containing 24,430 particles and displaying the highest resolution. The selected class was then refined using the 3D Autorefine tool, that allowed the generation of a map at 7.3 Å resolution.

The refined map was then improved and sharpened using the Post-processing tool. A crucial procedure to gain higher resolution in cryo-EM reconstruction is to create a mask covering the particle of interest and excluding the noisy contribution of the solvent. The step of post-processing improved the nominal resolution to 4.5 Å, generating a clearly more detailed map, where the secondary structures of the protein and the double helix of DNA with minor and major grooves were visible. The procedure is summarized in **Figure 4.6**.

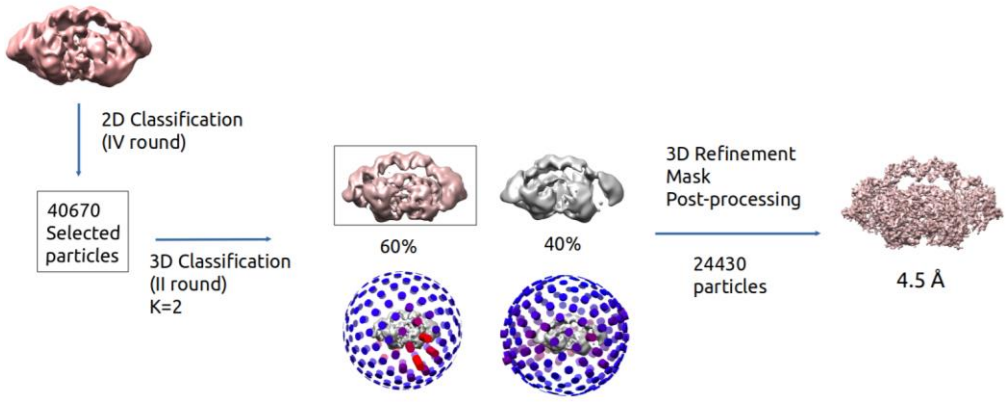


Figure 4.6 Analysis of the closed conformer: the route to the high-resolution map. The map obtained by the 3D classification was subjected to further 2D (K=50) and 3D (K=2) classifications. Among the resulting two classes, class number 1 was selected (shown in light red and squared). The map produced with refinement and post-processing steps is shown on the right. The angular distribution of particles projections is displayed in the same orientation as the relative map.

Additional steps of refinement can be performed on the post-processed maps with the aim of achieving further improvements in the resolution, including CTF-refinement and particle polishing procedures. CTF-refinement is useful to correct *a posteriori* the preliminary 3D reconstruction of the data for the optical aberrations and magnification anisotropy, which are strongly affected by the alignment of the magnetic lenses of the microscope, and to estimate a per-particle defocus. The latter represents a further improvement of the calculated CTF that was initially determined for the overall micrograph. The CTF-refinement algorithm re-estimates the defocus values for each particle. This optimization considers the possibility that ice distribution on the micrograph is not perfectly homogeneous and flat, resulting in slight shift of

the exact position of the single particle along the z-axis. In our case, several attempts of CTF-refinement did not yield higher-resolution reconstructions, so we decided to proceed only with polishing.

Particle polishing utilises a Bayesian approach to estimate and correct the beam-induced movement of single particles in the data set once a preliminary 3D-map reconstruction is generated, in a similar way to motion correction, that operates on the entire micrographs.

The application of a single round of Bayesian polishing procedure, followed by a further step of masked 3D refinement and post-processing, allowed us to reconstruct a map of PdxR-DNA complex in its closed conformation at a global resolution of 3.9 Å, without imposing any symmetry in refinement tasks. The FSC curve is showed in **Figure 4.7**.

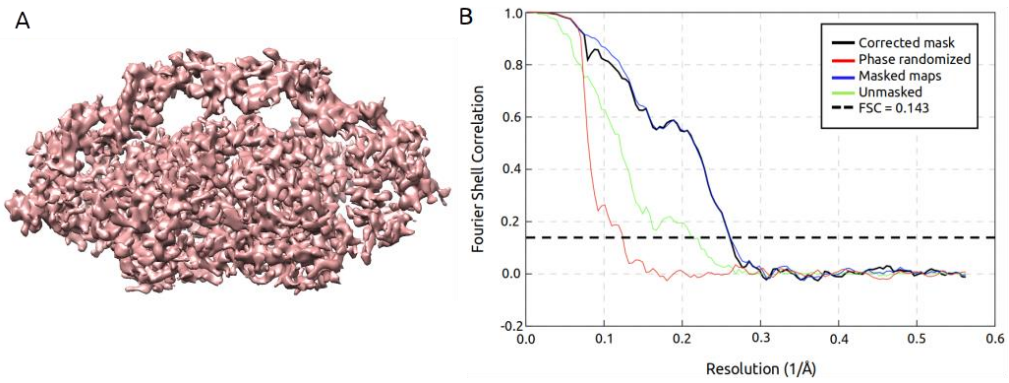


Figure 4.7 Cryo-EM map of closed PdxR-DNA complex with a C1 symmetry. A: The cryo-EM density reconstructed at 3.7 Å resolution. B: FSC curves calculated for independently refined half-maps.

As evidenced by the blue/red histograms shown in **Figure 4.5** and **4.6** under the different classes generated by 3D classification jobs, the angular distribution of particles projections reflects an inhomogeneous distribution of particles, with some views being less represented than others. More in details, the front views of the complex seem to be more populated than the top views. This is most likely due to a slight effect of preferential orientation which is adopted by the PdxR-DNA complex particles when they were deposited over the grid, and it represents a recurring problem in Cryo-EM. This phenomenon is principally caused by the adsorption of particles to the air/water interface or to the grid support in a preferred manner, preventing the random and uniform distribution of angular views of the particle projections, that results in a "smearing effect" of the 3D reconstruction, also slightly visible in our map of "closed" PdxR (**Figure 4.8**). A strategy usually employed in this case to tackle this phenomenon is to apply tilting during data collection (Tan et al., 2017). With our data, this problem was only partially solved since the collection was performed without tilting. However, the smearing effect, which is slightly visible after the refinement step, becomes less pronounced after post-processing, returning some regions of the map endowed with local resolution lower than others, mostly those around the more external portions of the DNA. All considered, our data were good enough to allow a good reconstruction of the density map that was used to build a reliable atomic model of the closed complex, as it will be better detailed later.

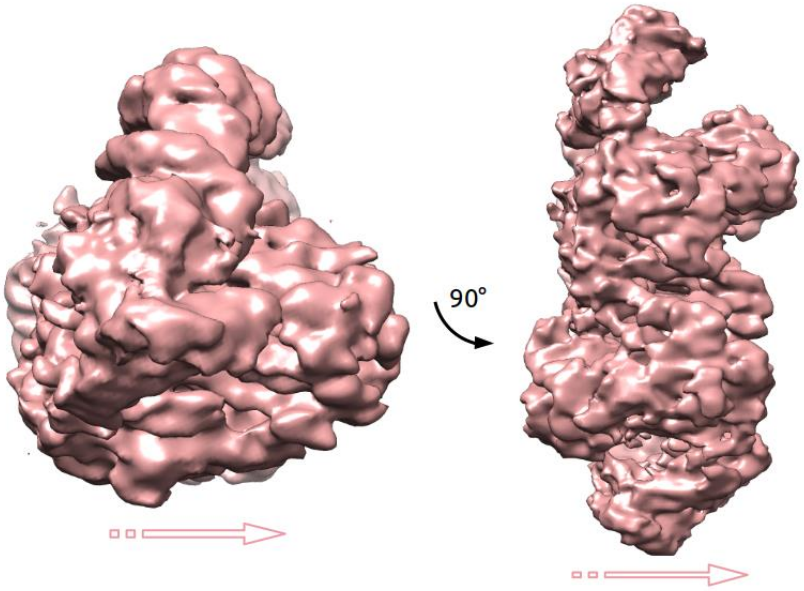


Figure 4.8 Preferred orientation in the closed PdxR map. Side and bottom views of the map generated with a masked refinement at 3.9 Å showing a slight "smearing effect", which direction is indicated by the pink arrow.

In single particle analysis it is possible to apply symmetry averaging during refinement, which is always helpful to boost the signal and increase the final resolution of the generated maps. PdxR was predicted to be structurally homologous to the domain-swap symmetric homodimer of *Bacillus subtilis* GabR, therefore, it should be endowed with an intrinsic C2 symmetry (Edayathumangalam et al., 2013; Tramonti et al., 2015). Nevertheless, the DNA binding could break the symmetry. However, the inspection of the reconstructed map in closed conformation revealed that the overall two-fold rotational symmetry (C2) of the protein is maintained also in complex with the DNA. Higher resolutions would have revealed further detail, such as DNA

nucleotides disposed in a non-symmetric manner, but at ~ 4 Å resolution such details are not distinguishable. Therefore, a further step of refinement was performed, this time imposing a C2 symmetry, using a reference map in which the symmetry axes were aligned with the main x,y,z axes of the coordinate system, according to RELION's conventions.

Starting from the Bayesian polished particles, a masked refinement and post-processing were run applying a C2 symmetry (**Figure 4.9**).

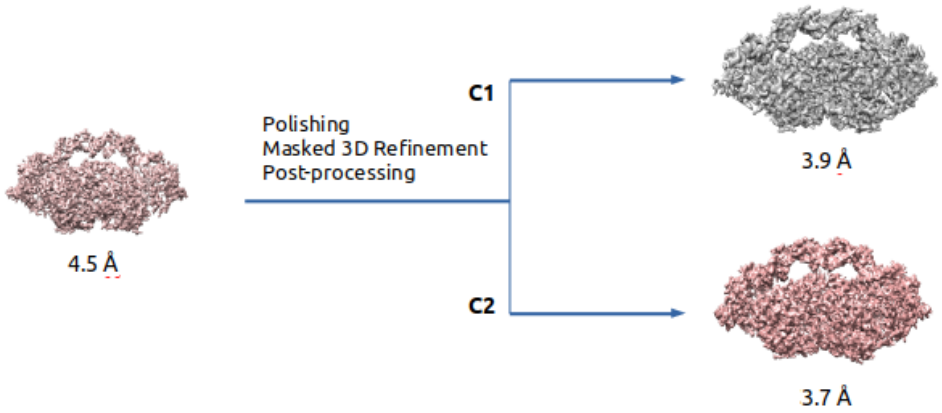


Figure 4.9 Analysis of the closed conformer. The map obtained from post-processing was subjected to a Bayesian polishing refinement. Further steps of masked refinement and postprocessing were performed both without (as discussed in 4.2.5; grey coloured map) and with a C2 symmetry (red coloured map).

This led to a map showing a global resolution of 3.7 Å, that did not improve after post-processing (**Figure 4.10**).

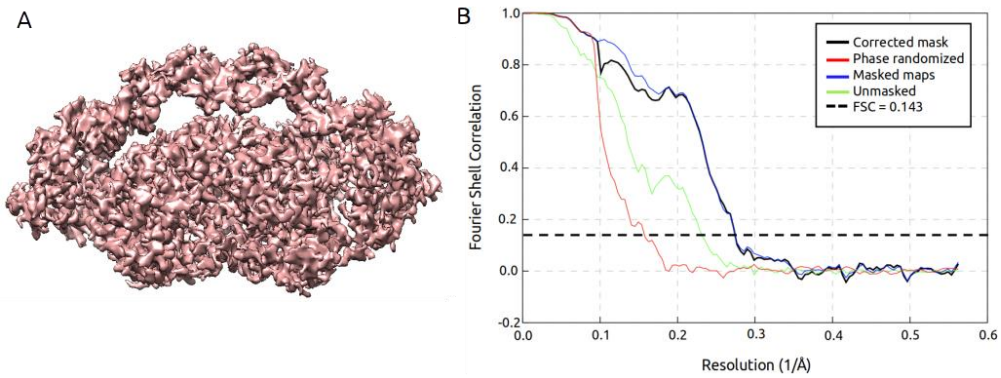


Figure 4.10 Cryo-EM map of closed PdxR-DNA complex with a C2 symmetry. A: The cryo-EM density reconstructed at 3.9 Å resolution. B: FSC curves calculated for independently refined half-maps.

Given the higher resolution, the map reconstructed at 3.7 Å was chosen to build the model of PdxR-DNA complex in the closed conformation.

4.2.6 Refinement of the open conformer

Similarly to what applied for the reconstruction of the closed conformer, the map displaying the PdxR-DNA complex in an open conformation (class 2 of the second round of 3D classification, showed in **Figure 4.5 B**), was subjected to further 2D (K=50) and 3D (K=2) classification rounds. The jobs carried out and the resulting maps are schematized in **Figure 4.11**.

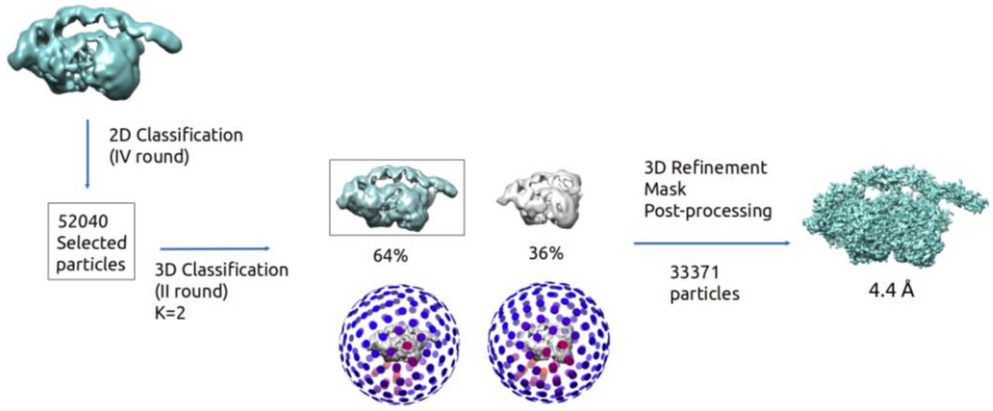


Figure 4.11 Analysis of the open conformer. The map obtained from the 3D classification were subjected to further cycles of 2D and 3D classifications that produced a better reconstructed class (n. 1) that was selected for refinement (shown in sea green and squared). The map produced after refinement and post-processing is shown on the right. The angular distribution plot of the different map orientations is displayed in the same orientation as the relative map.

The first refinement jobs led to a 4.4 Å resolution of the reconstructed density map, while particle polishing, masked refinement and post-processing improved the resolution of the final map up to a global resolution of 3.9 Å (**Figure 4.12**).

Contrarily to the closed conformer, in the open one the conformation adopted by the DNA and the absence of one of the wHTH domains produced an overall asymmetric structure. Therefore, in this case the C1 symmetry was kept in all the refinement jobs.

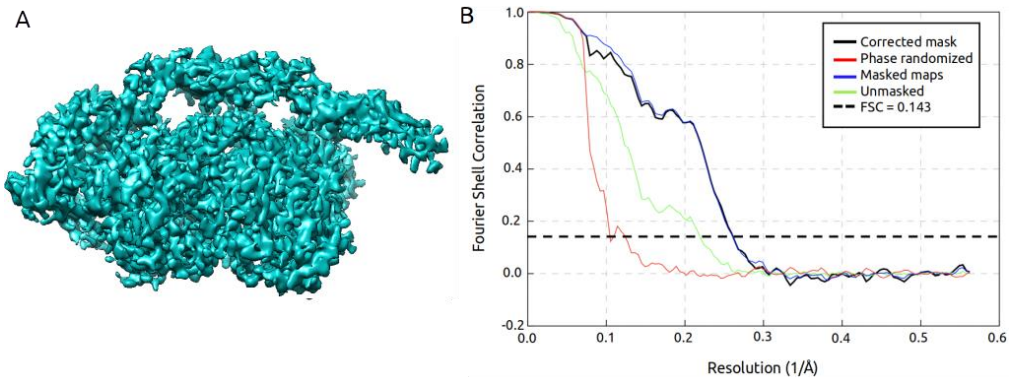


Figure 4.12 Cryo-EM map of the open PdxR-DNA complex. A: The cryo-EM density reconstructed at 3.9 Å resolution. B: FSC curves calculated for independently refined half-maps.

4.2.7 Local resolution estimation

A possible limitation of the global resolution estimation through FSC method is that the procedure does not consider local variations in resolution, due to processing errors or data heterogeneity, especially in the analysis of macromolecular complexes. The analysis of the local fluctuation of the resolution in the individual areas of the object of analysis guides the construction of a 3D atomic model into Cryo-EM maps, avoiding misinterpretation of not well-defined densities.

We employed Local Resolution tool implemented in RELION 3.1 to investigate how the resolution is distributed throughout the density map for both closed and open reconstructions of PdxR-DNA complex.

Local resolution estimation in the closed conformer

The first map of the closed conformer, refined without imposing symmetry, showed an overall resolution of 3.9 Å. In **Figure 4.13**, the final map is coloured according to local resolution estimated by Local Resolution in RELION, with a colour gradient ranging from red to blue representing the region with higher and lower resolution, respectively.

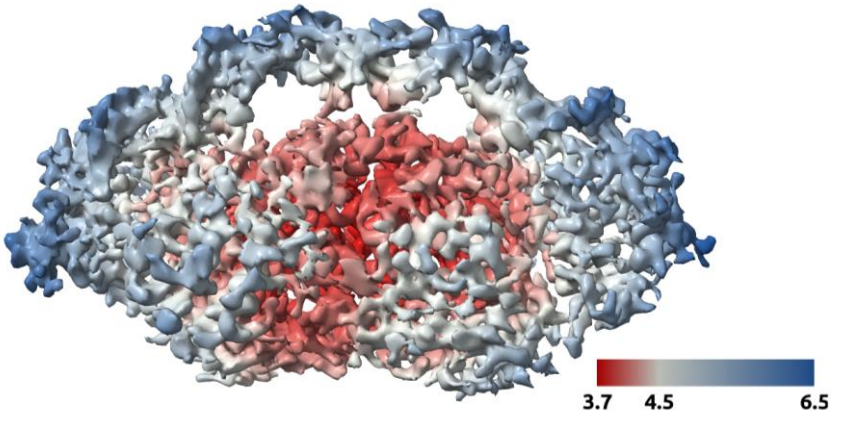


Figure 4.13 Local resolution of the closed PdxR-DNA map with C1 symmetry. The final C1-closed map is coloured according to local resolution, with a gradient ranging from red to blue representing the region with higher and lower resolution, respectively.

The distribution of resolution values over the map shows that the central core of PdxR containing the aspartate aminotransferase (AAT)-like domains represents the region reconstructed with the highest resolution (≤ 4 Å), as well as the protein-DNA interacting region, between 4 and 4.5 Å, both at the level of the wHTH and of the AAT-like domains. Indeed, in these regions, secondary

structures and bulky side chains of interacting residues are clearly visible. This is crucial to allow a more precise model building and to give insight into the interaction between the DNA bases or phosphate backbone and amino acid residues of PdxR.

Portions of the complex that are more exposed to the solvent or endowed with higher degree of flexibility, such as one of the wHTH domains and the external regions of DNA and of the AAT-like domains (coloured in blue), represent the area of the complex reconstructed with the lowest resolution.

The imposition of the C2 symmetry to the closed conformer map was not only useful to increase the global resolution but allowed also improving the resolution locally. However, similarly to the C1-closed map, the highest resolution values are registered in the central area, covering both part of the AAT-like domain and the DNA in the interacting regions, and it decreases moving from the centre to the external portions of the complex. In this map, the density of one of the wHTH domain was better defined and allowed a more reliable model building.

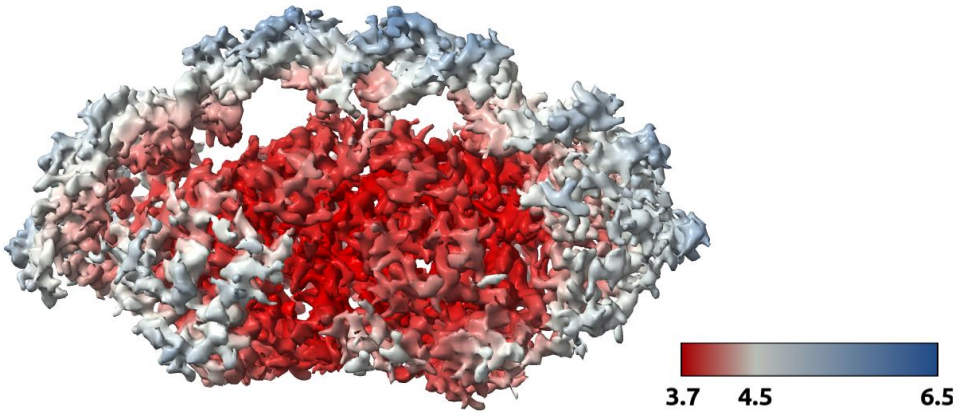


Figure 4.14 Local resolution of the closed PdxR-DNA map with C2 symmetry. Map is coloured according to the estimated local resolution values.

Local resolution estimation in the open conformer

Similarly to the closed map, the local resolution of the open conformer tends to be higher around the AAT-like dimer and in the DNA-binding regions, both at the AAT-like and the wHTH domains. However, the higher number of particles used for its reconstruction and their more homogeneous angular distribution allowed the reconstruction of a density map of the protein more defined and clearer than the one obtained for the closed one. Despite this, the double strand DNA was reconstructed at a slightly lower resolution, with the lowest values registered in the region that is detached from the protein core. An explanation for this phenomenon could be attributed to the higher mobility and flexibility of the DNA where it is not anchored to PdxR wHTH domain.

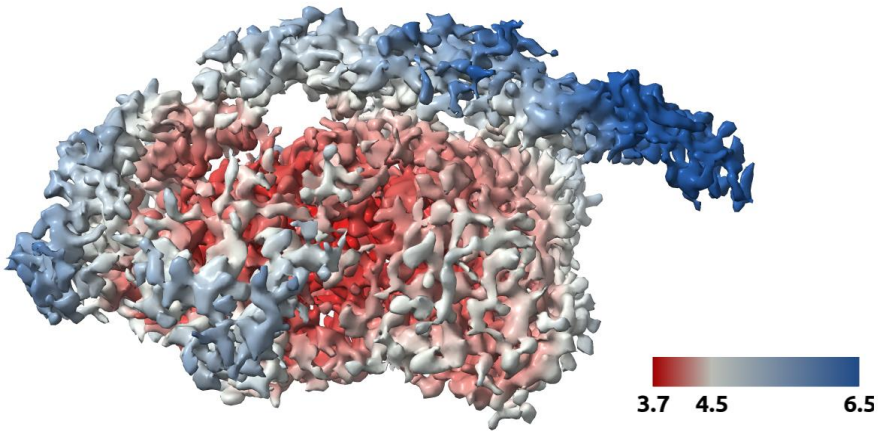


Figure 4.15 Local resolution of the open PdxR-DNA reconstruction. Map is coloured according to the estimated local resolution values.

4.3 Model building and Refinement

All density maps produced by the single particle analysis were sharpened using the Autosharpen tool in Phenix (Liebschner et al., 2019), that improved map contour and connectivity (**Figure 4.16 A and 4.16 B**).

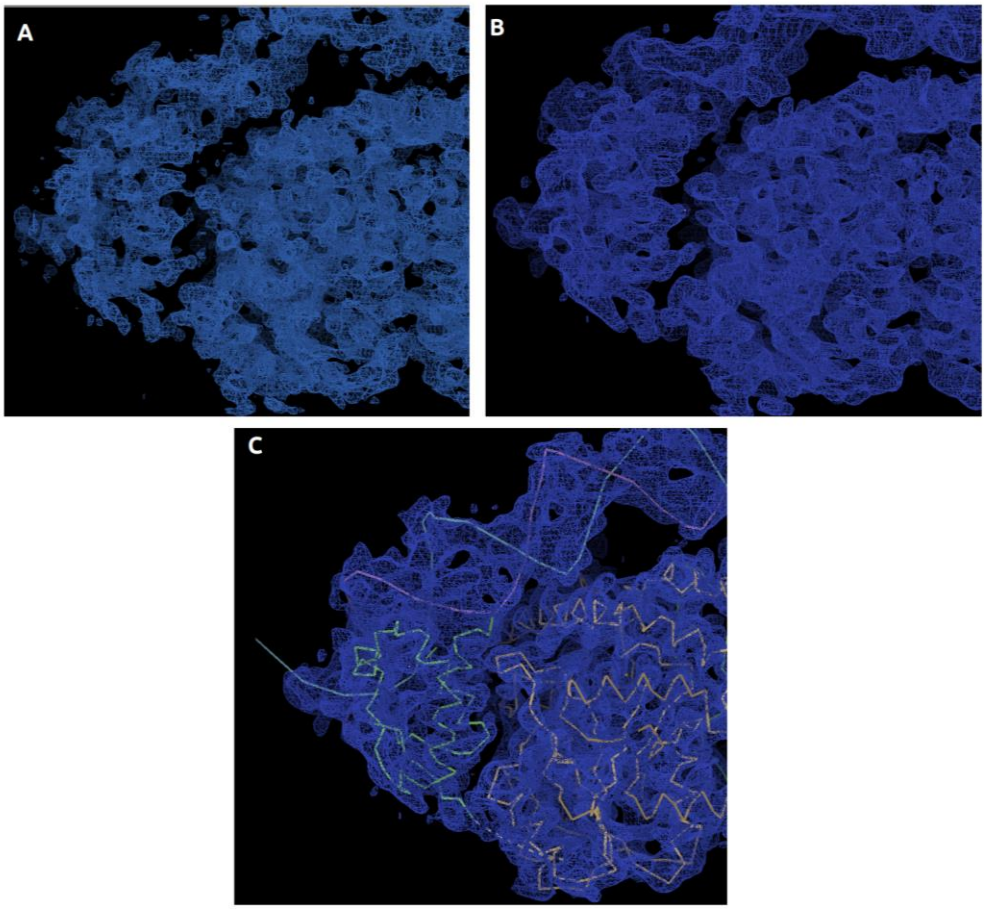


Figure 4.16 PdxR-DNA density map improvement and model fit. Panels A and B show the closed map before and after the sharpening procedure performed by the Autosharpening tool of Phenix. **C:** Atomic model of PdxR bound to double strand DNA modelled into the density map. All maps contoured at 5 σ are visualized in Coot.

Even though we could not rely on previously determined atomic structure of the full-length PdxR, it was possible to manually reconstruct an initial model for PdxR-DNA complex by taking advantage of the experimental and

computational data available on both GabR and PdxR. More in details, the initial atomic model was obtained by assembling the crystal structure of the AAT-like domain dimer in the apo state (data not published) with the wHTH domains obtained by homology modelling (Tramonti et al.,2015, **Figure 1.14**). The cryo-EM map of the C2-closed conformation was initially used to guide the reconstruction of the full-length protein structure. To this purpose, the AAT-like and the wHTH domains were rigid body fitted to the map by using UCSF Chimera (Pettersen et al., 2004). The rigid body fit of the AAT-like domain revealed an incorrect handedness of the reconstructed density map. Therefore, prior to model fit, the handedness of both the closed and the open cryo-EM reconstructions were corrected by using the flip handedness command of UCSF Chimera. Due to the absence of clear density, the linkers connecting the AAT-like domain with the wHTH domains of both dimers were removed from the model. The obtained initial model produced for PdxR is reported in **Figure 4.17**.

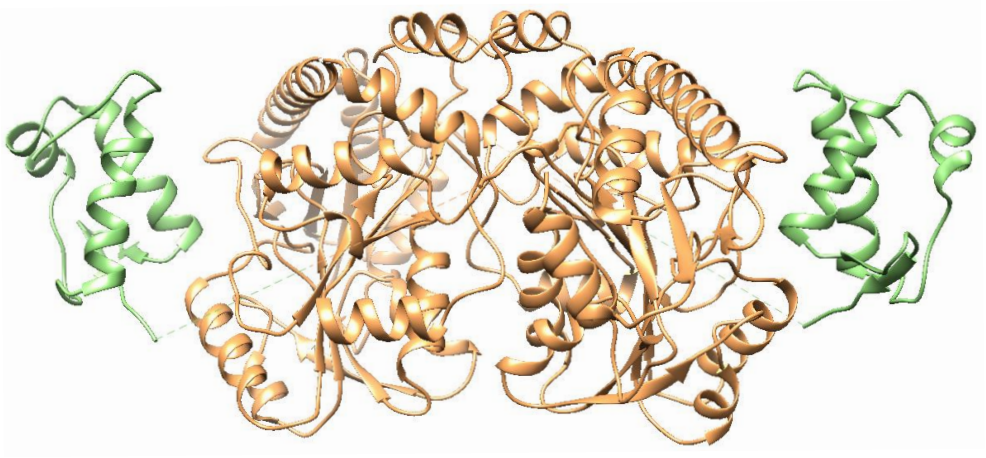


Figure 4.17 PdxR assembled 3D model that was initially used for rigid body fitting cryo-EM maps of the complex with DNA. The crystal structure of the AAT-like domain dimer of PdxR is represented in orange ribbons; the wHTH domains extracted by the homology model are shown in green ribbons.

The atomic model of DNA was initially generated by Coot as a 48-bp B-DNA straight fragment. The rigid body fit of this model to the density map did not produced satisfying fitting given the accentuate curvature defined by the electron density of the DNA, most likely induced by the binding with PdxR. Therefore, the DNA was fitted as follow: 1. the straight fragment of DNA was approximately fitted as a rigid body into the density map; 2. the DNA was refined and partially manually rebuilt, following the major and minor grooves clearly visible in the electron density. A detail of the final fit is showed in **Figure 4.16 C**.

A major problem in DNA fitting was the identification of the correct directionality adopted by the double strand with respect to PdxR dimer. In fact, while the direct and inverted repeats (motifs 1 and 3), targets of the PdxR

wHTH domains, have the same sequence with inverted orientation, the bridging sequence between the two motifs is not palindromic. As a result, the intrinsic asymmetry of the oligonucleotide sequence only allows the DNA to arrange in a unique direction with respect to the protein. Nevertheless, the quality of the DNA in the C2-closed map was not enough to clearly distinguish all the single nucleotide bases of the fragment, jeopardizing a correct assignment of the DNA direction. For this reason, a model for DNA was reconstructed by using the open map, that showed higher resolution features at the level of the protein-anchored portion up to the half of the DNA segment. The DNA building and fitting were guided by the identification of the densities corresponding to tracts of repeated bases (*i.e.,* the A/T rich tracts). As an example, **Figure 4.18** shows a tract of the double strand DNA where the density can only accommodate a repetition of pyrimidine bases in one strand and purine bases on the other, but not vice versa. Following this procedure, in the open map the DNA was fitted with motif 1 bound to the visible wHTH domain, while motif 3 corresponds to the portion of DNA detached from the protein, where the wHTH was missing.

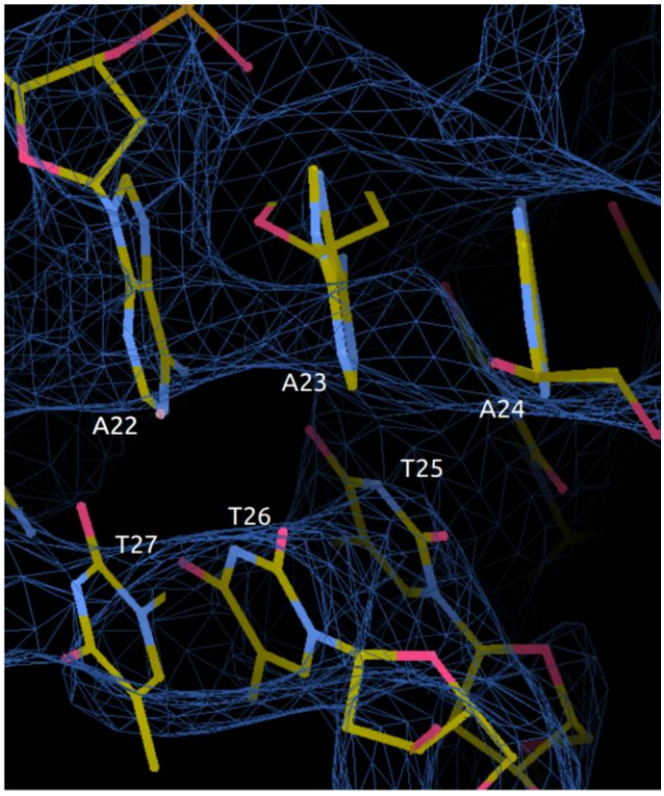


Figure 4.18 Close up view of the electron density map of the open PdxR-DNA conformer where nucleotides of the A/T-rich region were fitted. The thicker tract of the density well accommodates purines, while the thinner is well fitted by pyrimidine bases. The map is contoured at 5 σ and visualized by Coot.

Once all pieces were assembled, the atomic model of PdxR-DNA, without one of the wHTH domain, was rigid body refined against the open map and manually adjusted to better model secondary structural elements and the amino acid side chains where the density allowed manual reconstruction. The reconstructed model of DNA was also used to fit the corresponding density in

the closed map. Rigid body fit, manual building and refinement were carried out to reconstruct the final atomic model of the closed PdxR-DNA complex.

4.4 Cryo-EM analysis of *Bacillus clausii* PdxR in complex with its target DNA revealed the presence of two different conformational states

The cryo-EM structures of PdxR-DNA obtained are consistent with the predicted homology 3D model (Tramonti et al., 2015) and confirmed the domain-swap dimeric arrangement of the protein.

In addition, single particle analysis allowed the *in silico* isolation (and the reconstruction) of the two most abundant populations of the complex showing different conformations. In the closed conformation, PdxR is trapped in its canonical domain-swap arrangement. The simultaneous binding of the two wHTH domains to the double strand DNA compacts the structure of the complex, inducing a pronounced bending of the DNA, that forms an arc of ~ ~114°, and confers an overall C2-symmetry to the PdxR-DNA system. Conversely, in the open conformation, the complex adopts a more relaxed structure, with the DNA tail at motif 3 detached from the protein core and one of the wHTH domains most likely dissociated from the AAT-like domain of the adjacent monomer and not adopting a fixed position, hence its absence from the electron density map (**Figure 4.19**). In the open form, the total angle of curvature of the DNA is about 132°, *i.e.*, ~20° less pronounced than in the closed conformation.

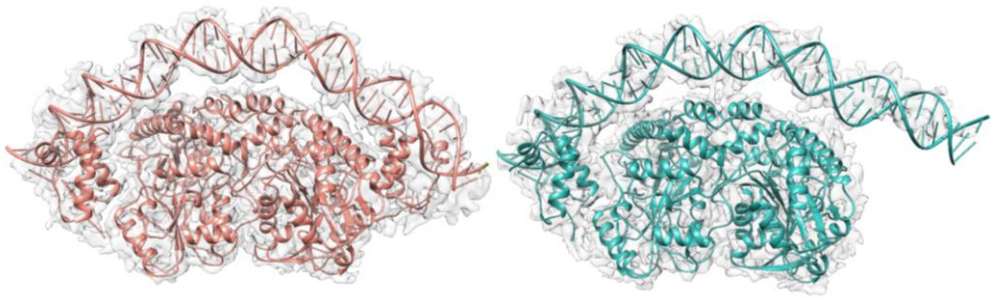


Figure 4.19 The overall structure of PdxR-DNA complex in its closed and open conformations fitted in their relative cryo-EM maps (light grey). Closed and open conformers are represented in red and sea green cartoons, respectively.

4.4.1 PdxR-DNA: closed conformation

The intergenic region target of PdxR is a ~120-bp DNA segment containing *pdxR* and *pdxST* promoters and the TF binding sites. In our cryo-EM experiments we used a 48-bp DNA fragment, covering the two direct and the inverted repeats predicted as the PdxR binding sites (**Figure 4.20**).

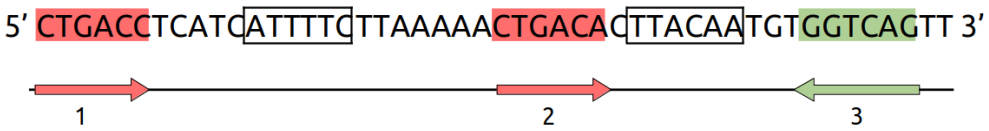


Figure 4.20 Schematic representation of the 48-bp oligonucleotide used as PdxR target DNA. The oligonucleotide comprises the two direct repeats (sequence and arrows highlighted in red) and the inverted repeat (sequence and arrow highlighted in green). The positions of the -10 and -35 hexamers are boxed.

As predicted, in its closed conformation PdxR binds the target DNA sites anchoring the first direct (motif 1) and its palindromic inverted repeat (motif 3) with the two wHTH domains. The DNA molecule is largely bent to allow the simultaneous contact with the PdxR DNA-binding domains, and it spans longitudinally over the protein surface, establishing further contacts with the AAT-like domains on a region that in GabR was called “positive ridge” (Edayathumangalam et al., 2013).

Therefore, PdxR binds its target DNA through three main regions that comprise: the two wHTH domains, in interaction with motif 1 and 3, and the positive ridge of the AAT-like domain that contacts the phosphate groups of the DNA backbone located at the site where the DNA fragment reaches its maximum curvature. Several sites of interaction that we identified in these regions involves PdxR residues also found in GabR and previously predicted to contribute to its DNA binding. **Figure 4.21** reports the sequence alignment of *B.clausii* PdxR and *B. subtilis* GabR, where the common residues involved in DNA binding are highlighted.

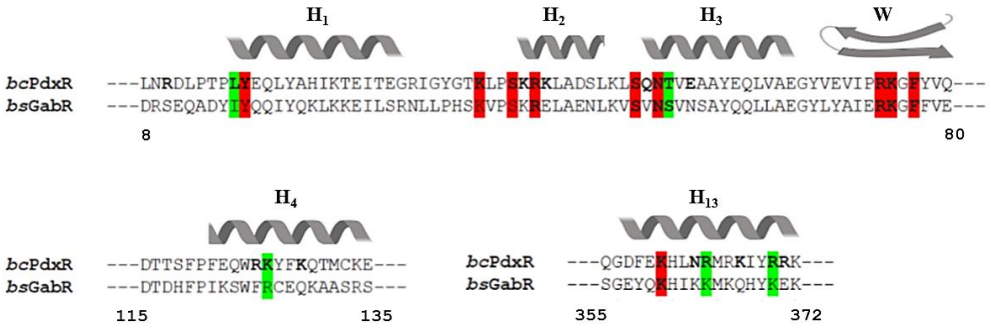


Figure 4.21 Sequence alignment of *B. clausii* PdxR and *B. subtilis* GabR. Partial structure-based sequence alignment of the regions that in PdxR contact the DNA was made using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo). Symbols of secondary structure elements of PdxR are shown on the top and marked with H (helices) and W (wing). Residues in black bold indicate amino acids that interact with DNA in PdxR and are predicted as DNA-binding sites in GabR. Fully conserved residues are displayed on a red background; conservative substitutions are displayed with green background.

A detailed list of the contacts identified in the closed structure is given in **Tables 6.1, 6.2, 6.3 and 6.4** (see Appendix). The most relevant interactions will be here described and discussed.

Similarly to other HTH transcription factors, PdxR recognizes its target DNA by means of its wHTH domain, conserved among the GntR family. The wHTH interacts with the DNA cognate site, *i.e.*, motif 1, by following the typical binding mode described for GntRs: the helices H_2 and H_3, the latter called "recognition helix", invade the DNA major groove, while the wing (W) contacts the minor groove (**Figure 4.22**).

This DNA-binding mode closely resembles the one found in FadR-DNA crystal structure, being the helices and the wing positioned in a conserved orientation with respect to the nucleic acid.

.

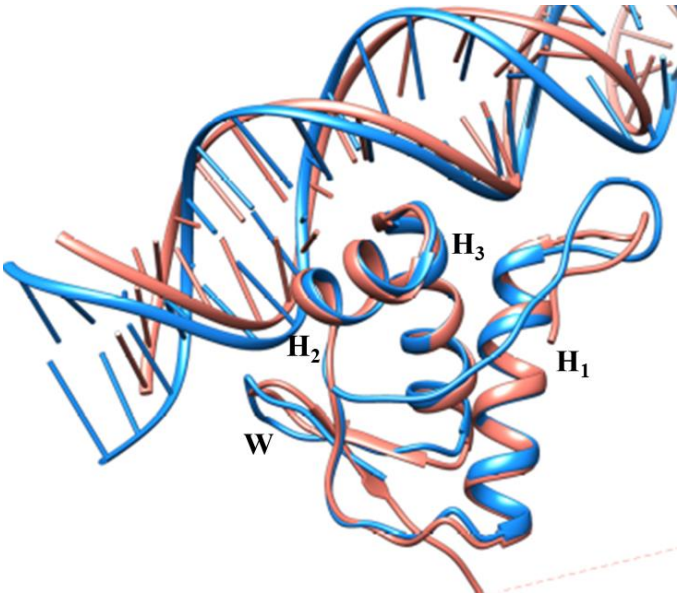


Figure 4.22 Close up view of the wHTH domain in the closed conformer of PdxR-DNA complex (red) superimposed to the FadR-DNA structure (pdb code: 1H9T; blue). The wHTH domain and the DNA molecule of the two structures are displayed in cartoons.

Three residues of the wHTH domain of PdxR were predicted and proposed as involved in DNA interaction, namely Arg43 and Ser52 located on the HTH motif, and Lys75 on the wing. The cryo-EM structure of PdxR-DNA complex confirmed these residues to be involved in the interactions with the consensus sequences at motif 1 and motif 3 (**Tables 6.1 and 6.2**). In addition, other contacts formed by flanking residues, *i.e.*, Lys42 and Lys44 on H_2, Gln53, Ans54 and Glu57 on H_3, Arg74 and Phe77 on the wing, contribute to the interaction with the DNA cognate sequences (**Figure 4.23**).

Therefore, we identified three crucial DNA-binding segments at the wHTH domain that form an extensive network of contacts with specific bases exposed

to the major groove and with sugars and/or phosphates of the backbone exposed to the minor groove formed at the PdxR specific DNA binding sites (**Table 6.1 and 6.2**).

Other amino acids at the wHTH domain contribute to the DNA binding by interacting with the phosphate backbone: Arg10, Leu16, and Tyr17 on H_1, Lys38 and Ser41 on the H_1-H_2 loop.

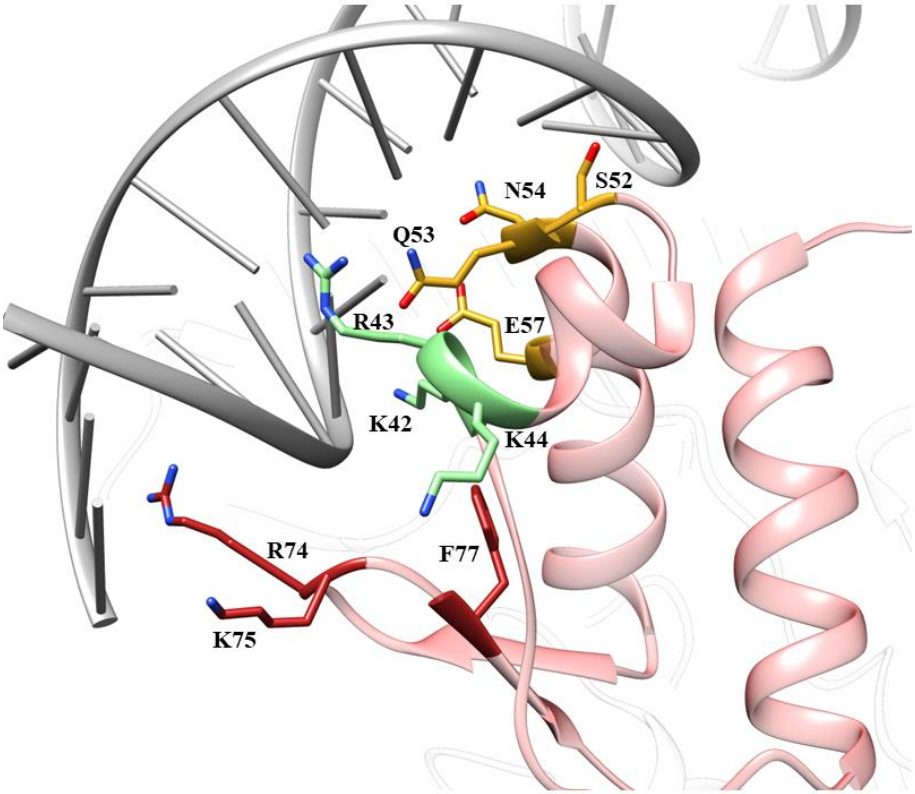


Figure 4.23 DNA-binding residues of the wHTH domain of the closed PdxR-DNA. Amino acids that interact with the DNA are highlighted and labelled. Residues on the helices H_2 and H_3, invading the major groove, are displayed in green and yellow sticks, respectively. Residues on the wing, of which R74 and K75 contact the minor groove, are displayed as red sticks.

Besides the DNA-binding residues on the wHTH domains, the final map reconstructed through single particle analysis showed clear density corresponding to the side chains of positively charged residues distributed at the top of the AAT-like domains which forms electrostatic interactions with DNA phosphate backbone. More specifically, we identified: i) Arg125, Lys126 and Lys129 on H_4 located at the interface between the two large subdomains; ii) Lys360, Asn363, Arg364, Lys367, Arg370, Arg371 at the H_{13} (**Figure 4.24**). These residues span longitudinally over the protein-DNA central interface and are positioned at sites whose polarity is conserved among the members of the MocR subfamily.

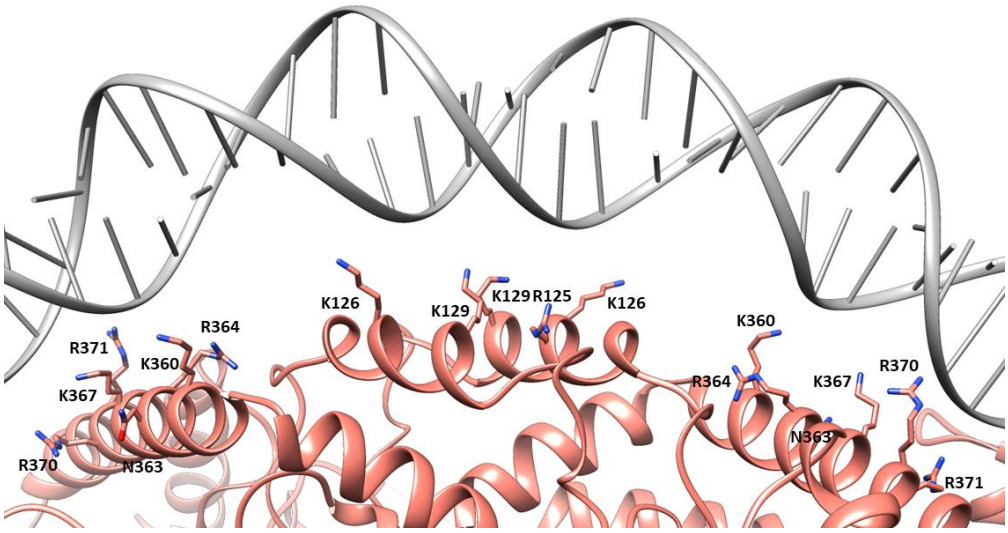


Figure 4.24 Amino acid residues on the AAT-like domains interact with the phosphate backbone of the DNA. Secondary structure elements are represented in red cartoons. DNA-binding residues are depicted in sticks and labelled. DNA is shown as grey cartoon.

Of relevance is the interactions established by Lys129 of both PdxR dimers that allows the long and flexible side chains of these basic residues to enter the narrow minor groove found at the site of maximum curvature of the DNA (**Figure 4.25**). The narrow minor groove is formed as consequence of the pronounced bending of the DNA segment upon PdxR binding. Such deformation, that enhanced the negative electrostatic potential of specific sites of DNA, has been described as a general structural tool that confers specificity to protein-DNA interactions occurring at the level of minor grooves (Rohs et al. 2009). Therefore, the interaction of PdxR at the maximum of the curvature of DNA represents a specific contact even if it is mediated by non-specific interactions. The specificity is provided by a non-canonical geometry adopted by the DNA when it is bound to PdxR.

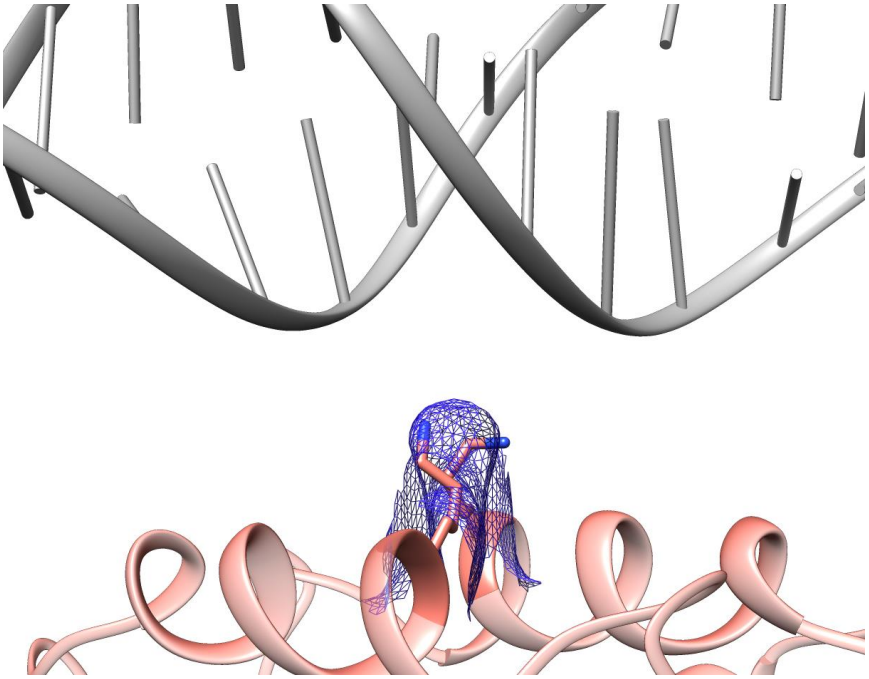


Figure 4.25 Insertion of Lys129 in the DNA narrow minor groove. Lys129 of both PdxR monomers is represented as red sticks and the cryo-EM density is shown as blue mesh, contoured at 5 σ. Secondary structure elements are depicted as light red ribbons. DNA is represented as grey cartoon.

Another structural feature that is worth mentioning is the PLP binding site at the AAT-like domain, that we identified in the PdxR-DNA cryo-EM structure. Indeed, as already mentioned, the PdxR C-terminal domain contains the effector binding pocket, which is lined by two subdomains, the larger one containing the PLP-binding invariant lysine, number 309 in PdxR, that forms a Schiff-base linkage with PLP, and the smaller one that interacts with its α-carboxylate group.

A partially defined electron density was detected nearby Lys309 in both monomers of the closed PdxR-DNA, which is consistent with a molecule of PLP covalently bound to Lys309 (**Figure 4.26**). In the cavity identified, PLP is surrounded by and contacts residues that are conserved in GabR and in fold type I PLP-dependent enzymes (Edayathumangalam et al., 2013; Catazaro et al., 2014). More in details, i) the nitrogen atom of PLP pyridine ring interacts with the side chain of Asp276 on β_3, a conserved residue also found in AATs and in GabR, that maintains the correct protonation state of the cofactor (Edayathumangalam et al., 2013); ii) the hydroxyl group at the C3 of the pyridine ring contacts Tyr202 on the β_5-H_7 loop, Phe247 on the second α-turn and Tyr278 on β_8; iii) the phosphate group is involved in hydrogen bonds or salt bridges with Ser308 on the third α-turn, Arg316 and Ser318 on β_{10}; iv) the imine nitrogen forms an hydrogen bond with Thr112 on the β_3-H_4 loop.

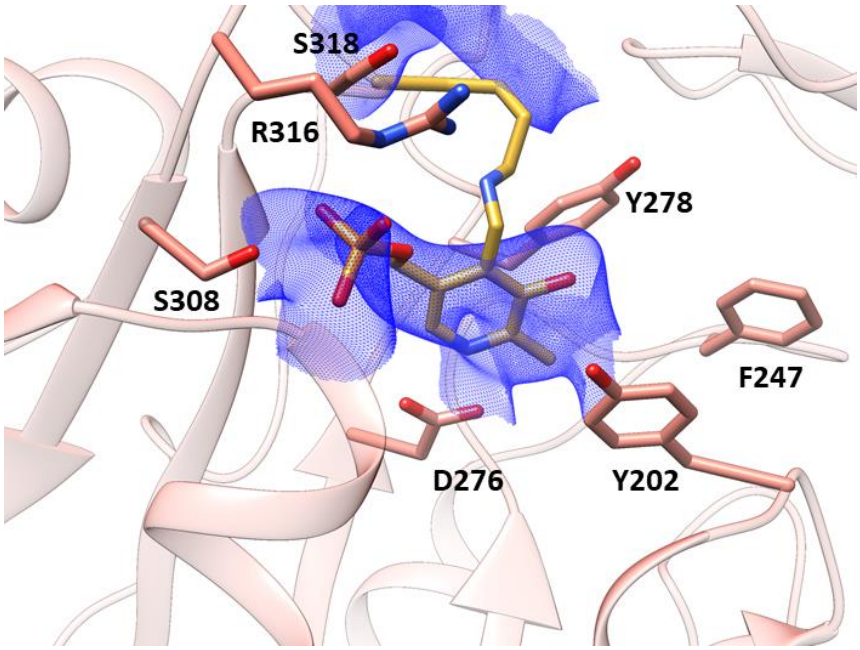


Figure 4.26 PLP bound in the AAT-like domain of the closed PdxR-DNA complex. Close up view of the PLP-binding pocket in the closed conformation. Secondary structure elements are shown in light red. PLP is represented as gold sticks and cryo-EM density (blue mesh, contoured at 5 σ) is shown around the cofactor. Residues involved in the interaction with PLP are shown in stick representation and labelled. For the sake of simplicity, only the PLP-binding pocket of the monomer A is shown.

4.4.2 PdxR-DNA: open conformation

In the open conformation, binding to DNA is achieved through a single wHTH domain that, according to our reconstruction, interacts with the first direct repeat (motif 1). Similarly to what observed in the closed conformation, the wHTH domain follows the very same mode of DNA binding, with H_2 and H_3 helices contacting the major groove and the wing contacting the minor one at the PdxR consensus sequence. Also in this state, an important contribution to

the complex stabilization is provided by contacts established by the basic residues at the ridge of the AAT-like domain of PdxR with the central portion of the DNA, where it experiences the maximum bending. The protein-DNA contacts identified in the open conformer both at the wHTH and at the AAT-like domains are not dissimilar to those found in the closed structure and are listed and detailed **in Tables 6.5, 6.6 and 6.7** (see Appendix).

No density was found for the wHTH domain that in the closed conformation contacts the inverted repeat (motif 3) of the DNA. The absence of such interaction, that provides a strong structural constrain to the final conformation assumed by the double strand DNA, causes the dissociation of one tail of the DNA segment from the AAT-like domain core and changes the angle of the overall DNA curvature of about 20 ° (**Figure 4.27**). The quality of the density map surrounding the protein-detached tail of DNA suffers of the enhanced conformational freedom of the segment, as it becomes lower.

Analogously, we believe that the absence of density defining one of the wHTH domains in the open conformer is due to the intrinsic limitations of single particle analysis in reconstructing highly flexible portions of macromolecules. In a state where only one wHTH domain is bound to DNA, the second wHTH domain may samples various conformations, being endowed with a high degree of flexibility which is conferred by the long linker that connects it to the AAT-like core. Therefore, in the absence of binding to DNA, the inherent dynamics of the linker-wHTH portion, that may explore multiple conformational states, hampered the cryo-EM map reconstruction of that site.

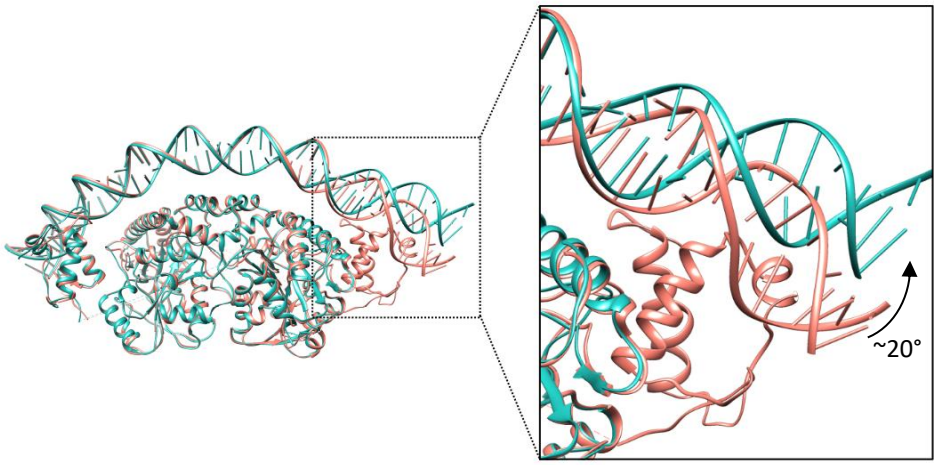


Figure 4.27 Superimposition of the closed and the open conformers of PdxR-DNA complex. Closed and open PdxR-DNA complexes are represented in red and sea green cartoons, respectively. The change in the DNA overall curvature which is due to the absence of the interaction with the second wHTH domain observed in the open complex is highlighted and zoomed in the inset.

This observation is supported by previous studies performed on PdxR and GabR that provide a physiological explanation. It was proposed that the flexible linker allows the wHTH domain to slide away from the AAT-like protein core regulating the interaction of the TF with the target DNA repeats (Tramonti et al., 2015; Milano et al., 2017; Frezzini et al., 2019), governing the transcription mechanism. In fact, the conformational change that might involve the linker-wHTH portion likely governs the binding to the regulatory elements in the *pdxR-pdxST* operon, and subsequently the molecular mechanism of the activation/repression of the transcription.

We can exclude that the open conformer is a result of a partially PLP bound form. Indeed, we found a discrete density in the PLP-binding pocket that

corresponds to one molecule of PLP covalently bound to Lys309 of both dimers (**Figure 4.28**).

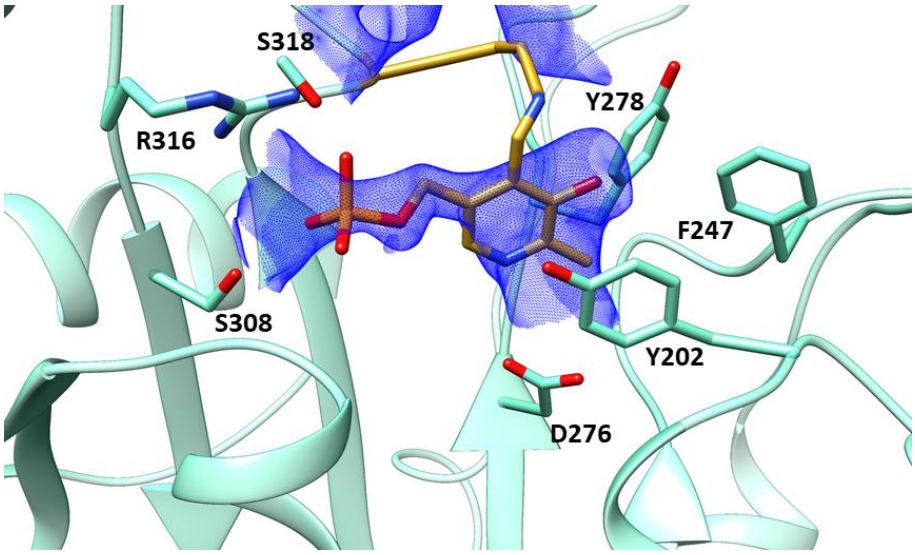


Figure 4.28 PLP bound in the AAT-like domain of the open PdxR-DNA complex. PLP is represented as gold sticks and its cryo-EM density is shown as blue mesh, contoured at 5 σ. Secondary structure elements are represented in sea green cartoons and residues nearby PLP are shown in sticks and labelled. For the sake of simplicity, only the PLP-binding cleft of the monomer A is shown.

4.5 Structural predictions revealed the intrinsic curvature of PdxR target DNA

The control of gene expression relies on the recognition of specific DNA sequences by TFs. A key element underlying the recognition between TFs and their cognate sequences is the so-called base readout (or direct readout; Seeman et al., 1976). The recognition of the unique chemical signatures of the

DNA bases predominantly occurs through the formation of hydrogen bonds and salt bridges between amino acid side chains (Arg, Asn, Gln, Lys, Ser) and base pairs exposing donors and acceptor groups in major and minor grooves, but also nonpolar interaction can be established (Harrison and Aggarwal, 1990). However, the base readout alone is not sufficient for proteins to discriminate DNA target sites. A fundamental role in the protein-DNA recognition mechanism is played by the shape and deformability of the nucleic acid (shape readout or indirect readout). A deviation from a straight B-DNA structure allows proteins to establish contacts also with distal DNA regions, conferring further specificity (Perez-Martin and De Lorenzo, 1997).

It is noteworthy that the shape readout has a significant influence on the base recognition from a protein, since the bases exposed in major and minor grooves are different depending on the conformation of the DNA (Harteis and Scheinder, 2014).

Thus, the stability of a protein-DNA interaction is maximized by the interplay of sequence recognition and conformational flexibility of both partners.

The deformation of the DNA shape can be localized in a specific site or the overall shape of a DNA region can be altered with respect to an ideal B-form.

A local shape variation usually involves a single base pair step, like in DNA kinks, or small DNA tracts with alterations in groove width, such as narrow minor grooves that are generally present in AT-rich sequences (Rohs et al., 2009).

An example of global shape distortion is DNA bending, which produces a curvature on the entire sequence. The curvature of a DNA segment is sequence-dependent since every nucleotide tract has an intrinsic propensity to bend in a certain direction (Hagerman, 1990). For example, contiguous A or T tracts are

known to generate curved structures, but also dinucleotides like AG or GC can contribute to alter DNA linearity (Bolshoy et al., 1991).

DNA bendability is not a static phenomenon and its interaction with proteins, such as TFs, relies on a certain degree of conformational dynamics. In fact, if a sequence-dependent preformed conformation of a DNA fragment is necessary for protein recognition, it can be further distorted and stabilized by the binding of a protein, in a way that it optimizes the exposure of the interacting surfaces of both partners (Rohs et al., 2009). This is a crucial mechanism that TFs exploit to exert their regulatory function.

The regulation of gene expression requires the correct assembly of the molecular machinery involved in transcription and both the intrinsic and protein-induced bendability of DNA plays a crucial role in the achievement of the adequate geometry of nucleoprotein complexes. In fact, the correct architecture of promoter regions is essential for TFs to recognize and discriminate the target sequence, to recruit interacting proteins and favour the opening of the DNA double helix (Perez-Martin and De Lorenzo, 1997; Van Der vliet, 1993).

To examine the role of the shape readout in PdxR-DNA recognition, we studied the intrinsic bendability of the 48-bp fragment by means of an *in silico* analysis performed by a DNA curvature analysis tool (C. Gohlke, www.lfd.uci.edu/~gohlke/dnacurve/). This prediction tool generates a 3D model of DNA calculating its global structure from the nucleotide sequence. Therefore, we provided the DNA sequence, and we chose the algorithm based on the AA wedge model to calculate a 3D structure showing the bendability of the fragment when it is not bound to the protein.

The resulting 3D-model of DNA revealed some extent of intrinsic propensity to bend of the fragment, displaying a slight overall bending that brings the two

cognate sites (motifs 1 and 3) toward each other (**Figure 4.29 A)**. In the unbound state, the total angle of curvature of the DNA is about 165°. It is possible to identify a region in which the curvature is more accentuated, which corresponds to an adenine-thymine rich tract (5' TTTTCTTAAAAA 3'), which, as already mentioned, is known to be a curvature-inducing feature (Hizver et al., 2001). Alternative 3D DNA structures were also generated, following algorithms based on different models (**Figure 4.29 B**). Nevertheless, all the resulting structures curve in the same direction, although showing different extent of bending.

Therefore, combining the information obtained by the cryo-EM structures of PdxR-DNA complex with the structural prediction of the inherent bending of the 48-bp oligonucleotide, it turns out that when PdxR binds the target DNA, it induces a conformational change in the double helix that imposes a further bending to its intrinsic curvature (**Figure 4.30**).

The specific interaction established between one wHTH domain of PdxR and motif 1 of the bent target DNA gives rise to the formation of an initial protein-DNA complex (open conformation) that further compacts when two wHTH domains simultaneously bind to motif 1 and 3 of the cognate DNA (closed conformation; **Figure 4.30 B and D**).

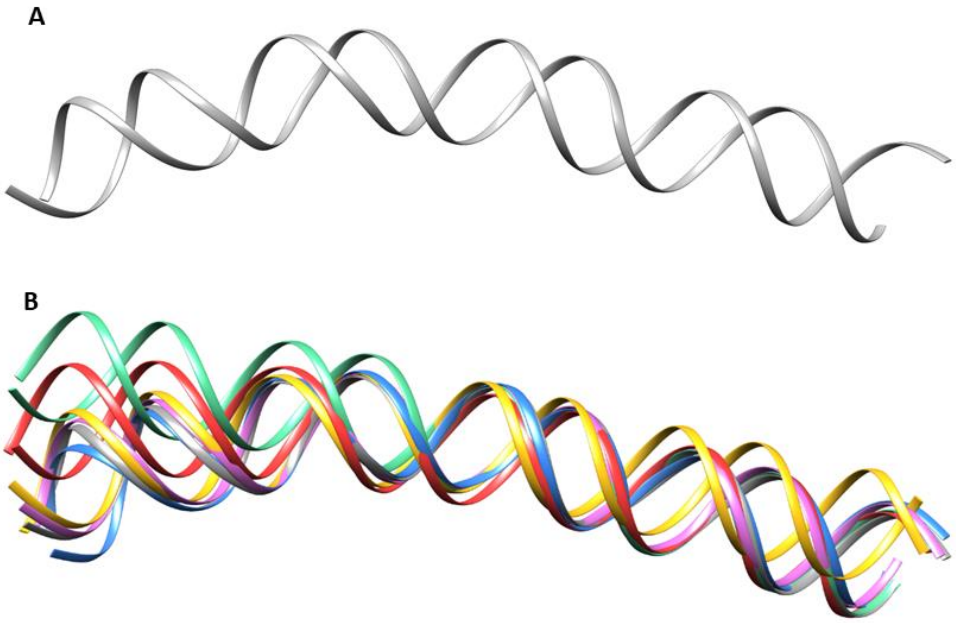


Figure 4.29 3D structures of the PdxR target DNA backbone generated using the DNA curvature analysis tool (C. Gohlke, www.lfd.uci.edu/~gohlke/dnacurve/). A: structural prediction of the DNA based on the AA wedge model. B: AA wedge-based 3D structure (grey) is superimposed to alternative structures generated using the following models: Bolshoy and Trifonov (blue), Cacchione & De Santis (magenta), Calladine & Drew (green), DNAse I consensus (yellow), nucleosome positioning (red).

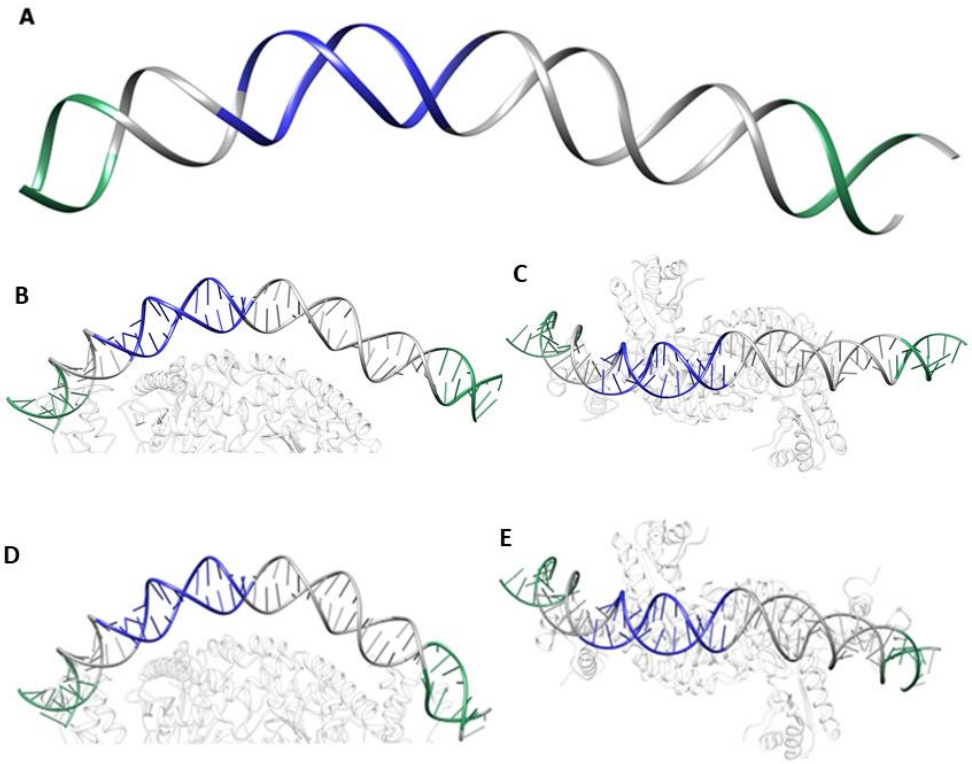


Figure 4.30 The curvature of PdxR target DNA. A: The predicted curvature of the 48-bp DNA containing the target sequences of PdxR. The 3D structure of the DNA backbone was generated using the AA wedge model provided by the DNA curvature analysis online tool (C. Gohlke, www.lfd.uci.edu/~gohlke/dnacurve/). B: Highlight on the DNA molecule in the PdxR-bound form (front view), as it was built in the open conformer model. C: Top view of the open conformer structure, showing the overall shape induced by the binding of one wHTH domain. D: Front view of the DNA in the closed PdxR-DNA structure. E: Top view of the closed PdxR-DNA structure. Double strand DNA are depicted in grey cartoons. PdxR cognate binding sites are highlighted and coloured in green; the AT-rich region of the fragment is coloured in blue. Protein is represented in white transparent cartoon.

It is possible that the conformational change induced by binding of the first wHTH domain of PdxR to DNA, that increases its curvature with respect to the unbound state of about 30°, might facilitate the anchoring of the second wHTH domain to the target inverted repeat. In this configuration, the simultaneous binding of the two wHTH domains can only occur on opposite sites, as observed in the closed structure, given the head-to-tail disposition of the two PdxR monomers. As a consequence, the DNA further bends of ~ 20° and slightly curves following the opposite position of the wHTH domains. It assumes an "S" shape, which is visible examining the top view of the protein-DNA closed complex. In the open conformer, a similar arrangement is not occurring, being the DNA anchored to one wHTH (**Figure 4.30 C and E**).

A local alteration of the canonical DNA shape is represented by the minor groove narrowing, which is correlated with the presence of AT-rich tracts (Rohs et al., 2009). This phenomenon is a direct consequence of the overall deviation from an ideal B-DNA form, typical of the PdxR target DNA in the unbound form. The predicted structural model revealed the presence of some narrow minor grooves around the site of maximum curvature which is AT-rich. Only one of them is exposed to the concave side of the DNA where PdxR binds. As discussed in paragraph 4.4.1, this atypical minor groove confers specificity to that site and constitute a key contact for lysines n. 129 at the interface of AAT-like domains (**Figure 4.25**). Notably, this local alteration is present even in the unbound state, as determined from the structural prediction. Indeed, we estimated that the angle of the groove changes of about 10° with respect to an ideal straight form. This observation reinforces the role of the DNA conformation in PdxR binding, representing an example of the shape readout in protein-DNA recognition.

To summarize, the DNA target of PdxR possesses an intrinsic and sequence-dependent propensity to bend. Once the first wHTH domain binds motif 1, inducing a conformational change to the DNA, such propensity could favour the reorienting of motif 3 in the position that allow the recognition of the specific bases by the second wHTH domain, imposing further curvature.
With the aim to prove the role of base and shape readout in PdxR-DNA binding mechanism, we produced modified forms of the 48-bp fragment where both the target sequences and the overall shape of the segment were altered. The change in PdxR binding capability was analysed.

4.6 The DNA sequence and shape dictate the binding specificity with PdxR

In order to understand the relevance of sequence specificity and DNA conformation in the mechanism of recognition of PdxR target sites, we designed different fragments that introduce variations in the target DNA. Two types of DNA were designed, i) to specifically alter the sequence of the two inverted cognate sites bound by PdxR, ii) to disrupt the intrinsic bendability of the fragment. The first sequence produced contains modified nucleotides at motifs 1 and 3, to impair the binding to the target direct and inverted repeat (M13). To exclude that the substitution in the cognates binding repeats introduced alterations in the DNA overall shape, a 3D structure of the M13 sequence was calculated (C. Gohlke, www.lfd.uci.edu/~gohlke/dnacurve/).
Two DNA sequences were designed to alter the intrinsic bendability of PdxR target DNA on the basis of the predicted models. The nucleotide sequence bridging the two repeats was modified in one mutant to achieve the maximum curvature possible with the 48-bp fragment of interest. In the second DNA

sequence, the nucleotide substitutions introduced aimed at totally abolishing the DNA curvature, generating a straight B-form. The two fragments were called “bent” and “straight”, respectively (**Figure 4.31 A and B**).

All DNA sequences are reported in **Table 3.2 (**Materials and Methods**)**. Their predicted 3D structures are shown in **Figure 4.31**.

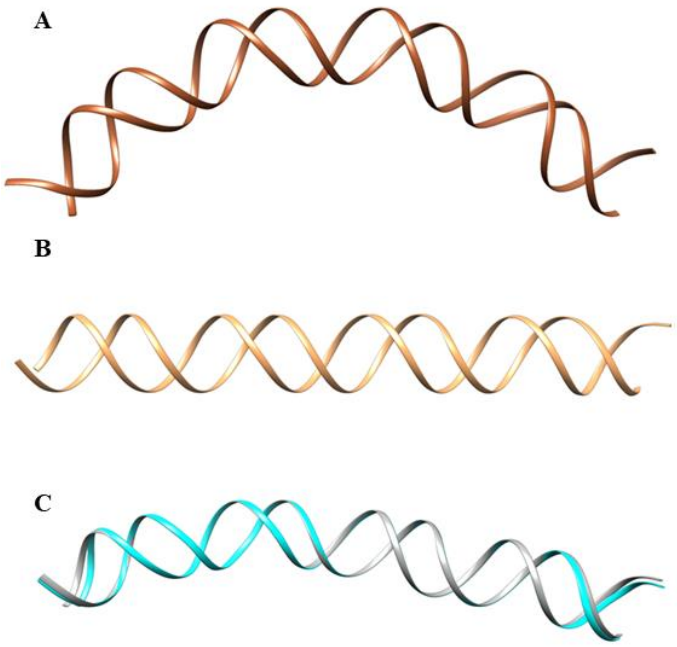


Figure 4.31 Predicted 3D structures of the mutated sequences of the PdxR target DNA in which alterations in the overall curvature (A and B) and in the binding motifs (C) have been introduced. A: structure of the bent mutant (maximum curvature achievable); B: straight mutant. C: the superimposition of the wt DNA form (grey) and of the M13 mutant confirmed that the mutations did not alter the overall conformation of the DNA.

The change in affinity of the DNA modified forms for PdxR compared to the wild type fragment (wt) was measured using Electrophoretic Mobility Shift Assays (EMSA).

EMSA were performed with a DNA concentration of 10 nM (wt and the three mutants) and varying the concentration of purified PdxR from 25 nM to 800 nM. The addition of PdxR to the DNA determines the formation of an upper band corresponding to a protein-DNA complex (black arrows in **Figure 4.32**), whose density increases with protein concentration.

In all the cases, the modifications introduced at the cognate sites and in the interspacing region of the fragment impair to some extent the PdxR-DNA binding compared to the wt (K_{Dapp} = 600±38 nM), as highlighted by the decrease in complex affinity. The K_{Dapp} measured for the M13 mutant is 1000±110 nM, while is 910±71 nM and 900±82 nM for the bent and the straight form, respectively.

Taken together, these results indicate that both the DNA base and the shape readout influence the mechanism followed by PdxR to recognize its DNA target sites, even if none of the designed mutant largely reduce or totally abolish the protein-DNA interaction.

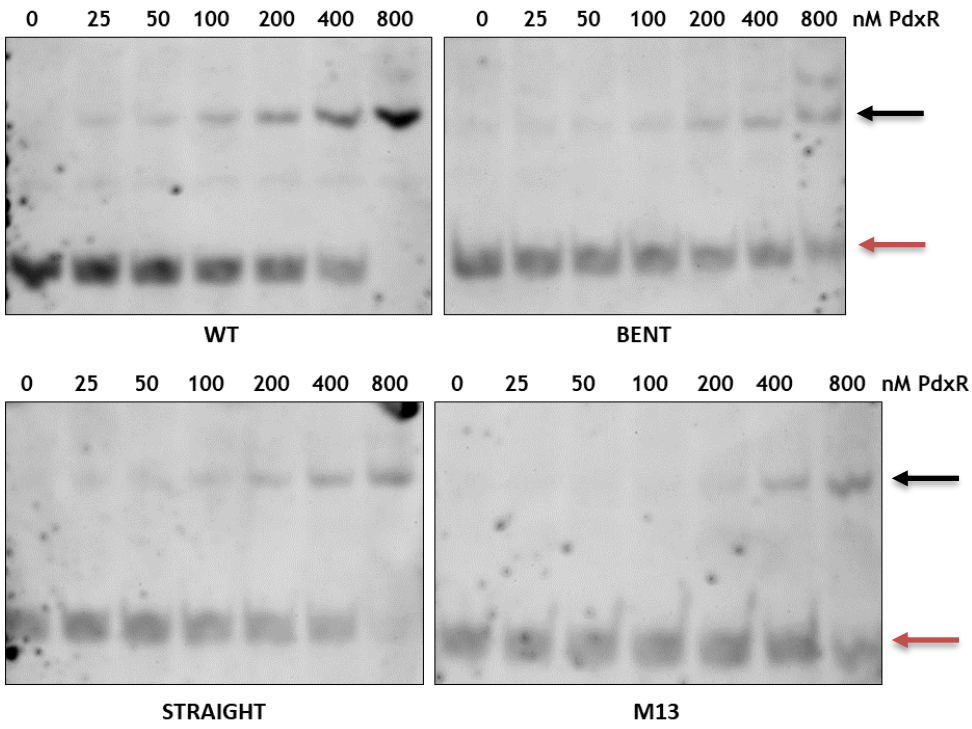


Figure 4.32 EMSA analyses of the interaction between PdxR and the different forms of the target DNA. EMSA was carried out with increasing protein concentrations and with a fixed DNA concentration (10 nM). For each PdxR form, a single representative experiment is shown. The table on the right reports the $K_{D\ app}$ values estimated from at least three independent experiments. Red and black arrows indicate the bands corresponding to free DNA and PdxR-DNA complex, respectively.

Of note, at high PdxR concentrations, and especially when complexed with the "bent" mutant, a slower band is visible which may be due to the formation of higher order complexes.

5. Conclusions and Future Perspectives

The regulation of gene expression in bacteria is the result of several intricate and finely tuned mechanisms to produce the right molecular machinery for a plethora of processes encompassing developmental plasticity, homeostasis and environmental adaptation.

Transcription factors (TFs) represent the main players in these networks, since in bacteria the control of gene expression relies mainly on transcriptional modulation.

The target of this work belongs to the MocR subfamily which is part of the GntR family of TFs. MocR represents an understudied group of regulators involved in pyridoxal 5'-phosphate (PLP) and amino compound metabolic pathways, broadly distributed among Eubacteria.

The molecular mechanism at the basis of MocR transcriptional regulation has not been clarified yet, given the absence of structural information of TF-DNA complexes. The member of the MocR family which has been better characterized is GabR from *Bacillus subtilis.* However, the only structure available for a GntR TF in complex with DNA is the one of FadR from *Escherichia coli* which is not a member of the MocR group.

Our studies aimed at filling this gap, solving the cryo-EM structure of *Bacillus clausii* PdxR in complex with its target DNA. The MocR regulator PdxR

intervenes in the *de novo* biosynthesis of PLP, acting on genes encoding the PLP-synthase subunits, as well as its own gene. The target regulon encompasses the *pdxR-pdxST* intergenic region (~120 bp) and its characterization highlighted the presence of two direct and one inverted repeat representing the PdxR binding sites. The affinity for the repeats changes according to the presence or the absence of bound PLP, which is the PdxR effector ligand that regulates its transcriptional activity.

Cryo-EM structure of PLP-bound PdxR revealed that the winged helix-turn-helix (wHTH) and the aspartate aminotransferase (AAT)-like domain are associated in a domain-swap homodimeric assembly, confirming the prediction based on the crystal structure of the homologous GabR.

We observed two different conformational states of PdxR-DNA complex, a closed and an open one. In the "closed" one, the domain-swap homodimeric assembly simultaneously binds the first direct repeat and the inverted one (also called motifs 1 and 3 in this thesis) through the wHTH domains, whereas in the "open" state only one of the wHTHs is bound to DNA. The second wHTH does not have a fixed position and it is not present in the cryo-EM map. Consequently, the inverted repeat is not bound to PdxR and it shows an enhanced mobility.

The cryo-EM maps were reconstructed at 3.7 Å (closed state) and 3.9 Å resolution (open state), with a higher local resolution at the level of the AAT-like core and of the region of protein-DNA contacts that allowed us to model PLP as an internal aldimine and to identify PdxR amino acids and nucleotide bases which are involved in protein-DNA interaction.

We observed that the wHTH domains contact the DNA grooves at motifs 1 and 3 in a conserved manner, as already noticed in the structure of FadR-DNA complex, and we identified several residues forming specific interactions with

DNA bases. Moreover, on the surface of the AAT-like dimer are present residues that anchor the DNA to the protein, establishing electrostatic contacts with the phosphate groups, as predicted in Tramonti et al. in 2015. The consistency of our cryo-EM structures with data reported in literature validates our 3D models.

Both the closed and open structures revealed that when the DNA is bound to PdxR it induces a pronounced curvature on the nucleic acid.

The *in silico* analysis of the DNA sequence highlighted that the PdxR target fragment has an intrinsic bendability in the unbound form, conferred by the AT-rich tracts located at the bridging segment between the first direct repeat and the indirect repeat.

To investigate the role of the DNA sequence and conformation in the complex formation and stability, a sequence with modifications of motifs 1 and 3 and two different DNA sequences with altered bendability (one straight and one bent), were produced and analysed for their affinity toward PdxR by electrophoretic mobility shift assay (EMSA). The lower stability of the PdxR-DNA complex calculated using oligos with altered DNA sequences highlighted that, although complex formation was not abolished, the mutations at consensus sequences and variation of the original DNA curvature altered the overall stability of the complex, suggesting that both base recognition and shape readout orchestrate the specific PdxR-DNA interaction.

The contribute of both base and shape readout to protein-DNA recognition was previously inferred from functional data for *Bacillus subtilis* GabR, and a kinetic model for TF-DNA binding was proposed (**Figure 5.1**). Al-Azyoud et al. hypothesized an initial step in which GabR binds the first repeat of its target DNA with a single wHTH domain. This step was supposed to be only driven by base readout interactions, independently from the DNA curvature.

Following this model, the binding of the second wHTH likely occurs when the DNA assumes a conformation which correctly orients the second repeat, forming a more stable complex still endowed with a certain degree of dynamics.

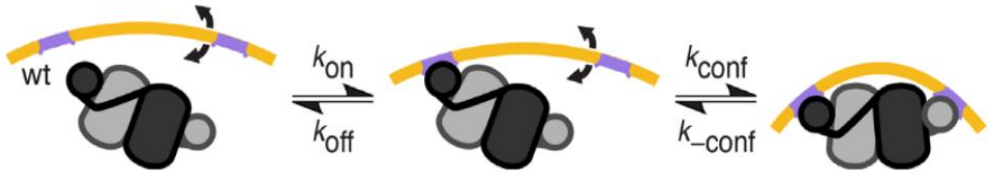


Figure 5.1 Schematic representation of the kinetic model proposed for GabR-DNA interaction. In the first step one wHTH domain binds the first DNA repeat independently of the bridging sequence. The second step depends on the conformational variations of the DNA and occurs when the conformation adopted by DNA allows the second repeat to be oriented correctly for the binding of the free wHTH domain (Adapted from Al-Zyoud et al., 2016).

Our cryo-EM data on PdxR provides the first direct structural evidence for this model, since we captured an open conformation with a wHTH detached by the inverted repeat.

Thus, the open conformer should represent an intermediate state preliminary to PdxR full binding to the first direct and the inverted repeat (closed state). This conformer is favored in the PLP-bound form of PdxR, when the transcription of the pdxST genes needs to be repressed.

Our data also indicate that in the mechanism of PdxR-DNA binding a conformational change is experienced by both the DNA and the protein.

One advantage of using cryo-EM is the possibility to detect more than one conformation of the macromolecule or complex of interest in a context, *i.e.*, the vitrified ice, that is representative of the distribution of states in solution. Indeed, particle classification allows to isolate and treat different states separately. This advanced methodology allows us to study flexible and sometimes transient complexes that usually are not amenable to crystallization. Our results represent a fundamental step to clarify the molecular mechanism of the MocR TFs, since we describe the PdxR-DNA interaction and its dynamics. These structures represent a model for the entire subfamily, given the predicted structural homology of the different members.

We are in the process of extending the single-particle analysis, including data recently collected on the 300kV Krios microscope at the ESRF facility, to explore the possibility to gain higher-resolution and more detailed reconstructions. Moreover, in collaboration with Prof. Anita Scipioni and Prof. Roberto Contestabile of Sapienza University, we plan to design new DNA sequences inserting modifications in different groups of nucleotides of the target sequences and in the bridging sequence to locally/globally alter DNA conformation. We also plan to mutate PdxR residues on the AAT-like domain to evaluate the role of the positively charged surface on the protein core in DNA binding, coupled to the DNA bending.

A clear understating of the regulatory activity of PdxR and more in general of MocR may shed light on different interesting aspects of bacterial metabolism and pathogenicity, since this group of TFs are known to be involved in the regulation of many metabolic pathways, such as host adaptation by bacteria. Inhibition of vital mechanisms could block bacterial proliferation and the development of latent infections. Therefore, systems regulating the production

of vitamin B_6, such as PdxR, may be promising drug targets for the treatment of bacterial infections.

6. Appendix

Table 6.1 Closed PdxR-DNA contacts on the wHTH domain, monomer A

PdxR	Group	Location	**DNA**	Group	Location	Distance (Å)
Arg10	NH1	turn	A39’	PO_4	backbone	3.6
	NH2	turn	A39’	PO_4	backbone	3.6
	N	turn	C5	PO_4	backbone	4.4
	O	turn	C5	PO_4	backbone	3.3
Leu16	N	H_1	A39’	PO_4	backbone	3.5
Tyr17	OH	H_1	G38’	PO_4	backbone	3.6
	ring	H_1	A39’	PO_4	backbone	3.9
Lys38	NZ	turn	A4	PO_4	m	4.0
Ser41	OG	turn	A4	PO_4	M	3.8
	OG	turn	C5	PO_4	M	4.2
Lys42	NZ	H_2	C6	PO_4	M	3.1
	NZ	H_2	C6	ring	M	3.7
Arg43	NE	H_2	A4	N7	M	3.0
	NE	H_2	A4	N6	M	3.7
	NE	H_2	C5	N4	M	3.0
	NH2	H_2	A4	N6	M	3.4
	NH2	H_2	G44’	O6	M	3.4
	NH2	H_2	C5	N4	M	2.8
Lys44	NZ	H_2	A4	PO_4	M	3.9
Ser52	OG	turn	T40’	PO_4	M	3.2
	OG	turn	G41’	PO_4	M	4.4
Gln53	NE2	H_3	C6	N4	M	4.0
	OE1	H_3	C5	N4	M	3.2
Asn54	ND2	H_3	T40’	O4	M	3.3
	ND2	H_3	G41’	O6	M	2.7
	ND2	H_3	C8	N4	M	3.7
	OD1	H_3	A42’	N6	M	3.7
	OD1	H_3	T7	O4	M	3.6
Glu57	OE1	H_3	C6	N4	M	4.2
Arg74	NE	wing	A47’	O4’	m	3.1
	NH2	wing	C46’	O3’	m	3.5
	NH1	wing	C46’	O4’	m	3.8
Lys75	NZ	wing	A47’	N3	m	5.0
	NZ	wing	G48’	O4’	m	4.2
	N	wing	C5	PO_4	m	4.4
	O	wing	C5	PO_4	m	3.3
Phe77	ring	wing	C5	PO_4	M	4.0

The symbol ’ indicates the 3’ → 5’ strand of the DNA
M = major groove
m = minor groove
backbone = edge between major and minor groove

Table 6.2 Closed PdxR-DNA contacts on the wHTH domain, monomer B

PdxR	Group	Location	**DNA**	Group	Location	Distance (Å)
Leu16	N	H_1	G39	PO_4	backbone	4.0
Tyr17	OH	H_1	T38	PO_4	backbone	5.0
	N	H_1	G39	PO_4	backbone	4.8
	ring	H_1	G39	PO_4	backbone	3.1
Lys38	NZ	turn	T4’	PO_4	backbone	4.2
Ser41	OG	turn	T4’	PO_4	m	2.9
Lys42	NZ	H_2	A6’	PO_4	M	4.8
Arg43	NH2	H_2	G5’	PO_4	M	3.6
	NH2	H_2	G5’	O5’	M	4.0
	NH2	H_2	G5’	O4’	M	3.2
	NH2	H_2	G5’	N2	M	3.5
	NH1	H_2	G5’	PO4	M	3.1
	NH1	H_2	G46	N2	M	3.8
	NH1	H_2	T4’	O2	M	2.8
	NH1	H_2	A45	N3	M	3.0
	NE	H_2	G5’	O4’	M	3.0
	NE	H_2	G5’	O5’	M	3.5
	NE	H_2	G5’	PO_4	M	3.6
	NE	H_2	T4’	O2	M	4.1
	NE	H_2	A45	N3	M	3.6
	NE	H_2	G5’	ring	M	3.5
Lys44	NZ	H_2	T4’	PO_4	M	3.8
	NZ	H_2	C3’	PO_4	M	3.9
	NZ	H_2	C3’	O5’	M	4.0
Ser52	OG	turn	T40	PO_4	M	3.5
Gln53	OE1	H_3	A6’	N7	M	3.1
	OE1	H_3	G5’	N7	M	4.3
	NE2	H_3	A6’	PO_4	M	3.7
Asn54	OD1	H_3	C7’	N4	M	3.1
	OD1	H_3	A6’	N7	M	3.8
	OD1	H_3	A6’	N6	M	5.0
	ND2	H_3	C7’	N4	M	3.3
	ND2	H_3	T40	O4	M	3.8
	ND2	H_3	G39	N7	M	3.8
Thr55	OG1	H_3	G39	PO_4	M	3.6
Glu57	OE2	H_3	A6’	O5’	M	3.3
	OE2	H_3	A6’	PO_4	M	3.6
	OE1	H_3	A6’	N7	M	3.9
	OE1	H_3	A6’	N9	M	4.3
Arg74	NE	wing	T47	O3’	m	3.3
	NH2	wing	T47	PO_4	m	5
Lys75	NZ	wing	G5’	PO_4	m	3.6
	NZ	wing	A6’	PO_4	m	4.1

The symbol ’ indicates the 3’ → 5’ strand of the DNA
M = major groove
m = minor groove
backbone = edge between major and minor groove

Table 6.3 Closed PdxR-DNA contacts on the AAT-like domain, monomer A

PdxR	Group	Location	**DNA**	Group	Location	Distance (Å)
Arg125	NH1	H_4	T26	PO_4	backbone	4.8
Lys126	NZ	H_4	A28	PO_4	M	4.4
Lys129	NZ	H_4	T26	PO_4	m	4.9
Lys360	NZ	H_{13}	A16’	PO_4	M	4.7
Asn363	ND2	H_{13}	A16’	PO_4	m	4.5
Arg364	NH2	H_{13}	A17’	PO_4	M	4.7
Lys367	NZ	H_{13}	A17’	PO_4	m	4.6
Arg370	NH1	H_{13}	A37	PO_4	backbone	3.5
	NH2	H_{13}	A37	PO_4	backbone	3.2
Arg371	NH1	H_{13}	A36	PO_4	backbone	3.6
	NH2	H_{13}	A36	PO_4	backbone	2.9

Table 6.4 Closed PdxR-DNA contacts on the AAT-like domain, monomer B

PdxR	Group	Location	**DNA**	Group	Location	Distance (Å)
Lys126	NZ	H_4	T28’	PO_4	M	3.7
Lys129	NZ	H_4	T27’	PO_4	m	3.7
Lys360	NZ	H_{13}	T16	PO_4	M	2.8
Asn363	ND2	H_{13}	T16	PO_4	m	4.3
Arg364	NH1	H_{13}	C17	PO_4	M	3.8
Lys367	NZ	H_{13}	A36’	O3’	m	3.5
Arg370	NH1	H_{13}	T37’	PO_4	M	3.0
Arg371	NH1	H_{13}	A36’	PO_4	backbone	3.0
	NH2	H_{13}	A36’	PO_4	backbone	3.5

The symbol **’** indicates the 3’ → 5’ strand of the DNA
M = major groove
m = minor groove
backbone = edge between major and minor groove

Table 6.5 Open PdxR-DNA contacts on the wHTH domain, monomer A

PdxR	Group	Location	**DNA**	Group	Location	Distance (Å)
Leu16	N	H_1	A39’	PO_4	backbone	3.1
Tyr17	OH	H_1	G38’	PO_4	backbone	5.0
	ring	H_1	A39’	PO_4	backbone	5.0
Lys38	NZ	turn	A4	PO_4	m	5.0
Ser41	OG	turn	A4	PO_4	M	4.0
	OG	turn	C5	PO_4	M	3.5
Lys42	NZ	H_2	C5	PO_4	M	3.1
	NZ	H_2	C5	ring	M	3.9
	NZ	H_2	C6	PO_4	M	4.6
	NZ	H_2	C6	ring	M	3.9
Arg43	NH2	H_2	A4	N6	M	5.0
	NH2	H_2	C5	N4	M	2.6
	NH2	H_2	A4	N7	M	4.5
	NH1	H_2	G44’	O6	M	4.8
	NH1	H_2	G43’	O6	M	4.3
Lys44	NZ	H_2	A4	PO_4	M	3.0
	NZ	H_2	G3	PO_4	M	4.2
Ser52	OG	turn	T40’	O4	M	4.4
	OG	turn	G41’	O6	M	4.4
	OG	turn	G41’	N7	M	2.7
Gln53	NE2	H_3	T7	O4	M	2.9
	OE1	H_3	C6	N4	M	2.9
	OE1	H_3	A42’	N6	M	2.8
Asn54	ND2	H_3	T40’	O4	M	3.6
	ND2	H_3	G41’	O6	M	2.6
	ND2	H_3	C8	N4	M	3.4
	ND2	H_3	A9	N6	M	4.4
	ND2	H_3	A42’	N6	M	4.2
	ND2	H_3	T7	O4	M	4.4
	OD1	H_3	A39’	N6	M	5.0
Glu57	OE1	H_3	T7	O4	M	4.8
	OE1	H_3	C6	N4	M	4.6
Arg74	NE	wing	C5	O4’	m	4.9
	NH2	wing	C46’	O3’	m	4.2
	NH2	wing	C46’	O4’	m	4.6
	NH1	wing	A47’	PO_4	m	5.0
	NH1	wing	A47’	O4’	m	2.5
	NH1	wing	G3	N2	m	4.0
	NH1	wing	C46’	O2	m	3.3
Lys75	NZ	wing	A47’	N3	m	4.9
	NZ	wing	G48’	O4’	m	4.2
	N	wing	A4	O3’	m	4.9
	O	wing	A4	O3’	m	4.6
Phe77	ring	wing	C5	PO_4	M	5.0

The symbol **’** indicates the 3’ → 5’ strand of the DNA
M = major groove
m = minor groove
backbone = edge between major and minor groove

Table 6.6 Open PdxR-DNA contacts on the AAT-like domain ridge, monomer A

PdxR	Group	Location	**DNA**	Group	Location	Distance (Å)
Arg125	NH1	H_4	T26	PO_4	backbone	4.4
Lys126	NZ	H_4	A28	PO_4	M	4.4
	NZ	H_{13}	G27	O5’	M	4.9
Lys129	NZ	H_{13}	G27	PO_4	m	2.9
Lys360	NZ	H_{13}	A17’	PO_4	M	3.6
Arg364	NH2	H_{13}	A17’	PO_4	M	4.6
Lys367	NZ	H_{13}	A36	PO_4	m	4.6
Arg370	NH1	H_{13}	A37	PO_4	backbone	5.0
	NH2	H_{13}	A36	PO_4	backbone	2.4
Arg371	NH1	H_{13}	C35	PO_4	m	4.0
	NH2	H_{13}	A36	PO_4	m	3.1

Table 6.7 Open PdxR-DNA contacts on the AAT-like domain ridge, monomer B

PdxR	Group	Location	**DNA**	Group	Location	Distance (Å)
Arg125	NH1	H_4	T27’	PO_4	backbone	4.5
Lys126	NZ	H_4	T28’	PO_4	M	4.7
Lys129	NZ	H_4	T26	PO_4	m	5.0
	NZ	H_4	T27’	PO_4	m	5.0
Lys360	NZ	H_{13}	T16	PO_4	M	4.0
Arg364	NH1	H_{13}	C17	PO_4	M	4.9
Asn363	ND2	H_{13}	T16	PO_4	m	4.1
Lys367	NZ	H_{13}	T16	O3’	m	4.2
	NZ	H_{13}	C17	PO_4	m	4.6
Arg370	NH1	H_{13}	T37’	PO_4	backbone	3.3
Arg371	NH1	H_{13}	A36’	PO_4	M	4.8

The symbol **’** indicates the 3’ → 5’ strand of the DNA
M = major groove
m = minor groove
backbone = edge between major and minor groove

7. References

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.; Liu, D.; Lee, L. K.; Böcking, T. Binding of Transcription Factor GabR to DNA Requires Recognition of DNA Shape at a Location Distinct from Its Cognate Binding Site. *Nucleic Acids Research* **2016**, *44* (3), 1411–1420. https://doi.org/10.1093/nar/gkv1466.

Amidani, D.; Tramonti, A.; Canosa, A. V.; Campanini, B.; Maggi, S.; Milano, T.; Salvo, M. L. di; Pascarella, S.; Contestabile, R.; Bettati, S.; Rivetti, C. Study of DNA Binding and Bending by Bacillus Subtilis GabR, a PLP-Dependent Transcription Factor. *Biochimica et Biophysica Acta - General Subjects* **2017**, *1861* (1), 3474–3489. https://doi.org/10.1016/j.bbagen.2016.09.013.

Aravind, L.; Anantharaman, V. HutC/FarR-like Bacterial Transcription Factors of the GntR Family Contain a Small Molecule-Binding Domain of the Chorismate Lyase Fold. *FEMS Microbiology Letters* **2003**, *222* (1), 17–23. https://doi.org/10.1016/S0378-1097(03)00242-8.

Aravind, L.; Anantharaman, V.; Balaji, S.; Babu, M. M.; Iyer, L. M. The Many Faces of the Helix-Turn-Helix Domain: Transcription Regulation and Beyond. *FEMS Microbiology Reviews* **2005**, *29* (2), 231–262. https://doi.org/10.1016/j.femsre.2004.12.008.

Badis, G.; Berger, M. F.; Philippakis, A. A.; Talukder, S.; Gehrke, A. R.; Jaeger, S. A.; Chan, E. T.; Metzler, G.; Vedenko, A.; Chen, X.; Kuznetsov, H.; Wang, C.-F.; Coburn, D.; Newburger, D. E.; Morris, Q.; Hughes, T. R.; Bulyk, M. L. *Diversity and Complexity in DNA Recognition by Transcription Factors*.

Balleza, E.; López-Bojorquez, L. N.; Martínez-Antonio, A.; Resendis-Antonio, O.; Lozada-Chávez, I.; Balderas-Martínez, Y. I.; Encarnación, S.; Collado-Vides, J. Regulation by Transcription Factors in Bacteria: Beyond Description. *FEMS Microbiology Reviews* **2009**, *33* (1), 133–151. https://doi.org/10.1111/j.1574-6976.2008.00145.x.

Barad, B. A.; Echols, N.; Wang, R. Y. R.; Cheng, Y.; Dimaio, F.; Adams, P. D.; Fraser, J. S. EMRinger: Side Chain-Directed Model and Map Validation for 3D Cryo-Electron Microscopy. *Nature Methods* **2015**, *12* (10), 943–946. https://doi.org/10.1038/nmeth.3541.

Belitsky, B. R. Bacillus Subtilis GabR, a Protein with DNA-Binding and Aminotransferase Domains, Is a PLP-Dependent Transcriptional Regulator. *Journal of Molecular Biology* **2004**, *340* (4), 655–664. https://doi.org/10.1016/j.jmb.2004.05.020.

Belitsky, B. R. Role of Pdxr in the Activation of Vitamin B6 Biosynthesis in Listeria Monocytogenes. *Molecular Microbiology* **2014**, *92* (5), 1113–1128. https://doi.org/10.1111/mmi.12618.

Belitsky, B. R.; Sonenshein, A. L. GabR, a Member of a Novel Protein Family, Regulates the Utilization of γ-Aminobutyrate in Bacillus Subtilis. *Molecular Microbiology* **2002**, *45* (2), 569–583. https://doi.org/10.1046/j.1365-2958.2002.03036.x.

Bervoets, I. and Charlier, D. Diversity, versatility and complexity of bacterial gene regulation mechanisms: Opportunities and drawbacks for applications in synthetic biology. *FEMS Microbiology Reviews* **2019**, Vol. 43, Issue 3, pp. 304–339. https://doi.org/10.1093/femsre/fuz001

Bolshoy, A.; Mcnamarat, P.; Harringtont, R. E.; Trifonov, E. N. *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* **1991**; Vol. 88, pp 2312–2316. doi: 10.1073/pnas.88.6.2312.

Bramucci, E.; Milano, T.; Pascarella, S. 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. *Biochemical and Biophysical Research Communications* **2011**, *415* (1), 88–93. https://doi.org/10.1016/j.bbrc.2011.10.017.

Brennan, R. G.; Matthews, B. W. The Helix-Turn-Helix DNA Binding Motif. *Journal of Biological Chemistry* **1989**, *264* (4), 1903–1906. https://doi.org/10.1016/s0021-9258(18)94115-3.

Brown, A.; Long, F.; Nicholls, R. A.; Toots, J.; Emsley, P. and Murshudov, G. Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. *Acta crystallographica*. **2015**, Section D, Biological crystallography, 71(Pt 1), 136–153. https://doi.org/10.1107/S1399004714021683.

Browning, D. F.; Busby, S. J. W. The Regulation of Bacterial Transcription Initiation. *Nature Reviews Microbiology* **2004**, *2* (1), 57–65. https://doi.org/10.1038/nrmicro787.

Carroni, M. and Saibil, H. R. Cryo electron microscopy to determine the structure of macromolecular complexes. *Methods* **2016**, Vol. 95, pp. 78–85 https://doi.org/10.1016/j.ymeth.2015.11.023.

Catazaro, J.; Caprez, A.; Guru, A.; Swanson, D.; Powers, R. Functional Evolution of PLP-Dependent Enzymes Based on Active-Site Structural Similarities. *Proteins: Structure, Function and Bioinformatics* **2014**, *82* (10), 2597–2608. https://doi.org/10.1002/prot.24624.

Christen, P., Mehta, P. K. From Cofactor to Enzymes. The Molecular Evolution of Pyridoxal-5'-Phosphate-Dependent Enzymes **2001,** *Chem Rec*; Vol. 1, pp 436–447. doi: 10.1002/tcr.10005.

Di Salvo, M. L.; Contestabile, R.; Safo, M. K. Vitamin B 6 Salvage Enzymes: Mechanism, Structure and Regulation. *Biochimica et Biophysica Acta - Proteins and Proteomics* **2011**, *1814* (11), 1597–1608. https://doi.org/10.1016/j.bbapap.2010.12.006.

Dubochet, J.; Adrian, M.; Chang, J.J.; Homo, J.C.; Lepault, J.; McDowall, A.W.; Schultz, P. Cryo-Electron Microscopy of Vitrified Specimens. *Quarterly Reviews of Biophysics* **1988**, *21* (2), 129–228. https://doi.org/10.1017/S0033583500004297.

Edayathumangalam, R.; Wu, R.; Garcia, R.; Wang, Y.; Wang, W.; Kreinbring, C. A.; Bach, A.; Liao, J.; Stone, T. A.; Terwilliger, T. C.; Hoang, Q. Q.; Belitsky, B. R.; Petsko, G. A.; Ringe, D.; Liu, D. Crystal Structure of Bacillus Subtilis GabR, an Autorepressor and Transcriptional Activator of GabT. *Proceedings of the National Academy of Sciences of the United States of America* **2013**, *110* (44), 17820–17825. https://doi.org/10.1073/pnas.1315887110.

El Qaidi, S.; Yang, J.; Zhang, J. R.; Metzger, D. W.; Bai, G. 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. *Journal of Bacteriology* **2013**, *195* (10), 2187–2196. https://doi.org/10.1128/JB.00041-13.

Eliot, A. C.; Kirsch, J. F. Pyridoxal Phosphate Enzymes: Mechanistic, Structural, and Evolutionary Considerations. *Annual Review of Biochemistry* **2004**, *73*, 383–415. https://doi.org/10.1146/annurev.biochem.73.011303.074021.

Emsley, P.; Cowtan, K. Coot: Model-Building Tools for Molecular Graphics. *Acta Crystallographica Section D: Biological Crystallography* **2004**, *60* (12 I), 2126–2132. https://doi.org/10.1107/S0907444904019158.

Fitzpatrick, T. B.; Amrhein, N.; Kappes, B.; Macheroux, P.; Tews, I.; Raschle, T. Two Independent Routes of de Novo Vitamin B6 Biosynthesis: Not That Different after All. *Biochemical Journal* **2007**, *407* (1), 1–13. https://doi.org/10.1042/BJ20070765.

Frezzini, M.; Guidoni, L.; Pascarella, S. Conformational Transitions Induced by γ-Amino Butyrate Binding in GabR, a Bacterial Transcriptional Regulator. *Scientific Reports* **2019**, *9* (1), 1–15. https://doi.org/10.1038/s41598-019-55581-1.

Fujita, Y.; Fujita, T. The Gluconate Operon Gnt of Bacillus Subtilis Encodes Its Own Transcriptional Negative Regulator (GntR Protein/Insertional Mutagenesis/GntkR Mutation/Sugar Acid Catabolism). *Genetics* **1987,** Vol. 84, pp 4524–4528.

Gallegos, M.-T.; Michan, C.; Ramos, J. L. The XylS/AraC Family of Regulators; *Nucleic Acids Res*.**1993**; Vol. 21, pp 807–810. doi: 10.1093/nar/21.4.807.

Gokhale, R.S.; Khosla, C. Role of linkers in communication between protein modules. *Curr Opin Chem Biol.* **2000**, Feb;4(1):22-7. doi: 10.1016/s1367-5931(99)00046-0.

Grishin, N. V.; Phillips, M. A.; Goldsmith, E. J. Modeling of the Spatial Structure of Eukaryotic Ornithine Decarboxylases; *Protein Science*, **1995**; Vol. 4, pp 1291–1304. https://doi.org/10.1002/pro.5560040705.

Hagerman, P. J. Sequence-directed curvature of DNA. *Nature*; **1990**, 321**,** 449–450. https://doi.org/10.1038/321449a0.

Harrison, S. C.; Aggarwal, A. K. DNA Recognition by Proteins with the Helix-Turn-Helix Motif. *Annual Review of Biochemistry* **1990**, *59*, 933–969. https://doi.org/10.1146/annurev.bi.59.070190.004441.

Harteis, S.; Schneider, S. Making the Bend: DNA Tertiary Structure and Protein-DNA Interactions. *International Journal of Molecular Sciences* **2014**, *15* (7), 12335–12363. https://doi.org/10.3390/ijms150712335.

Haydon, D. J.; Guest, J. R. A New Family of Bacterial Regulatory Proteins. *FEMS Microbiology Letters* **1991**, *79* (2–3), 291–296. https://doi.org/10.1111/j.1574-6968.1991.tb04544.x.

Henikoff, S.; Haughn, G. W.; Calvo, J. M.; Wallace, J. C. A Large Family of Bacterial Activator Proteins. *Proceedings of the National Academy of Sciences of the United States of America* **1988**, *85* (18), 6602–6606. https://doi.org/10.1073/pnas.85.18.6602.

Hizver, J.; Rozenberg, H.; Frolow, F.; Rabinovich, D.; Shakked, Z. DNA Bending by an Adenine-Thymine Tract and Its Role in Gene Regulation. *Proceedings of the National Academy of Sciences of the United States of America* 2001; Vol. 17, pp 8490–8495. https://doi.org/10.1073/pnas.151247298.

Hoskisson, P. A.; Rigali, S. Chapter 1 Variation in Form and Function. The Helix-Turn-Helix Regulators of the GntR Superfamily. *Advances in Applied Microbiology* **2009**, *69*, 1–22. https://doi.org/10.1016/S0065-2164(09)69001-8.

Jain, D.; Nair, D. T. Spacing between Core Recognition Motifs Determines Relative Orientation of AraR Monomers on Bipartite Operators. *Nucleic Acids Research* **2013**, *41* (1), 639–647. https://doi.org/10.1093/nar/gks962.

Jansonius, J. N. Structure, Evolution and Action of Vitamin B6-Dependent Enzymes. *Current Opinion in Structural Biology* **1998**, *8* (6), 759–769. https://doi.org/10.1016/S0959-440X(98)80096-1.

Jochmann, N.; Götker, S.; Tauch, A. Positive Transcriptional Control of the Pyridoxal Phosphate Biosynthesis Genes PdxST by the MocR-Type Regulator PdxR of Corynebacterium Glutamicum ATCC 13032. *Microbiology* **2011**, *157* (1), 77–88. https://doi.org/10.1099/mic.0.044818-0.

Kirsch, J.F.; Eichele, G.; Ford, G.C.; Vincent, M.G.; Jansonius, J.N.; Gehring, H.; Christen, P. Mechanism of action of aspartate aminotransferase proposed on the basis of its spatial structure. *J Mol Biol.* **1984**, 5;174(3):497-525. doi: 10.1016/0022-2836(84)90333-4.

Liao, S.; Bitoun, J. P.; Nguyen, A. H.; Bozner, D.; Yao, X.; Wen, Z. T. Deficiency of PdxR in Streptococcus Mutans Affects Vitamin B6 Metabolism, Acid Tolerance Response and Biofilm Formation. *Molecular Oral Microbiology* **2015**, *30* (4), 255–268. https://doi.org/10.1111/omi.12090.

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.; Moriarty, N. W.; Oeffner, R. D.; Poon, B. K.; Prisant, M. G.; Read, R. J.; Richardson, J. S.; Richardson, D. C.; Sammito, M. D.; Sobolev, O. V.; Stockwell, D. H.; Terwilliger, T. C.; Urzhumtsev, A. G.; Videau, L. L.; Williams, C. J.; Adams, P. D. Macromolecular Structure Determination Using X-Rays, Neutrons and Electrons: Recent Developments in Phenix. *Acta Crystallographica Section D: Structural Biology* **2019**, *75*, 861–877. https://doi.org/10.1107/S2059798319011471.

Mattison, K.; Oropeza, R.; Kenney, L. J. The Linker Region Plays an Important Role in the Interdomain Communication of the Response Regulator OmpR. *Journal of Biological Chemistry* **2002**, *277* (36), 32714–32721. https://doi.org/10.1074/jbc.M204122200.

McPhalen, C. A.; Vincent, M. G.; Picot, D.; Jansonius, J. N.; Lesk, A. M.; Chothia, C. Domain Closure in Mitochondrial Aspartate Aminotransferase. *Journal of Molecular Biology* **1992**, *227* (1), 197–213. https://doi.org/10.1016/0022-2836(92)90691-C.

Milano, T.; Angelaccio, S.; Tramonti, A.; Salvo, M. L. D.; Contestabile, R.; Pascarella, S. Structural Properties of the Linkers Connecting the N- and C-Terminal Domains in the MocR Bacterial Transcriptional Regulators. *Biochimie Open* **2016**, *3*, 8–18. https://doi.org/10.1016/j.biopen.2016.07.002.

Milano, T.; Gulzar, A.; Narzi, D.; Guidoni, L.; Pascarella, S. 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* **2017**, *12* (12), 1–21. https://doi.org/10.1371/journal.pone.0189270.

Morozov, A. V.; Siggia, E. D. Connecting Protein Structure with Predictions of Regulatory Sites. *Proceedings of the National Academy of Sciences of the United States of America* **2007**, *104* (17), 7068–7073. https://doi.org/10.1073/pnas.0701356104.

Nardella, C.; Barile, A.; Salvo, M. L. di; Milano, T.; Pascarella, S.; Tramonti, A.; Contestabile, R. Interaction of Bacillus Subtilis GabR with the GabTD Promoter: Role of Repeated Sequences and Effect of GABA in Transcriptional Activation. *FEBS Journal* **2020**, *287* (22), 4952–4970. https://doi.org/10.1111/febs.15286.

Nguyen, C. C.; Saier, M. H. Phylogenetic, Structural and Functional Analyses of the LacI-GalR Family of Bacterial Transcription Factors. *FEBS letters* **1995**; Vol. 377, pp 98–102. https://doi.org/10.1016/0014-5793(95)01344-X.

Okuda, K.; Ito, T.; Goto, M.; Takenaka, T.; Hemmi, H.; Yoshimura, T. Domain Characterization of Bacillus Subtilis GabR, a Pyridoxal 50-Phosphate-Dependent Transcriptional Regulator. *Journal of Biochemistry* **2015**, *158* (3), 225–234. https://doi.org/10.1093/jb/mvv040.

Okuda, K.; Kato, S.; Ito, T.; Shiraki, S.; Kawase, Y.; Goto, M.; Kawashima, S.; Hemmi, H.; Fukada, H.; Yoshimura, T. Role of the Aminotransferase Domain in Bacillus SubtilisGabR, a Pyridoxal 5′-Phosphate-Dependent Transcriptional Regulator. *Molecular Microbiology* **2015**, *95* (2), 245–257. https://doi.org/10.1111/mmi.12861.

Orlova, E. v.and Saibil, H. R. Structural analysis of macromolecular assemblies by electron microscopy. *Chemical Reviews* **2019**, Vol. 111, Issue 12, pp. 7710–7748. https://doi.org/10.1021/cr100353t

Pabo, C. O.; Sauer, R. T. Transcription Factors: Structural Families and Principles of DNA Recognition. *Annual Review of Biochemistry* **1992**, *61*, 1053–1095. https://doi.org/10.1146/annurev.bi.61.070192.005201.

Park, S. A.; Park, Y. S.; Lee, K. S. Crystal Structure of the C-Terminal Domain of Bacillus Subtilis GabR Reveals a Closed Conformation by γ-

Aminobutyric Acid Binding, Inducing Transcriptional Activation. *Biochemical and Biophysical Research Communications* **2017**, *487* (2), 287–291. https://doi.org/10.1016/j.bbrc.2017.04.052.

Parra, M.; Stahl, S.; Hellmann, H. Vitamin B6 and Its Role in Cell Metabolism and Physiology. *Cells* **2018**, *7* (7), 84. https://doi.org/10.3390/cells7070084.

Pérez-Martín, J. and de Lorenzo, V. Clues and consequences of DNA bending in transcription. *Annu Rev Microbiol.* **1997**; 51:593-628. doi: 10.1146/annurev.micro.51.1.593.

Peterson, E. A. and Sober H. A. Preparation of Crystalline Phosphorylated Derivatives of Vitamin B6. *Journal of the American Chemical Society* **1954**, 76 (1), 169-175. doi: 10.1021/ja01630a045

Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera - A Visualization System for Exploratory Research and Analysis. *Journal of Computational Chemistry* **2004**, *25* (13), 1605–1612. https://doi.org/10.1002/jcc.20084.

Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Meng, E. C.; Couch, G. S.; Croll, T. I.; Morris, J. H.; Ferrin, T. E. UCSF ChimeraX: Structure Visualization for Researchers, Educators, and Developers. *Protein Science* **2021**, *30* (1), 70–82. https://doi.org/10.1002/pro.3943.

Raschle, T.; Arigoni, D.; Brunisholz, R.; Rechsteiner, H.; Amrhein, N.; Fitzpatrick, T. B. Reaction Mechanism of Pyridoxal 5′-Phosphate Synthase: Detection of an Enzyme-Bound Chromophoric Intermediate. *Journal of Biological Chemistry* **2007**, *282* (9), 6098–6105. https://doi.org/10.1074/jbc.M610614200.

Rigali, S.; Derouaux, A.; Giannotta, F.; Dusart, J. Subdivision of the Helix-Turn-Helix GntR Family of Bacterial Regulators in the FadR, HutC, MocR, and YtrA Subfamilies. *Journal of Biological Chemistry* **2002**, *277* (15), 12507–12515. https://doi.org/10.1074/jbc.M110968200.

Rigali, S.; Schlicht, M.; Hoskisson, P.; Nothaft, H.; Merzbacher, M.; Joris, B.; Titgemeyer, F. Extending the Classification of Bacterial Transcription Factors beyond the Helix-Turn-Helix Motif as an Alternative Approach to Discover New Cis/Trans Relationships. *Nucleic Acids Research* **2004**, *32* (11), 3418–3426. https://doi.org/10.1093/nar/gkh673.

Rohou, A.; Grigorieff, N. CTFFIND4: Fast and Accurate Defocus Estimation from Electron Micrographs. *J Struct Biol.* **2015,** Nov;192(2):216-21. doi: 10.1016/j.jsb.2015.08.008.

Rohs, R.; West, S. M.; Sosinsky, A.; Liu, P.; Mann, R. S.; Honig, B. The Role of DNA Shape in Protein-DNA Recognition. *Nature* **2009**, *461* (7268), 1248–1253. https://doi.org/10.1038/nature08473.

Rosenthal, P. B.; Henderson, R. Optimal Determination of Particle Orientation, Absolute Hand, and Contrast Loss in Single-Particle Electron Cryomicroscopy. *Journal of Molecular Biology* **2003**, *333* (4), 721–745. https://doi.org/10.1016/j.jmb.2003.07.013.

Rossbach, S.; Kulpa, D. A.; Rossbach, U.; Bruijn, F. J. Molecular and Genetic Characterization of the Rhizopine Catabolism (MocABRC) Genes of Rhizobium Meliloti L5-30. *Mol Gen Genet* **1994**; Vol. 245, pp 11–24. doi: 10.1007/BF00279746.

Scheres, S. H. W. RELION: Implementation of a Bayesian Approach to Cryo-EM Structure Determination. *Journal of Structural Biology* **2012**, *180* (3), 519–530. https://doi.org/10.1016/j.jsb.2012.09.006.

Scheres, S. H. W. Processing of Structurally Heterogeneous Cryo-EM Data in RELION. *Methods in Enzymology* **2016**, Vol. 579, pp. 125–157. https://doi.org/10.1016/bs.mie.2016.04.012.

Scheres, S. H. W. and Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nature Methods* **2012**, Vol. 9, Issue 9, pp. 853–854. https://doi.org/10.1038/nmeth.2115.

Schneider, G.; Käck, H.; Lindqvist, Y. The Manifold of Vitamin B6 Dependent Enzymes. *Structure* **2000**, *8* (1), 1–6. https://doi.org/10.1016/S0969-2126(00)00085-X.

Seeman, N. C.; Rosenberg, J. M.; Rich, A. Sequence Specific Recognition of Double Helical Nucleic Acids by Proteins. *Proceedings of the National Academy of Sciences of the United States of America* **1976**, *73* (3), 804–808. https://doi.org/10.1073/pnas.73.3.804.

Strohmeier, M.; Raschle, T.; Mazurkiewicz, J.; Rippe, K.; Sinning, I.; Fitzpatrick, T. B.; Tews, I. Structure of a Bacterial Pyridoxal 5′-Phosphate Synthase Complex. *Proceedings of the National Academy of Sciences of the United States of America* **2006**, *103* (51), 19284–19289. https://doi.org/10.1073/pnas.0604950103.

Suvorova, I. A.; Korostelev, Y. D.; Gelfand, M. S. GntR Family of Bacterial Transcription Factors and Their DNA Binding Motifs: Structure, Positioning and Co-Evolution. *PloS one* **2015**, *10* (7), e0132618. https://doi.org/10.1371/journal.pone.0132618.

Tan, Y. Z.; Baldwin, P. R.; Davis, J. H.; Williamson, J. R.; Potter, C. S.; Carragher, B.; Lyumkis, D. Addressing Preferred Specimen Orientation in Single-Particle Cryo-EMthrough Tilting. *Nature Methods* **2017**, *14* (8), 793–796. https://doi.org/10.1038/nmeth.4347.

Tramonti, A.; Fiascarelli, A.; Milano, T.; Salvo, M. L. D.; Nogués, I.; Pascarella, S.; Contestabile, R. Molecular Mechanism of PdxR - A Transcriptional Activator Involved in the Regulation of Vitamin B6 Biosynthesis in the Probiotic Bacterium Bacillus Clausii. *FEBS Journal* **2015**, *282* (15), 2966–2984. https://doi.org/10.1111/febs.13338.

Tramonti, A.; Nardella, C.; Salvo, M. L. di; Pascarella, S.; Contestabile, R. The MocR-like Transcription Factors: Pyridoxal 5′-Phosphate-Dependent Regulators of Bacterial Metabolism. *FEBS Journal* **2018**, *285* (21), 3925–3944. https://doi.org/10.1111/febs.14599.

Van Aalten, D. M. F.; Dirusso, C. C.; Knudsen, J.; Wierenga, R. K. Crystal Structure of FadR, a Fatty Acid-Responsive Transcription Factor with a Novel Acyl Coenzyme A-Binding Fold. EMBO Journal 2000, 19 (19), 5167–5177. https://doi.org/10.1093/emboj/19.19.5167.

van der Vliet, P. C.; Verrijzer, C.P. Bending of DNA by transcription factors. *Bioessays* **1993**, Jan;15(1):25-32. doi: 10.1002/bies.950150105.

Wiethaus, J.; Schubert, B.; Pfänder, Y.; Narberhaus, F.; Masepohl, B. The GntR-like Regulator TauR Activates Expression of Taurine Utilization Genes in Rhodobacter Capsulatus. *Journal of Bacteriology* **2008**, *190* (2), 487–493. https://doi.org/10.1128/JB.01510-07.

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.; Jain, S.; Lewis, S. M.; Arendall, W. B.; Snoeyink, J.; Adams, P. D.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C. MolProbity: More and Better Reference Data for Improved All-Atom Structure Validation. *Protein Science* **2018**, *27* (1), 293–315. https://doi.org/10.1002/pro.3330.

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.; Liu, D. PLP and GABA Trigger GabR-Mediated Transcription Regulation in Bacillus Subtilis via External Aldimine Formation. *Proceedings of the National Academy of Sciences of the United States of America* **2017**, *114* (15), 3891–3896. https://doi.org/10.1073/pnas.1703019114.

Xu, Y.; Heath, R. J.; Li, Z.; Rock, C. O.; White, S. W. The Fad·RDNA Complex. Transcriptional Control of Fatty Acid Metabolism in Escherichia Coli. *Journal of Biological Chemistry* **2001**, *276* (20), 17373–17379. https://doi.org/10.1074/jbc.M100195200.

Young, M. A.; Beveridge, D. L. Molecular Dynamics Simulations of an Oligonucleotide Duplex with Adenine Tracts Phased by a Full Helix Turn. *J Mol Biol* **1998**, Vol. 281, pp 675–687. doi: 10.1006/jmbi.

Zheng, S. Q., Palovcak, E., Armache, J. P., Verba, K. A., Cheng, Y.; Agard, D. A. MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature Methods* **2017**, Vol. 14, Issue 4, pp. 331–332. https://doi.org/10.1038/nmeth.4193.

8. List of Publications

- Exertier C, Milazzo L, **Freda I**, Montemiglio LC, Scaglione A, Cerutti G, Parisi G, Anselmi M, Smulevich G, Savino C, Vallone B. Proximal and distal control for ligand binding in neuroglobin: role of the CD loop and evidence for His64 gating. Sci Rep. 2019 Mar 29;9(1):5326. doi: 10.1038/s41598-019-41780-3.

- Milazzo L, Exertier C, Becucci M, **Freda I**, Montemiglio LC, Savino C, Vallone B, Smulevich G. Lack of orientation selectivity of the heme insertion in murine neuroglobin revealed by resonance Raman spectroscopy. FEBS J. 2020 Sep;287(18):4082-4097. Doi: 10.1111/febs.15241. Epub 2020 Feb 27.

- Parisi G*, **Freda I***, Exertier C*, Cecchetti C, Gugole E, Cerutti G, D'Auria L, Macone A, Vallone B, Savino C, Montemiglio LC. Dissecting the Cytochrome P450 OleP Substrate Specificity: Evidence for a Preferential Substrate. Biomolecules. 2020 Oct 6;10(10):1411. doi: 10.3390/biom10101411.

- Cerutti G, Gugole E, Montemiglio LC, Turbé-Doan A, Chena D, Navarro D, Lomascolo A, Piumi F, Exertier C, **Freda I**, Vallone B, Record E, Savino C, Sciara G. Crystal structure and functional characterization of an oligosaccharide dehydrogenase from Pycnoporus cinnabarinus provides insights into fungal breakdown of lignocellulose. Biotechnol Biofuels. 2021 Jul 22;14(1):161. doi: 10.1186/s13068-021-02003-y.

*** Equally contributing authors**