UNIVERSITÀ DEGLI STUDI DI PARMA

Department of Life Sciences
Laboratories of Biochemistry and Molecular Biology

PhD Program in
Biochemistry and Molecular Biology

XXVIII cycle

Biochemical characterization and DNA binding properties of the MocR family member GabR, a PLP-dependent transcription factor

Coordinator:
Prof. Andrea Mozzarelli

Tutor:
Prof. Claudio Rivetti

PhD student:
Davide Amidani

2013-2015

Abstract

GabR is a chimeric transcriptional regulator belonging to the novel MocR/GabR family, characterized by a short N-terminal helix-turn-helix DNA-binding domain and a long C-terminal effector binding and/or oligomerization domain connected by a 29 amino acids flexible linker. The C-terminal domain is structurally homologous to the fold-type I aminotransferases, a group of enzymes involved in amino acids metabolism that use pyridoxal 5'-phosphate (PLP) as a cofactor. In the presence of γ-aminobutyrate (GABA) and PLP, GabR activates the transcription of the *gabT* and *gabD* genes, which encode two enzymes involved in GABA metabolism. GabR crystalized as a head-to-tail domain-swap homodimer and binds two hexameric direct repeats (ATACCA) separated by a 34 bp spacer at the *gabTD* promoter.

This thesis reports the biochemical, structural and DNA binding properties of *Bacillus subtilis* GabR. Spectroscopic analysis indicates that GabR binds PLP as an internal aldimine and reacts with GABA to form the external aldimine. GabR-*gabTD* binding reactions imaged employing Atomic Force Microscopy (AFM) suggest that both holo-GabR and apo-GabR bind the cognate DNA site as a dimer with *K*d of about 40 nM. Conversely, in presence of GABA conformational rearrangements of the nucleoprotein complexes, reduction of the DNA wrapping around GabR and a two-fold increase of the *K*d were observed, independently of the sequence of the *gabTD* promoter investigated.

To evaluate the role of the DNA topology on the binding of GabR, five and ten bp were inserted within the spacer region, separating the direct repeat sequences. The intrinsic bending of the *gabTD* promoter region was strongly modified by the insertion of five bp, whereas the insertion of ten bp caused only an increase of the distance between the GabR binding sequences. Interestingly, the insertion of five bp determined a significant loss of affinity, whereas the ten bp insertion had only a slight effect on binding.

Electrostatic potential analysis of GabR protein surface showed a positive groove that extends along one entire protein face. To investigate the relevance of this feature on DNA-binding, we constructed the GabR R129Q and K362-366Q mutants, in which the surface positive charge was weakened. DNA binding affinity decreased in both GabR mutants, particularly in GabR K362-366Q for which DNA binding was lost.

Together, our results suggest that the GabR binding at the *gabTD* promoter relies on the intrinsic DNA bending and the correct orientation of the two binding hexamers rather than on their distance. The positive GabR surface is also critical to accommodate and/or to favor DNA bending on the protein core. Accordingly, the GabR-*gabTD* binding could be a typical example in which specific binding site recognition, DNA topology and protein structural and superficial electrostatic features, are simultaneously required to assemble a stable protein-DNA complex.

Contents

Introduction

Results

Discussion

Introduction

Bacteria have lived on our planet for billions of years, during which time they have adapted to many fluctuating and harsh environments and have colonized virtually all available ecological niches[1]. Indeed bacterial cells are able to constantly sense environmental changes, keeping the homeostasis and cellular functions. Probably, this is the main demonstration that bacterial cells are not static, but fantastically dynamic life forms able to adapt and survive in a wide range of environmental conditions. Survival and adaptation capability of a cell are tightly connected to its genes and to the way the expression of the genes is regulated. Gene expression regulation consists of a wide range of mechanisms that are used by cells to increase or decrease the production of specific gene products, such as proteins, depending on life conditions. Because microorganisms live in instable environments, cells respond to abrupt environmental changes by launching gene expression programs that help to adjust the cellular physiology and metabolism to the new conditions and that protect against cell damage or death[1].

Transcription, and successively translation, are the steps that constitute the gene expression in prokaryotes and eukaryotes. Transcription is the process in which genes begin their expression route and the DNA copy into RNA molecules is performed primarily by the enzyme RNA polymerase. The successive step of gene expression consists in translation, in which the information contained into the RNA (messenger RNA, mRNA) nucleotides is converted into amino acid chains.

In transcription the RNA polymerase enzyme (RNAP) is one of the crucial players. Transcription initiation occurs through the specific interaction of the RNAP sigma factor with the promoter[2,3,4]. In bacteria the promoter core consists in two relatively conserved hexamers placed in -35 and -10 position with respect to the transcription start site +1. In addition to -35 and -10, other promoter elements such as the spacer between the -35 and the -10, the extended -10 located upstream of the -10 and the discriminator element from position -4 to -6 are involved in specific interactions with four conserved domains of the sigma factor, called determinant regions[2,4].

These interactions between RNAP and the promoter assemble the initial closed complex (RP_c) which, through conformational rearrangements in both molecules, isomerizes to the catalytically active RNAP– DNA open complex (RP_o)[3,5], where *ex-novo* 5'-3' RNA synthesis takes place, following the Watson-Crick complementary base pairing

mechanism. Transcription ends when the gene/s are completely transcribed and RNAP recognizes the terminator sequences. Specific molecular mechanisms[6,7] promote the RNA synthesis suspension, transcription complex dissociation and the release of RNA from the RNAP.

Gene transcription regulation

During the cell lifetime, many genes coding for different proteins and/or RNA are not expressed at any time. The survival of an organism is deeply connected to its ability to express combinations of genes depending on the biological context[1]. This is demonstrated, especially in genes coding for enzymes involved in nutrients metabolism. In fact, it is necessary that these genes have to be express when specific nutrients are available in the environment, to obtain energy from their metabolism. Many molecular mechanisms in gene expression are operative from the gene to its product, however it is not surprising that a lot of them act at the beginning of the transcription. Indeed, this represents the main point in transcription regulation, providing both energetic advantages to the cell (mRNA are synthetized only when a gene product function is required for survival) and rigor in gene expression control.

Depending on their sequence some promoters can be weakly bound by the RNAP and the initial transcription steps (transition from RP_c to RP_o) can occur. Genes controlled by these promoters are expressed at a basal level and their expression is also defined constitutive. This is the case of the housekeeping genes that are typically required for the maintenance of basic cellular function, and are expressed under normal physiological conditions.

Many other genes have to be expressed in highly specific moments, as a function of the surrounding environmental conditions. To satisfy this cellular need for adaptation, the existence of an efficient transcription regulation device is crucial. This mechanism is based on regulatory proteins, known as transcription factors, capable to repress or activate the transcription of target genes in response to an environmental or cellular triggers[4,8].

Typical example in transcription repression is represented by the *Lac repressor*, a 155 kDa tetrameric protein that binds to three operator sites within the operon, repressing the transcription of the *lacZYA* genes (Lac operon) coding for enzymes involved in lactose metabolism. Lac repressor binds the promoter region of the Lac operon in absence of

lactose. On the contrary, the presence of lactose in the environment results in the production of allolactose, a side product of the β-galactosidase (*lacZ*) which binds to the repressor and releases it from the operator sites, inducing the transcription of the lac genes[9].

Some transcription factors can also operate as activator. Activators generally function by reducing the free energy for either to stabilize the initial RNAP-promoter complex or to accelerate the transition to the RP_o.

In principle there are two ways in which an activator can promote gene transcription: by altering the conformation of the promoter / promoting the transition to the RP_o (allosteric mechanism) or by interacting directly with the RNAP (RNAP recruitment mechanism)[4].

The MerR family transcription factor represents the best understood example in which the promoter topology is modified by a regulatory protein. Target promoters for these activators have a nonoptimal spacing between the -35 and -10 elements witch hinder RNAP binding. This hindrance is overcome by the activator, MerR or BmtR for instance, which causes a twist in the spacer that results in the promoter -10 element being brought into register with the -35 element, thereby triggering transcription initiation[10].

Paradigmatic of the recruitment mechanism is the catabolite activator protein CAP. In presence of cAMP, CAP acts by binding to specific DNA sites in or near target promoters and enhancing the ability of RNAP to bind and initiate transcription at more than one hundred promoters. Depending on the promoter region engaged by CAP[11] (class, I, II and III CAP promoter were characterized), the interactions with the RNAP are different; usually CAP-RNAP interactions occurred in αCTD domain, sigma factor domain 4 and the αNTD domain of RNAP.

Transcription factors

Transcription factors are essential for the regulation of gene expression and are widespread among prokaryotes and eukaryotes[12]. It means that transcription factors are critical to control almost the overall vital processes such as DNA replication, transcription, DNA recombination, DNA damage repair, response to environment and many others.

In order to precisely regulate gene expression, transcription factors must distinguish the target DNA sequence from the entire genome and bind it depending on the environmental

or intracellular conditions. This biological recruitment is reflected in transcription factor structure, generally consisting in a DNA-binding domain and an signal sensing domain that senses external signals and transmits the information to the rest of the molecule. Usually, in signal sensing domain hydrophobic motives are exposed to the environment and they are employed in transcription factor dimerization. Depending on the nature of the dimerization domain and the abundance of individual transcription factors, homodimers or heterodimers can form. This structural architecture adds an extra element of complexity and versatility in DNA binding function. Indeed, the majority of the dimeric transcription factors have two DNA-binding domains involved in DNA interaction and recognition of palindromic sequences. Even if exceptions to the palindromic sequence recognition exist in the large transcription factor family, this DNA binding fashion is widely represented and characterized[9,13,14,15] (Fig.1). These DNA-binding proteins, bind sequences of ≈ 15-20 nucleotides long, in which half sequence is recognized by the DNA-binding domain of one monomer. This suggests that transcription factors dimerization is a way to increase both the DNA-binding specificity and strength, as a consequence of the longer DNA sequence recognition.

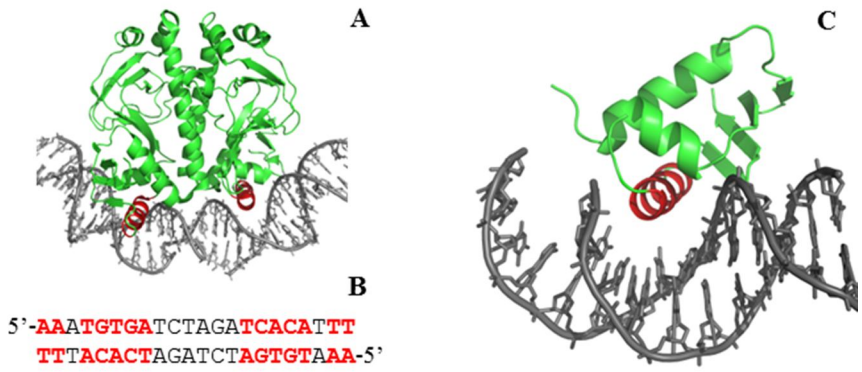

Fig. 1. (A) *E.Coli* Catabolite Activator Protein (CAP) is an example of dimeric transcription factor that recognizes palindromic sequences through the HTH DNA binding domain. One subunit of CAP interacts with one half of the DNA site while the other subunit binds the other half of the DNA site. The majority of CAP-DNA interactions are mediated by the HTH DNA binding motif present in each subunit of CAP. **(B)** CAP interacts with 22 bp; positions at which CAP exhibits strong sequence preferences are indicated in red. **(C)** CAP HTH motif; the probe α helix (in red) is tilted in the major groove to recognize specific DNA bp.

Interactions between the protein amino acids and the nucleic acid bases take place through hydrogen bonds, ionic bonds, Van Der Waals interactions and hydrophobic interactions. These bonds are very weak (1-7 Kcal/mol), however the formation of many of them during the specific protein-DNA interaction promotes the assembly of stable nucleoprotein complexes.

The binding between a DNA-binding protein and the DNA is based on the chemical and structural complementarity between the two molecular surfaces.

The specific protein-DNA interaction can occur in the major and the minor groove of the DNA through the formation of hydrogen bonds and the recognition of DNA base non-polar groups by some amino acids of the protein folded to compose the DNA-binding domain. This mechanism of interaction is indicated as direct or base readout[16,17]. However, not only the specific sequences recognized by a DNA-binding protein are critical to support the interaction, but also the topology and the mechanical properties of the DNA have to be considered. Indeed, in shape or indirect readout binding mechanism, many proteins take advantage from the sequence-dependent flexibility or intrinsic bending of the DNA double helix to specifically recognize their binding site at the DNA at lower free energy cost[18].

The determination of high quality structures of DNA-binding proteins complexed with DNA have provided insight into stereochemical principles of binding, including how particular base sequences are recognized and how the DNA topology is modified by binding. Based on the structural information about several DNA-binding proteins, Luscombe and colleagues have sorted in detail the DNA-binding domains[19]: helix-turn-helix domains (HTH) including the "winged" HTH domains, zinc-coordinating domains, zipper-type domains, β sheet and β hairpin/ribbon are the main groups of the classification.

In spite that all listed DNA-binding domains are represented in prokaryotes, the majority of the interactions with specific DNA sequences are carried out by helix-turn-helix DNA-binding domain[20,21], in which a bundle of three to six α helix provides the stabilizing hydrophobic core, although the motive is also traditionally defined as a 20 amino acid segment of two almost perpendicular α helix connected by a four residues β turn. This motif binds in the major groove employing the recognition (or probe) α helix which specifically interacts with the binding sequences. However, to perfectly accommodate

inside the major groove, it is necessary that the α helix is tilted, because of its parallel orientation in respect to the DNA double helix axis is not ideal for fitting the major groove shape[18] (Fig.1). A variant of the HTH domain, is the "winged" HTH, which is characterized by the presence of a third α helix and adjacent β sheet in the DNA-binding motif. The DNA recognition occurs as in the regular HTH, and the extra elements provide additional contacts with the DNA backbone.

GntR family

First described in 1991 by Haydon and Guest and named after the gluconate- operon repressor in *Bacillus subtilis*, GntR transcription factors are comprised of a N-terminal HTH DNA-binding domain and a C-terminal effector-binding and/or oligomerization domain (E-b/O) linked together[22].

The hallmark of this family consists in the HTH domain (Fig.1) that shows strong similarities among all members of the group whereas the C-terminal domains do not. Nevertheless, C-terminal domain is critical in DNA binding regulation as well. Indeed, the binding of effector molecules in C-terminal domain can impose oligomerization or conformational changes into the protein and steric constraints on the DNA-binding domain necessary to correctly orient the HTH-motif for the specific DNA interaction[23].

C-terminal heterogeneity was limited to four structural domain topologies which underwent a genic fusion event during the evolution to a common HTH DNA-binding domain[23]. In agreement with different type of C-terminal domains, the GntR transcription factor family can be divided into four main subfamilies (FadR, HutC, MocR, and YtrA) and two minor subfamilies (AraR and PlmA).

About 40% of known GntR members are grouped into the FadR subfamily. The C-terminal domain consists of six / seven α helix and the average domain length is about 150 – 170 amino acids. FadR members are involved in the regulation of oxidized substrates related to amino acids or emerging from the central metabolism, or at the crossroads of various metabolic pathways, such as glycolate (GlcC), galactonate (DgoR), pyruvate (PdhR), lactate (LldR), or gluconate (GntR)[23].

In HutC subfamily the C-terminal domain is about 170 amino acids long and consist of both α helix and β sheet. HutC subfamily comprises 31% of GntR members. C-terminal

domain is organized with the same fold of chorismate lyase. Therefore, HutC members could bind small effector molecules, such as histidine, fatty acids, sugars and alkylphosphonates, similarly to chorismate lyase. Other HutC proteins are involved in the regulation of N-acetylglucosamine utilization and in conjugative plasmid transfer in various *Streptomyces* species[22].

The MocR subfamily is significantly different from the other GntR members, because its C-terminal domain has an average length of about 350 amino acids and homology with fold-type I pyridoxal 5'-phosphate (PLP)-dependent aminotransferase proteins[22]. Aminotransferase enzymes catalyse the transamination reactions in several physiological processes, such as synthesis and degradation of amino acids and nitrogen metabolism. Usually, fold-type I aminotransferases members, are organized in head-to tail homodimers or higher-order oligomers with two active sites per dimer[24]. Probably, dimerization also occurs in MocR subfamily members such as in *Bacillus subtilis* GabR[25].

The YtrA subfamily regroups 6% of GntR members that possess a very short C-terminal domain (about 50 amino acids) with two α helix. Small dimensions of C-terminal domain suggest that dimerization is necessary for the sensing mechanism of the effector molecules. Most genes of the YtrA-subfamily form operons with ATP-binding cassette (ABC) transporters[23]. Minor PlmA subfamily is exclusively composed of GntR members from cyanobacterial species, whereas the AraR subfamily has a C-terminal domain homologous to that of the GalR/LacI family and controls expression of genes encoding transporters and enzymes involved in the uptake and utilization of L-arabinose and arabinose-containing polysaccharides, xylose and galactose in *Firmicutes*[22].

MocR transcription factors

MocR family members are chimeric proteins characterized by an exceptional length of the C-terminal domain, homologous to that of the fold type I family of PLP-dependent enzymes, coupled with the HTH DNA-binding domain. From an evolutionary and functional point of view this is very interesting. However it still remain an open question, why both DNA-binding function and enzymatic activity are maintained within the same molecule during the evolution. Moreover, MocR family is indicated to be the only GntR family whose members are capable of binding direct repeat sequences on the DNA[23].

Even if MocR proteins are composed of a C-terminal domain belonging to the aminotransferase family, they have followed the GntR transcription factors evolutionary route. In fact, MocR members are widespread among eubacteria, even if with heterogeneity. Genomic distribution studies suggest that the MocR presence in microorganisms is highly correlated to the size of the genome, and consequently, to the complexity of the metabolic and regulatory pathways. Apparently, the minimum genome size compatible with the presence of MocR genes is around 2 – 2.5 Mb; additionally, it has to be considered that microorganisms with similar genome size possess different number of MocR protein members[26].

Following the Goldsmith classification, PLP-dependent enzymes can be grouped into five different categories on the basis of their amino acid sequence comparisons, secondary structure elements and availability of three-dimensional structural information[27].

MocR C-terminal domain belong to the fold-type I family PLP-dependent enzymes, also called aspartate aminotransferases (AAT) family. These enzymes are catalytically active as homodimers or, in some cases, as larger complexes. Each aminotransferase-like domain folds in a large domain, characterized by the presence of seven β sheet, and a small domain comprising the C-terminal part of the chain, which folds into a three / four-stranded β sheet covered with helices on one side. Small domain structure is completed by the N-terminal part of the chain composed by variable secondary structure among the fold-type I members. Residues belonging to both domains are involved in the active site formation, PLP and reaction substrate binding. In lacking the reaction substrate in the active site, the aldehyde group of PLP interacts with the amino group of the catalytic lysine, while the phosphate of the cofactor is anchored to a large domains α helix and the aromatic ring is packed against the neighbouring β strands.

Aminotransferase enzymes catalyze the transamination reaction, consisting in the transfer of an α-amino group from an α-amino acid to an α-ketoacid employing the PLP as cofactor (Fig. 2). Many different transaminases are known and they are generally of broad specificity for amino acids substrates[24].

In addition to transamination, fold-type I enzymes can catalyze many other reactions taking advantage from both its structural arrangement and the interaction with the cofactor PLP.

Indeed, PLP is the biologically active form of vitamin B6 and arguably represents the most versatile cofactor in biology; it is not surprising that PLP-dependent enzymes are unrivaled in the diversity of reactions that they catalyze[24].

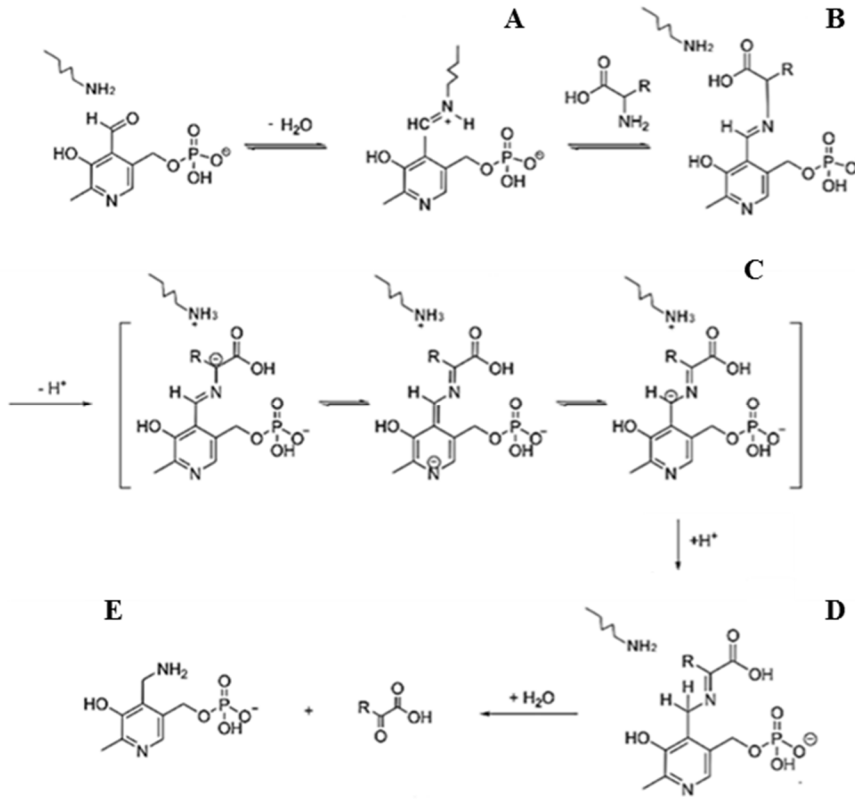

Fig. 2. Transamination reaction steps. **(A)** In the resting state, the PLP aldehyde group forms the Schiff base, or internal aldimine, with the amino group of the catalytic lysine. **(B)** When a specific amino acid is in the active site its amino group displace the lysine to form a new Schiff base, or external aldimine. **(C)** Next, a basic residue of the enzyme abstracts the α-hydrogen of the external aldimine to produce anionic intermediates, including the quinonoid intermediate. **(D)** The abstracted α-hydrogen is added to C4′ of the cofactor to form a ketimine intermediate. **(E)** Hydrolysis of the ketimine intermediate produces keto acid and the pyridoxamine 5′-phosphate (PMP), completing the first half-transamination. The second half-transamination consists in the formation of a product amino acid from the substrate keto acid with the recovery of PLP from PMP, proceeds via the reverse course of the first half-transamination

PLP-dependent enzymes are associated with biochemical pathways that involve amino compounds, in particular amino acid substrates. The reactions carried out by the PLP-dependent enzymes on amino acids, in addition to transamination, consist in

decarboxylation, interconversion of L- and D-amino acids (racemization), and removal (elimination) or replacement of chemical groups bound at the β- or γ-carbon. All PLP-dependent enzymes have the external aldimine formation in common, and it is from this chemical intermediate that the different reaction types diverge and take place[28].

However, vitamin B6 derivatives, that include pyridoxine, pyridoxal, pyridoxal 5'-phosphate and pyridoxamine, are not strictly correlated to catalysis because of many other important biological functions they catalyze. Indeed, B6 may function as antioxidant interacting with singlet molecular oxygen[28], as resistance factors to biological stress in plants[29] and as important virulence factors[30].

PLP is also relevant in gene transcription regulation. Variation in PLP concentration can determine modulatory effects on steroid-induced gene expression in eukaryotic cells[31]. Also in prokaryotes PLP can be crucial in gene transcription regulation, as well demonstrated in functional characterization of two transcription factors belonging to the MocR family: *Listeria monocytogenes*, *Corynebacterium glutamicum*, *Bacillus clausii*, and *Streptococcus pneumoniae* PdxR[32,33,34,35] and *Bacillus subtilis* GabR[36,37]. In the case of the transcription factor PdxR, PLP is responsible for the control of *pdxST* genes coding for PLP synthase, whereas, employing a PLP-dependent mechanism, GabR controls the *gabT* and *gabD* genes expression, both involved in γ aminobutyrate (GABA) metabolism.

Bacillus subtilis GabR

GabR is a transcription factor belonging to the GntR family and the MocR/GabR subfamily. As the MocR family, GabR is widespread in eubacteria.

Bacillus subtilis GabR is a chimeric protein that comprises a N-terminal winged HTH DNA-binding domain and a C-terminal domain characteristic of the fold type I PLP-dependent aminotransferase family. Two domains are connected by a flexible linker 29 amino acids long. GabR crystallizes as a homodimer with the two monomers organized in a head-to-tail domain-swap arrangement[25] (Fig. 3).

GabR C-terminal domain has sequence similarity with functional aminotransferases, such as *Thermus thermophilus* α-aminoadipate aminotransferase. However, some relevant residues for the transamination activity were replaced in GabR. This is the main reason why GabR cannot operate the overall transamination of its substrate GABA. In effect, the

reaction stops at the formation of the external aldimine between GABA and PLP[37].

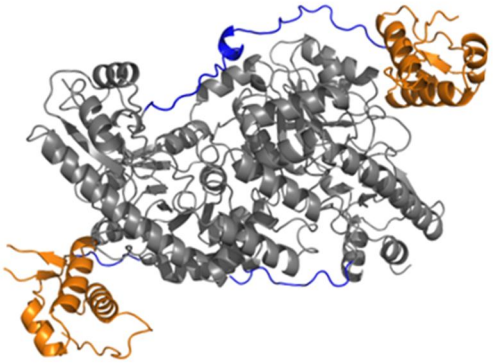

Fig. 3. Quaternary structure of GabR. GabR is a head-to-tail swap homodimer; C-terminal domains are shown in grey (AT-fold), flexible linkers in blue and the N-terminal HTH DNA binding domains in orange.

GabR physiological role consists in the regulation of the GABA metabolism, because GabR controls the expression of the operon *gabTD* that allows the utilization of the extracellular GABA as nitrogen and carbon source.

B.subtilis GABA metabolism is carried out by *gabT* and *gabD* genes that encode two enzymes for an alternative route of glutamate biosynthesis using GABA. GabT is a fold-type I aminotransferase that catalyzes the transamination of GABA to form the products succinic semialdehyde and glutamate. The metabolic pathway is completed by GabD, a NAD-dependent dehydrogenase that converts succinic semialdehyde to succinate.

Despite the inability to catalyze the transamination of GABA, this molecule is crucial in GabR activity as transcription factor. *In vitro* data, demonstrated that the simultaneous interaction of GabR with the ligands PLP and GABA was necessary to promote the *gabT* and *gabD* transcription[37,36]. This suggests that GabR operates as a PLP- and GABA-dependent transcriptional activator. Moreover, GabR is also able to regulate its own transcription, behaving like a GABA-independent negative autoregulator.

The specific interaction of GabR-DNA occurs with an extended DNA region of 47 bp that overlaps the −35 and −10 elements of the *gabTD* and *gabR* promoters, respectively[36] (Fig. 4). In the GabR protected region, two ATACCA direct sequences, separated by 34 bp considering the centers of the repeats, appear to be essential for GabR binding. The

interaction with direct repeats is unconventional in prokaryote homodimeric transcription factors, however the typical quaternary structure of the MocR family members is capable to accommodate this particular interaction[23]. Additionally to GabR, the MocR member *Rhodobacter capsulatus* TauR, that is involved in taurine metabolism, specifically recognizes direct repeats[38].

Based on the GabR crystal structure, Edayathumangalam and colleagues postulated two GabR-DNA binding models: in the first one, a dimer of dimers binds at the promoter, whereby only one HTH domain of each dimer contacts one of the two ATACCA direct repeats. In the second model, a single GabR dimer binds to its recognition site on the promoter, and both the HTH domains are involved in interaction with the direct repeats[25].

Employing different approaches, evidences concerning the GabR-DNA binding stoichiometry were obtained[37,39]; GabR binds the DNA as a dimer, in agreement with the second binding model described above. The interaction of GABA and PLP with GabR does not significantly modify the binding affinity with the *gabTD* promoter.

Given that the GABA and PLP binding does not completely cancel the GabR affinity for the specific DNA binding region, it is possible that the transcription activation of the *gabTD* operon is supported by GABA-dependent conformational changes in GabR. About this point clear information is still lacking, however it has been hypothesized that GABA rearranges the three-dimensional organization either the GabR aminotransferase-like domain or the flexible linker region[25,37,39].

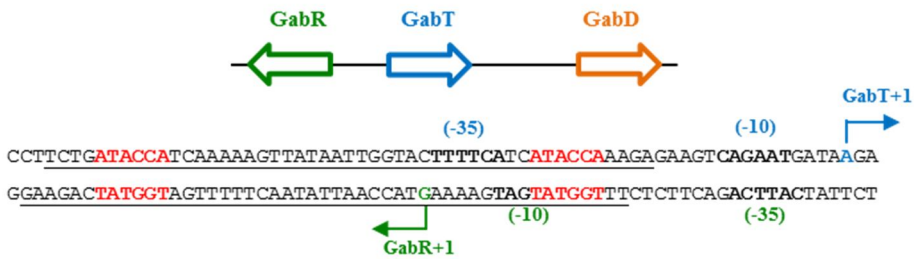

Fig. 4. Schematic representation and DNA sequence of the *gabR* and *gabTD* regulatory region. The *gabR* gene is located upstream of the *gabTD* operon and is divergently transcribed. Direct repeat hexamers recognized by GabR are indicated in red. The -35, -10 and +1 positions of *gabR* and *gabT* genes are in bold and marked in green and blue, respectively. The 47 DNA bp protected by the GabR binding are underlined.

Visualization and analysis of single molecules and protein-DNA complexes using Atomic Force Microscopy

Since the Atomic Force Microscopy (AFM) development in 1986 by Gerd Binnig and Christoph Gerber, this microscopic technique is widely used to study many biological samples. Both, easy sample preparation and the acquisition of three-dimensional high resolution images make of AFM a powerful experimental strategy to analyze single molecules or single nucleoprotein complexes in physiological conditions[40].

The sample deposited onto dedicate supports, such as mica or silicon wafer, is gently scanned by a tiny tip mounted at the end of a flexible cantilever. The tip-sample interaction is monitored by reflecting a laser off the back of the cantilever into photodiode detector (Fig. 5). A feedback loop maintains a constant oscillation amplitude of the tip on the sample surface by moving the scanner vertically at every x,y data point. In this way, the AFM tip gently interacts with the deposited sample without damaging it. Three-dimensional reconstruction of the sample topography (at nanometer lateral and sub-angstrom vertical resolution) is provided by the recording and elaboration of the x, y and z coordinates for each scanned point of the sample surface. Therefore, the analysis of AFM digital images provides detailed and real structural information of single molecules or molecular complexes, concerning their height, perimeter, surface area and volume[41].

However, AFM is employed not only in structural characterization of molecules, but also to determine stoichiometry, binding specificity and affinity of protein (transcription factors, RNAP, restriction enzymes for instance) - DNA complexes[40,42,43].

AFM is a single molecule approach that allows to determine where on the DNA the protein is bound and to distinguish between specific and nonspecific protein-DNA interactions. Additionally, the different molecular species which compose a protein-DNA binding reaction (free protein, free DNA and protein-DNA complexes) are clearly noticeable and representative of those in solution when deposited onto the mica support (Fig. 5). This means that the deposition on the surface does not bias the molecular population by selectively binding one or the other species[44]. Despite other approaches are usually preferred for the characterization of the protein-DNA interactions[45,46], the analysis of a single set of AFM images can simultaneously provide information concerning stoichiometry, binding architecture (height, surface and volume of the complexes) and

thermodynamic parameters (dissociation constant and Gibbs free energy) of a protein-DNA interaction[43].

By counting the different molecular components of a protein-DNA binding reaction, it is possible to obtain a statistically meaningful value for the promoter occupancy. This is defined as the percent ratio between the number of complexes with a protein specifically bound at the promoter site and the total number of DNA molecules (bound and unbound). The promoter occupancy, in conjunction with the protein and DNA concentration used in assembly of the binding reaction, can be employed to determine the microscopic dissociation constant and the correspondent ΔG value of the protein-DNA binding.

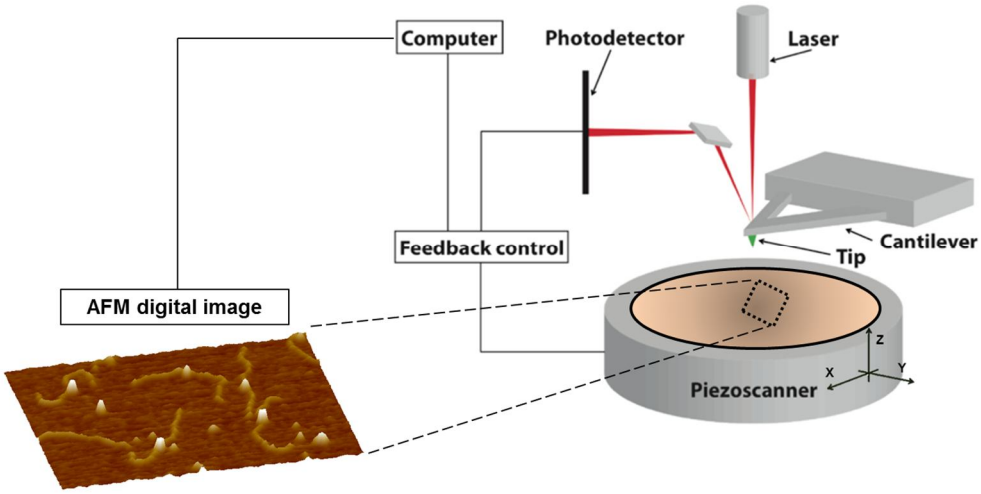

Fig. 5. AFM working principle. AFM relies on the measurement of different interaction forces between atoms of the surface and atoms of the tip. The tip is fixed under a flexible cantilever. The sample is put on a piezoelectric ceramic material controlling movements in the three directions x, y, z. When the sample is brought near the tip, interaction forces between the tip and the sample cause a deflection of the cantilever. A laser beam, reflected on the back side of the cantilever and directed on a photodetector allows to measure the cantilever deflection. Because for each point of the scanned sample the three coordinates are recorded, an image of the sample surface can be recreated. The AFM analysis of a protein-DNA binding reaction provide images in which the DNA fragments and the protein molecules appear as fibril-like structures and globular-like domains, respectively.

Results

Three-dimensional structure of GabR is not enough to obtain precise information about the way in which GabR interacts with the *gabTD* promoter region. Besides the biochemical characterization of the protein, different binding conditions were employed to resolve the GabR-DNA binding architecture. In all binding conditions tested, GabR-DNA binding stoichiometry, geometric features of the nucleoprotein complexes and the microscopic dissociation constant values were evaluated. Additionally, the role of GABA, PLP, the topologic features of the *gabTD* promoter and finally the structural and electrostatic properties of GabR, are discussed to complete the framework in which the interaction between the protein and the *gabTD* promoter region takes place.

Spectroscopic properties of GabR

Additionally to the typical proteins peak at 280 nm, produced by the excitation of aromatic amino acids, holo-GabR showed absorption maxima at 340 and 425 nm (Fig. 6A). The peaks were representative of the internal aldimine equilibrium between enolimine and ketoenamine tautomers respectively[47]. However, in some typical PLP-dependent functional enzymes, the ketoenamine tautomer shows a peak at 412 nm[48,49], whereas the absorption shift toward higher wavelength is often ascribed to external aldimine formation. To assess whether the internal or external aldimine were responsible of the GabR absorption at 425 nm, sodium borohydride ($NaBH_4$) reduction and successive dialysis in 6M urea were carried out. If PLP is involved in internal aldimine formation, the cofactor is covalently bound to the catalytic lysine after reduction and protein denaturation. In this condition the PLP cannot be released from the protein, providing only a maximum absorption at about 340 nm. As reported in Fig. 6A, reduced and denatured GabR showed an absorption at 340 nm, suggesting that PLP interacted with the catalytic lysine to form the internal aldimine.

Through UV-visible spectroscopy, we examined the affinity between holo-GabR and GABA. 8 μM holo-GabR was titrated with GABA using a concentration range from 0 to 35 mM. Addition of GABA to GabR caused a decrease in the absorption peak at 425 nm and an increase at 340 nm (Fig. 6B). This spectrum is attributed to the external aldimine formation between holo-GabR and GABA. By measuring the absorption change recorded

at 420 nm, a dissociation constant value of 2 ± 0.2 mM was estimated, in agreement with other published data[37].

Spectroscopic characterization was also performed on apo-GabR, which was produced through the reaction of holo-GabR and L-cysteine. In fact, the formation of the thiazolidine adduct between L-cysteine and PLP and the further dialysis step promote the PLP removal from the GabR active site. Contrary to holo-GabR, apo-GabR UV-visible spectra did not show the PLP absorption contributions at 340 and 425 nm, demonstrating complete PLP removal from the GabR active site (Fig. 6A).

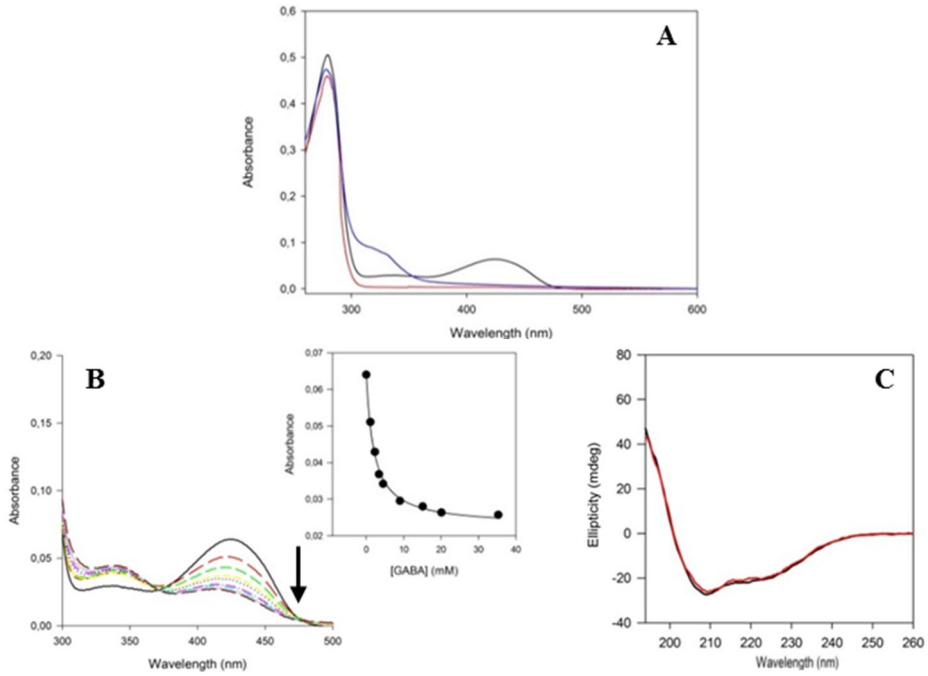

Fig. 6. (A) UV-visible spectrum of holo-GabR (black line). $NaBH_4$ treatment of holo-GabR (blue line) reduces the Schiff base to yield a stable substituted amine. Subsequent dialysis under denaturing conditions does not lead to the release of the bound cofactor that absorbs at around 340 nm, while the peak at 422 nm disappears. UV-visible spectrum of apo-GabR (red line) does not show PLP absorption contribution. **(B)** Titration of holo-GabR with GABA (from 1 to 35 mM) leads to a decrease of the peak centered at 422 nm (dependence of the GabR maximum absorption at 422 nm as a function of the GABA concentration, right graph) and to an increase in the absorption at 330 nm, suggesting the formation of an external aldimine. **(C)** Circular Dichroism spectra of holo-GabR and apo-GabR (black and red lines, respectively).

Circular Dichroism (CD) spectroscopy was employed to estimate the holo- and apo-GabR protein secondary structure content. The spectra were collected in the wavelength range 260-195 nm and presented a band at 208 nm more pronounced than that at 222 nm, suggesting a typical fold of mixed α/β proteins. Spectra deconvolution[50,51] indicated 44% α helix and 13% β sheet content (Fig. 6C) in both holo- and apo-GabR.

GabR is a stable dimer in solution

To evaluate the GabR oligomeric state in solution, size-exclusion chromatography was performed. The predominant elution peak corresponding to the holo-GabR dimeric form (112 KDa) is reported in Fig. 7A. Moreover, a minor peak, partially overlapping the predominant one, may represent a monomeric holo-GabR fraction (56 KDa). The measure did not reveal other peaks representative of the holo-GabR tetrameric form or higher oligomerization states. The result suggested that holo-GabR is a stable dimer in solution, similarly to other aminotransferase fold-type I family members[24]. This quaternary structure is supported by the interaction between the aminotransferase-like domains of two monomers, arranged in a head-to-tail manner[25].

AFM is a leading technique in 3D-high resolution imaging of biological molecules in physiological conditions[40]. The sample deposited onto dedicate supports, mica in our case, is gently scanned by a tiny tip to form a three-dimensional reconstruction of the sample topography. Therefore, the analysis of AFM digital images can provide detailed structural information of single protein molecules concerning their height, perimeter, surface area and volume[41]. We used this microscopic approach to structurally characterized holo- and apo-GabR, to gain insights concerning their geometric properties and oligomeric state.

Holo- and apo-GabR resemble globular-like particles of different dimensions when deposited onto the mica support (Fig. 7C-D). The volume of each single protein molecule was measured and two volume frequency peaks were obtained for both proteins. Comparing these data with a protein molecular weight/AFM protein volume calibration curve, it is deducible that holo- and apo-GabR are organized in monomers and dimers (Fig. 7E). Since size exclusion chromatography data suggests that the holo-GabR is a stable dimer in solution, it is probable that the electrostatic surface of mica promotes the dissociation of the proteins from the dimeric to the monomeric form. Conversely, GabR

maintained the dimeric structure onto mica when it was treated with glutaraldehyde in crosslinking reaction. Under these conditions, the volume peak corresponding to the monomeric form was lacking, whereas the dimeric population was clearly represented. A small peak representing the GabR tetrameric form was also seen.

Because the structural characterization of holo- and apo-GabR did not reveal significant differences between the two GabR forms, in agreement with the crystallographic data, we conclude that PLP is not critical for the GabR folding[25]. However, the higher ratio between the monomeric and dimeric populations detected in apo-GabR compared to the holo-form (Fig. 7E), suggested that PLP supports the interactions between the aminotransferase-like domains, necessary to the GabR oligomerization.

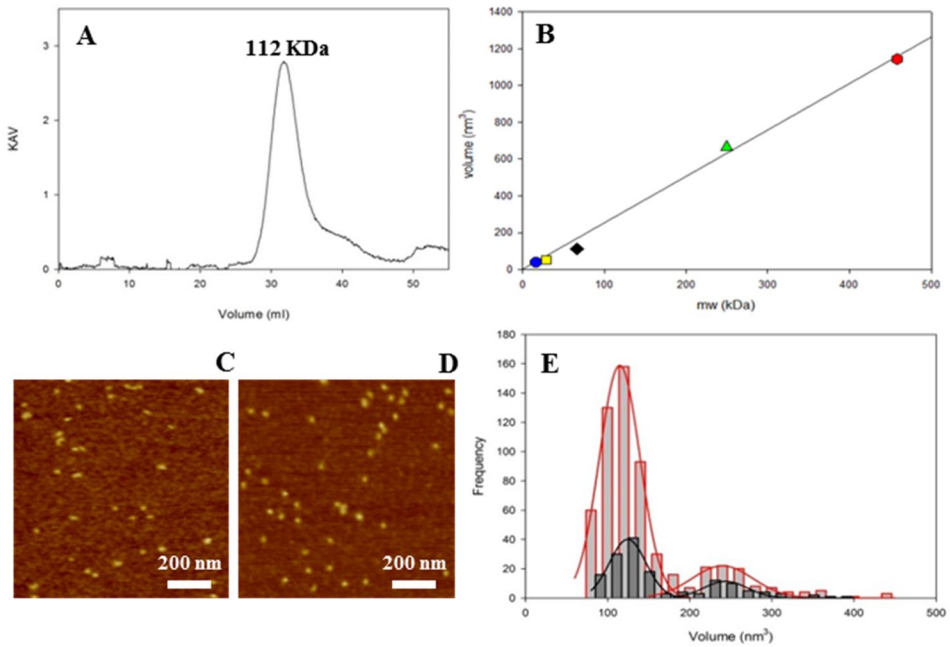

Fig. 7. (A) Size exclusion chromatography elution profile of GabR: the predominant peak is representative of the GabR dimer, 112 KDa. **(B)** Calibration line to correlate molecular weight and AFM volume of GabR and GabR-DNA complexes. *Equus caballus* myoglobin (17 kDa), bovine pancreas DNase I (30 kDa), bovine serum albumin (66.5 kDa), bovine liver catalase (250kDa) and *Escherichia col*i RNA polymerase-σ70 (458 kDa) were used to obtain the calibration line. **(C-D)** AFM images of holo- and apo-GabR. **(E)** Volume distribution of holo-GabR (black) and apo-GabR (red); two volume peaks at ≈125 and ≈250 nm^3 represent the GabR monomeric and dimeric forms respectively.

Investigation of the protein-DNA binding properties using AFM

In our study the AFM analysis was extensively employed to investigate the GabR- *gabTD* promoter interaction under different experimental conditions. DNA fragments used in the experiments and harboring *gabTD* promoter were 800 bp long. We have considered the interaction to be specific when 480 and 320 bp separate the GabR binding site from the DNA ends of the fragment respectively (Fig. 8A) and a 0.65 ratio between the DNA short arm (320 bp) and the DNA long arm (480 bp) is detectable. We determined the promoter occupancy, microscopic dissociation constant (*K*d) and the correspondent ΔG values of the GabR-DNA binding reactions. Moreover, a detail geometric characterization of the nucleoprotein complexes (height, perimeter, surface area and volume) were carried out to estimate the GabR-DNA stoichiometry, complexes conformational changes and variation of DNA binding architecture depending on the experimental conditions. In addition, supplementary evidences concerning the DNA binding architecture were obtained through the measure of the DNA contour length of both specifically bound and unbound DNA molecules. This parameter provides information about the extent of the interaction between GabR and DNA, suggesting for possible changes in complexes three-dimensional organization.

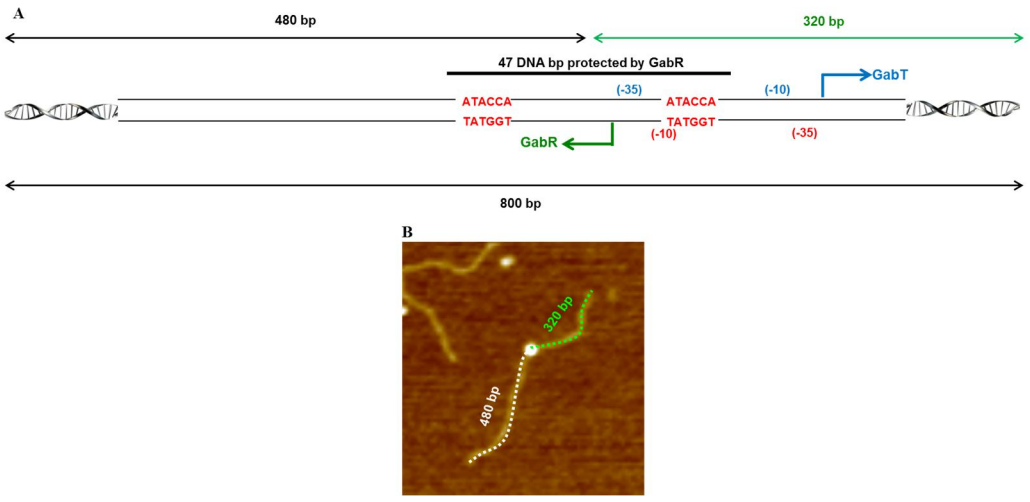

Fig. 8. (A) Schematic representation of the 800bp DNA fragment harboring *gabTD* promoter used in AFM analysis. Direct repeats recognized by GabR are marked in red. The mid-point of the DNA region protected by GabR is located 480 bp from 5' (DNA long arm) and 320 bp from the 3' (DNA short arm). **(B)** Representative AFM image of a specific GabR-*gabTD* nucleoprotein complex.

High affinity interaction between holo-GabR and the gabTD promoter is impaired by GABA

The expression of the *gabTD* operon is tightly connected to the transcription regulator activity of GabR. *In vitro* transcription assay demonstrated that the simultaneous interaction of GabR with its ligands, GABA and PLP, promotes the *gabT* and *gabD* gene expression[36,37].

To assess the role of GABA in holo-GabR *gabTD* promoter interaction we evaluated the binding stoichiometry, the binding affinity and structural features of the nucleoprotein complexes.

First, we investigated the holo-GabR *gabTD* binding in the absence of GABA (Fig. 9A).

For each DNA fragment bound by GabR we measured the DNA contour length, the GabR binding specificity with the promoter and the geometric parameters of the nucleoprotein complexes.

Our first aim was to define the holo-GabR-*gabTD* promoter binding stoichiometry. Two binding models have been postulated on the basis of the GabR crystal structure: in first one, two GabR dimers bind at the promoter, whereby only one HTH domain of each dimer contacts one of the two ATACCA direct repeats. In the second model, a single GabR dimer binds to its recognition site on the promoter, involving both the HTH domains in interaction with the direct repeats[25].

In agreement with the second binding model, published data demonstrated that GabR interacts as a dimer with the *gabTD* promoter [37,39], protecting 47 DNA bp and assembling a nucleoprotein complex of about 145 KDa.

To define GabR-*gabTD* binding stoichiometry, the volume of all specific complexes imaged by AFM was measured and the volume distribution showed a mean value of 260 ± 68 nm^3 (Fig. 9C). In our calibration curve, the volume mean value matched with a molecular weight of about 125 kDa, less than the expected molecular weight of the nucleoprotein complex composed by a single GabR dimer. Thus, our results support the hypothesis that holo-GabR binds at the *gabTD* promoter in a dimeric form. Indeed, none of the complexes has a volume of about 600 nm^3, corresponding to that of a hypothetic nucleoprotein complex composed of two GabR dimers (≈250 kDa).

Given that the mica surface promoted GabR dissociation to the monomeric form, DNA binding reactions were also investigated after glutaraldehyde crosslinking. Under these conditions, no significant differences in volume and geometric parameters of the complexes, with respect of the same data collected in uncrosslinked GabR-DNA binding reactions, were detected. These evidences demonstrate that the interaction with DNA is sufficiently strong to protect the GabR dimeric structure from dissociation.

To test the effect of GABA on the holo-GabR *gabTD* promoter interaction, saturating concentration of the ligand (10 mM for 15' at 25°C) was preincubated with holo-GabR, prior to DNA incubation (Fig. 9B). In this case, nucleoprotein complexes volume showed a mean value of 276 ± 83 nm^3 (Fig. 9D), similar to that obtain in binding experiments performed in the absence of GABA, suggesting that the ligand does not affect the DNA binding stoichiometry of holo-GabR.

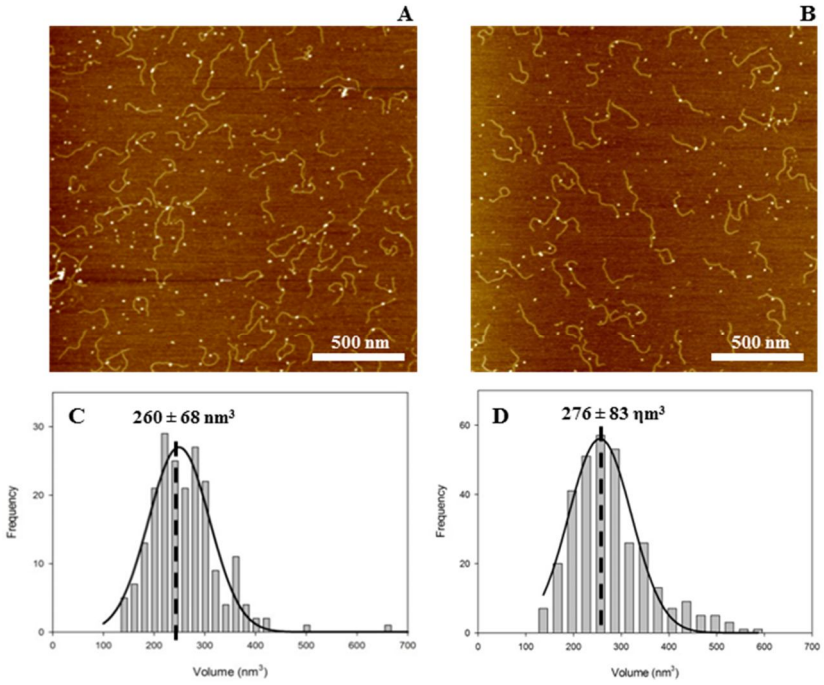

Fig. 9. AFM images of holo-GabR-*gabTD* promoter complexes obtained in the absence **(A)** and presence of GABA respectively **(B)**. **(C-D)** Volume distribution of holo-GabR-*gabTD* complexes formed in the absence **(C)** and presence of GABA **(D)**.

We estimated a *K*d of 41 ± 2 nM and a promoter occupancy of 28.9 ± 1.4% concerning the interaction between holo-GabR and the *gabTD* promoter in the absence of GABA. On the contrary, a *K*d of 92 ± 2 nM and a promoter occupancy of 15.5 ± 0.3% were calculated in the binding reactions assembled with 10 mM GABA. Moreover, the negative effect of GABA on the holo-GabR DNA binding affinity was also reflected by the corresponding ΔΔG of +2 kJ/mol.

Geometric characterization of specific complexes revealed some significant morphologic differences depending on GABA. As summarized in table 1, GABA promoted the formation of complexes smaller in maximum height and wider in surface and perimeter compared to those viewed in the absence of GABA. This observation suggests that without GABA the DNA compaction around the protein was higher than that detected with the ligand. The result was also confirmed by the contour length measurements (Fig. 10). In the absence of GABA the DNA contour length distribution showed a mean value of 253.7 ± 0.6 nm, 6.7 ± 0.4 nm shorter than the contour length detected with GABA: 260.4 ± 1.1 nm. AFM structural data demonstrate that GABA promotes morphologic rearrangements in nucleoprotein complexes. Unfortunately, AFM image resolution is not sufficiently high to detect precise GabR conformational changes induced by GABA. However geometric parameters of complexes and DNA contour length data may suggest that the ligand supports alternative GabR-*gabTD* binding architecture essential for transcription activation of *gabT* and *gabD* genes.

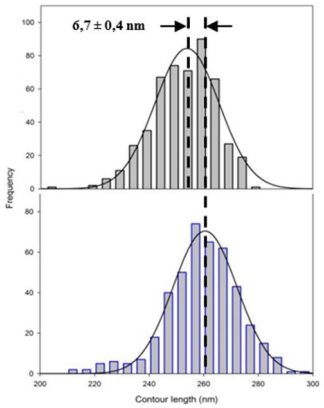

Fig. 10. DNA contour length distributions of specific holo-GabR complexes assembled onto a 800 bp long DNA fragment harboring the *gabTD* promoter, in the absence (top panel, grey) and presence of 10 mM GABA (bottom panel, blue).

Table 1. summary of the geometric features of the complexes and computed dissociation constants and free energies of associations.

Binding conditions	Height (nm)	Perimeter (nm)	Area (nm^2)	Volume (nm^3)	DNA contour length (nm)	Promoter occupancy (%)	Kd (nM)	ΔG (kJ/mol)
holo-GabR + *gabTD*	1.7 ± 0.4	53 ± 7	275 ± 56	260 ± 68	253.7 ± 0.6	28.9 ± 1.4	41 ± 2	-42.0 ± 0.1
holo-GabR + *gabTD* + 10 mM GABA	1.5 ± 0.4	59 ± 8	324 ± 67	276 ± 83	260.4 ± 1.1	15.5 ± 0.3	92 ± 2	-40.0 ± 0.1

PLP does not affect the GabR binding to DNA

DNA binding stoichiometry, binding affinity and structural features of nucleoprotein complexes, in the absence and presence of GABA were investigated to gain insight concerning the relevance of PLP in GabR-DNA binding function (Fig. 11A-B).

Apo-GabR DNA binding stoichiometry was evaluated by measuring the volume of specific complexes. Similarly to holo-GabR, it was obtained a volume mean value of 284 ± 77 nm^3, again suggesting that apo-GabR interacts with the promoter as a dimer. The presence of 10 mM GABA did not modify the binding stoichiometry, as indicated by a volume mean value of 323 ± 103 nm^3.

However, geometric characterization of the apo-GabR-*gabTD* promoter complexes showed significant changes in morphology induced by GABA. In effect, increasing in complexes perimeter and surface area were detected in the presence of the ligand (Table 2).

In addition, DNA binding architecture changes were showed by DNA contour length analysis. DNA compaction of 5.5 ± 0.9 nm was recorded without GABA (Fig. 11C), suggesting a more relaxed interaction between the apo-GabR and the promoter in the presence of the ligand.

Negative effect of GABA in apo-GabR DNA binding properties was highlighted through the microscopic dissociation constant measurements. In apo-GabR-DNA binding reactions we estimated *K*d values of 44 ± 4 nM and 121 ± 3 nM in the absence and presence of GABA respectively and corresponding ΔΔG of +2.6 kJ/mol. Since *K*d values calculated in apo- and holo-GabR DNA binding reactions were not significantly different (corresponding ΔΔG value of +0.1 kJ/mol), PLP is not required to support the GabR-DNA interaction, although it is directly involved in transcription activation of the *gabTD* operon. Moreover, because of the GABA effects on DNA binding affinity were also detectable in apo-GabR, it seems that PLP is not crucial in GABA sensing mechanism. It means that

GABA impairs the GabR binding affinity at *gabTD* promoter independently of the external aldimine formation with PLP.

Relating the thermodynamic parameters calculated in binding reactions between holo-GabR-*gabTD* and apo-GabR-*gabTD* in the presence of GABA, it is possible to evaluate the influence of PLP and GABA in GabR-DNA complex assembly. In these binding conditions a ΔΔG of 2.7 kJ/mol was obtained. Since the sum of the ΔΔG that separately describe the effect of GABA (+2.6 kJ/mol) and PLP (0.1 kJ/mol) on GabR-DNA binding provides the "global" ΔΔG of 2.7 kJ/mol, we can conclude that the ligands influence the GabR transcriptional regulation activity by an independent mechanism.

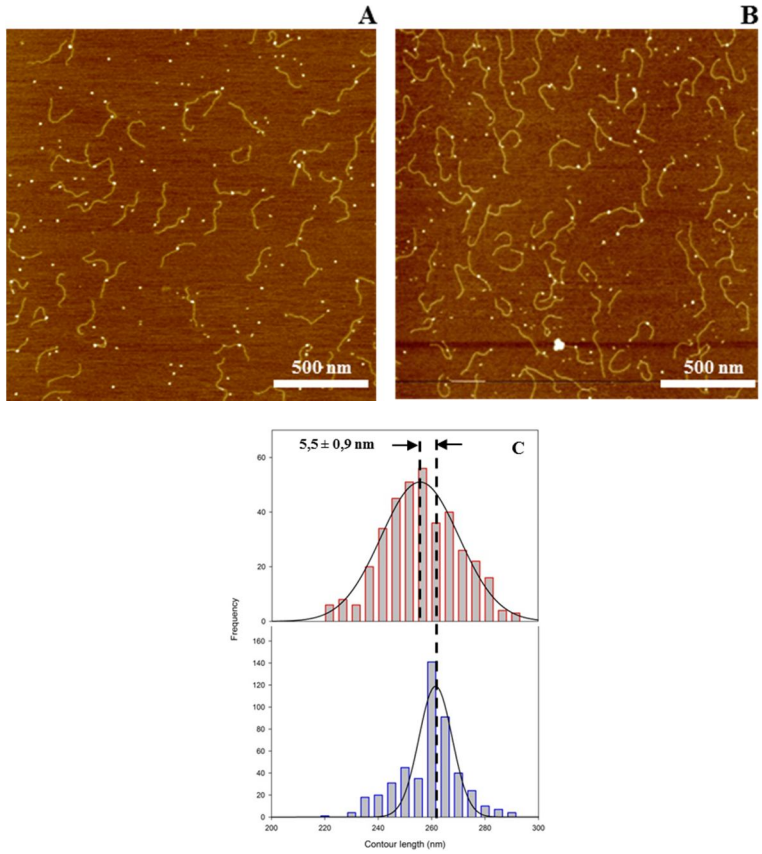

Fig. 11. AFM images of apo-GabR-*gabTD* promoter complexes obtained in the absence (**A)** and presence of GABA **(B)** respectively. **(C)** DNA contour length distributions of specific apo-GabR complexes assembled onto a 800 bp long DNA fragment harboring the *gabTD* promoter, in absence (top panel, grey) and presence of 10 mM GABA (bottom panel, blue).

Table 2. Comparison between holo- and apo-GabR. Summary of the geometric features of the complexes and computed dissociation constants and free energies of associations.

Binding conditions	Height (nm)	Perimeter (nm)	Area (nm^2)	Volume (nm^3)	DNA contour length (nm)	Promoter occupancy (%)	Kd (nM)	ΔG (kJ/mol)
holo-GabR + *gabTD*	1.7 ± 0.4	53 ± 7	275 ± 56	260 ± 68	253.7 ± 0.6	28.9 ± 1.4	41 ± 2	-42.0 ± 0.1
holo-GabR + *gabTD* + 10 mM GABA	1.5 ± 0.4	59 ± 8	324 ± 67	276 ± 83	260.4 ± 1.1	15.5 ± 0.3	92 ± 2	-40.0 ± 0.1
apo-GabR + *gabTD*	1.9 ± 0.5	52 ± 6	276 ± 48	284 ± 77	254.8 ± 1.6	27.7 ± 2.1	44 ± 4	-41.9 ± 0.2
apo-GabR + *gabTD* + 10 mM GABA	1.9 ± 0.5	61 ± 8	333 ± 69	323 ± 103	260.3 ± 1.8	12.3 ± 0.3	121 ± 3	-39.3 ± 0.1

The role of direct and inverted repeats in GabR binding

Published data regarding the GabR-DNA binding mechanism, indicated that the protein interacts with the *gabTD* promoter via two direct repeat hexamers (ATACCA) placed in the intergenic region between the *gabR* gene and the *gabTD* operon[39,36]. DNA binding architecture of GabR appears intriguing for two reasons: i) in a wide number of cases, the interaction between head-to-tail homodimeric transcriptions factor and their cognate DNA binding sites takes place at palindromic repeat sequences[15,9], ii) the three-dimensional positioning of the GabR HTH DNA binding domains indicated by the crystallographic structure, it is not ideal to simultaneously accommodate two direct binding sites[25].

In the light of these observations, we focused on the *gabTD* promoter sequence, to look for other possible GabR DNA binding modes involving alternative sequences.

Analysis of the *gabR-gabTD* promoter region of different *Bacillus* species showed that the ATACCA direct repeats were conserved (Fig. 12), as well as the TGGTAC sequence placed in the spacer region, which has inverted orientation compared to that of the ATACCA. This inverted sequence was investigated to evaluate its possible role in GabR-*gabTD* promoter interaction. Employing site specific mutagenesis, we introduced mutations into the direct repeats and inverted sequence. Three mutants of the *gabTD* promoter composed of different combinations of GabR binding sequences were obtained (Fig. 13). Next, GabR binding affinity for the mutant promoters was studied by AFM measurements (Table 3).

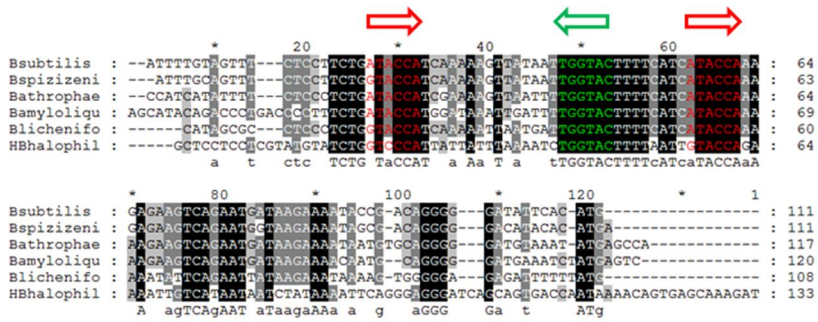

Fig. 12. Alignment of the *gabTD* promoter region of six *Bacillus* species: *B. spizizenii*, *B. athrophaeus*, *B. amyloliquefaciens*, *B. licheniformis* and *Halobacillus halophilus*. Direct repeat hexamers and the inverted sequence are marked in red and green, respectively

The microscopic *K*d regarding the interaction between holo-GabR with the wild-type *gabTD* promoter was of 41 ± 2 nM. Holo-GabR was incubated with the *gabTD sub1* promoter, which was lacking the TGGTAC inverted sequence. In this binding reaction a *K*d of 49 ± 2 nM was calculated, similar to the value determined in the interaction with the wild-type promoter. On the contrary, significant increasing of the *K*d, up to 138 ± 8 nM, was detected on *gabTD sub2* promoter, composed of the inverted sequence and one direct repeat. *K*d values obtained in binding experiments on *gabTD sub1* and *gabTD sub2* promoters, suggested that the simultaneous presence of both direct repeat sequences is crucial for GabR interaction. Moreover, these evidences were supported by the third binding analysis performed on *gabTD sub3* promoter. The presence of only one direct sequence induced a pronounced increasing of the *K*d of 10-fold (464 ± 45 nM) compared to the binding occurred on the wild-type *gabTD*. On the other hand, this was a surprising result because it suggests that the inverted sequence has a specific role in GabR interaction. Indeed, if the inverted sequence does not play any role in interaction with GabR, the estimated *K*d in binding reactions assembled with holo-GabR and *gabTD sub2* / *gabTD sub3* should have be comparable, however, a relevant binding affinity gap was revealed.

Binding experiments performed on mutant *gabTD* promoters, demonstrated that the simultaneous presence of the ATACCA direct repeats represents the main request for nucleoprotein complexes assembling, through their engage with both the GabR HTH DNA binding domains. However, the inverted sequence has a role in GabR-DNA interaction; probably, it is recognized by alternative GabR region different from the HTH domains.

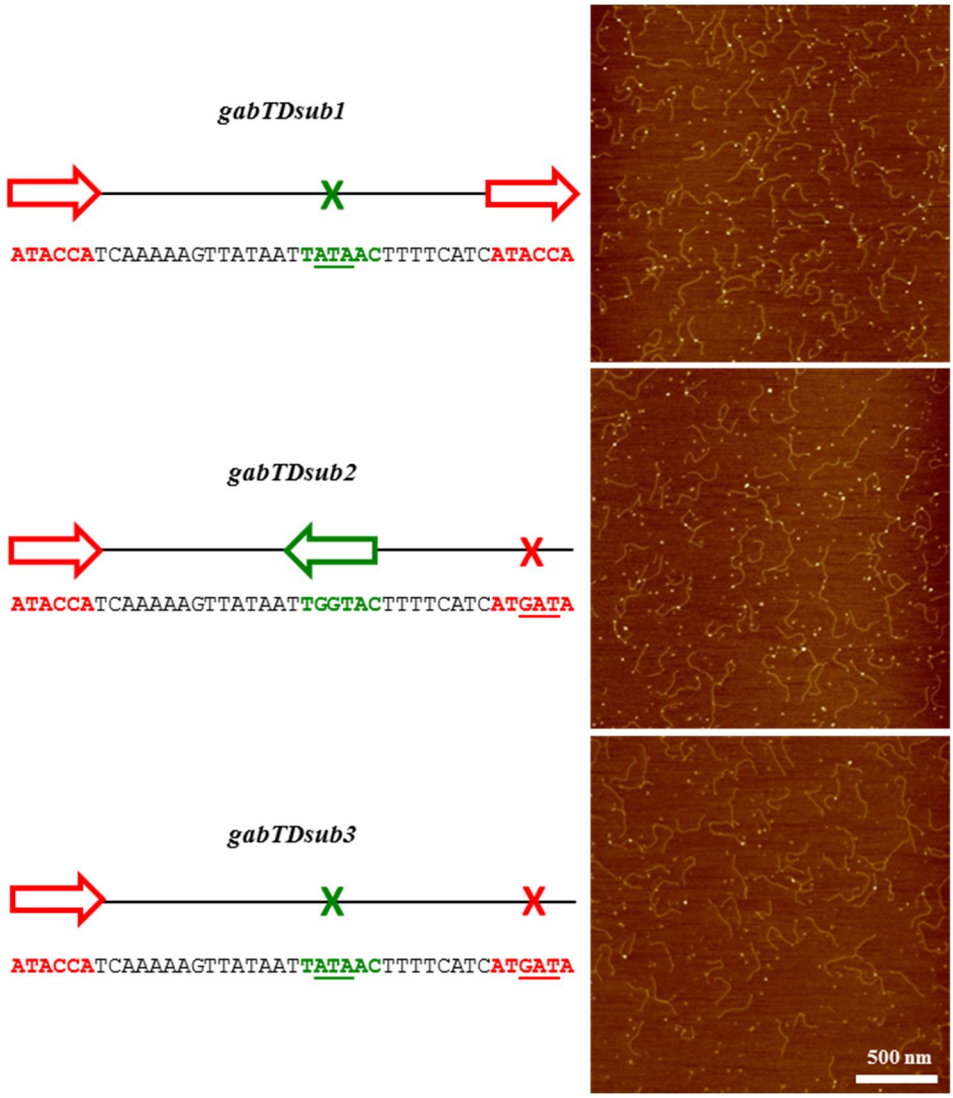

Fig. 13. Schematic representation and DNA sequences of mutant *gabTD* promoters. Direct repeat hexamers and the inverted sequence are marked in red and green, respectively. Mutations obtained through site-specific mutagenesis are underlined. On the right, representative AFM images of holo-GabR complexes formed on the mutant promoters.

GabR binding affinity for three mutant *gabTD* promoters was also measured in the presence of 10 mM GABA. A *K*d increase of about 2.5-fold with respect of the same measure performed in absence of the ligand was recorded in all binding experiments performed.

These interesting evidences indicated that the postulated GabR conformational changes induced by GABA, impaired the DNA binding affinity independently of the *gabTD* promoter sequence. In the light of these results, we hypothesized that the effects of GABA are not only restricted to structural modification in HTH DNA binding domains, but also reflected on the entire GabR molecule.

Table 3. Comparison among holo-GabR interactions at different *gabTD* promoters. Summary of the geometric features of the complexes and computed dissociation constants and free energies of associations

Binding conditions	Height (nm)	Perimeter (nm)	Area (nm^2)	Volume (nm^3)	Promoter occupancy (%)	Kd (nM)	ΔG (kJ/mol)
holo-GabR + *gabTD*	1.7 ± 0.4	53 ± 7	275 ± 56	260 ± 68	28.9 ± 1.4	41 ± 2	-42.0 ± 0.1
holo-GabR + *gabTD* + 10 mM GABA	1.5 ± 0.4	59 ± 8	324 ± 67	276 ± 83	15.5 ± 0.3	92 ± 2	-40.0 ± 0.1
holo-GabR + *gabTDsub1*	1.5 ± 0.5	62 ± 8	361 ± 75	278 ± 88	26.1 ± 0.8	49 ± 2	-41.6 ± 0.1
holo-GabR + *gabTDsub1* + 10 mM GABA	1.3 ± 0.5	63 ± 6	374 ± 58	252 ± 66	12.6 ± 0.7	118 ± 6	-39.4 ± 0.1
holo-GabR + *gabTDsub2*	1.6 ± 0.3	62 ± 5	372 ± 52	290 ± 77	11.0 ± 0.7	138 ± 8	-39.0 ± 0.1
holo-GabR + *gabTDsub2* + 10 mM GABA	1.7 ± 0.6	57 ± 6	308 ± 50	267 ± 94	5.2 ± 0.4	313 ± 27	-36.6 ± 0.1
holo-GabR + *gabTDsub3*	1.7 ± 0.4	61 ± 6	361 ± 57	290 ± 85	3.6 ± 0.3	464 ± 45	-36.0 ± 0.2
holo-GabR + *gabTDsub3* + 10 mM GABA	1.5 ± 0.4	59 ± 6	328 ± 56	251 ± 78	1.6 ± 0.2	1044 ± 136	-34.0± 0.3

The gabTD DNA topology is critical to support a stable interaction with GabR

The *gabTD* promoter is characterized by a rich A/T content in particular in the spacer region that separates the direct repeats recognized by GabR. In these region, 22 out of a total of 29 bp in the spacer region were A or T. DNA fragments rich in A/T are associated with both considerable DNA bending and narrower minor groove than that typically observed in B-DNA molecules[52].

In light of this observation, we investigated the relevance of the *gabTD* promoter topology in GabR binding. In particular we focused on the promoter bending and the positioning of the GabR binding sites on the DNA double helix.

Firstly, we predicted the global three-dimensional structure of the *gabTD* promoter, employing "Christoph Gohlke DNA curvature analysis" software[53]. The prediction attributes to the promoter a significant intrinsic bending (Fig. 14), in agreement with its high A/T content. Moreover, the ATACCA direct repeats recognized by GabR are

separated by 35 bp, thus exposing corresponding chemical determinants on opposite DNA faces.

To evaluate the role of these particular topological arrangements with respect to the GabR binding mechanism, we constructed two *gabTD* mutants with insertions of 5 bp (half DNA turn) and 10 bp (one DNA turn) in the spacer region. The insertion of 5 bp caused relevant alteration on the promoter topology. More specifically, in addition to the increased distance between the direct repeats, the GabR binding sites were exposed on the same DNA face and the intrinsic bending was strongly impaired because the insertion of 5 bp provides to the promoter an inverted curvature (Fig. 14).

On the contrary, the insertion of 10 bp determined only an increase of the distance between the GabR binding sequences, whereas both opposite exposition of direct repeats on the DNA faces and the DNA bending matched with the wild-type *gabTD* promoter arrangement (Fig. 14).

Holo-GabR binding affinity for the *gabTD+5* and *gabTD+10* mutant promoters was evaluated. Compared to the wild-type *gabTD* promoter, the holo-GabR interaction with the *gabTD+5* was substantially compromised; in this case, *K*d value of 498 ± 42 nM and a corresponding promoter occupancy of 3.3 ± 0.3% were estimated.

Conversely and surprisingly, holo-GabR binding affinity for *gabTD+10* promoter was partially recovered. Indeed, *K*d of 57 ± 2 nM and promoter occupancy of 22.6 ± 0.7% appeared only slightly different in respect of the same parameters calculated in the wild-type *gabTD* promoter (Table 4).

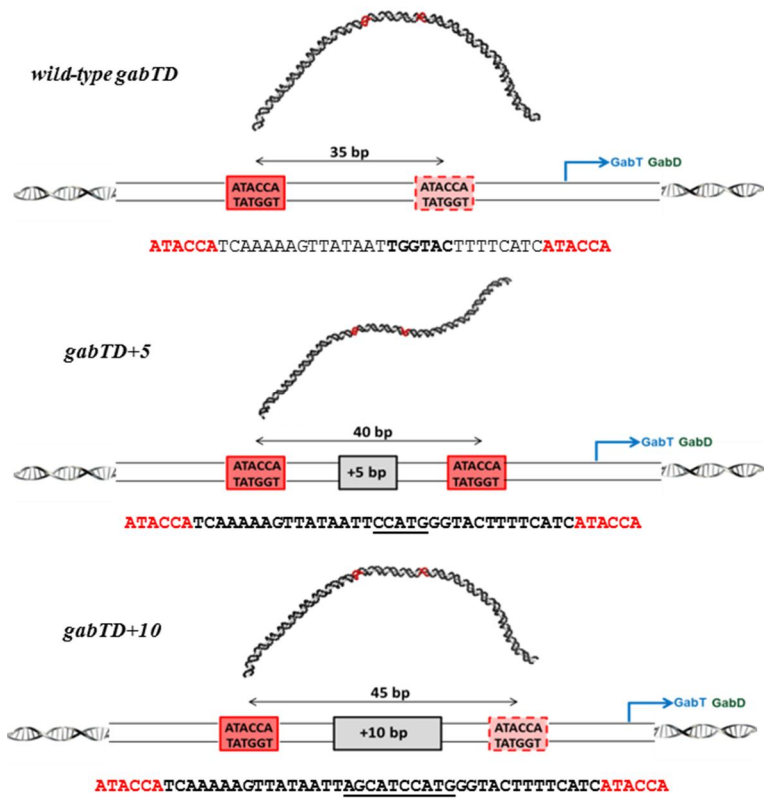

Fig. 14. DNA topology of wild-type *gabTD*, *gabTD+5* and *gabTD+10* promoters. Bending prediction of the promoters are reported in image; the wild-type *gabTD* and the *gabTD+10* show a significant in-phase DNA bending which is lost in *gabTD+5* as well as the opposite exposition of the direct repeats recognized by GabR (marked in red). The 5 and 10 bp inserted in the spacer region are underlined in promoter sequences

These evidences highlighted that both GabR binding site orientation and DNA bending are crucial for the GabR interaction with the *gabTD* promoter, whereas the distance between the direct repeats is not so crucial in GabR binding because the insertion of 10 bp did not substantially affect the holo-GabR affinity for the promoter. Geometric characterization of the specific complexes assembled on the *gabTD+10*, suggested the GabR interacts with the mutant promoter as a dimer, nevertheless the binding architecture was different. In particular we observed that an increased distance between the GabR binding sites promoted the formation of shorter and wider complexes (Fig. 15) compared to those measured on the wild-type promoter, suggesting a possible "GabR elongation" necessary to compensate the addition of one DNA helix turn. It is possible that the long and flexible

linker connecting the HTH DNA binding domain to the AT-fold domain could be directly involved in this accommodation mechanism.

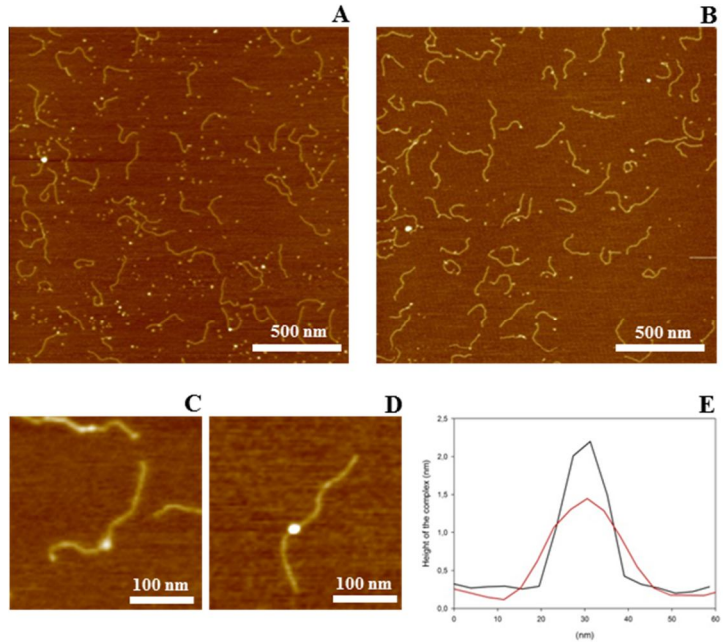

Fig. 15. Representative AFM images of holo-GabR complexes assembled on the mutant promoters *gabTD+5* **(A)** and *gabTD+10* **(B)**. Specific complexes formed between holo-GabR and the *gabTD+10* promoter **(C)** and wild-type *gabTD* promoter **(D)**. Comparison of the profiles of the holo-GabR-*gabTD+10* (red line) and holo-GabR-wild-type *gabTD* (black line) complexes **(E)**.

Table 4. Comparison among holo-GabR interactions at different *gabTD* promoters. Summary of the geometric features of the complexes and computed dissociation constants and free energies of associations

Binding conditions	Height (nm)	Perimeter (nm)	Area (nm^2)	Volume (nm^3)	Promoter occupancy (%)	Kd (nM)	ΔG (kJ/mol)
holo-GabR + *gabTD*	1.7 ± 0.4	53 ± 7	275 ± 56	260 ± 68	28.9 ± 1.4	41 ± 2	-42.0 ± 0.1
holo-GabR + *gabTDmut5*	1.5 ± 0.6	65 ± 10	379 ± 89	307 ± 121	3.3 ± 0.3	498 ± 42	-35.8 ± 0.2
holo-GabR + *gabTDmut10*	1.2 ± 0.3	62 ± 9	350 ± 72	253 ± 56	22.6 ± 0.7	57 ± 2	-41.2 ± 0.1

Differently to homologous aminotransferases, GabR has a positive surface

We next focused on the GabR surface electrostatic features to obtain clues about the possible role of other GabR regions involved in DNA binding mechanism.

As depicted in Fig. 16, GabR dimer showed a positive groove that extends along one entire protein face, connecting the N-terminal HTH DNA binding domains placed at the far ends

of GabR. The positive groove is mainly determined by lysine residues of the AT-fold domain. However, this particular electrostatic arrangement was not detected in other functional aminotransferase enzymes. Because of GabR is a DNA binding protein, it is conceivable that the positive groove is directly involved in interactions with the *gabTD* promoter.

To prove this hypothesis we decreased the GabR surface positive charge producing two GabR mutants. Employing site specific mutagenesis and substituting arginine and lysine with glutamine residues, GabR single and double mutants were obtained (GabR R129Q and GabR K362-366Q). Because GabR binds the DNA as an homodimer, the positive groove of the two mutants loses two and four charged amino acids respectively.

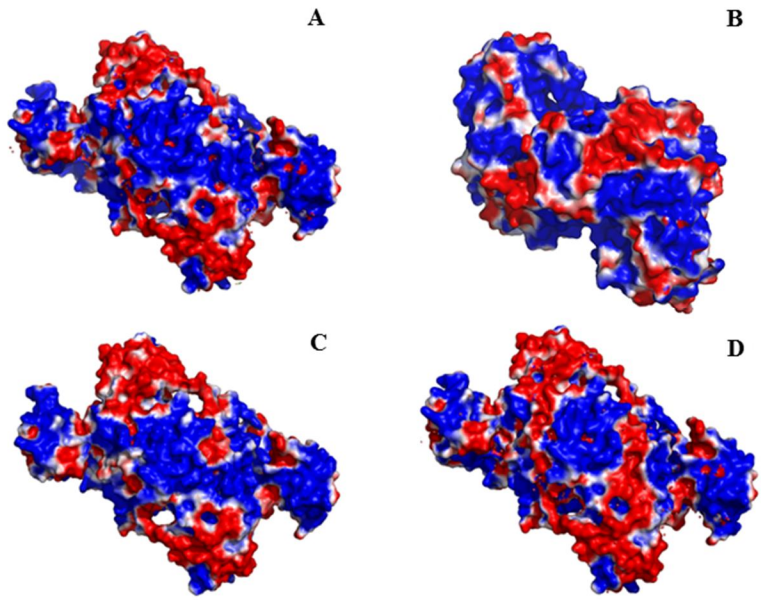

Fig. 16. (A) Electrostatic potential surface map of *Bacillus subtilis* GabR, **(B)** *Thermus thermophilus* 2-aminoadipate aminotransferase (homologous of the GabR AT-fold domain), **(C)** GabR R129Q and **(D)** GabR K362-366Q. Positive and negative potentials are shown in blue and red, respectively. GabR positive groove extends along the entire protein face, however the substitution of lysine residues reduces the positive potential of this region.

GabR mutants reproduce the spectroscopic and structural properties of the wild-type

GabR mutants were biochemically and structurally characterized before proceeding with the DNA binding affinity study. UV-visible analysis did not indicate spectroscopic

differences in GabR mutants with respect to wild-type GabR. Typical GabR internal aldimine maxima absorption were observed at 340 and 425 nm (Fig. 17A). GabR mutants were also titrated with GABA using a concentration range from 0 to 35 mM. GABA determined the decrease of the absorption at 425 nm and the increase of that at 340 nm suggesting for the external aldimine formation. Similarly to the *K*d calculated in wild-type GabR-GABA interaction, in GabR R129Q and GabR K362-366Q were estimated *K*d values of 2.7 ± 0.7 mM and 3.5 ± 0.8 mM respectively.

CD spectra did not show significant differences in GabR mutants secondary structure comparted to that of the wild-type protein. To determine protein stability, thermal unfolding experiments were also carried out employing CD. Thermal denaturation curves showed a similar behavior in wild-type GabR and GabR mutants, which appeared completely unfolded at about 65°C (Fig. 17B-C).

AFM structural characterization demonstrated that the GabR mutants volume distribution appeared comparable to that obtained for the wild-type GabR (Fig. 17D).

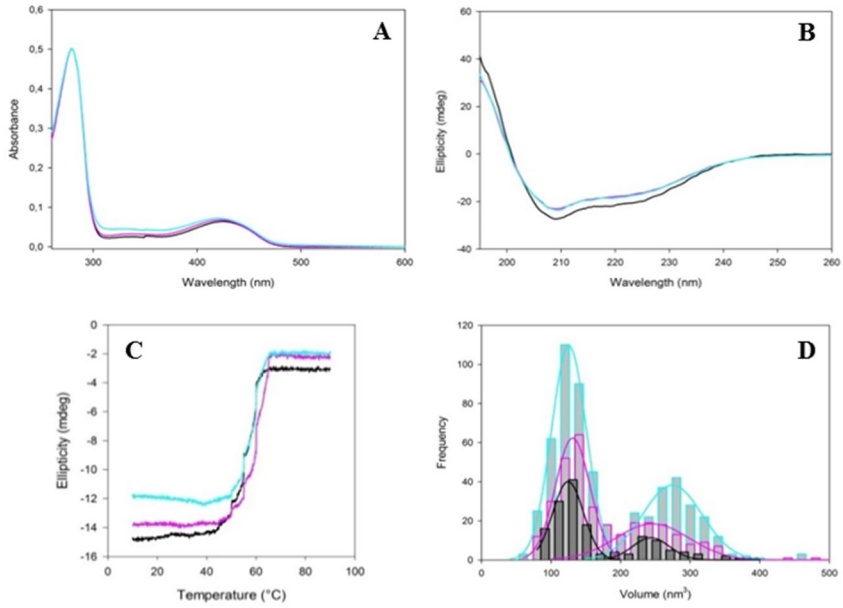

Fig. 17. Biochemical and structural characterization of GabR R129Q (pink) and GabR K362-366Q (cyan). Wild-type holo-GabR (black) is reported in each plot as reference. **(A)** UV-visible spectroscopy and **(B-C)** CD measurements did not show significant differences among the three GabR forms. **(D)** AFM structural analysis revealed two volume peaks at ≈125 and ≈250 nm^3 in GabR mutants, suggesting their dissociation from the dimeric to the monomeric form onto mica.

GabR mutants bind the promoter with low affinity

DNA-binding affinity of GabR mutants were evaluated after their incubation with the wild-type *gabTD* promoter (Fig. 18). Recalling that a *K*d value of 41 ± 2 nM was calculated in the interaction between GabR and the *gabTD* promoter, in the case of the GabR mutants the DNA-binding affinity was reduced. Indeed, we estimated a *K*d of 81 ± 6 nM and a corresponding promoter occupancy of 25.1 ± 1.3%, for the interaction between GabR R129Q and the promoter. Unexpectedly, the DNA-binding capability of GabR K362-366Q was dramatically compromised, as demonstrated by both *K*d and promoter occupancy values, quantified in 1219 ± 95 nM and 4.4 ± 0.3%, respectively (Table 5). These evidences demonstrated that the substitution of four positive residues from the GabR dimer surface impairs the DNA binding mechanism, suggesting that positive GabR face is directly involved in *gabTD* promoter recognition, as well as the N-terminal HTH DNA binding domains.

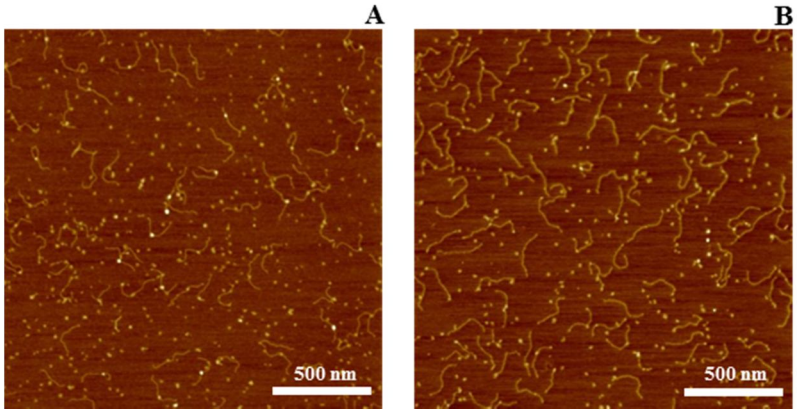

Fig. 18. Representative AFM images of the complexes formed by GabR R129Q **(A)** and GabR K362-366Q **(B)** with the *wild-type gabTD* promoter. The positive surface of GabR is critical to support the specific interaction at the promoter as demonstrated by the low number of the GabR K362-366Q-DNA complexes scored in images.

Table 5. Comparison between the GabR R129Q and GabR K362-366Q interactions at the wild-type *gabTD* promoter. Summary of the geometric features of the complexes and computed dissociation constants and free energies of associations.

Binding conditions	Height (nm)	Perimeter (nm)	Area (nm^2)	Volume (nm^3)	Promoter occupancy (%)	Kd (nM)	ΔG (kJ/mol)
holo-GabR + *gabTD*	1.7 ± 0.4	53 ± 7	275 ± 56	260 ± 68	28.9 ± 1.4	41 ± 2	-42.0 ± 0.1
GabR R129Q + *gabTD*	1.3 ± 0.5	65 ± 10	390 ± 96	264 ± 85	25.1 ± 1.3	81 ± 6	-40.3 ± 0.2
GabR K362-366Q + *gabTD*	1.4 ± 0.5	64 ± 8	364 ± 68	267 ± 100	4.4 ± 0.3	1219 ± 95	-33.6 ± 0.2

Structural model of the GabR-DNA interaction

Published data, in conjunction with the evidences obtained in this study, were used to guide the GabR-*gabTD* docking simulation employing HADDOCK software.

To identify putative GabR DNA binding residues, Edayathumangalam and colleagues investigated *E.coli* FadR, because its HTH DNA-binding domain is homologous to that of GabR and FadR crystal structure complexed with the target DNA is available. Moreover, FadR binds TCTGGT inverted repeat hexamers, similar to the direct repeats recognized by GabR. Simulation of the interaction between one copy of the GabR HTH domain and the sequence recognized by FadR provided an identical binding architecture to that observed in FadR-DNA crystal structure[25]. This result suggests that GabR residues, R43, S52 and K75, are good candidates for the specific recognition of the ATACCA direct repeats. In particular, GabR R43 might interact with one of the guanine in its direct repeat[25].

Because there are evidences that GabR binds DNA as a dimer and that the topology of the promoter DNA is an important determinant of the specific recognition, we performed the docking simulation using a GabR dimer and a DNA fragment harboring the *gabTD* promoter region with an intrinsic curvature of 90°, uniformly distributed between the two direct repeats. Notably, docking simulations performed using a straight DNA fragment did not converge. Figure 19 shows the docking simulation with the best score obtained by defining residue R43 as “active residue” interacting with the ATACCA direct repeat hexamer. As expected the DNA spacer region bends against the protein core making an extended contact that involves the two HTH domains and the intervening surface of the two AT domains of the dimer where the positive groove was located. The α2 helix of the HTH motif invades the major groove to specifically recognize one guanine of the hexamer,

while residues of the α3 interact with the phosphate backbone. On the contrary, the orientation of the other HTH domain is not ideal to perform the same interaction with the second direct repeat sequence. However, during the GabR-DNA complex assembly, it possible that the long and flexible linker, that connects the HTH domain to the AT-fold domain, promotes three-dimensional rearrangements necessary to correctly interact with the direct repeat sequence.

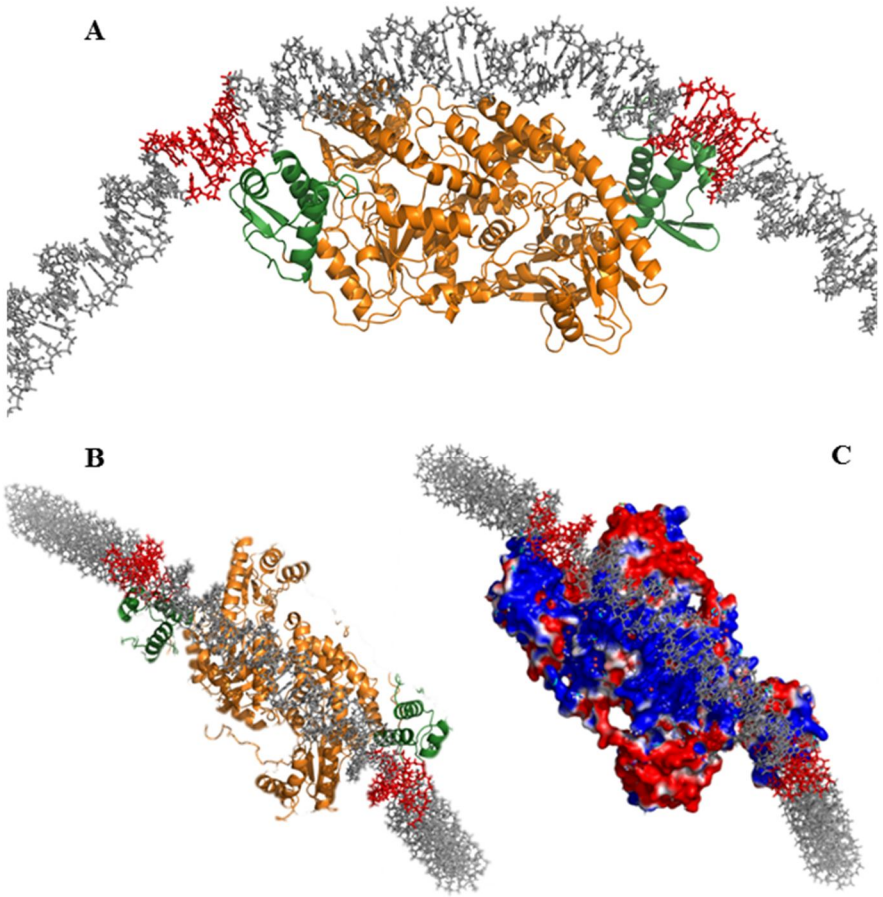

Fig. 19. GabR-DNA interaction model. The AT-fold domains of the GabR dimer are shown in orange, whereas the HTH DNA-binding domain in green. The bent DNA is shown in sticks and the direct repeat hexamers specifically recognized by GabR are marked in red. **(A)** Front view and **(B)** top view of the GabR-DNA complex. **(C)** Top view of the complex in which the GabR dimer is shown in red and blue depending on the electrostatic potential of the surface.

Importantly, this model suggest that several hydrogen bonds can be formed among the GabR residues R129, K133, K362, K365, K366 and DNA phosphates (Fig. 20). In particular, K133 appears to point into the DNA minor groove of the inverted sequence TGGTAC which we have identified as an element directly involved in the interaction. It should be pointed up that none of the AT domain residues was defined as "active residue" in the docking simulation.

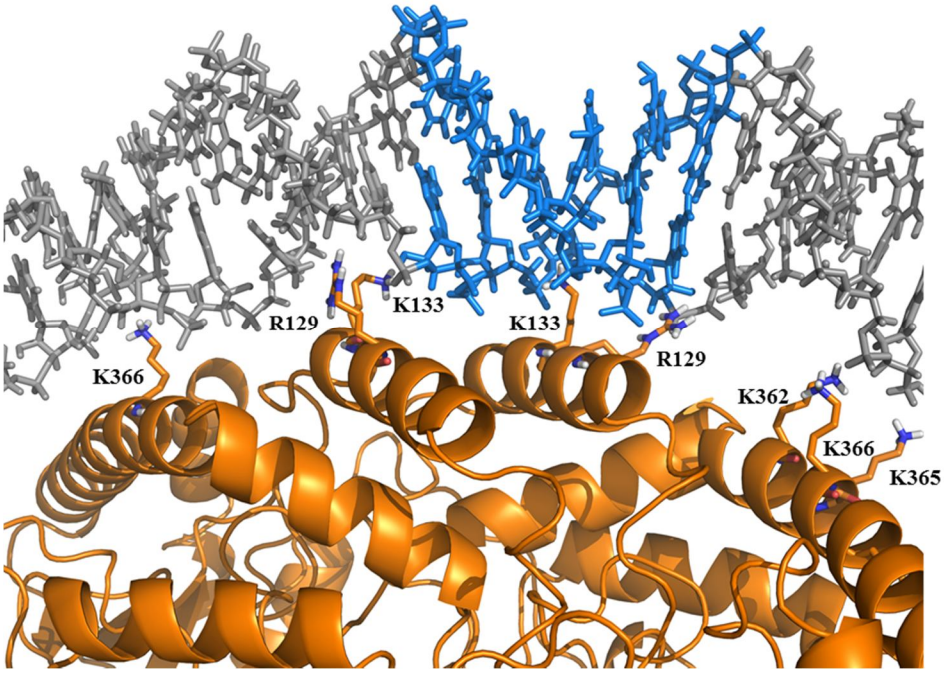

Fig. 20. GabR residues involved in hydrogen bonds formation with the bent DNA. The TGGTAC inverted sequence is marked in blue. These interactions between GabR and the DNA occur with the spacer DNA region which separates the direct repeat sequences specifically recognized by the GabR HTH domains.

Binding conditions	Total DNA molecules	N° specific complexes	N° aspecific complexes	Height (nm)	Perimeter (nm)	Area (nm^2)	Volume (nm^3)	Promoter occupancy (%)	Kd (nM)	ΔG (kJ/mol)
holo-GabR + *gabTD*	1641	496	281	1.7 ± 0.4	53 ± 7	275 ± 56	260 ± 68	28.9 ± 1.4	41 ± 2	-42.0 ± 0.1
holo-GabR + *gabTD* + 10 mM GABA	2738	429	302	1.5 ± 0.4	59 ± 8	324 ± 67	276 ± 83	15.5 ± 0.3	92 ± 2	-40.0 ± 0.1
apo-GabR + *gabTD*	1334	371	214	1.9 ± 0.5	52 ± 6	276 ± 48	284 ± 77	27.7 ± 2.1	44 ± 4	-41.9 ± 0.2
apo-GabR + *gabTD* + 10 mM GABA	2958	471	352	1.9 ± 0.5	61 ± 8	333 ± 69	323 ± 103	12.3 ± 0.3	121 ± 3	-39.3 ± 0.1
holo-GabR + *gabTDmut5*	961	33	57	1.5 ± 0.6	65 ± 10	379 ± 89	307 ± 121	3.3 ± 0.3	498 ± 42	-35.8 ± 0.2
holo-GabR + *gabTDmut10*	532	124	93	1.2 ± 0.3	62 ± 9	350 ± 72	253 ± 56	22.6 ± 0.7	57 ± 2	-41.2 ± 0.1
holo-GabR + *gabTDsub1*	1471	386	198	1.5 ± 0.5	62 ± 8	361 ± 75	278 ± 88	26.1 ± 0.8	49 ± 2	-41.6 ± 0.1
holo-GabR + *gabTDsub1* + 10 mM GABA	1984	253	164	1.3 ± 0.5	63 ± 6	374 ± 58	252 ± 66	12.6 ± 0.7	118 ± 6	-39.4 ± 0.1
holo-GabR + *gabTDsub2*	1988	225	184	1.6 ± 0.3	62 ± 5	372 ± 52	290 ± 77	11.0 ± 0.7	138 ± 8	-39.0 ± 0.1
holo-GabR + *gabTDsub2* + 10 mM GABA	2086	113	133	1.7 ± 0.6	57 ± 6	308 ± 50	267 ± 94	5.2 ± 0.4	313 ± 27	-36.6 ± 0.1
holo-GabR + *gabTDsub3*	3086	116	149	1.7 ± 0.4	61 ± 6	361 ± 57	290 ± 85	3.6 ± 0.3	464 ± 45	-36.0 ± 0.2
holo-GabR + *gabTDsub3* + 10 mM GABA	1733	26	53	1.5 ± 0.4	59 ± 6	328 ± 56	251 ± 78	1.6 ± 0.2	1044 ± 136	-34.0 ± 0.3
GabR mut1 + *gabTD*	2012	486	285	1.3 ± 0.5	65 ± 10	390 ± 96	264 ± 85	25.1 ± 1.3	81 ± 6	-40.3 ± 0.2
GabR mut2 + *gabTD*	2252	101	130	1.4 ± 0.5	64 ± 8	364 ± 68	267 ± 100	4.4 ± 0.3	1219 ± 95	-33.6 ± 0.2

Table 6. GabR-*gabTD* interaction; summary of the results obtained by the AFM analysis under different experimental conditions. "Total DNA molecules" represent the total number of free DNA molecules, specific and nonspecific complexes counted. We considered specific complexes those with a 0.65 ± 0.1 ratio between the DNA short and long arms. All specific complexes imaged were structurally characterized and the arithmetic mean of the height, perimeter, surface area and volume values measured is reported. Promoter occupancy, microscopic dissociation constant *K*d and corresponding ΔG values describe the GabR binding affinity.

Discussion

Bacillus subtilis GabR is a transcription factor belonging to the MocR/GabR family. As other MocR members, GabR is widespread in eubacteria and composed of an N-terminal HTH DNA binding domain and a fold type I PLP-dependent aminotransferase-like C-terminal domain. The physiological role of GabR consists in the expression regulation of the *gabT* and *gabD* genes, that are critical in GABA metabolism. The GabR interaction with PLP and GABA and the binding on direct repeat sequences at the *gabTD* promoter, switch on the *gabT* and *gabD* genes transcription, whereas in the absence of both ligands, GabR acts alone as a negative transcription autoregulator.

So far, crystallographic evidences provided the dimeric structure of GabR[25], while the role of the ligands GABA and PLP in GabR genic regulation mechanism of the operon *gabTD* is demonstrated through *in vitro* transcription studies[36,37]. However, many questions about the GabR-DNA binding mode still remain unclear, because only models of interaction have been postulated.

Employing different binding conditions, we investigated the relevance of structure and surface electrostatic properties of GabR, of both *gabTD* promoter sequence and topology and the role of the ligands GABA and PLP, with respect to the GabR-DNA interaction mechanism.

GabR is a dimer in solution and binds PLP as an internal aldimine

In this thesis we analyzed four GabR forms: wild-type GabR in holo- and apo- form and the GabR mutants R129Q and K362-366Q. After recombinant expression and purification different GabR were spectroscopically and structurally characterized. Except apo-GabR, all GabR forms showed the internal aldimine peaks centered at 340 and 425 nm and representative of the enolimine and ketoenamine tautomers respectively. Usually the ketoenamine tautomers maximum absorption is detectable at 412 nm, nevertheless it is possible to observed the shift of this peak toward higher wavelength depending on the neighboring structure of the active site[54]. All GabR forms were organized in dimeric structures in solution (≈112 KDa) that dissociate to monomers when deposited onto mica during the AFM analysis. Monomeric and dimeric GabR organizations were indicated by two volume distributions. Because of molecules can deposit onto mica in several diverse

three-dimensional orientations, the same structural organization of the protein can be represented through different volume values[41].

In spite of the structural features matched among the GabR forms, volume analysis showed that the ratio between the total number of monomers and dimers was about two-fold higher in apo-GabR than that measured in the wild-type or mutants holo-GabR forms. Even if the GabR secondary structure was not affected by the presence of PLP, AFM data suggested that the presence of two molecules of the cofactor per GabR dimer can contribute to stabilize the protein oligomerization[55] in conjunction with the strong interaction between the AT-fold domains arranged in a head-to-tail manner.

PLP cannot act as an enzymatic cofactor, because the lacking of critical residues at the GabR active site does not support the catalysis of the overall GABA transamination reaction[25,37]. Therefore, it is thought that both the PLP and the AT-domain conservation during evolution is only due to their role in GabR dimerization and regulation of the GabR transcription factor activity[56].

Dimeric GabR binds two direct repeat hexamers at the gabTD promoter

We characterized the interaction between GabR and the *gabTD* promoter region from different point of view and under different experimental binding conditions. Our evidences demonstrated that holo-GabR interacts as a dimer with the *gabTD* promoter, in agreement with other recent findings obtained using different techniques[37,39]. In addition we observed that the GabR-DNA stoichiometry did not change independently of the presence of PLP in GabR active site, and the *gabTD* sequence and topology. Therefore, our results support the binding model postulated by Edayathumangalam and colleagues, in which a GabR dimer specifically recognized two ATACCA direct repeats involving both the HTH DNA biding domains[25]. It has been suggested that this DNA binding mode could be shared by other regulators of the MocR family[23]; for instance the MocR member *Rhodobacter capsulatus* TauR also recognizes two pairs of almost perfect direct repeats[38]. Nevertheless, the GabR crystallographic data suggested that the three-dimensional orientation of the HTH DNA binding domains was not ideal to bind direct repeats and additionally, it is unusual that an homodimeric transcription factor specifically recognizes direct repeats[38,57]. Because an inverted sequence TGGTAC was revealed in the spacer region between the direct repeats,

we called into question the GabR binding mechanism and the possibility that not only the ATACCA hexamers are involved in interaction but also the inverted sequence. Interestingly, the MocR family member *Bacillus clausii* PdxR, can recognize direct and inverted repeats on the *pdxST* promoter with comparable affinity[34]. Using mutational analysis we separately evaluated the role of the sequences in GabR interaction. Our data showed that the higher GabR affinity for the *gabTD* promoter was detected in the presence of the two direct repeats. However, even if partially, the inverted repeat supports the interaction between GabR and the promoter. These observations suggest that not only the HTH domains are involved in GabR-DNA binding but also other GabR region could be critical for the interaction with the promoter.

Effects of PLP and GABA on the GabR-DNA binding properties

The study of the interaction of holo- and apo-GabR with the wild-type *gabTD* indicated that both proteins have a similar binding affinity for the promoter. The microscopic dissociation constants are comparable in binding reactions assembled with the two protein forms, suggesting that the PLP is not required to support the GabR-DNA binding. Conversely, when the same measures were performed in the presence of GABA, the *K*d values for both holo- and apo-GabR increased of about 2.5 fold compared to those obtained in the absence of GABA. Published *in vitro* transcription results[36,37], demonstrate that both GABA and PLP are directly involved in *gabT* and *gabD* genes expression, however our results indicate that the interaction of GabR with the *gabTD* promoter is affected by GABA also in the absence of PLP, suggesting that the formation of an external aldimine is not required for GabR-DNA binding regulation. Similar evidences were also collected in *Rhodobacter capsulatus* TauR, which binds to the *tpa* promoter independently of the presence of PLP and its effector taurine[38]. Although PLP does not affect GabR-DNA binding properties, in the light of the *in vitro* transcription data, PLP could be exploit its action by regulating the molecular mechanisms of transcription initiation, such as RNAP recruitment or the transition from the closed to the open complex.

Each specific complex assembled with holo- and apo-GabR on the *gabTD* promoter, in the presence and absence of GABA, was morphologically characterized. We detected a decrease in the maximum height and an increase in perimeter and surface area of the complexes in the presence of GABA. It has been postulated that GABA may induce

structural rearrangements of GabR; in particular, the three-dimensional orientation of the HTH DNA binding domains supported by the flexibility of the linkers, could represent the key mechanism that promotes the *gabTD* operon transcription[25]. However, our results indicate that the interaction between GabR and GABA causes changing in DNA binding affinity and complexes morphology independently of the promoter sequence. It should be noted that the GabR binding affinity for the mutant promoter lacking one direct repeat (*gabTDsub2*) was impaired by GABA although one of the HTH domain was likely not involved in a specific interaction with DNA. Therefore, it is possible that the effect of GABA is not restricted to the three-dimensional modifications of the GabR HTH DNA binding domains involved in specific recognition of the promoter sequences. Conversely, conformational rearrangements induced by GABA could occur in the GabR AT-fold domains, in agreement with the structural behavior of some fold-type I aminotransferases upon binding of their specific substrates[37]. Interestingly, it has been reported that the GabR ability to activate the *gabTD* operon transcription was lost in the absence of G253 which acts as a junction between the AT-fold large and small domains, supporting their mobility[58].

The boundaries of the GabR-DNA binding site on the *gabTD* promoter region were identified using DNA footprinting[36]. GabR protects 47 bp that overlaps the -35 position of the *gabTD* promoter and the -10 and +1 positions of the *gabR* promoter. Binding of GabR represses *gabR* transcription, but in the presence of GABA and PLP GabR activates the *gabTD* operon. It has been hypothesized that GABA can promote conformational rearrangements in GabR, necessary to switch on the *gabTD* operon transcription. However, many questions still remain open: given that the GabR binding overlaps the -35 position of the *gabTD* promoter, how can RNAP recognize this critical element? Can GABA induce conformational changes of GabR to improve the accessibility of the -35 promoter region? Alternatively, can the interaction of the RNAP with the promoter be improved by GabR? This recruitment mechanism has already been described for the transcription factor CAP which binds at different promoters (known as class II promoters), by overlapping the -35 element[11]. CAP provides multiple interactions with the RNAP necessary to stabilize the complex and to promote transcription activation, regardless of the RNAP interaction with the -35 element. Our DNA contour length measurements show that the GabR complexes

formed with the *gabTD* promoter in the presence of GABA are less wrapped around the protein and the DNA contour length increases of about 6 nm. Notably, 6 nm of DNA correspond to 19 DNA bp, which is the distance that separates the -35 element from the right boundary of the GabR protected region. Taken together, these data suggest that GABA promotes a more relaxed interaction between the promoter DNA and GabR, probably important to expose the -35 element to RNAP. Our working hypothesis is that GABA triggers a conformational rearrangement into the AT-fold domains, such to weaken the interaction with the DNA, without compromising the simultaneous interaction of both HTH domains with the direct repeat hexamers. Therefore, the accessibility to the -35 positon cannot be ascribable to the loss HTH domain affinity for the right direct sequence. In that case similar *K*d values should have been measured for holo-GabR binding at *gabTD sub2* promoter (lacking the right direct repeat) and holo-GabR binding at the wild-type *gabTD* promoter in the presence of GABA. Nevertheless, *K*d differed significantly, 138 ± 8 nM and 92 ± 2 nM respectively, demonstrating that the right direct repeat recognition is involved in GabR-DNA interaction also in presence of GABA.

gabTD topology and GabR surface properties are important determinants of the protein DNA interaction

The ability of DNA binding proteins to specifically recognize DNA sequences is a hallmark of biological regulatory processes**.** Nucleotide specific interaction often involved the formation of hydrogen bonds between the base pairs and amino acid residues primarily into the DNA major groove. However, this "direct readout" mechanism it is not the only one involved in protein-DNA interactions. In particular, protein-DNA recognition is also connected to the intrinsic three-dimensional features of the molecules, in conjunction with their propensity to carry out conformational adjustment necessary to facilitate a stable interaction ("indirect readout" mechanism)[59,60].

The high A/T content of the GabR DNA binding region confers to the *gabTD* promoter an intrinsic bending and the narrowing of the minor groove. Such of A-tracts periodicity, detected in nucleosomal DNA as well, is responsible of this topological features[52,61]. In addition, the distance of 35 bp between the direct repeat sequences exposes the hexamers on opposite faces of the DNA double helix.

The relevance of the *gabTD* promoter topology in the interaction with GabR was studied employing mutational approach. Our findings shows that the insertion of 5 bp of DNA between the GabR binding hexamers significantly affects *gabTD* topology and strongly compromised the GabR binding affinity. On the contrary, the insertion of 10 bp, that has the effect to increase the distance between the direct repeats but not the phase, recovered part of the GabR binding affinity.

With the aim to better comprehend the GabR-DNA interaction, we also focused our attention on the GabR surface properties. Differently from other homologous aminotransferase enzymes, several lysine residues form a positive groove on the GabR AT-domains surface.

To evaluate the role of this positive surface in the interaction with the DNA, we used site specific mutagenesis to obtain two GabR mutants, R129Q and K362-366Q, with reduced positive charge. For both GabR mutants we observed a decrease in affinity toward the *gabTD* promoter. In particular for GabR K362-366Q the DNA binding affinity was dramatically impaired, suggesting that the positive groove on the AT-domains surface is an important determinant of the DNA interaction. Interestingly, the study of chimeric protein composed of the GabR HTH DNA binding domain and the *Thermus thermophilus* 2-aminoadipate aminotransferase domain (which lacks the positive surface), substituting the GabR C-terminal AT-domain, showed that the binding at the *gabTD* promoter was apparently lower compared to that measured on the wild-type GabR[56]. The relevance of this feature in GabR-DNA binding function was also suggested by the conservation of the superficial positive charge in *Rhodobacter capsulatus* TauR, which shares quaternary structure, direct repeats recognition and probably the entire DNA binding architecture with GabR[38].

Three-dimensional binding model of the GabR-gabTD interaction

In the light of our findings and published data, we generated a structural binding model for the GabR-*gabTD* promoter interaction, in which both direct and indirect readout mechanisms are crucial to the binding.

Dimeric GabR binds two ATACCA direct repeats through both the N-terminal HTH DNA binding domains, even though GabR crystallographic data suggest that the HTH domains orientation is not suitable to accommodate the interaction with the two direct repeats.

However, both the weak interaction at the interface between the HTH- and AT-domains and the flexible and long linker that connects them, can promote the HTH domain reorientation necessary to correctly house it at the direct repeat hexamer[25].

In our binding model, the intrinsic bending of the *gabTD* promoter is a key point. In fact, mutations designed to modify such DNA bending strongly compromise the GabR binding affinity (see *gabTD+5*).

Likewise, the GabR structural and electrostatic properties as well as the *gabTD* promoter topology, are important for the assembly of a stable nucleoprotein complex. Mutations designed to weaken the positive charge on the AT-domains surface dramatically reduce the GabR binding affinity for the promoter (see GabR K362-366Q). Therefore, the positive GabR face is thought to interact with the concave side of bent DNA and interact with it.

In *gabTD* promoter the A-, T-T and A-T tracts facilitate the bending and narrowing of the minor groove. However, despite a pronounced intrinsic bending already occurred, GabR binding induces conformational changes in the *gabTD* promoter, which become more compact and bent[39], increasing the negative potential into the minor groove. Following a mechanism nucleosome-like[62], the positive residues on the GabR AT-domains surface may penetrate into the minor groove providing a crucial stabilizing interaction.

Conclusions

The results presented in this thesis describe the biochemical and structural investigation of the *Bacillus subtilis* transcription factor GabR.

UV-visible spectroscopy was employed to characterized the GabR-PLP interaction and to measure the binding affinity of the effector molecule GABA. PLP forms an internal aldimine in GabR active site, that switches to external aldimine in the presence of GABA. This chemical transition was monitored through the change in the absorbance at 425 nm and 340 nm, as a function of the GABA concentration.

AFM was employed to investigate the GabR-*gabTD* promoter interaction from a structural and thermodynamic point of view. The analysis of geometric features of the nucleoprotein complexes imaged by AFM, allowed us to determine the GabR-DNA binding stoichiometry and binding architecture. By measuring the height, perimeter, surface area and volume of the complexes, we observed that GabR specifically binds the promoter as a dimer under all binding conditions tested. However, conformational rearrangements of the complexes were observed in the presence of GABA. We hypothesize that these structural changes are key molecular mechanism triggered by GABA that activates transcription of the operon *gabTD*.

Because in AFM images the different molecular species which compose the GabR-DNA binding reactions (free protein, free DNA and protein-DNA complexes) are easily discernable and because the discrimination between specific and nonspecific nucleoprotein complexes is feasible, we determined the microscopic *K*d for the GabR-DNA complex.

We estimated *K*d values in the nanomolar range, which increase about 2.5 fold in the presence of GABA, independently of the *gabTD* promoter sequence. Given that effects of GABA on complex structure and GabR-DNA binding affinity were observed under all the binding conditions investigated in this study, it is possible that GABA determines GabR structural rearrangements not restricted to the HTH DNA binding domains.

GabR specifically recognizes two direct repeat hexamers through its HTH DNA binding domains, however we found that the promoter topology is also important to support the protein-DNA interaction. In fact, mutations of the promoter sequence that alter its intrinsic bending and modify the phase of the direct repeat sequences on the DNA double helix, strongly impair the GabR-DNA binding.

Moreover, the positively charged groove found on the surface of GabR, is directly involved in interaction with the DNA by contacting and bending the nucleic acid between the HTH DNA binding domains. Accordingly, mutations affecting the GabR positively charged surface (GabR R129Q and K362-366Q) significantly reduced the binding affinity between the GabR mutants and the promoter.

Relying on these experimental evidences we propose a structural model for the binding of GabR to DNA, in which the *gabTD* promoter region bend around the positive protein core while the HTH DNA binding domains specifically recognize the direct repeat hexamers. This arrangement makes of GabR an example of how DNA topology, superficial features of the protein and the specific binding to the DNA are equally relevant to support a stable protein-DNA interaction.

Finally, the structural evidences collected in this thesis will be useful for finding better conditions for the crystallization of GabR-DNA complexes and for the determination of an high-resolution structure of the complex.

Materials and Methods

Cloning, expression and purification of recombinant GabR

The genomic *Bacillus subtilis* strain WB800N was purified in accordance with the procedure reported by Kalia and collagues[63].

The coding sequence of GabR was amplified by PCR using Taq DNA polymerase with primer forward 5'-AATATAAAGGTCTCAAATGGATATCACGATTACACTC-3' and primer reverse 5'ATAATTGGTCTC AGCGCTATCCCCTGTAACGGGG-3' in standard reaction conditions. The amplified GabR gene was digested with BsaI (isoschizomer of Eco31I) restriction enzyme and ligated to a linearized pASK-IBA3 plus vector (IBA, Göttingen, Germany) with compatible ends. The new pASK-GabR construct is such to add the *Strep-tag II*, a short peptide of eight amino acids to the C-terminus of GabR provided by the sequence of the pASK-IBA3 plus vector. The integrity of the construct was confirmed by DNA sequencing. The pASK-GabR construct was used to transform *E.coli* BL21(DE3) competent cells for protein expression.

10 ml of an O/N culture (30 ml, 37°C) of *E.coli* BL21(DE3) cells harboring the pASK-GabR were used to inoculate 500 ml of Luria-Bertani medium containing 100 μg ml^{-1} ampicillin. Cells were grown for 3 h at 37°C until a OD_{600} of 0.6 was reached. GabR expression was induced with the addition of 200 ng ml^{-1} anhydrotetracycline and the culture was grown at 30°C for another 3 h. Culture was harvested at 4°C and the cellular pellet was resuspended in pH 8 lysis-buffer containing 50 mM NaH_2PO_4, 300 mM NaCl, 0.2 mM PLP, 1mM DTT, 1mM EDTA and 0.2 mM protease inhibitor PMSF (Sigma Aldrich). 1g of cellular pellet was resuspended in 10ml of lysis buffer. Cell lysis was performed by sonication on ice (10 sonication cycles, 15''pulses with 1'30''intervales). The cell homogenate was centrifuged at 13.000 x g for 45' at 4 °C and the resultant supernatant was subjected to the Strep-Tactin column (IBA, Göttingen, Germany) pre-equilibrated with 5 bed volume of lysis-buffer. Then, the column was washed using 5 bed volume of pH8 wash-buffer, composed of 50 mM NaH_2PO_4, 300 mM NaCl, 1mM DTT and 1mM EDTA. GabR was eluted with 3 ml of pH 8 elution buffer consisting of 50 mM NaH_2PO_4, 300 mM NaCl, 1mM DTT,1mM EDTA and 2.5 mM desthiobiotin. The concentration of purified GabR was determined by UV-visible spectroscopy in the wavelength range 600-260 nm, (ε_{280} = 53.290 M^{-1} cm^{-1}). 10X PLP molar excess was then added to the GabR solution that was dialyzed O/N at 4°C, in pH 8 storage buffer consisting

of 50 mM NaH_2PO_4, 300 mM NaCl and 1mM EDTA. Final GabR concentration and purity were determined through UV-visible spectra and SDS-PAGE respectively. Glycerol to a final concentration of 15% v/v was added and GabR was stored at -80°C.

Preparation of gabR apo-form

Apo-GabR was prepared by incubating 100 µl of 140 µM purified holo-GabR with 100 µl of 100 mM L-cysteine at room temperature for 15'. The PLP removal from the GabR active site was monitored through UV-visible spectra in the wavelength range 600-260 nm. O/N dialysis in storage buffer at 4°C and addition of glycerol to a final concentration of 15% v/v were performed prior to permanently store the apo-protein at -80°C.

Site direct mutagenesis, expression and purification of mutant GabR forms

The R129Q and K362-366Q GabR mutants were obtained using 50 ng of pASK-GabR construct as template in QuickChange kit from Stratagene (Santa Clara, CA, USA) and two couples of complementary oligonucleotides containing the mutation as primers (0.5 µM each). In R129Q the primer forward was 5'-CTGGTTCCAGTGCGAGCAAAAAG-3' and the primer reverse was 5'-CTTTTTGCTCGCACTGGAACCAG-3', whereas in the case of K362-366Q the primer forward was 5'-GAATATCAGCAGCATATAAAACA AATGAAGC-3' and the primer reverse was 5'-GCTTCATTTGTTTTATATGCT GCTGATATTC-3'. The integrity of the constructs were confirmed by DNA sequencing prior to transform *E.coli* BL21(DE3) cells. Strategy of expression and purification of the mutant GabR forms was in accordance with the procedure described for the wild-type GabR.

Spectroscopic measurements

Spectroscopic analysis were performed at room temperature in pH 8 storage buffer composed by 50 mM NaH_2PO_4, 300 mM NaCl and 1mM EDTA with Varian CARY400 spectrophotometer. Dissociation constants of holo- and mutant GabR-GABA binding were calculated from saturation curves by monitoring the absorbance variation at 420 and 340 nm depending on the GABA concentration. 8 µM GabR forms were titrated with

increasing GABA concentration (from 0 to 35 mM) at room temperature in HEPES 100 mM pH 7.5 and 100 mM NaCl.

CD spectra were collected using JASCO J-715 spectropolarimeter. Measurements were performed in far-UV (wavelength range form 260-195 nm) in 20 mM NaH_2PO_4 pH 7.5 and at room temperature, except thermal denaturation measurements. Protein secondary structure determination was carried out by DichroWeb[50,51]. Each spectrum is the average of three measurements and is subtracted of the buffer contribution.

Size exclusion chromatography

Gel filtration was performed on Ultrogel AcA44 resin (exclusion limit 200 kDa, operating range 17-175 kDa, column volume 63 ml and void volume 20.4 ml). Study of the oligomerization state of GabR was carried out with 500 µl of GabR 0.35 mg ml^{-1} at 4°C and at a flow rate of 0.2 ml min^{-1} in pH 8 buffer composed by 50 mM NaH_2PO_4 and 300 mM NaCl. The elution profiles were obtained from absorbance at 280 nm. The system was calibrated using markers of different molecular weight: dextran (2000 kDa), aldolase (158 kDa), conalbumine (75 kDa), ovalbumin (44 kDa) and myoglobin (17 kDa).

Cloning, mutagenesis and production of gabTD promoter

The *gabTD* promoter region *Bacillus subtilis* WB800N strain was amplified by PCR from the position -244 to +169 with respect to the transcription start site of the *gabT* gene. The amplicon was digested with EcoRI restriction enzyme and ligated to a linearized pNEB193 vector (New England BioLabs) with compatible ends to obtain the pNEB-*gabTD* construct. Mutant promoters used to characterize the GabR DNA binding properties were obtained by site-directed mutagenesis using primer extension for base insertion. Flanking primers, complementary to the ends of the target sequence and internal primers, containing the mis-matched bases to insert or mutate, were used. During the first round of PCR two reactions were assembled using one flanking and one internal primers. The products were then mixed for the second PCR reaction, that was assembled with the flanking primers. The complementary ends of the products hybridized in this PCR reaction to create the final product, which contains the mutated internal sequence[64]. Table 7 reports the primers used to obtain the pNEB-*gabTD* mutated constructs. The mutated amplicons were digested with

EcoRI and ligated into a linearized pNEB193 vector. In Table 8 are reported the pNEB-*gabTD* mutated constructs.

Table 7: primers used to amplify and to mutate *gabTD* promoter regions. Sequences underlined indicated the insertion or substitution of bp in *wild-type gabTD* promoter region

Constructs obtained	Primers	Sequences 5' – 3'
pNEB – *gabTD*	GabTD-For (*) GabTD-Rev (*)	ATATGAATTCTCCGGCCATCCAGATC ATATGAATTCTGAATTT ACGCTGACC
pNEB - *gabTD+5*	GabTD mut5bp-For GabTD mut5bp-Rev	ATGAAAAGTACCCATGGAATTATAACTTTTTGATGGTATC ATAATTCCATGGGTACTTTTCATCATACCAAAG
pNEB - *gabTD+10*	GabTD mut10bp-For GabTD mut10bp-Rev	AAGTACCCATGGATGCTAATTATAACTTTTTGATGGTATC TAATTAGCATCCATGGGTACTTTTCATCATAC
pNEB - *gabTDsub1*	GabTDsub1-For GabTDsub1-Rev	GATGAAAAGTTATAATTATAACTTTTTG CAAAAAGTTATAATTATAACTTTTCATC
pNEB - *gabTDsub2*	GabTDsub2-For GabTDsub2-Rev	GACTTCTCTTTATCATGATGAAAAG CTTTTCATCATGATAAAGAGAAGTC
pNEB - *gabTDsub3* (**)	GabTDsub2-For GabTDsub2-Rev	GACTTCTCTTTATCATGATGAAAAG CTTTTCATCATGATAAAGAGAAGTC

(*): GabTD-For and GabTD-Rev were also used as flanking primers in site-direct mutagenesis procedure.
(**): the construct pNEB- *gabTDsub3* was produced using the mutant pNEB- *gabTDsub1* construct as template.

Table 8: list of the *gabTD* promoters employed in this study. The promoter sequences between the direct repeats hexamers are reported below. The direct repeat sequences (wild-type ATACCA or mutated hexamers) are underlined.

Constructs	DNA sequence between the direct repeats (5'-3')
Wild-type pNEB-gabTD	ATACCATCAAAAAGTTATAATTGGTACTTTTCATCATACCA
pNEB-*gabTD+5*	ATACCATCAAAAAGTTATAATTCCATGGGTACTTTTCATCATACCA
pNEB-*gabTD+10*	ATACCATCAAAAAGTTATAATTAGCATCCATGGTACTTTTCATCATACCA
pNEB-*gabTDsub1*	ATACCATCAAAAAGTTATAATTATAACTTTTCATCATACCA
pNEB-*gabTDsub2*	ATACCATCAAAAAGTTATAATTGGTACTTTTCATCATGATA
pNEB-*gabTDsub3*	ATACCATCAAAAAGTTATAATTATAACTTTTCATCATGATA

The 803, 808 and 813 bp long DNA fragments used in AFM experiments were obtained by PCR using Taq DNA polymerase and primers: forward 5'-TTGGCGGGTGTCGGGGCTG and reverse 5'-CACAGGAAACAGCTATGACC. All DNA fragments were gel purified and electroeluted using an Elutrap apparatus (Schleicher & Schuell, Keene NH). The recovered DNA was phenol/chloroform extracted, ethanol precipitated and resuspended in 5 mM Tris-HCl pH 8. The DNA concentration was determined by absorbance at 260 nm.

Construction of the MW vs volume calibration curve

The MW vs volume calibration curve used to investigate both the GabR structure and the stoichiometry of the GabR-DNA complexes, was obtained with five globular proteins of known MW: *Equus caballus* myoglobin (17 kDa), Bovine pancreas DNase I (30 kDa), Bovine Serum Albumin (66,5 kDa), Bovine liver catalase (250kDa) and *Escherichia coli* RNA polymerase-σ70 (458 kDa). Each protein was diluted in deposition buffer (4 mM HEPES pH 8, 10 mM NaCl, 2 mM $MgCl_2$) at a concentration of 10-15 nM and deposited onto freshly-cleaved mica for 2 min. The mica disk was then rinsed with milliQ water dried with nitrogen. AFM images of 512×512 pixels were collected with a scan size of 1 μm at a scan rate of 2.5 lines per second. In the case of GabR imaging, the protein was cross-linked with 10 mM glutaraldehyde for 2' prior of deposition. Volume of globular features was measured using the "*Zero basis*" volume algorithm of the Gwyddion software (v2.38). Data were analyzed and graphed using SigmaPlot.

AFM of GabR-DNA complexes

Protein-DNA complexes were assembled using 20 nM DNA and 175 nM protein (holo-, apo- or mutant-GabR) in buffer (100 mM HEPES pH 8, 500 mM KCl and 20 mM $MgCl_2$) and incubating the reaction at 25 °C for 15 minutes. When required, GabR was preincubating with 10 mM GABA at 25 °C for 15' prior to add the DNA. The reaction was diluted ten times in deposition buffer (4 mM HEPES pH 7.4, 10 mM NaCl, 2 mM $MgCl_2$) and 20 μl were deposited onto freshly-cleaved mica for 2 minutes before the surface was rinsed with MillQ water and dried with nitrogen. AFM imaging was carried out "in air" with a Nanoscope IIIA (Digital Instruments, Santa Barbara, CA, USA) microscope equipped with the E scanner and commercial silicon cantilevers (MikroMasch Tallinn,

Estonia) operating in tapping mode. Images of 512×512 pixels were collected with a scan size of 2 μm at a scan rate of 2.5 lines per second.

Images analysis

DNA contour length measurements was performed as described in (REF) using the following contour length estimator: L = (0.963ne + 1.362no) x S/W, where "ne" and "no" are the number of even and odd chain codes respectively, S is the image scan size (2000 nm), W is the image width (512 pixels). Position of the DNA bound GabR was manually selected by clicking the center of the protein with the mouse. The DNA contour length from the protein to the nearest end was defined short-arm while the DNA contour length from the protein to the further end was defined long-arm. Complexes were classified as "specific" when the short arm/long arm ratio was of 0.65 ± 0.1. Outside this range of arm ratio the complexes were classified as "non-specific".

The volume of GabR-DNA specific complexes was measured using Matlab. The complex boundaries were outlined with the free-hand tool and the mean height of the boundary pixels was used as reference background. The volume of the complex was computed by multiplying the surface area of the pixels within the boundary by their average pixel height relative to the reference background.

Dissociation constant determination from AFM images

Since both a clear distinction between specific and nonspecific complexes is feasible in AFM images and the deposition of a protein-DNA binding reaction onto mica reflect the molecular organization in solution[44], the determination of the microscopic dissociation constant of promoter complexes is possible.

Scoring the total number of DNA molecules bound and unbound by GabR, the specific and nonspecific complexes, the promoter occupancy and the microscopic dissociation constants of protein-DNA binding reactions can be estimated[43].

Below is reported the equation which we have used to determine the *K*d:

$$K_{\mathrm{d}} \approx \frac{(1 - O_{\mathrm{SP}}) \times ([\mathrm{P}] - [\mathrm{D}] \times O_{\mathrm{Fragment}})}{O_{\mathrm{SP}}}$$

O_{SP} is the fractional promoter occupancy i.e. the ratio between the number of specific complexes and the total number of DNA molecules in the pool of images. $O_{Fragment}$ is the average number of RNAP bound per DNA which is obtained dividing the total number of complexes (specific and non-specific) by the total number of DNA molecules in the pool of images. [P] is the total RNAP concentration and [D] is the total DNA concentration. Note that the term ([P] – [D] × $O_{Fragment}$) represents a better way to determine the free GabR concentration.

ΔG is related to the binding reaction dissociation constant *K*d and was calculated as reported in the equation below:

$$\Delta G = RT \ln K_d$$

Computation of electrostatic potential map

The PDB2PQR server (http://nbcr-222.ucsd.edu/pdb2pqr_2.0.0/) was used to compute the PQR file from the PDB protein structure coordinates. AMBER forcefield and PROPKA were used to calculate the electrostatic potential at pH 8. The electrostatic potential map was mapped on the protein structure with the PyMOL Plugin APBS 2.1.

GabR-DNA docking

Docking protocol consists of rigid-body docking, semi-flexible refinement stage and final refinement in explicit solvent. Docking simulations were carried out using the GabR crystallographic structure (PDB id: 4N0B) and a 96 bp bent DNA fragment generated using the 3D-DART server (http://haddock.chem.uu.nl/dna)[65]. The DNA sequence corresponded to that from positions -60 to +36 of the *gabTD* promoter with a global DNA bending of 90° uniformly distributed between the two ATACCA direct repeats. GabR and DNA were docked using the data-driven docking program HADDOCK 2.1[66]. Docking was performed by defining R43, S52 and K75 of each GabR monomer as “active residues” interacting with 5’-ATACCA direct repeats.

Bibliography

1. López-Maury, L., Marguerat, S. & Bähler, J. Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. *Nat. Rev. Genet.* **9,** 583–593 (2008).

2. van Hijum, S. A. F. T., Medema, M. H. & Kuipers, O. P. Mechanisms and Evolution of Control Logic in Prokaryotic Transcriptional Regulation. *Microbiol. Mol. Biol. Rev.* **73,** 481–509 (2009).

3. Saecker, R. M., Record, M. T. & deHaseth, P. L. Mechanism of Bacterial Transcription Initiation: RNA Polymerase - Promoter Binding, Isomerization to Initiation-Competent Open Complexes, and Initiation of RNA Synthesis. *J. Mol. Biol.* **412,** 754–771 (2011).

4. Lee, D. J., Minchin, S. D. & Busby, S. J. W. Activating Transcription in Bacteria. *Annu. Rev. Microbiol.* **66,** 125–152 (2012).

5. Robb, N. C. *et al.* The transcription bubble of the RNA polymerase-promoter open complex exhibits conformational heterogeneity and millisecond-scale dynamics: Implications for transcription start-site selection. *J. Mol. Biol.* **425,** 875–885 (2013).

6. Farnham, P. J. & Platt, T. Rho-independent termination: Dyad symmetry in DNA causes RNA polymerase to pause during transcription in vitro. *Nucleic Acids Res.* **9,** 563–577 (1981).

7. Richardson, J. P. & Avenue, E. K. Loadin Rho to Terminate Transcription. **114,** 157–159 (1999).

8. Balleza, E. *et al.* Regulation by transcription factors in bacteria: beyond description: Figure 1. *FEMS Microbiol. Rev.* **33,** 133–151 (2009).

9. Bell, C. E. & Lewis, M. A closer view of the conformation of the Lac repressor bound to operator. *Nat. Struct. Biol.* **7,** 209–214 (2000).

10. Brown, N. L., Stoyanov, J. V., Kidd, S. P. & Hobman, J. L. The MerR family of transcriptional regulators. *FEMS Microbiol. Rev.* **27,** 145–163 (2003).

11. Busby, S. & Ebright, R. H. Transcription activation by catabolite activator protein (CAP). *J. Mol. Biol.* **293,** 199–213 (1999).

12. Price, M. N., Dehal, P. S. & Arkin, A. P. Orthologous Transcription Factors in Bacteria Have Different Functions and Regulate Different Genes. *PLoS Comput. Biol.* **3,** e175 (2007).

13. Parkinson, G. *et al.* Structure of the CAP-DNA Complex at 2.5 Å Resolution: A Complete Picture of the Protein-DNA Interface. *J. Mol. Biol.* **260,** 395–408 (1996).

14. Albright, R. A. & Matthews, B. W. How Cro and lambda-repressor distinguish between operators: the structural basis underlying a genetic switch. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 3431–6 (1998).

15. Lawson, C. L. *et al.* Catabolite activator protein (CAP): DNA binding and transcription activation. *October* **14,** 10–20 (2009).

16. Rohs, R. *et al.* Origins of specificity in protein-DNA recognition. *Annu. Rev. Biochem.* **79,** 233–269 (2010).

17. Zhang, Y., Xi, Z., Hegde, R. S., Shakked, Z. & Crothers, D. M. Predicting indirect readout effects in protein-DNA interactions. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 8337–41 (2004).

18. Xiong, Y. & Sundaralingam, M. Protein–nucleic acid interaction: major groove recognition determinants. *eLS* 1–8 (2001). doi:10.1038/npg.els.0003138

19. Luscombe, N. M., Austin, S. E., Berman, H. M. & Thornton, J. M. An overview of the structures of protein-DNA complexes. *Genome Biol.* **1,** REVIEWS001 (2000).

20. Brennan, R. G. & Matthews, B. W. The Helix-Turn-Helix DNA Binding Motif. *J. biol. Chem* **264,** 22–25 (1989).

21. Aravind, L., Anantharaman, V., Balaji, S., Babu, M. & Iyer, L. The many faces of the helix-turn-helix domain: Transcription regulation and beyond. *FEMS Microbiol. Rev.* **29,** 231–262 (2005).

22. 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* **10,** e0132618 (2015).

23. 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. *J. Biol. Chem.* **277,** 12507–12515 (2002).

24. Eliot, A. C. & Kirsch, J. F. PYRIDOXAL PHOSPHATE ENZYMES: Mechanistic, Structural, and Evolutionary Considerations. *Annu. Rev. Biochem.* **73,** 383–415 (2004).

25. Edayathumangalam, R. *et al*. Crystal structure of Bacillus subtilis GabR, an autorepressor and transcriptional activator of gabT. *Proc. Natl. Acad. Sci*. **110,** 17820–17825 (2013).

26. 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. *Biochem. Biophys. Res. Commun.* **415,** 88–93 (2011).

27. Schneider, G., Käck, H. & Lindqvist, Y. The manifold of vitamin B6 dependent enzymes. *Structure* **8,** R1–R6 (2000).

28. Toney, M. D. Controlling reaction specificity in pyridoxal phosphate enzymes. *Biochim. Biophys. Acta* **1814,** 1407–18 (2011).

29. Ehrenshaft, M., Bilski, P., Li, M. Y., Chignell, C. F. & Daub, M. E. A highly conserved sequence is a novel gene involved in de novo vitamin B6 biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 9374–9378 (1999).

30. Belitsky, B. R., Gustafsson, M. C., Sonenshein, a L. & Von Wachenfeldt, C. An lrp-like gene of Bacillus subtilis involved in branched-chain amino acid transport. *J. Bacteriol.* **179,** 5448–5457 (1997).

31. Tully D, B., Allgood V, E. & Cidlowski J, A. Modulation of steroid receptor-mediated gene expression by vitamin B-6. *Faseb J.* **8,** 343–349 (1994).

32. Belitsky, B. R. Role of pdxr in the activation of vitamin B6 biosynthesis in Listeria monocytogenes. *Mol. Microbiol.* **92,** 1113–1128 (2014).

33. 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* **157,** 77–88 (2011).

34. Tramonti, A. *et al*. Molecular mechanism of PdxR - a transcriptional activator involved in the regulation of vitamin B_6 biosynthesis in the probiotic bacterium *Bacillus clausii*. *FEBS J.* **282,** 2966–2984 (2015).

35. El Qaidi, S., Yang, J., Zhang, J.-R., Metzger, D. W. & Bai, G. The vitamin B_6 biosynthesis pathway in Streptococcus pneumoniae is controlled by pyridoxal 5'-phosphate and the transcription factor PdxR and has an impact on ear infection. *J. Bacteriol.* **195,** 2187–96 (2013).

36. Belitsky, B. R. Bacillus subtilis GabR, a Protein with DNA-binding and Aminotransferase Domains, is a PLP-dependent Transcriptional Regulator. *J. Mol. Biol.* **340,** 655–664 (2004).

37. Okuda, K. *et al.* Role of the aminotransferase domain in Bacillus subtilis GabR, a pyridoxal 5′-phosphate-dependent transcriptional regulator. *Mol. Microbiol.* **95,** 245–257 (2015).

38. Wiethaus, J., Schubert, B., Pfander, Y., Narberhaus, F. & Masepohl, B. The GntR-like regulator TauR activates expression of taurine utilization genes in Rhodobacter capsulatus. *J. Bacteriol.* **190,** 487–493 (2008).

39. Al-zyoud, W. A. *et al.* Binding of transcription factor GabR to DNA requires recognition of DNA shape at a location distinct from its cognate binding site. 1–10 (2015). doi:10.1093/nar/gkv1466

40. Alessandrini, A. & Facci, P. AFM: a versatile tool in biophysics. *Meas. Sci. Technol.* **16,** R65–R92 (2005).

41. Li, Y. L., Meng, Y. F., Zhang, Z. M. & Jiang, Y. Detecting the oligomeric state of Escherichia coli MutS from its geometric architecture observed by an atomic force microscope at a single molecular level. *J. Phys. Chem. B* **118,** 9218–9224 (2014).

42. Doniselli, N. *et al.* New insights into the regulatory mechanisms of ppGpp and DksA on Escherichia coli RNA polymerase-promoter complex. *Nucleic Acids Res.* 1–14 (2015). doi:10.1093/nar/gkv391

43. Yang, Y., Sass, L. E., Du, C., Hsieh, P. & Erie, D. A. Determination of protein-DNA binding constants and specificities from statistical analyses of single molecules: MutS-DNA interactions. *Nucleic Acids Res.* **33,** 4322–34 (2005).

44. Rivetti, C., Guthold, M. & Bustamante, C. Scanning force microscopy of DNA deposited onto mica: equilibration versus kinetic trapping studied by statistical polymer chain analysis. *J. Mol. Biol.* **264,** 919–932 (1996).

45. Hellman, L. M. & Fried, M. G. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nat. Protoc.* **2,** 1849–1861 (2007).

46. Jacques, D. a. & Trewhella, J. Small-angle scattering for structural biology - Expanding the frontier while avoiding the pitfalls. *Protein Sci.* **19,** 642–657 (2010).

47. Hideyuki, M., Ohno, T. & Medina, A. Spectroscopic Analysis of Recombinant Rat Histidine Decarboxylasel Enzyme Purification Procedures-The General Observations on Purified Preparations After. **132,** 433–439 (2002).

48. Bettati, S. *et al.* Role of pyridoxal 5'-phosphate in the structural stabilization of O-acetylserine sulfhydrylase. *J. Biol. Chem.* **275,** 40244–51 (2000).

49. Storici, P. *et al.* Structures of gamma-aminobutyric acid (GABA) aminotransferase, a pyridoxal 5'-phosphate, and [2Fe-2S] cluster-containing enzyme, complexed with gamma-ethynyl-GABA and with the antiepilepsy drug vigabatrin. *J. Biol. Chem.* **279,** 363–73 (2004).

50. Whitmore, L. & Wallace, B. A. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res.* **32,** W668–W673 (2004).

51. Whitmore, L. & Wallace, B. A. Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases. *Biopolymers* **89,** 392–400 (2008).

52. Geggier, S. & Vologodskii, A. Sequence dependence of DNA bending rigidity. *Proc. Natl. Acad. Sci.* **107,** 15421–15426 (2010).

53. Bolshoy, a, McNamara, P., Harrington, R. E. & Trifonov, E. N. Curved DNA without A-A: experimental estimation of all 16 DNA wedge angles. *Proc. Natl. Acad. Sci. U. S. A.* **88,** 2312–2316 (1991).

54. Raman, M. C. C. *et al.* The External Aldimine Form of Serine Palmitoyltransferase STRUCTURAL , KINETIC , AND SPECTROSCOPIC ANALYSIS OF THE WILD-TYPE ENZYME AND. **284,** 17328–17339 (2009).

55. Mishra, V., Ali, V., Nozaki, T. & Bhakuni, V. Biophysical characterization of Entamoeba histolytica phosphoserine aminotransferase (EhPSAT): role of cofactor and domains in stability and subunit assembly. *Eur. Biophys. J.* **40,** 599–610 (2011).

56. Okuda, K. *et al*. Domain characterization of Bacillus subtilis GabR, a pyridoxal 5′-phosphate-dependent transcriptional regulator. *J. Biochem.* **158 ,** 225–234 (2015).

57. Lee, K. a. Dimeric transcription factor families: it takes two to tango but who decides on partners and the venue? *J. Cell Sci.* **103 (Pt 1,** 9–14 (1992).

58. Belitsky, B. R. & Sonenshein, A. L. GabR, a member of a novel protein family, regulates the utilization of gamma-aminobutyrate in Bacillus subtilis. *Mol. Microbiol.* **45,** 569–583 (2002).

59. Mondal, M., Choudhury, D., Chakrabarti, J. & Bhattacharyya, D. Role of indirect readout mechanism in TATA box binding protein-DNA interaction. *J. Comput. Aided. Mol. Des.* 283–295 (2015). doi:10.1007/s10822-014-9828-x

60. Rohs, R. *et al*. The role of DNA shape in protein-DNA recognition. *Nature* **461,** 1248–1253 (2009).

61. Satchwell, S. C., Drew, H. R. & Travers, A. A. Sequence periodicities in chicken nucleosome core DNA. *J. Mol. Biol.* **191,** 659–675 (1986).

62. Richmond, T. J. & Davey, C. A. The structure of DNA in the nucleosome core. *Nature* **423,** 145–150 (2003).

63. Kalia, a, Rattan, a & Chopra, P. A method for extraction of high-quality and high-quantity genomic DNA generally applicable to pathogenic bacteria. *Anal. Biochem.* **275,** 1–5 (1999).

64. Reikofski, J. & Tao, B. Y. Polymerase chain reaction (PCR) techniques for site-directed mutagenesis. *Biotechnol. Adv.* **10,** 535–547 (1992).

65. van Dijk, M. & Bonvin, A. M. J. J. 3D-DART: A DNA structure modelling server. *Nucleic Acids Res.* **37,** 235–239 (2009).

66. Van Dijk, M., Van Dijk, A. D. J., Hsu, V., Rolf, B. & Bonvin, A. M. J. J. Information-driven protein-DNA docking using HADDOCK: It is a matter of flexibility. *Nucleic Acids Res.* **34,** 3317–3325 (2006).

Other research projects

Applications of Atomic Force Microscopy (AFM) and Scanning Electron Microscopy (SEM) to the study of lung surfactants and Dry Powder formulations for inhalation

Pulmonary surfactant (PS) is a complex lipo-protein mixture, synthesized, assembled and secreted into the alveolar spaces by type II pneumocytes[1]. PS exists in all air-breathing vertebrates, although with different composition. Nevertheless, the composition of mammalian PS is similar among diverse species[2,3], approximatively 90% and 10% by weight in lipids and proteins, respectively. The lipid fraction consists mainly of phospholipids (90-95%) and neutral lipids, such as cholesterol (5-10%). Phosphatidylcholine represents ≈ 80% of the total phospholipids fraction, whereas the remaining 20% are unsaturated phospholipids. Four surfactant proteins are associated with PS and consist in hydrophilic large multimeric SP-A and SP-D and small hydrophobic SP-B and SP-C[4].

Operative surfactant system is critical in lung function because it reduces the surface tension at the air–liquid interface of the alveolus, avoiding the alveolar collapse and reducing the work of breathing. In addition, PS provides an efficient barrier against environmental insults, such as the inhalation of pathogens and atmospheric nanoparticles[3,]. Therefore, it is not surprising that PS absence, deficiency or inactivation are tightly connected to sever respiratory diseases, such as the neonatal respiratory distress syndrome (RDS), the major cause of mortality in preterm babies, and acute respiratory distress syndrome (ARDS) that can affect patients of any age and occurs as a rapid onset of respiratory failure. Supplementation of exogenous natural or synthetic surfactant into the patient's lung represents a therapeutic approach routinely performed in RDS patients[1,4].

Even if active surfactant clinical formulations are available today, the molecular mechanism of the surfactant system assembly and the role of the surfactant proteins and lipid membranes in respiratory modulation are not fully understood.

Employing a multidisciplinary approach, we structurally and functionally characterized natural and synthetic surfactant clinical formulations produced by Chiesi Farmaceutici, Parma.

AFM imaging provided high resolution insight concerning the morphology and the three dimensional organization of the surfactant preparations investigated "in air".
Through the Langmuir-Blodgett technique (LB), were studied the physical, chemical and mechanical properties of surfactant monolayer at the air-water interface. The isotherm curves obtained by the dynamic compression at near-zero surface tensions, similarly to the lung physiological conditions, describe the compressibility and the tridimensional transitions of the surfactant films[5].
These evidences were also corroborated by the morphological characterization of the surfactant domains organized at the air-water interface. At a specific surface pressure values reached during the compression, portions of surfactant film were transferred to suitable support and investigated from a structural point of view by AFM[6]. In addition, a coupled approach LB-fluorescence microscopy[7] allowed to *real-time* monitor the surfactant film phase-transitions and the surfactant domain morphologic changes at the air-water interface depending on the compression.

The most classical formulation approach employed in the delivery of dry powder pharmaceuticals for inhalation is the use of carrier of different composition blended with active pharmaceutical ingredient (API). The role of the carrier is to bind the API powder and allow the respiratory inhalation[8].
The mechanism of drug-carrier blend formation and/or liberation is not clearly understood, but it is known that the physic-chemical characteristics of the carrier are important. For example, roughness, shape, dimension and surface energy of the carrier are just a few variables that have been indicated to influence the aerosol performance[9].
Using AFM and SEM, the topologic parameters of different carrier samples were investigated.

This work is a part of an ongoing collaborative Research Project with the Department of Analytics and Early Formulation of Chiesi Farmaceutici, Parma.
Because all the collected data are subject to industrial property rights, they will not be discussed in this PhD thesis.

Bibliography

1. Perez-Gil, J. & Weaver, T. E. Pulmonary surfactant pathophysiology: current models and open questions. *Physiology (Bethesda).* **25,** 132–141 (2010).

2. Lang, C. J. *et al.* Dipalmitoylphosphatidylcholine is not the major surfactant phospholipid species in all mammals. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **289,** R1426–39 (2005).

3. Blanco, O. & Pérez-Gil, J. Biochemical and pharmacological differences between preparations of exogenous natural surfactant used to treat Respiratory Distress Syndrome: role of the different components in an efficient pulmonary surfactant. *Eur. J. Pharmacol.* **568,** 1–15 (2007).

4. Zuo, Y. Y., Veldhuizen, R. a W., Neumann, a. W., Petersen, N. O. & Possmayer, F. Current perspectives in pulmonary surfactant - Inhibition, enhancement and evaluation. *Biochim. Biophys. Acta - Biomembr.* **1778,** 1947–1977 (2008).

5. Lee, K. Y. C. Collapse mechanisms of Langmuir monolayers. *Annu. Rev. Phys. Chem.* **59,** 771–791 (2008).

6. Jiao, X. *et al.* Atomic force microscopy analysis of rat pulmonary surfactant films. *Biophys. Chem.* **158,** 119–25 (2011).

7. Lipp, M. M., Lee, K. Y. C., Zasadzinski, J. a. & Waring, A. J. Design and performance of an integrated fluorescence, polarized fluorescence, and Brewster angle microscope/Langmuir trough assembly for the study of lung surfactant monolayers. *Rev. Sci. Instrum.* **68,** 2574 (1997).

8. Islam, N. & Cleary, M. J. Developing an efficient and reliable dry powder inhaler for pulmonary drug delivery - A review for multidisciplinary researchers. *Med. Eng. Phys.* **34,** 409–427 (2012).

9. Ooi, J., Traini, D., Hoe, S., Wong, W. & Young, P. M. Does carrier size matter? A fundamental study of drug aerosolisation from carrier based dry powder inhalation systems. *Int. J. Pharm.* **413,** 1–9 (2011).