Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

Biologia cellulare e molecolare

Ciclo XXXIII

Settore Concorsuale: 05/E2 - BIOLOGIA MOLECOLARE

Settore Scientifico Disciplinare: BIO/11 - BIOLOGIA MOLECOLARE

Insight into the molecular and functional determinants of HP1043, the essential master regulator of *Helicobacter pylori*.

Presentata da: Annamaria Zannoni

Coordinatore Dottorato

Prof. Giovanni Capranico

Supervisore

Prof. Vincenzo Scarlato

Co-supervisore

Prof. Davide Roncarati

Esame finale anno 2021

ABSTRACT

Helicobacter pylori is one of the most widespread and successful human pathogens since it colonizes the stomach mucosa of half of the population. Although it primarily causes chronic gastritis, 1% of the *H. pylori* carriers is expected to develop correlated gastric malignancies. Due to the increasing number of antimicrobial-resistant strains, *H. pylori* was included in the 2017 World Health Organization list of 12 pathogens that pose a major threat for humankind, to stimulate the development of novel antimicrobial strategies.

In this study, we propose as a new molecular target HP1043, an OmpR-like dimeric orphan response regulator that is essential for the viability of *H. pylori* as it orchestrates all the most important cellular processes. Amino acids involved in HP1043 dimerization and target DNA recognition were mutagenized and the mutant proteins were assayed for functionality to identify the most relevant determinants for HP1043 function. This experimental information, along with the HP1043 consensus binding motif and the available structural data, was used to guide an *in-silico* protein-DNA docking and generate a high-resolution structural model of the interacting HP1043 dimer and its target DNA. The model was experimentally validated and exploited to carry out a virtual screening of small molecule libraries, identifying 8 compounds potentially able to interfere with HP1043 function and likely block *H. pylori* infection.

The second line of research aimed at the characterization of the regulatory function of HP1043 and the tight mechanisms of regulation of the *hp1043* gene expression. In particular, we investigated the role of HP1043 as a transcriptional activator and proved that the regulator interacts directly with the housekeeping σ^{80} factor of the RNA polymerase. Moreover, we generated a conditional mutant *H. pylori* strain overexpressing a synthetic copy of the *hp1043* gene altered in nucleotide sequence yet encoding the wild type protein, achieving for the first time increased intracellular levels of HP1043. However, overexpression of HP1043 did not result in an upregulation of target genes transcription nor modulation of *hp1043* transcript levels, pinpointing the existence of multiple overlayed mechanisms of regulation that affect both protein levels and functionality as well as maintain steady the amount of *hp1043* transcript. Finally, we proposed that a mechanism of post-transcriptional regulation could depend on an antisense transcript to the *hp1043* gene which was validated in two different strains.

Taken together, the results of the project shed light on the regulatory mechanism of the essential regulator HP1043 and represented a further step towards the design of novel molecules able to block *H. pylori* infection.

CONTENT

1. INTRODUCTION

1.1 Helicobacter pylori

1.1.1 Epidemiology and infection

H. pylori is a spiral rod-shaped, microaerophilic, motile, Gram-negative bacterium, and human pathogen which colonizes the gastric mucosa. In 2015 *H. pylori* was esteemed to infect more than 50% of the population, with a higher prevalence in developing countries (Hooi et al., 2017). The infection is transmitted person-to-person via the oral-oral or oral-fecal route, and therefore its spread is facilitated by consumption of untreated water, overcrowding, and poor hygiene conditions. Transmission typically happens during early childhood, but *H. pylori* can persist for decades in its niche causing chronic active gastritis. Although mostly asymptomatic, *H. pylori* infection is strongly correlated with severe diseases of the upper gastrointestinal tract such as peptic and duodenal ulcer and represents a major risk factor for mucosa-associated lymphoid tissue lymphoma and gastric adenocarcinoma. For this reason, *H. pylori* is classified by the World Health Organization (WHO) as a group I carcinogen.

The successful infection and long-term colonization of the gastric niche rely on the concerted expression of a plethora of virulence factors by the pathogen. An in-depth description of these mechanisms is extensively revised in (Ansari and Yamaoka, 2019; Salama et al., 2013). Briefly, initial colonization of the stomach largely depends on bacterial urease activity that neutralizes the harsh acidic environment. Then, chemotaxis directed and flagellar-based motility, facilitated by the helical shape of the bacterium, allow the pathogen to penetrate the gastric mucus layer and reach the gastric epithelium. There, the expression of a wide variety of adhesins promotes colonization of the epithelial cells. Persistence depends on both reduced antigenicity of pathogen-associated molecular patterns such as lipopolisaccaride (LPS) and flagellin to escape recognition by the innate immune response and immunomodulation of the host adaptive immunity. Finally, *H. pylori* causes gastric disease due to the production of multiple cell pathogenicity factors. Among these, the secreted pore-forming VacA toxin and the CagA cytotoxin, which is translocated into host cells by the *cag*-type IV secretion system, act in combination to promote the production of pro-inflammatory cytokines, disrupt cell polarity, and cause tissue damage. The advantage for the bacterium to produce these host-detrimental determinants remains poorly understood and may involve the improved nutrient acquisition or enhanced dissemination.

H. pylori infections have been traditionally treated with a combination of a proton pump inhibitor and antimicrobials such as clarithromycin and amoxicillin or metronidazole. However, the standard triple therapy has been losing its efficacy, with eradication cure rates lower than 70%, mostly because of the increasing insurgence of antibiotic-resistant strains (Vianna et al., 2016). Treatment strategy has therefore shifted in favor of quadruple therapy as a first-line strategy in populations predominantly colonized with resistant strains (Fallone et al., 2016). However, *H. pylori* has proven to be able to circumvent the selective pressure of all antimicrobial chemotherapeutics currently in use by modifying its genome largely through the acquisition of point mutations (Alfizah et al., 2014; Haley and Gaddy, 2015; Vianna et al., 2016).

Because of the high prevalence of infection in the worldwide population as well as reduced efficacy of currently used therapies, *H. pylori* was introduced in 2017 WHO's first-ever published list of 12 antibiotic-resistant pathogens that pose the greatest threat to human health, and against which it is necessary to prioritize research and development of new antimicrobial strategies (Tacconelli et al., 2018).

Since an antimicrobial eradication is not expected to protect against reinfection and the high rates of antibiotic-associated side effects often determine poor compliance of patients, an effective vaccine against *H. pylori* remains a preferential option. Nonetheless, the development of a vaccine capable of preventing or eradicating *H. pylori* infection has proven extremely challenging and severely hampered by the lack of significant investment from large pharmaceutical companies (Sutton and Boag, 2019). Complications for the development of a vaccine against *H. pylori* stem from the fact that an optimal composition of protective antigens has not been identified as well as that although *H. pylori* infection gives rise to a substantial systemic and mucosal immune response, they usually result as nonprotective, since once the infection is established it is often life-persistent (Svennerholm and Lundgren, 2007).

1.1.2 Genome and regulatory functions

The ~1.7 Mbp genome of *H. pylori* has a low GC content (< 40%) and encodes approximately 1500 open reading frames (ORFs). A peculiarity of *H. pylori* is that it possesses a transcription machinery which is less complex with respect to that of other bacteria and only 17 transcriptional regulators. In *E. coli*, which is the prototype for Gram-negative organisms, the RNA polymerase (RNAP) core enzyme consists of 5 subunits ($\alpha_2\beta\beta'\omega$), and one out of 7 σ alternative sigma factors

that provides promoter recognition specificity, enabling transcription of different subsets of genes. In contrast, in *H. pylori* the β and β' subunits that form the catalytic core of the RNAP are fused as one continuous open reading frame, and only 3 σ factors are found within the chromosome (Scarlato et al., 2001). Among these, σ^{80} is the primary σ factor used for the transcription of most genes, while σ^{54} and σ^{28} are both involved in transcriptional regulation of flagellar genes (Niehus et al., 2004). Moreover, transcriptome analysis also revealed peculiarities in promoter organization of σ^{80} regulated genes. Differently from *E. coli* where promoters recognized by the housekeeping σ^{70} are characterized by two conserved boxes positioned 35 (TTGACA) and 10 (TATAAT) nucleotides upstream from the transcriptional start site (TSS), *H. pylori* σ^{80}-dependent promoters possess an extended -10 motif (tgnTAtaAT) and a periodic AT-rich motif upstream the -14 position in substitution to the -35 box (Sharma et al., 2010).

Besides the 3 σ factors, most of the other transcriptional regulators of *H. pylori* are two-component systems (TCSs), widespread signal transduction systems that rely on a phosphotransfer between a membrane-bound sensor histidine kinase (HK) and a cytosolic response regulator (RR) which elicits the cellular response. Typically, an external stimulus is detected by the HK sensory domain and triggers autophosphorylation of the HK transmitter domain on a histidine residue. The phosphate is then transferred to an invariant aspartate residue in the highly conserved N-terminal receiver domain of a cognate RR. This phosphorylation induces a conformational change of the RR that activates the variable DNA-binding output domain, modulating transcription of target genes (Nguyen et al., 2015). *H. pylori* uses TCSs to sense a diverse range of environmental stimuli and consequently adapt cellular functions such as motility (FlgS/FlgR), chemotaxis (CheA/CheY), acid acclimation (ArsS/ArsR), and copper resistance (CrdS/CrdR) (Foynes et al., 2000; Niehus et al., 2004; Pflock et al., 2006, 2007). The *H. pylori* genome also harbors two orphan RR, HP1021 and HP1043, which act independently of a cognate histidine kinase (Schär et al., 2005). HP1021 is involved in acetone metabolism and replication (Donczew et al., 2015), while HP1043 appears to be a master regulator of multiple key cellular processes (Pelliciari et al., 2017). Finally, the *H. pylori* genome encodes for 4 standalone regulators involved in metal ion homeostasis (Fur and NikR) and heat shock/stress response (HspR and HrcA) (Danielli and Scarlato, 2010; Roncarati and Scarlato, 2018).

The paucity of transcriptional regulators identified in *H. pylori* has been speculated to be a sign of reductive evolution, reflecting the high degree of adaptation of the bacterium to its constrained

niche in the human gastric mucosa as well as the lack of competition from other microorganisms (de Reuse and Bereswill, 2007). Nevertheless, recent analyses of microbial composition by 16S rRNA sequencing revealed the presence of a wide bacterial diversity in the human stomach (Bik et al., 2006; Mailhe et al., 2018; Yang et al., 2013). Additionally, the analysis of the *H. pylori* primary transcriptome using differential RNA-seq revealed massive antisense transcription uncorrelated with GC content and the expression of at least 60 small RNAs, an amount comparable to that of *E. coli,* considering the difference in genome size (Bischler et al., 2015; Sharma et al., 2010). Therefore, it is presumable that *H. pylori* compensates the lack of transcriptional regulators by relying on different mechanisms for its transcriptional regulation.

1.2   HP1043: an atypical orphan response regulator

HP1043, also referred to as HsrA (homeostatic stress regulator) (Olekhnovich et al., 2014), is classified as a member of the OmpR/PhoB family of RR transcription factors based on similarity to their DNA-binding domain, which is characterized by a winged helix-turn-helix (wHTH) motif. Phosphorylation by the cognate HK often induces dimerization of these regulators and increases their affinity for target DNA (Nguyen et al., 2015). However, HP1043 is an orphan RR, since an associated HK was not identified. HP1043 is also referred to as 'atypical' because its largely degenerate receiver domain lacks conserved residues that are required for phosphorylation and is capable of specific binding to its target genes in a phosphorylation independent manner (Delany et al., 2002; Müller et al., 2007; Schär et al., 2005).

1.2.1   Topology of HP1043

The structure of full-length HP1043 solved by NMR (Fig. 1) (Hong et al., 2004) shows that its molecular topology resembles that of the OmpR/PhoB family (Hong et al., 2007). HP1043 exists as a 2-fold symmetric dimer in solution in the absence of phosphorylation and has two functional domains: an N-terminal dimerization domain (HP1043N, residues 1-113) and a C-terminal DNA binding domain (HP1043_DBD, residues 121-223), connected by a flexible linker (residues 114-120).
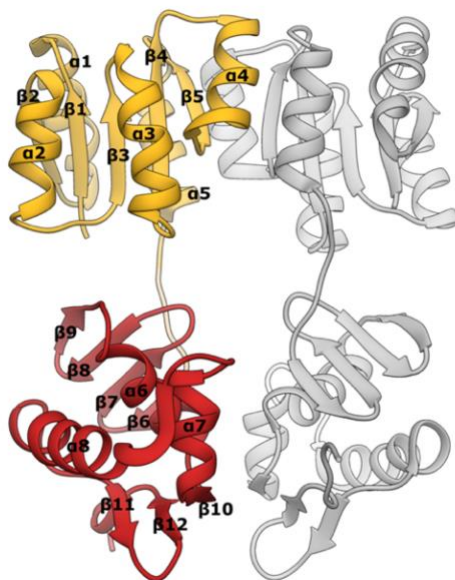
Figure 1. NMR structure and molecular topology of HP1043. The global fold of full-length HP1043 (PDB accession number 2HQR) is displayed as ribbon diagram and labeled with secondary structures. In monomer A, the N-terminal dimerization domain (residues 1-113) is shown in *yellow*, the flexible linker (residues 114-120) in *tan*, and the C-terminal DNA binding domain (residues 121-223) in *red*. Monomer B is depicted in *grey*.

The dimerization domain consists of an α/β/α sandwich fold, with a central 5-stranded parallel β-sheet surrounded by 5 α-helices. A crystal structure of the dimeric HP1043N is available (PDB accession code 2PLN). Dimerization of HP1043N relies on the α4-β5-α5 interface where three salt bridges are formed among the oppositely charged side chains of residues D93–R108, E83–R108, and R100–E83. Inter-subunit hydrogen bonds between residues consolidate ionic interactions (R112 side chain with D92 carbonyl oxygen, and R108 side chain with Y94 carbonyl oxygen), whereas hydrophobic interactions among residues V84, F87, A104, A107, and A111 further contribute to dimer stability (Fig. 2) (Jeong et al., 2013). The dimerization interface of HP1043 is similar to that of ArcA and PhoB proteins in their phosphorylation activated state, although only a few residues are conserved and HP1043 lacks a phosphoacceptor residue (Hong et al., 2007; Schär et al., 2005). The HP1043_DBD consists of a 4-stranded antiparallel β-sheet followed by a helix bundle of 3 α-helices. The helix-turn-helix DNA binding motif is formed by helices α7 and α8, with α8 being the recognition helix, while the winged loop is comprised between strands β11 and β12. Additionally, the α7-α8 loop is thought to be a transactivation loop that serves for interaction with σ[80]. Residues in both the wHTH motif and the hydrophobic core appear to be conserved in comparison with other members of the OmpR/PhoB family (Hong et al., 2007).
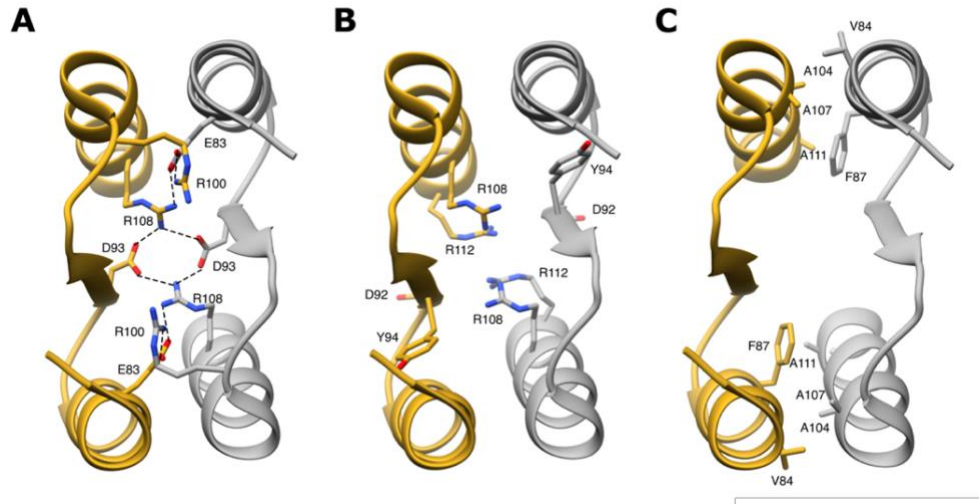
Figure 2. Interactions at the dimerization interface. The α4-β5-α5 interface of monomer A (*yellow*) and B (*grey*) is displayed as ribbon. (A) Interactions between oppositely charged residues. (B) Residues involved in hydrogen bond formation between HP1043 monomers. (C) Residues participating in hydrophobic interactions. Adapted from (Jeong et al., 2013).

1.2.2 HP1043 regulon

The HP1043 direct regulon was recently determined *in vivo* by using chromatin immunoprecipitation-sequencing (ChIP-seq) (Pelliciari et al., 2017). The analysis identified 37 highly reproducible binding sites, ~90% of which located in promoter regions, which is consistent with HP1043 functioning as a canonical transcriptional regulator. Nonetheless, 8% of the binding sites mapped within protein coding sequences and only 1 mapped in an intergenic region suggesting that HP1043 might have different regulatory functions (Visweswariah and Busby, 2015). Gene ontology analysis of the identified targets revealed that HP1043 exerts a pleiotropic function, directly regulating all the fundamental processes in the cell, including translation, transcription, replication, energy metabolism, and virulence (Fig. 3A). Although not included in this listing, a few experimentally validated HP1043 targets (*hp1043*, *tlpB*, *porG* and *nuoA*) are included in the full list of 107 identified peaks, hinting that HP1043 might have a broader targetome (Delany et al., 2002; Olekhnovich et al., 2013; Pelliciari et al., 2017).

HP1043 binding sites in promoter regions overlap the core promoter and are located directly upstream of the -10 box typical of housekeeping σ⁸⁰ regulated genes, overlapping two AT-rich stretch that substitutes the conserved -35 promoter element in *H. pylori*. This positioning is typical of transcriptional activators (Browning and Busby, 2016). To further support that HP1043 acts as

a positive regulator, there is experimental proof that HP1043 stimulates transcription from target promoters *in vitro* (Pelliciari, 2018), and HP1043 direct target genes resulted enriched among those up-regulated upon translational arrest *in vivo* (Pelliciari et al., 2017). Moreover, a mutant of HP1043 in its DBD that showed increased affinity for its targets *in vitro* also determined increase in transcription from regulated promoters *in vivo* (Olekhnovich et al., 2013).
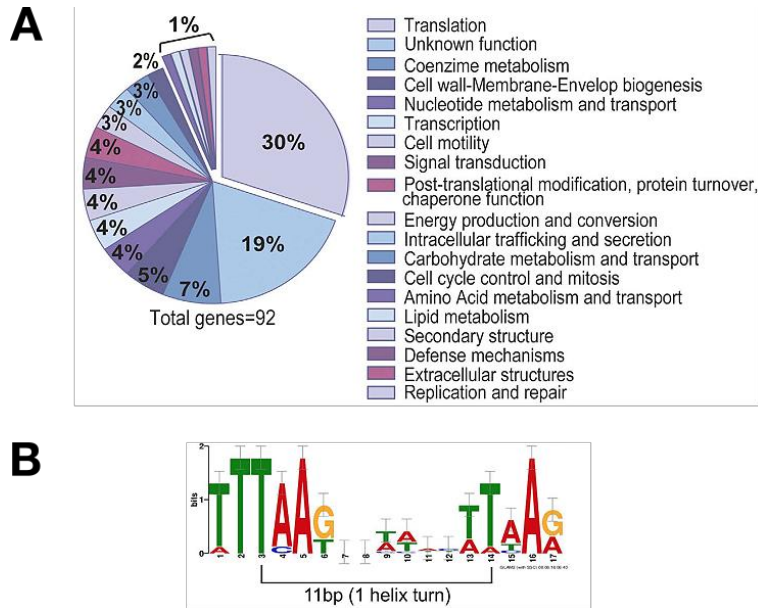
Figure 3. **Target gene ontology analysis (A) and consensus binding motif (B) of HP1043** as characterized by (Pelliciari et al., 2017). In A, genes associated with HP1043 binding sites are categorized based on their function. In B, sequence logo of HP1043 consensus binding sequence as derived by alignment of experimentally validated binding regions in target promoters. The height of each letter in the logo is a measure of the relative conservation of each base.

1.2.3   HP1043 consensus binding sequence

Alignment of experimentally validated target sequences was used to define the consensus binding sequence of HP1043, which consists of a direct repeat of two TTTAAG hexamers, the latter being less conserved, separated by a fixed length 5 bp spacer (Fig. 3B) (Pelliciari et al., 2017). The center of the two hemisites of the consensus is positioned 1 helix turn apart, so that the hexamers are exposed on the same face of the DNA double helix allowing recognition by a dimeric HP1043. This is further supported by the evidence that deletion of one of the two hemisites abolishes HP1043 binding to a target DNA *in vitro* (Fig. 4) (Hong et al., 2007; Pelliciari, 2018).

Experiments of scanning mutagenesis showed that sequence conservation is not the only parameter to consider for defining the relevance of each nucleotide in a target DNA for recognition by HP1043 (Pelliciari, 2018). In particular, in the case of the first hemisite mutation of nucleotides A1F, A2F and GF most compromised binding of HP1043 *in vitro*. In the second hemisite, although less conserved, mutation of 4 out of 6 nucleotides (T2S, T3S, A1S, A2S) hampered recognition by HP1043 (Fig. 4). A consensus based on relevance of nucleotides in HP1043 binding sequence might therefore be written as tttAAG-5bp-tTTAAg, where capital letters represent most important bases. These results also evidence an asymmetry in recognition of the two hemisites by HP1043.
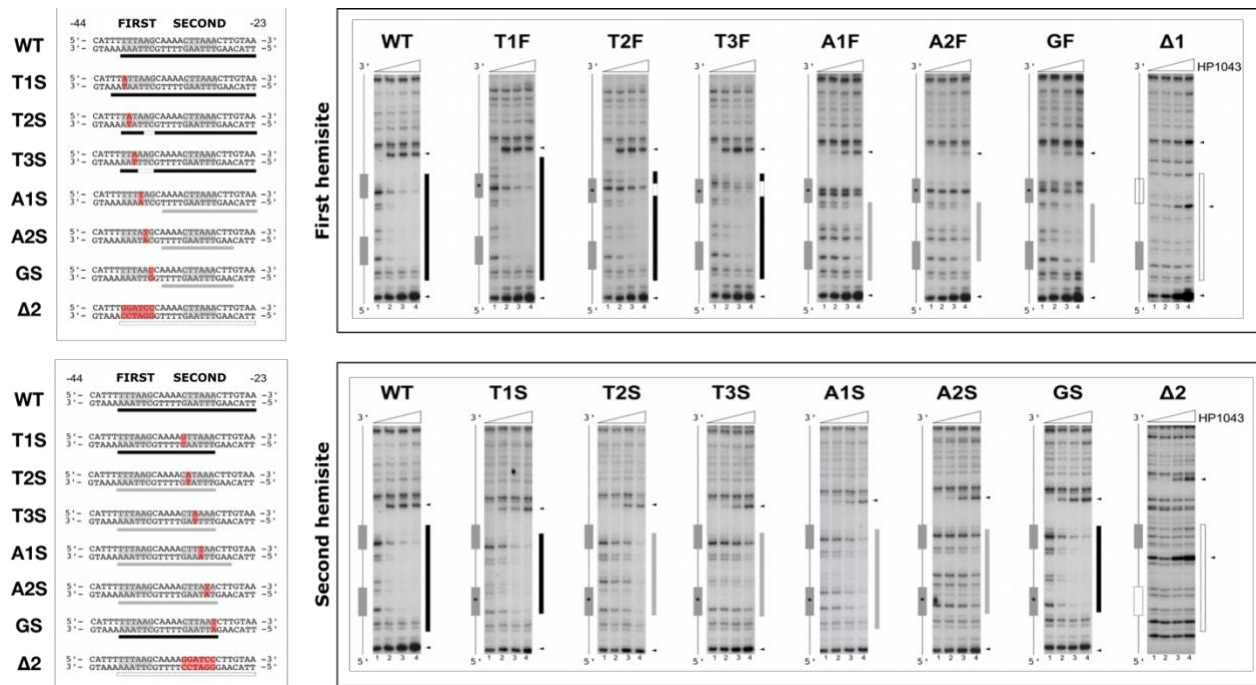
Figure 4: *In vitro* interaction of HP1043 with WT and mutant *P_{hp1227}* promoter probes. Scanning mutagenesis experiments in which each base of the consensus was substituted with the complementary one to maintain the overall GC content in the probe. On the left, a schematic representation of HP1043 binding site on the *P_{hp1227}* probe used for DNase I footprinting experiments. The hemisites of HP1043 consensus sequence are enclosed in *grey* boxes and mutations in the probe are highlighted in *red*. On the right, DNase I footprinting experiments performed with increasing amounts of dimeric HP1043 protein (0, 1.7, 6.6 and 13.3 μM). Grey boxes mark the position of the hemisites of the HP1043 consensus sequence on the probe and nucleotide mutations are indicated with an asterisk. For both panels, length of the bars under the sequences or on the left on *in vitro* binding assays indicate coverage of HP1043 protection of the probe from DNase I digestion, while color indicates strength of interaction, with *black* representing strong binding, *grey* intermediate binding and *white* loss of binding. Adapted from (Pelliciari, 2018).

1.2.4   HP1043 regulation

To date, not much is known about the regulation of HP1043 besides the fact that this transcription factor is essential for cell viability and undergoes a tight regulation inside the cell. In fact, not only the *hp1043* gene could not be deleted by allelic replacement unless a second copy of the gene was introduced in the chromosome (McDaniel et al., 2001; Schär et al., 2005), but also the amount of protein could not be modulated by addition of an extra copy of the wild-type gene under the control of an inducible or strong constitutive promoter (Delany et al., 2002; Müller et al., 2007), nor by interference with an antisense RNA (Croxen et al., 2007). The impossibility to modulate HP1043 protein levels severely hampered the possibility to study the mechanisms of its regulation.

The *hp1043* gene is transcribed from a σ^{80} promoter (P_{hp1043}) in a phase-dependent manner, reaching the highest levels of transcription during the exponential phase of growth. The amount of *hp1043* transcript halves already at the beginning of the stationary phase and is further reduced progressing to the late stationary phase (Delany et al., 2002). HP1043 binds to its own promoter suggesting an autoregulatory role (Delany et al., 2002). Interestingly, P_{hp1043} is the only characterized promoter where the consensus binding sequence for HP1043 maps on the non-coding DNA strand (Pelliciari et al., 2017). Nevertheless, HP1043 positively regulates transcription from its own promoter (Pelliciari, 2018).

The HP1043 receiver domain is strongly degenerate as residues of the acidic pocket that act as phosphorylation site in other response regulators of the OmpR/PhoB family are either mutated or deleted (Beier and Frank, 2000; McDaniel et al., 2001; Müller et al., 2007; Schär et al., 2005). However, mutation of these residues to the consensus (Schär et al., 2005) or replacement of *hp1043* with the gene encoding the orthologue HPMG440 from *Helicobacter pullorum,* which responds to phosphorylation by a cognate HK (Bauer et al., 2013), did not affect the function of HP1043, demonstrating that phosphorylation is not necessary for HP1043 in *H. pylori*. Based on *in vitro* binding experiments, the affinity of recombinant HP1043 for DNA appears weaker if compared to other regulators of *H. pylori* and to a cellular extract fraction enriched in the native protein, suggesting the existence of diverse post-translational modifications of HP1043 that are specific to *H. pylori* (Olekhnovich et al., 2013). However, how HP1043 senses and responds to environmental stimuli remains poorly defined.

In a work of Delany and colleagues, P_{hp1043} was not responsive to heat shock, osmotic shock, or acid shock, nor to treatment with a high concentration of iron or nickel, or with acetyl phosphate

as assessed by primer extension experiments in exponentially growing bacteria (Delany et al., 2002). On the contrary, a more recent report shows a significant increase in *hp1043* transcription in real-time PCR experiments following exposure of stationary cultures to urea, low pH, metals, biofilm, and AGS gastric epithelial cells (De la Cruz et al., 2017). These discrepancies might be due to differences in strain, experimental conditions, and methods of evaluation used. Finally, HP1043 might be indirectly involved in the response to oxidative stress, since treatment of growing cells with redox-active compounds determined a decrease in protein levels as well as reduction in transcription of HP1043 regulated genes (Olekhnovich et al., 2014).

2. AIM OF THE STUDY

Helicobacter pylori is one of the most widespread and successful human pathogens since it colonizes the stomach of half of the population. If untreated, *H. pylori* causes chronic gastritis which can remain clinically silent or evolve into more severe diseases such as atrophic gastritis, peptic ulcer, and MALT-lymphoma or gastric adenocarcinoma. Although only approximately 1% of *H. pylori* carriers is expected to develop associated malignancies, the worldwide dissemination, high prevalence, and life-long persistence of the infection, together with the increasing emergence of antibiotic-resistant strains pinpointed the necessity to develop novel antimicrobial strategies to eradicate the bacterium (Tacconelli et al., 2018).

A trending strategy in the field is represented by the targeting of essential bacterial transcriptional regulators (BTR) (González et al., 2018). This approach allows to obtain both a specific and multitargeting effect, as BTRs are usually soluble cytoplasmic proteins that lack a human counterpart, and often a single BTR regulates multiple genes. In the case of *H. pylori*, the transcriptional regulator HP1043 represents an ideal target. In fact, a recent study that dissected the direct regulon of HP1043 shed light on its function as a master regulator of all the fundamental processes of the cell, such as replication, transcription, translation, energy metabolism, and virulence (Pelliciari et al., 2017). As a result of the key role carried out by HP1043, the *hp1043* gene cannot be deleted and its expression proved to be tightly regulated as multiple attempts to modulate the intracellular concentration of HP1043 failed (Croxen et al., 2007; Delany et al., 2002; Müller et al., 2007).

In light of this background information, the presented Ph.D. thesis project consisted of two major research lines. The first aimed at the production of a computational model of the dimeric HP1043 interacting with a target DNA to be used, in the absence of co-crystals, to carry out a virtual screening of ligand libraries and identify compounds possibly able to interfere with the HP1043 regulatory activity. Also, an *in vivo* reporter system was set up to investigate the effects of the putative ligands on HP1043.

The second line of research was aimed at the characterization of the regulatory function and mechanisms of regulation of HP1043. This was addressed through the generation of a conditional *hp1043* overexpressing strain and the validation of an antisense transcript to the *hp1043* gene.

3. RESULTS

3.1 Molecular characterization of HP1043 interaction with DNA and dimerization and virtual screening for potential inhibitors

3.1.1 Residue conservation analysis of HP1043

Amino acid positions that are critical for maintaining the structure of a protein and/or its function often undergo evolutionary constraints and are therefore invariant within the protein family. Thus, the evolutionary conservation rate of a residue can be used to estimate its biological importance. To help determine the relevance of single HP1043 residues for dimerization and DNA recognition we exploited ConSurf, a web server that computes evolutionary conservation scores for each amino acid position and maps them onto the 3D structure of the protein (Ashkenazy et al., 2016). As from the ConSurf analysis, the dimerization domain (HP1043N) results poorly conserved, except for the α4-β5-α5 dimer interface (Fig 5B). This analysis, in combination with the high degree of degeneration of the phosphor-acceptor site in comparison with the receiver domain of components of the OmpR/PhoB family, further supports that HP1043 might represent a new subclass of response regulators in which the N-terminal domain exerts the oligomerization function but acts independently of phosphorylation (Hong et al., 2007). Whether the degenerate receiver domain is capable of responding to different signals remains to be defined.

The ConSurf calculations also demonstrate the high level of conservation of the winged helix-turn-helix (wHTH) DNA recognition motif, and the hydrophobic core which is necessary for correct folding and therefore functionality of the HP1043 DNA-binding domain (HP1043_DBD) (Fig 5B). Differently, the 4-stranded antiparallel β-sheet (β6-β9) at the interface with the N-terminal dimerization domain show lower conservation scores. This might reflect protein interactions that vary among bacteria. Crystal structures of other regulators of the OmpR/PhoB family in the unphosphorylated state show interdomain contacts occurring between the regulatory and the effector domain, suggesting a mechanism of mutual inhibition (Buckler et al., 2002; Friedland et al., 2007; Nowak et al., 2006; Robinson et al., 2003). In particular, in DrrD of *Thermotoga maritima* the interaction occurs between helix α5 of the dimer interface and the antiparallel β-sheet of the DBD and is supposed to be released upon phosphorylation dependent conformational changes (Buckler et al., 2002). Therefore, it is conceivable that differences in the antiparallel

β-sheet depend on diversities in the HP1043 N-terminal domain or reflect the fact that HP1043 does not respond to phosphorylation.
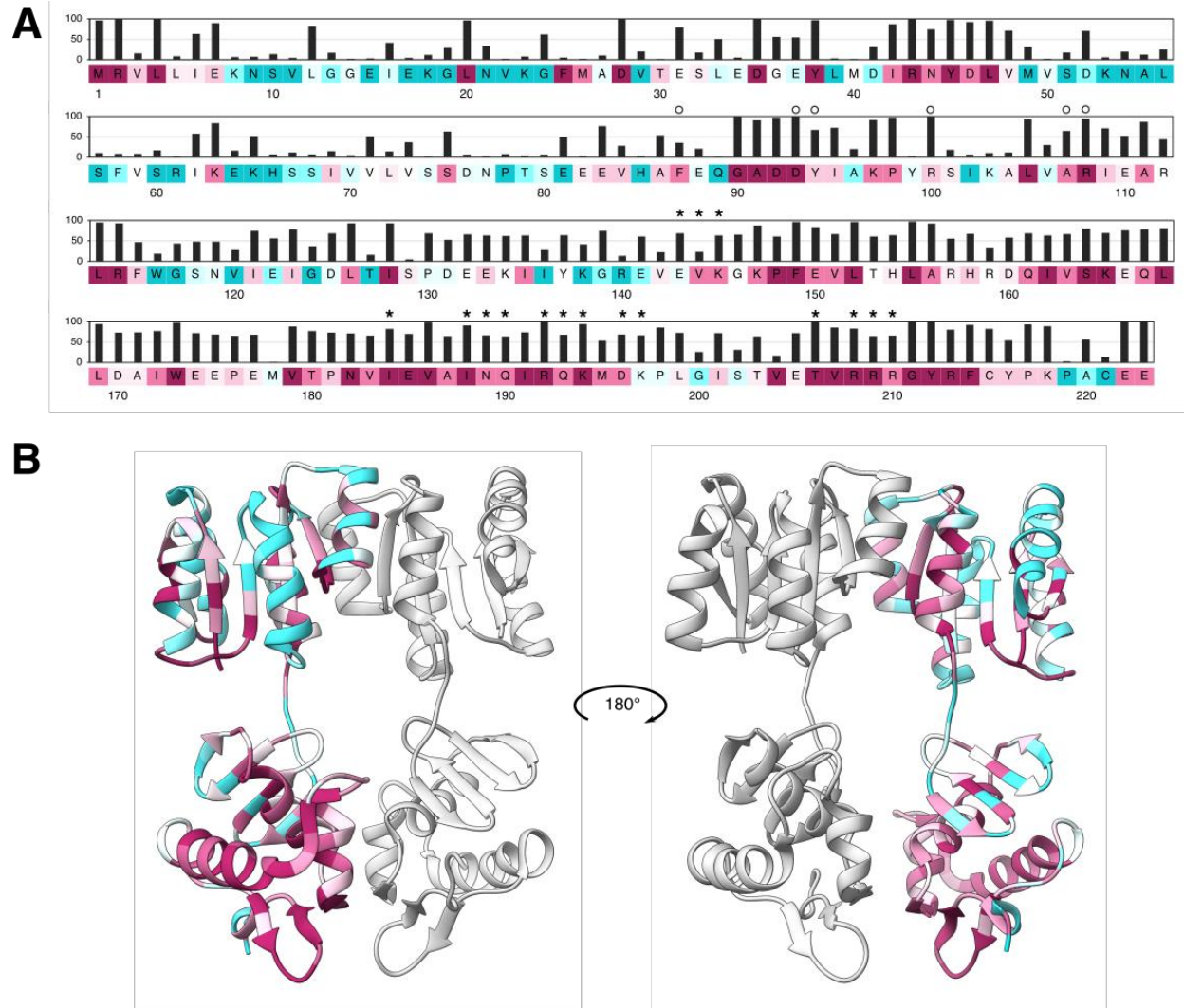
Figure 5. **Evolutionary conservation of HP1043 residues.** Conservation analysis performed using default settings of the ConSurf web server on the HP1043 protein sequence. Residues are colored depending on their calculated conservation scores against homologs, in a scale of nine grades ranging from the most variable positions (grade 1) colored *turquoise*, through intermediately conserved positions (grade 5) colored *white*, to the most conserved positions (grade 9) colored *maroon*. A) Protein sequence of HP1043. Above, relative frequency (in percentage) for the amino acid residue present in HP1043. Residues selected for mutagenesis are indicated with an *asterisk* for the DBD, and with an *open circle* for the dimerization interface. B) The global fold of HP1043 is shown as cartoon. Monomer A is colored by ConSurf scores, monomer B is depicted in *grey*.

3.1.2 Identification of amino acid residues fundamental for target DNA recognition by HP1043

The HP1043_DBD is characterized by a helix-turn-helix motif formed by helices α7 and α8, among which α8 is the recognition helix that is supposed to interact with the DNA major groove, and the β11-β12 winged loop that should interact with the DNA minor groove (Hong et al., 2007; Jeong et al., 2013) (Fig. 6A).

To assess which residues of the HP1043_DBD are fundamental for recognition of target promoters, we selected a few amino acids to mutate in both the functional elements of the wHTH motif (Fig. 6B) and used an alanine scanning mutagenesis approach to remove the function of targeted residues without altering the 3D structure of the protein. Wild-type and mutant *E. coli* purified recombinant HP1043 were then tested for their ability to bind *in vitro* a target promoter through DNase I in footprinting assays (Fig. 6C). The HP1043 binding site on *hp1227* promoter (P*hp1227*) was chosen as DNA probe since, among the already validated target promoters, it is the one that contains the HP1043 binding sequence most similar to the determined consensus (Pelliciari et al., 2017). As previously reported, the wild-type HP1043 binds directly upstream the -10 box of the promoter and specifically protects the region of P*hp1227* that spans from position -18 to -35 with respect to the transcriptional start site. In particular, partial protection of the probe is achieved already at the lowest concentration tested for the wild-type protein (0.6 μg of *E. coli* purified recombinant HP1043) (Fig. 6C, left panel). In our experimental settings, mutation of residues I188, R192, D196, and K197 in the α8 helix and of residues T206 and R208 in the β11-β12 winged loop completely abolished HP1043 binding, since effective DNA protection from DNase I digestion was not achieved even at a protein concentration 27 times higher than that required for partial protection of the probe by the wild-type protein (Fig. 6C, panel on the right). The result suggests that these residues are fundamental for recognition and stable DNA binding by HP1043. However, I188A and D196A mutations reduced solubility of the protein preparations and hampered the stability in thermal shift assays (Fig. 8D), suggesting that these residues might be critical for the correct folding of HP1043_DBD. For this reason, residues I188 and D196 were excluded from further analysis.

All the other identified residues show a high degree of conservation (Fig. 5A) supporting their relevance for the functionality of the wHTH motif. Among these, T206D and R208A mutants were previously reported to abolish HP1043 binding to DNA *in vitro*, confirming our result (Hong et

al., 2007), while the relevance of R192 and K197 was first estimated in this study. Also, in a previous work, the mutation of residues N189 and Q193 in the HP1043_DBD resulted in decreased DNA binding activity *in vitro* (Hong et al., 2007). However, in our experimental conditions, N189A and Q193A mutants showed a DNA binding affinity comparable to the wild type and therefore were not further investigated (Fig. 6C).
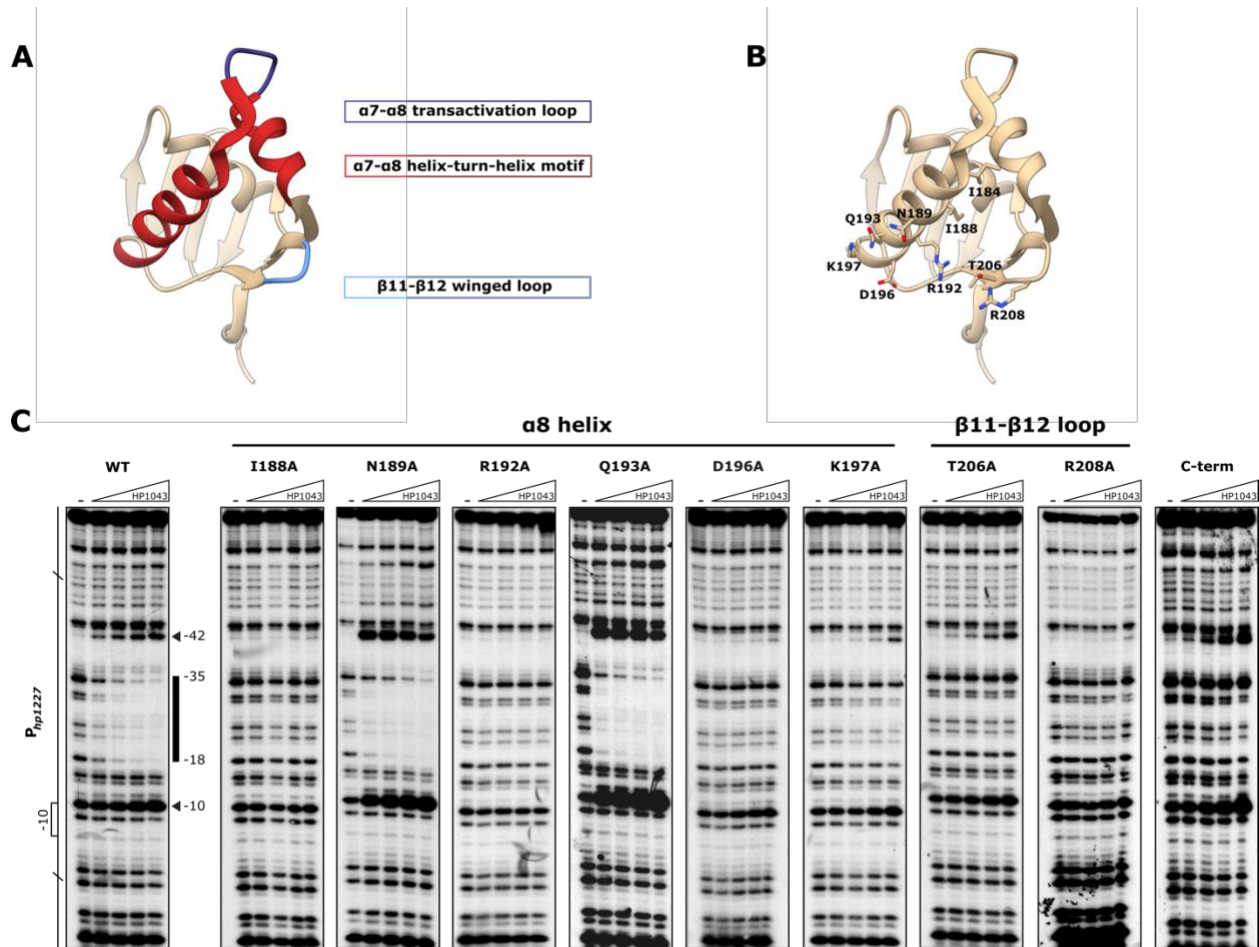
Figure 6. *In vitro* DNA binding of wild-type and mutant HP1043 to the P$_{hp1227}$ target promoter. A, B) DNA binding domain (DBD) of HP1043 depicted as ribbon. Functional elements of the wHTH motifs are highlighted in A, while residues selected for mutation into alanine are shown in B. C) DNase I footprinting experiments. A probe containing the HP1043 binding site in the *hp1227* promoter, radiolabelled on the coding strand, was incubated with increasing concentrations (0, 0.6, 2.0, 6.0, and 18.0 μg) of purified HP1043 wild-type and its mutants in the DNA-binding domain prior to digestion with DNase I. On the *left* of the autoradiographic film, a schematic representation of P$_{hp1227}$, where the -10 promoter element is depicted as an open box. On the *right*, the black box represents the region of DNase I protection on the target promoter, while the black arrowheads indicate DNase I hypersensitive sites.

3.1.3   Structural modeling of the HP1043-DNA interaction

In the absence of protein-DNA co-crystals, we took advantage of the experimental information gained in this study for the identification of the most relevant residues in both the HP1043_DBD and the P*hp1227* target promoter to guide a two-stage data-driven docking with the program HADDOCK 2.2 and produce a computational model of the HP1043-DNA interaction (in collaboration with Dr. F. Musiani, University of Bologna). In particular, protein residues R192, K197, T206, and R208 and nucleotides ttt<u>TTTAAG</u>caaaa<u>cTTAA</u>acttgta (in uppercase) were used as initial molecular restraints to guide the interaction. The amino-acid residues are the ones that proved to be fundamental for DNA binding, while the selected nucleotides are the most conserved residues of the consensus sequence, which also include the most relevant for interaction (ttt<u>AAG</u>-5bp-<u>tTTAAg</u>, in uppercase (Pelliciari, 2018)).

Due to the flexibility of the HP1043 interdomain linker, the computational analysis was performed on a library of 20 HP1043_DBDs derived from the available NMR structure of the full-length protein (PDB: 2HQR), by imposing interaction of one HP1043_DBD with the selected nucleotides of the first hemisite of the consensus sequence, and of another HP1043_DBD with the residues selected from the second hemisite. Results of the first docking round were manually screened, to select clusters in which the relative orientation of the HP1043_DBDs was compatible with the distances allowed by the linker between the DBD and the dimeric N-terminal domain. Interestingly, the only cluster that fitted this requirement showed a 180° rotation of the DBDs compared to the conformation of the dimeric HP1043 in solution, and this configuration appeared to be stabilized by reciprocal interaction between residues E143, V144, K145 at the interface of the two DBDs (Fig. 8). This cluster was selected to perform the second docking run and obtain the final model (Fig. 7) in which a dimeric HP1043 protein interacts with DNA both at the major and the minor groove through helices α8 and wings β11-β12, respectively, although the constraints used in the docking algorithm were not explicitly set to reward interactions with the minor groove. In particular, the interaction with the major groove of helices α8 with the two hexamers of the consensus binding sequence determines a contraction of the three adjacent minor grooves, in agreement with previous hydroxyl-radical footprinting experiments where three areas of protection were found on target probes, reflecting limited access to radical ions cleavages upon HP1043 binding (Pelliciari et al., 2017).

In the model, the two HP1043_DBDs are bound on the same side of the DNA molecule, in congruence with the 1-helix turn that distances the center of the two hexamers of the consensus binding sequence, in a head-to-head conformation which would explain the minor conservation of the second hemisite of the HP1043 consensus sequence. This modeled configuration of HP1043 differs from that of other dimeric regulators of the OmpR/PhoB family such as *E. coli* PhoB, and *Klebsiella pneumoniae* RstA and PmrA for which the sole DBDs or the full-length protein were co-crystalized with DNA in a head-to-tail conformation, coherently with the recognition of a direct repeat (Blanco et al., 2002; Li et al., 2014; Lou et al., 2015). Nonetheless, OmpR can recognize different DNA sequences and bind in either a head-to-tail or head-to-head orientation (Martínez-Hackert and Stock, 1997; Mizuno and Tanaka, 1997; Rhee et al., 2008). Such diversities in the way in which regulators of the same protein subfamily recognize target tandem DNA sequences were proposed to depend on the length and flexibility of the interdomain linker (Walthers et al., 2003).
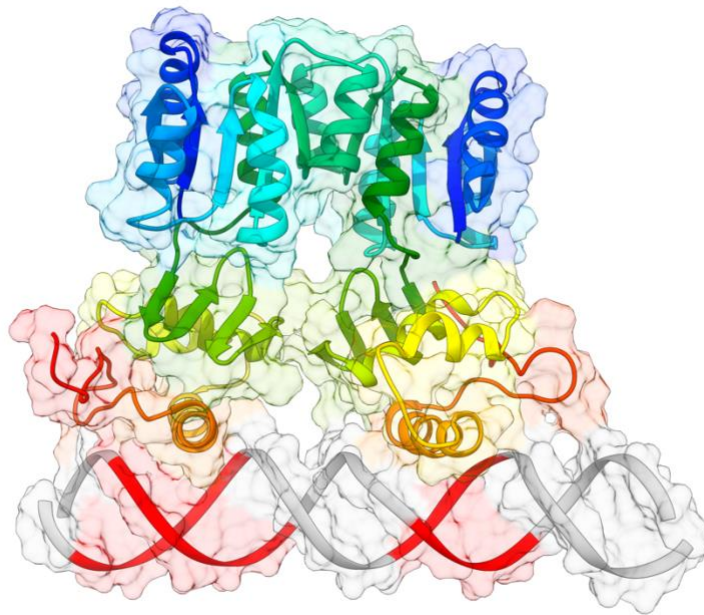
Figure 7. HP1043-P$_{hp1227}$ docking model. HP1043-DNA interaction model depicted as ribbon and surface. HP1043 monomers are colored in *rainbow*, and the DNA in *grey*, with the exception of the active residues, reported in *red*. HP1043_DBDs interact with DNA in a head-to-head conformation. Helices α8 of the wHTH motif contact the DNA major grooves which in turn are broadened, while the adjacent minor grooves are narrowed and recognized by the β11-β12 wings.

3.1.4 Validation of the HP1043-P$_{hp1227}$ docking model

To validate the model, we selected for each of the functional elements of the wHTH motif a couple of residues, one of which resulted to interact with DNA and a negative control of interaction (residues K194 and Q190 for the α8 helix, and residues R210 and R209 in the β11-β12 winged loop, respectively), to test by alanine scanning mutagenesis (Fig 8A, left). In congruence with the model, mutation of residues K194 and R210 partially (for K194) or completely (for R210) impaired protection of the probe by HP1043 in *in vitro* binding assays, while the proteins mutated in residues Q190 and R209 retained the ability to bind DNA (Fig. 8C). Moreover, to validate the putative EVK interaction at the interface of the DBDs, multiple mutants were produced (Fig. 8C). Mutation of the EVK triplet into AAA altered the thermal stability of the protein and was therefore excluded from analysis, while the EAA and AAE mutants exhibited a thermal unfolding profile similar to that of the wild-type, although the overall fluorescence signal is lower (Fig. 8D). Nonetheless, these mutants were not able to protect the target DNA in footprinting experiments. Since the EVK triplet is positioned in the loop connecting β8 and α6 which is exposed to the solvent in the HP1043 configuration resolved by NMR, the results of *in vitro* binding assays could support an involvement of the triplet in the stabilization of an HP1043 conformation that is functional for interaction with DNA. However, additional experiments are required to rule out the involvement of these residues in different intra- or inter-domain interactions that might be relevant for HP1043 function.

Additionally, DNA binding interference experiments were implemented to estimate the relative relevance of HP1043 interactions with minor and major grooves of the DNA (Fig. 8E-G). In particular, once the HP1043 concentration required to shift the target P$_{hp1227}$ probe was determined (~8 μM, Fig. 8E), *in vitro* binding reactions were performed in the presence of increasing concentrations of the double-stranded DNA (dsDNA) groove binders Netropsin and methyl green. Specifically, Netropsin is an oligopeptide that binds non-covalently to the minor groove of dsDNA preferably in AT-rich regions, while methyl green is a positively charged dye that forms electrostatic interactions with negatively charged phosphate radicals in the major groove of dsDNA. In our experimental settings, Netropsin interfered with HP1043 binding at nanomolar concentrations, and completely abolished protein-dependent specific shift of P$_{hp1227}$ at a concentration of 1 μM, which is almost one order of magnitude less than the HP1043

concentration, demonstrating the high relevance of HP1043 contacts at the DNA minor groove (Fig. 8F).
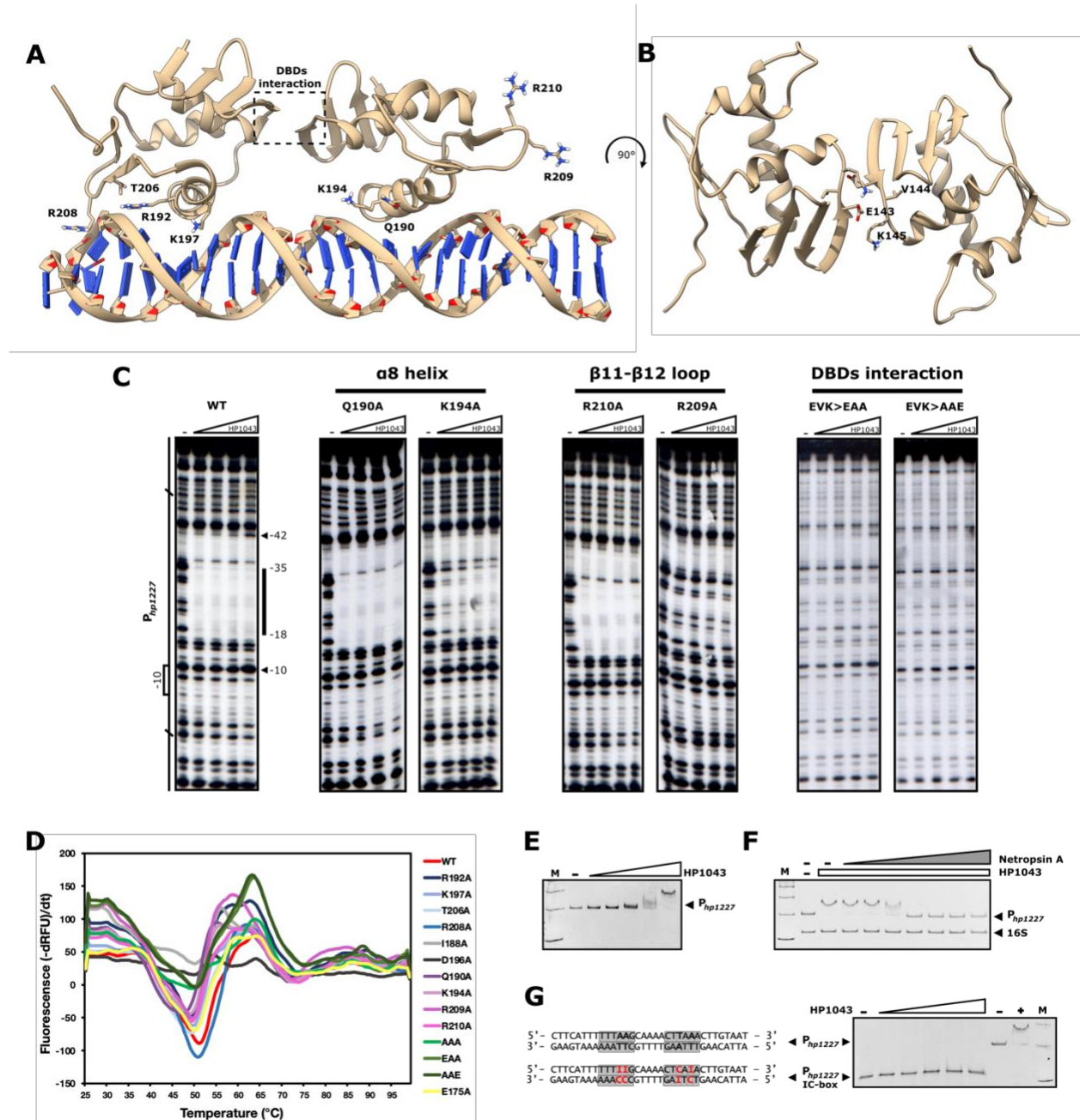
Figure 8. Validation of HP1043-P$_{hp1227}$ interaction model. A) HP1043_DBD-P$_{hp1227}$ computational model depicted as ribbon. On the *left,* amino acid residues used as restraints to guide the docking, on the *right,* residues tested to validate the model. The region of putative interaction between HP1043 DBDs is highlighted. B) View from above. Residues presumed to be involved in the HP1043 DBDs interaction are exposed. C) DNase I footprinting experiments performed to validate the docking model, where residues K194 and R209, but not amino acids Q190 and R210, interact with DNA. On the *right, in vitro* binding experiment to investigate the putative interaction among HP1043 DBDs, by mutation of residues 143-145. Experiments were executed as

described in Fig. 6. D) Representative first derivative curves of fluorescence obtained in a thermal stability shift assay performed on wild-type and mutant HP1043 proteins in the presence of SYPRO orange. The minimum of the curves corresponds to the protein melting temperature. E-G) Electrophoretic mobility shift assays (EMSAs) to validate HP1043 interaction with minor and major grooves of the DNA. In **E**, EMSA to estimate the concentration of HP1043 required to shift the target probe. Eighty pmol of a 100 bp probe encompassing the HP1043 binding site on the *hp1227* promoter were incubated with increasing concentrations of purified recombinant HP1043 (0, 1, 2, 4, 8, 16 μM). In **F**, EMSA in the presence of a fixed amount of HP1043 (8 μM) and increasing concentrations of the DNA minor groove binder Netropsin (0.001, 0.01, 0.1, 1, 10, 100, 500 μM). A 60 bp probe of the 16S rRNA gene was used as internal specificity control. In E, increasing amounts of HP1043 (0, 8, 12, 16, 20, 24 μM) were incubated with a P$_{hp1227}$ probe harboring IC-box substitutions in the sequences recognized by HP1043. As a control of reaction, 8 μM HP1043 was incubated with the wild-type P$_{hp1227}$ DNA. IC-box substitutions are highlighted in *red*, hexamers of the HP1043 consensus binding sequence are enclosed in *grey* boxes.

Oppositely, methyl green did not interfere with HP1043 binding even at a concentration 60 times higher than that of the protein (data not shown). Given that HP1043, similarly to other regulators of the OmpR/PhoB family, was proven to interact extensively with the major groove via the α8 recognition helix of the HTH motif, the result could be explained as a stronger affinity of the protein for the DNA compared to methyl green. To prove this hypothesis, we exploited the fact that hydrogen bonding on purine and pyrimidine rings in inosine-cytosine (I-C) pairs resemble A-T pairs in the minor groove and G-C couples in the major groove. Therefore, we tested the ability of HP1043 to bind an *hp1227* promoter probe harboring I-C box substitutions in A-T couples that previously proved to be most important for HP1043 recognition of its consensus binding sequence (Pelliciari, 2018) (Fig. 8G). Since HP1043 failed to bind and shift the I-C box substituted target promoter even at a protein concentration of 24 μM (Fig. 8G), we can conclude that in the previous experiment HP1043 out-competed methyl green for binding to the DNA and that both a major and minor groove readout are required for recognition and binding of target sequences by the regulator.
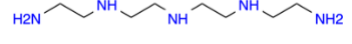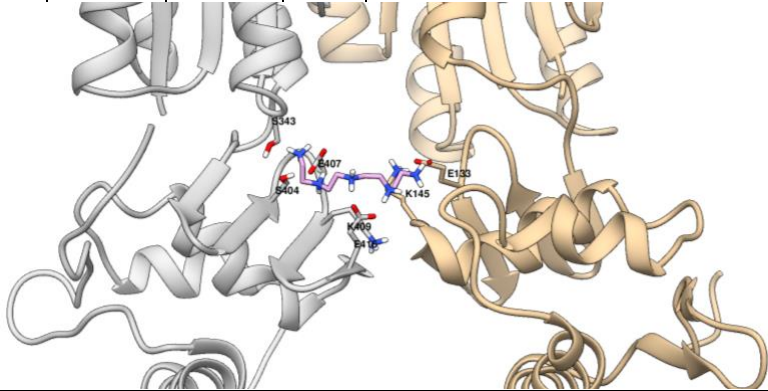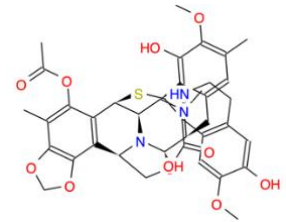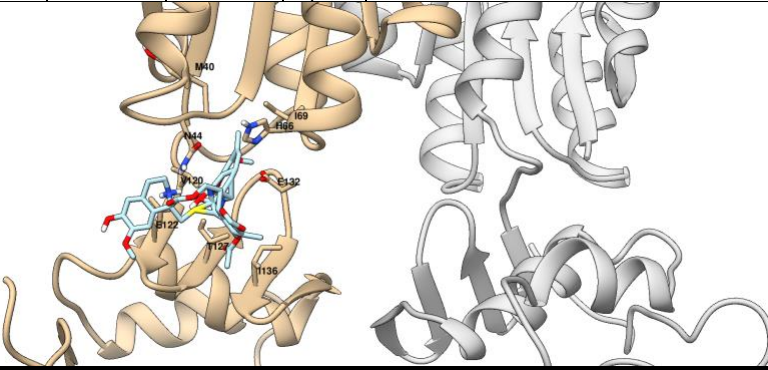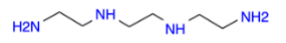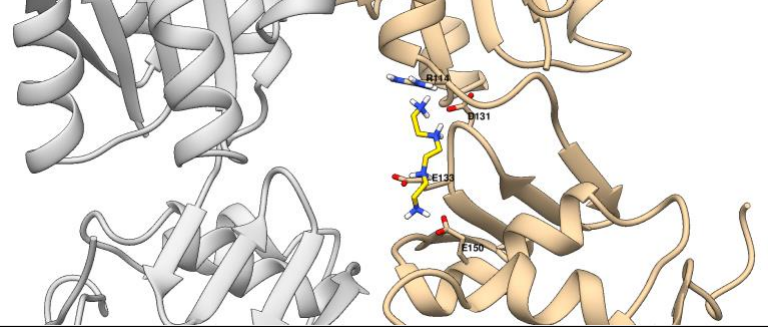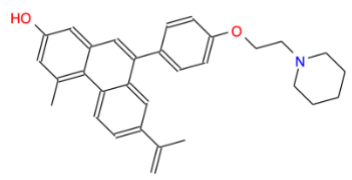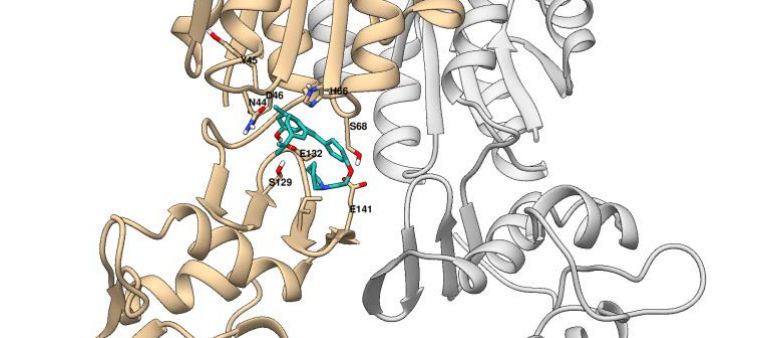
3.1.5 Virtual screening for potential HP1043 inhibitors

The produced molecular model of dimeric HP1043 bound to DNA was used to carry out a virtual screening for novel potential HP1043 inhibitors (in collaboration with Dr. F. Chiappori, CNR-ITB, Milano). About 18300 molecules were included in the screening and derived from 5 different datasets obtained from the ZINC15 database (Sterling and Irwin, 2015): substances targeting transcription factors (11025 molecules), Drug Bank approved molecules (2868), Food and Drug

Administration approved drugs (1585), Therapeutic Target Database approved drugs (2164), and the NIH clinical collection (669).

The first docking screening returned 180 putative ligands with an inhibitory constant (Ki) value in the pM order of magnitude. These molecules were undergone a second docking round, in order to identify a subset of best candidates. The re-docking procedure returned 323 conformations (multiple poses of the same ligand were allowed) with a Ki value in the pM order of magnitude. The 50 ligands that also displayed a cluster numerosity greater than 20 units can be considered interesting candidates and are listed in table 7. Among these, a subset of 8 ligands (reported below) bound at the interface between the dimerization and the DNA-binding domains, and/or between the two HP1043 chains, and therefore can be considered as good candidates for HP1043 also in the free form. The remaining molecules bound at the interface between protein and DNA can be considered as ligands with a lower specificity for the free form of HP1043. A selection of the most promising HP1043 binders available for purchase is currently under evaluation. Selected binders will be assayed for their ability to block HP1043 function *in vitro* (trough DNA-binding assays) and to inhibit the growth of *H. pylori* (both on agar plates and in liquid cultures).

Ligand (commercial name)	E_B	Cluster Nr.	Ki		Known targets/Pharmacological application/Chemical role
ZINC000022443609 (Plerixafor)	-23.84	54	3.37	aM	Chemokine receptors CXCR4, CCR2 Hematopoietic Stem Cell mobilizer

23

Ligand (commercial name)	E_B	Cluster Nr.	Ki		Known targets/Pharmacological application/Chemical role
ZINC000019363537 (Tetraethylenepentamine)	-15.57	34	3.86	pM	Metal-chelator

Ligand (commercial name)	E_B	Cluster Nr.	Ki		Known targets/Pharmacological application/Chemical role
ZINC000150338708 (Ecteinascidin)	-14.39	27	28.4	pM	Antitumor antibiotic
	-13.89	27	65.55	pM	

Ligand (commercial name)	E_B	Cluster Nr.	Ki		Known targets/Pharmacological application/Chemical role
ZINC19364225 (Trientine)	-13.84	51	71.87	pM	Copper chelating agent

Ligand (commercial name)	E_B	Cluster Nr.	Ki		Known targets/Pharmacological application/Chemical role
ZINC000013494069	-13.1	31	250.98	pM	Estrogen receptors ESR1/ESR2 --- not for sale ---
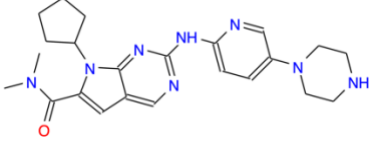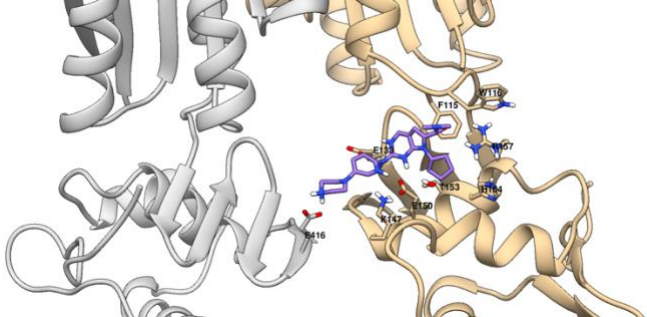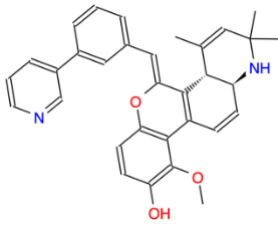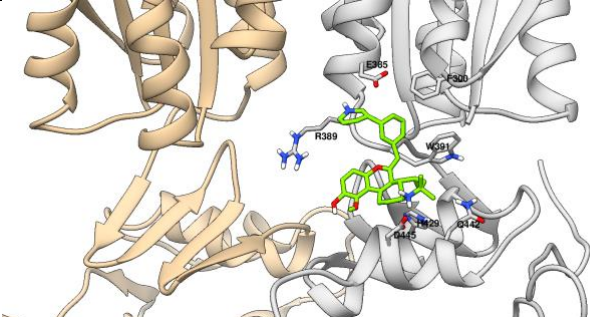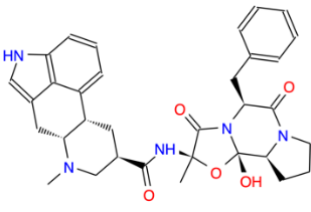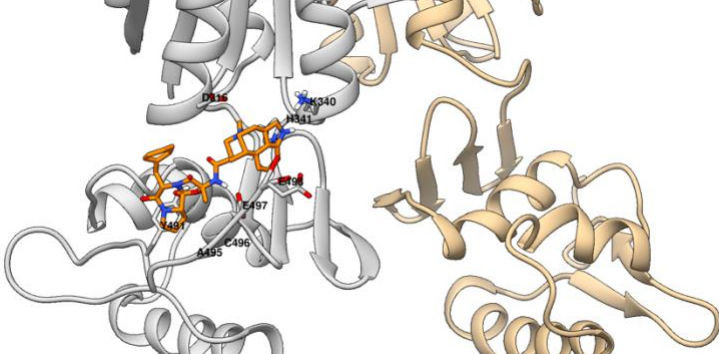
Ligand (commercial name)	E_B	Cluster Nr.	Ki		Known targets/Pharmacological application/Chemical role
ZINC000072316335 (Ribociclib)	-12.7	25	487.81	pM	Cyclin-dependent kinase (CDK) inhibitor
ZINC000028823510	-12.53	20	652.73	pM	Glucocorticoid receptor --- not for sale ---
ZINC000014880002 (Dihydroergotoxine)	-12.39	52	827.22	pM	Alpha-adrenergic antagonist, dopamine agonist

Table 1. Putative HP1043 ligands obtained from a virtual screening of a selection of small molecule libraries. The table reports the ZINC DB code of the ligand and commercial name, and calculated values for the parameter used for selection during the docking procedure (calculated energy of binding [E_B], inhibition constant [Ki], and numerosity [Cluster Nr]). For each ligand, the chemical structure and a representation of the best docking pose are reported. In the docking models, HP1043 chains are depicted as ribbon and chain A is colored in *tan*, chain B in *grey*. Residues within a 4 Å radius from the ligand atoms are displayed, as predicted interacting amino acids. Rosetta numbering is used, so that each residue in the model is univocally identified by a number (HP1043 chain A: residues 1-223, DNA chain B: 224-249, DNA chain C: 250-275, protein chain B: 276-298).

3.1.6 Identification of amino acid residues relevant for HP1043 dimerization

As resulted from a dynamic light scattering assay on the *E. coli* purified recombinant HP1043, the dimeric form of the regulator exists in solution in an equilibrium with its monomeric form (data not shown). Differently from other regulators of the OmpR/PhoB family for which it was possible to perform *in vitro* experiments with the sole DBDs and co-crystalize them bound to the DNA (Blanco et al., 2002; Lou et al., 2014; Rhee et al., 2008), the purified HP1043_DBDs were not able to protect a target promoter in DNase I footprinting experiments (Fig. 6C, rightmost panel), likely indicating that dimerization is a prerequisite for HP1043 stable binding to DNA. Therefore, a possible strategy to interfere with HP1043 function as transcriptional regulator is blocking the protein in its monomeric form. To investigate the relevance of different amino acid residues involved in inter-monomer interactions, we selected a few residues per each of the three types of interactions that are present at the dimer interface: D93, R100, and R108 participating in salt bridges, Y94 and R108 involved in H-bond formation, and F87, A107 and A111 forming hydrophobic interactions (Figure 2). A disruptive mutation approach was used, leading to the substitution of charged residues with non-polar ones and vice versa, in order to maximize the perturbation of dimerization. We then tested the ability of mutants in the dimerization domain (HP1043N) to form dimers *in vivo* by exploiting a bacterial two-hybrid system (BACTH, Euromedex) which relies on the functional reconstitution of the *Bordetella pertussis* adenylate cyclase enzyme only in case of dimerization of the fused HP1043Ns. Adenylate cyclase produces cAMP, which in combination with the CAP protein activates the expression of endogenous genes that can be used as a reporter, such as *lacZ*. The β-galactosidase activity was therefore measured to estimate the capability of HP1043N mutants to dimerize in comparison to the wild-type protein (Figure 9). Results show that only mutation of residues involved in ionic interactions and H-bond formation (D93, Y94, A108) abolish dimer formation, confirming a previous hypothesis that ionic interactions might serve as a driving force for initial dimerization, whereas hydrophobic interactions stabilize the dimer (Hong et al., 2007).

In contrast, a different approach was used for mutation of F87. This residue is more often a leucine residue in other response regulator proteins of the OmpR/PhoB family (result of ConSurf analysis, data not shown) and a previous study reports that the F87L mutant is soluble and exists as a monomer by gel filtration (Hong et al., 2007). For this reason, we decided to include also the F87L

mutant in our experiment. Contrary to what was expected, our results indicate that F87L retains the ability to dimerize *in vivo* and was therefore excluded from further analysis.
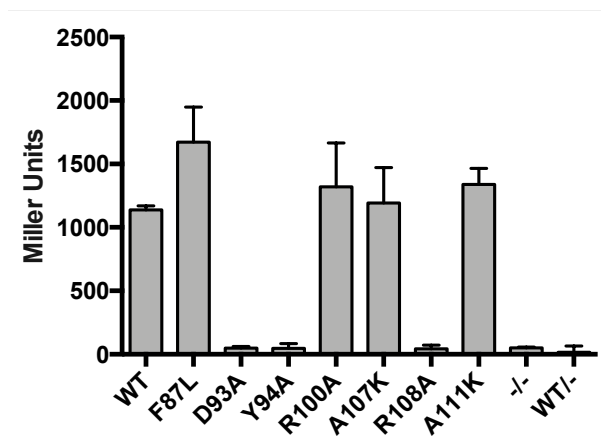
Figure 9: **Detection of *in vivo* interactions between HP1043N mutants using the BACTH system.** LacZ activity of strains carrying wild-type HP1043N (WT), mutant HP1043N or no fused protein (-) was determined from three independent replicates and is given in Miller Units as a measure of the efficiency of adenylate cyclase functional complementation.

To investigate the DNA binding ability of mutants that failed dimerization, we attempted to overexpress the D93A, Y94A, and R108A variants of the full-length HP1043. Unfortunately, the *E. coli* bacterial cultures set up for the production of the D93A and R108A mutants exhibited severe growth defects, hinting at a grade of toxicity of the mutant proteins, while the post-induction control of the Y94A mutant culture showed the accumulation of proteins of lower molecular weight than the one expected for HP1043. Throughout the entire study, the pTrc-HisA plasmid was used for the cloning and overexpression of wild-type and mutant HP1043 in the *E. coli* DH5α host. In the pTrc-HisA plasmid, the gene encoding the protein of interest is put under the control of the strong hybrid P$_{trc}$ promoter, which is repressed by the constitutive expression of the LacI protein and induced by the addition IPTG (isopropyl-β-D-thiogalactoside) to the culture medium. Although the system promotes the reach of high yields of purified proteins, the leakiness in expression also in the absence of the inducer is problematic in the case of toxic proteins. The observed toxicity might be due to the accumulation already in the starter culture of misfolded proteins, which in the case of the Y94A mutant are subjected to protein degradation. In support of this, the D93N and R108E mutants previously investigated by Hong and colleagues were reported to purify in the insoluble fraction (Hong et al., 2007), while mutation of the Y94 residue has not yet been reported. Therefore, a different expression strategy would be required to pursue the

expression of these HP1043 mutants, such as the use of a more tightly regulated promoter (*e.g.* the arabinose promoter P_{BAD}), to avoid basal transcription, and the protease deficient BL21 *E. coli* strain as host.

Nonetheless, since also the full-length protein functionally complemented the *B. pertussis* adenylate cyclase enzyme (data not shown), the BACTH system could be used as a second-line validation step in the HP1043 ligands screening, to investigate ligand-induced conformational changes of HP1043 *in vivo*.

3.2 Investigation of HP1043 regulatory function and regulation

3.2.1 HP1043 interacts with the housekeeping σ factor of the RNA polymerase

To stimulate transcription, HP1043 is likely to interact with the RNA polymerase (RNAP) enzyme. Since HP1043 binding sites on target promoters are located directly upstream the -10 box, and overlap the AT-rich region that substitutes the -35 box in *H. pylori* (Pelliciari et al., 2017), the regulator might act as a bacterial class II transcriptional activator, which most often interact with region 4 of the housekeeping σ factor or the N-terminal domain of the α subunit of the RNAP (Browning and Busby, 2016). To investigate the HP1043-RNAP interaction, we purified the full-length σ^{80} factor and the region 4 of σ^{80} ($\sigma^{80.4}$), and the isolated N-terminal (αNTD) and C-terminal (αCTD) domains of the α subunit of the *H. pylori* RNAP. As expected, $\sigma^{80.4}$ expressed as inclusion bodies due to the elevated hydrophobicity of the domain, while the other recombinant proteins purified in the soluble fraction and were therefore tested for interaction with HP1043 in a dot-blot assay. In this experiment (Fig. 10 B), an equivalent molar amount of the native protein preparations was spotted on a nitrocellulose membrane and incubated with the *E. coli* purified HP1043. Then, the α-HP1043 antiserum and an HRP-conjugated secondary antibody were used for detection, so that a signal was produced only in the case of an interaction between the spotted proteins and HP1043.
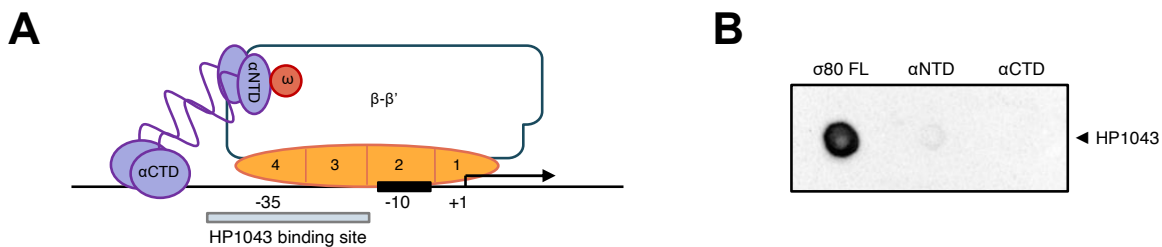
Figure 10. **HP1043 interacts with *H. pylori* σ^{80} factor.** A) Representation of the *H. pylori* housekeeping RNA polymerase holoenzyme (RNAP) on a typical promoter region. The α subunit of RNAP is depicted in *purple*, the β-β' subunit in *blue*, the ω subunit in *red* and σ^{80} factor in *orange*. Functional subdomains of the α subunit and of the σ^{80} factor are also highlighted. In the promoter, the transcriptional start site (+1) is represented as a bent arrow, the -10 sequence element as a *black* box, and HP1043 binding site as *grey* box. B) Result of a protein dot blot assay where the native full-length σ^{80} protein factor (σ^{80} FL), N-terminal and C-terminal domains of the α subunit (αNTD and αCTD respectively) of the *H. pylori* RNAP were spotted in equimolar concentrations on a nitrocellulose membrane and revealed by interaction with the HP1043 purified protein and specific antibodies.

Our results indicate that HP1043 interacts primarily with the σ^{80} factor. The result is coherent with the mechanism of activation carried out by the *E. coli* PhoB, which binds as a dimer to promoters in which the -35 box is substituted by one or multiple copies of an 11-bp direct repeat (*pho box*). The regulator was proven to interact with region 4 of the housekeeping σ^{70} factor via the flexible transactivation loop connecting helices $\alpha 2$ and $\alpha 3$ of the HTH DNA-binding motif (Blanco et al., 2011; Canals et al., 2012; Makino et al., 1996), thus enhancing transcription from promoters which lack the canonical -35 element. On the other hand, an interaction with the α subunit of the RNAP was previously postulated for OmpR, since mutations in the *rpoA* gene specifically affected transcriptional control by the regulator (Slauch et al., 1991). However, no significant signal deriving from a specific HP1043 interaction was detected in correspondence of the spotted αNTD and αCTD protein preparations in our experimental settings.

3.2.2   Construction of an HP1043 overexpressing strain

HP1043 is essential for the viability of *H. pylori* and therefore the *hp1043* gene cannot be deleted, limiting the possibility to investigate its regulatory function. To overcome this problem, Delany and colleagues produced a recombinant *H. pylori* strain in which a second copy of the *hp1043* gene and its ribosome binding site were introduced in the *vacA* locus under the control of the iron inducible P$_{pfr}$ promoter, in an attempt to overexpress HP1043 (Delany et al., 2002). The addition of 1 mM FeSO$_4$ did not affect transcription from P$_{hp1043}$ in the wild-type strain but determined an 11-fold increase in transcription of the additional *hp1043* in the recombinant strain after 15 min of treatment. However, HP1043 protein levels were not increased, nor were the transcript levels of its targets (*hp1043* and *tlpB*), pinpointing the existence of a tight mechanism of regulation of the HP1043 expression happening at the post-transcriptional level. To investigate whether this regulation happens at the transcript or protein level, we produced a recombinant *H. pylori* strain harboring an additional copy of the *hp1043* gene that was altered in nucleotide sequence but encoded a wild-type HP1043 protein (the gene is indicated as *hp1043S*), under the control of the P$_{pfr}$ promoter and ribosome binding site (Fig. 11 A). We then analyzed variations in the transcription of HP1043 target genes (*hp1043* itself, *hp1227*, and *hp1296*) and total HP1043 protein levels after induction with 1mM FeSO$_4$. A 15-fold increase in *hp1043S* transcription was obtained after 15 min from induction and was maintained above 10-fold after 1 h (Fig. 11 C). Diversely from the experiment of Delany and colleagues, HP1043 protein levels did increase in

our recombinant strain, reaching a peak at 45 min after treatment with iron and then diminishing at 1 h (Fig. 11 B). Nonetheless, no significant variation in the transcription of HP1043 targets was obtained (Fig. 11 C). Different experimental settings such as variations in iron concentration, the timing of sampling, or optical density of the culture at the time of induction did not give significantly different results (data not shown).
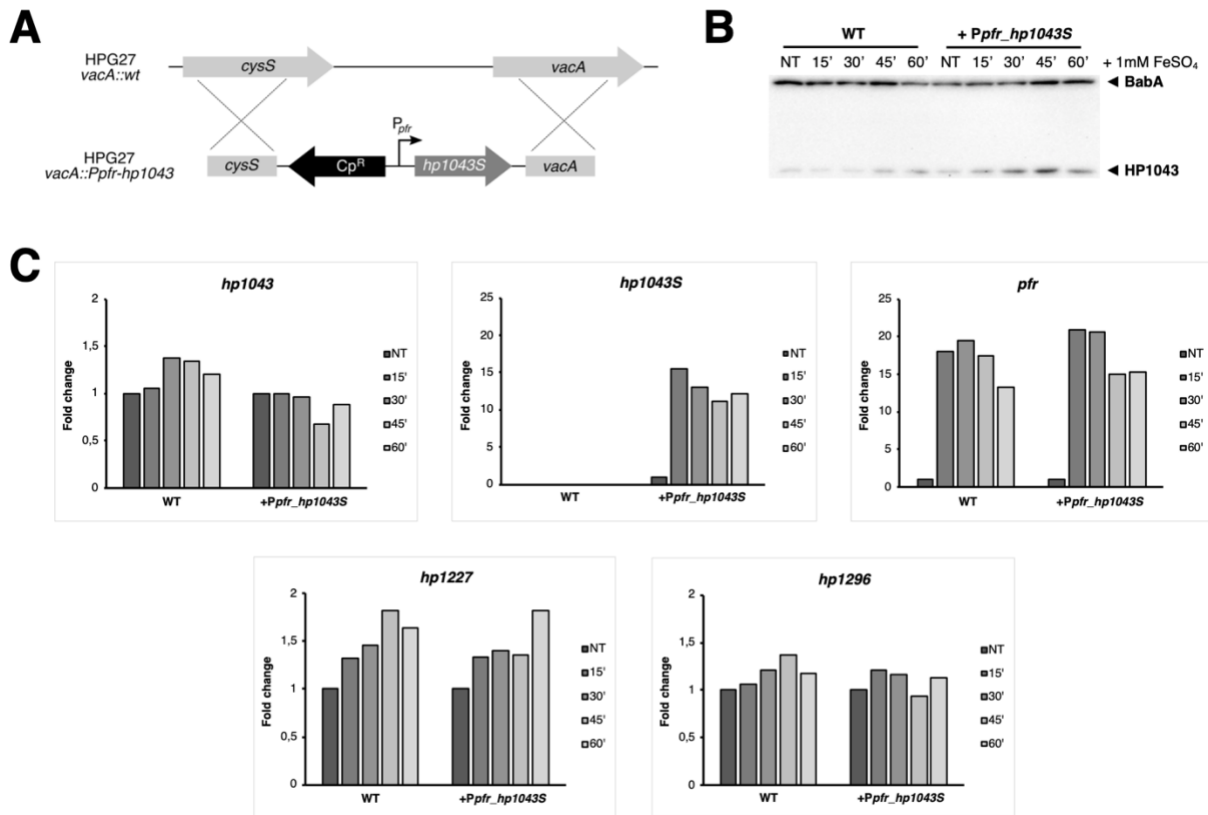
Figure 11. Increased HP1043 protein levels do not result in upregulation of target genes transcription. A) Schematic representation of the *vacA* locus of *H. pylori* G27 wild-type and recombinant strain harboring the *hp1043S* gene under the control of the iron inducible *pfr* promoter obtained by double homologous recombination and selected by chloramphenicol resistance. The *hp1043S* gene encodes the wild-type HP1043 protein but differs in nucleotide sequence from the *hp1043* gene. In B-C) Representative results of an HP1043S induction experiment. Samples were taken before (NT) and after (15, 30, 45, 60 min) induction with 1 mM $FeSO_4$ and processed for protein and RNA extraction. B) Western blot analysis of HP1043 protein levels in comparison with BabA. C) Transcript levels of *hp1043S* and HP1043 target genes (*hp1043*, *hp1227* and *hp1296*) were analysed by Real-Time PCR and normalized to the *16S* rRNA gene. The mRNA abundance of the endogenous *pfr* gene was monitored as positive control of iron induction. Fold change was calculated following the ΔΔCt method. Results were confirmed in other similar time-course experiments.

Since the HP1043 protein amount was downregulated after 1 h from induction of *hp1043S*, a protease might be involved in the post-translational regulation of HP1043 cellular concentration

(Müller et al., 2007; Olekhnovich et al., 2014). Moreover, the fact that *hp1043* mRNA levels were not upregulated in response to increased levels of HP1043 nor downregulated to compensate the elevated HP1043 amount indicates the presence of a further transcriptional or post-transcriptional layer of regulation that maintains steady the levels of *hp1043* mRNA.

3.2.3 Validation of a putative regulatory transcript antisense to the *hp1043* gene

The primary transcriptome analysis of the *H. pylori* 26695 strain revealed the presence of a plethora of sRNAs and massive antisense transcription, indicating that the bacterium uses riboregulation for its gene expression control (Sharma et al., 2010). Moreover, the differential RNA-sequencing (dRNA-seq) approach used allowed to discriminate between primary and processed transcripts, by parallel sequencing of cDNA library pairs: a library of untreated total bacterial RNA (-TEX), and another one enriched for primary transcripts by treatment with terminator exonuclease which degrades 5'-P but not 5'-PPP RNAs (+TEX). Analysis of annotated transcriptional start sites (TSSs) in the *hp1043* locus revealed the presence of an antisense transcript that overlaps the terminal third of the *hp1043* gene (HPnc5480, here denoted as *as1043*) (Fig. 13 B). Interestingly, *hp1043* and *as1043* transcripts show contiguous 5' processed ends in -TEX libraries (Fig. 12).

To validate the existence of the *as1043* and the processed transcripts, we investigated their conservation among strains. To this aim, we set up a pilot experiment in which multiple primer extension (PE) reactions were carried out to identify TSSs and cleaved fragments of both *hp1043* and *as1043* transcripts in the 26695 and G27 *H. pylori* strains (Fig. 13). Sequence alignment shows that the *hp1043* promoter region is highly conserved among the two strains, as the HP1043 binding site and the extended -10 element are identical, demonstrating the relevance of these sequences for the regulation of *hp1043* expression (Fig. 13 A). Although the sequence at the TSS is not conserved, the position of the first nucleotide of the two *hp1043* transcripts as derived from PE is the same. In the case of the regulatory region of *as1043*, the 26695 strain carries a 5 bp deletion in between the extended TG and the -10 box of the promoter in respect to the G27 strain, which explains the 5 bp shift in the determined positions of the TSSs (Fig. 13 C). Nonetheless, the two transcripts show the same pattern in PE (Fig. 13 D-E).
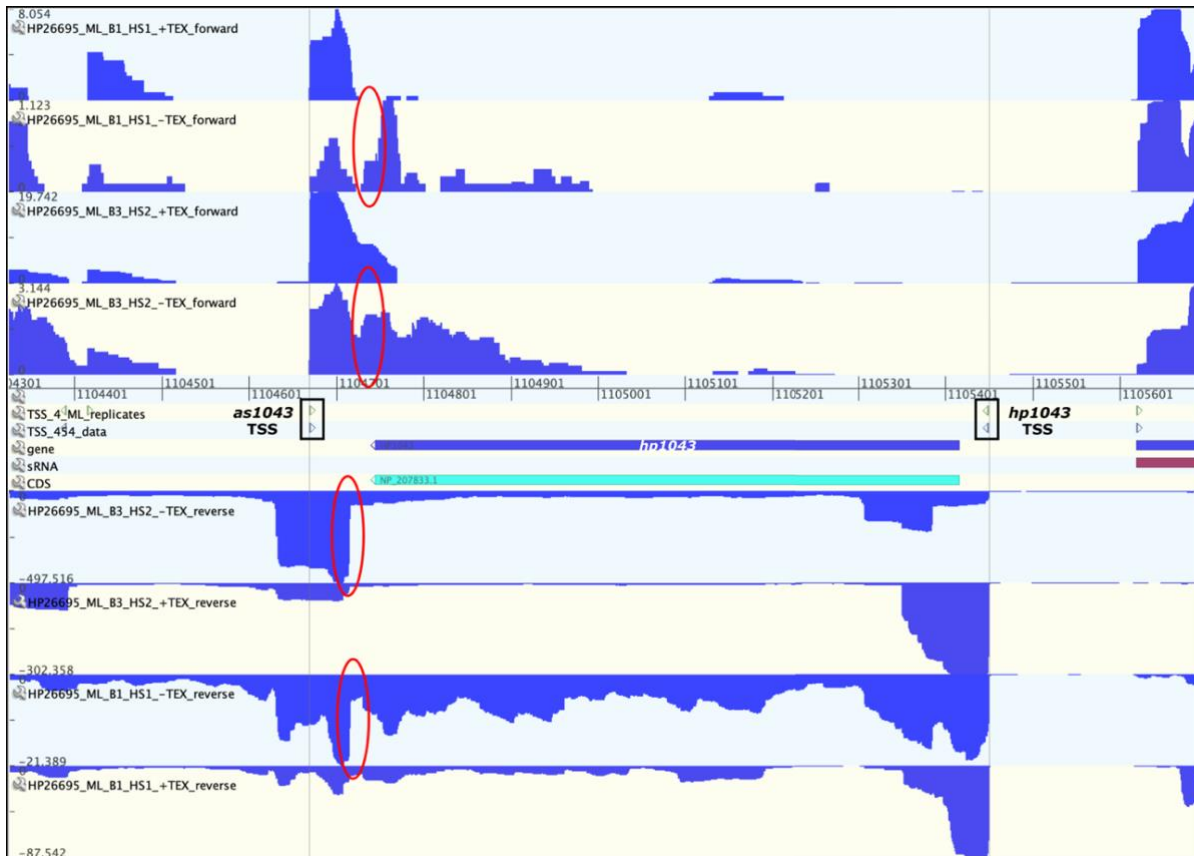
Figure 12: **View of the *hp1043* locus in ±TEX cDNA libraries from a dRNA-seq experiment** (Bischler et al., 2015). Focus on the *hp1043* locus in cDNA libraries from two distinct differential RNA-sequencing (dRNA-seq) experiments as showed in the online genome browser (Bischler et al., 2015). +TEX libraries are enriched in 5'-PPP primary transcripts, while both 5'-P processed and 5'-PPP primary transcripts are present in the -TEX libraries. Transcriptional start sites (TSSs) for *hp1043* and *hpnc5480* (*as1043*) are highlighted with *black* boxes, while 5' processed ends in the -TEX libraries for both *hp1043* and *as1043* are enclosed in *red* ovals.

As expected from the dRNA-seq data, contiguous 5'-termini corresponding to processed transcripts were found for both *hp1043* and *as1043* RNAs in the 26695 strain (Fig. 13 C, red reverse triangles indicated as 1 and A, and Fig. E). Additional 5'-ends (indicated with dashed reverse triangles in Fig. 13 C) closer to the primer used in the PE reactions were enriched and appeared as saturated bands on the autoradiographic films (Fig. 13 E). Given the close proximity of the determined 5'-ends for these processed transcripts (5 bp), we postulate that they result from a cleavage operated on the *hp1043-as1043* mRNA duplex by a double-strand-specific endoribonuclease (ds-endoRNase), and that the additional downstream fragments are stable byproducts of a subsequent degradation by a 5'-to-3' exoribonuclease (5'-3'exoRNase).
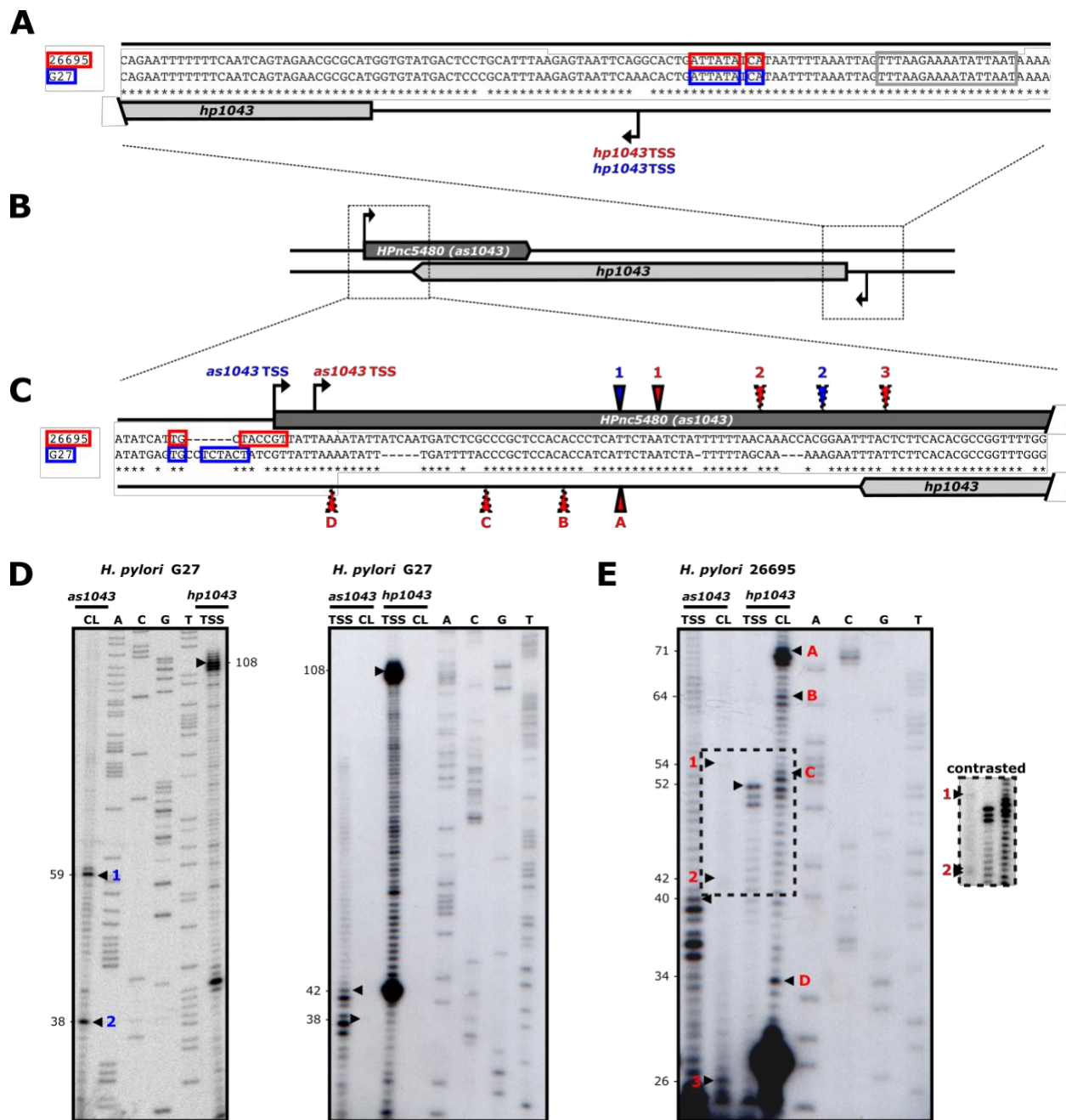
Figure 13: **Primer extension analysis for the validation of an antisense transcript to *hp1043*.** Representation of the regulatory regions of *hp1043* (A) and *HPnc5480* (*as1043*) (C) and the *hp1043* locus (B). In A, C) An alignment of DNA sequences corresponding to the + strand of the genome is reported for *H. pylori* strains 26695 and G27. Extended -10 promoter elements are highlighted. Transcriptional start sites (TSSs) derived from primer extension experiments are reported as bent arrows, while 5'-ends of processed transcripts are indicated with reverse triangles. Symbols are color coded depending on the strain (26695 in *red*, G27 in *blue*). HP1043 binding site in the *hp1043* promoter is enclosed in a *grey* box. D-E) Primer extension analysis in the G27 and 26695 H. pylori strains. In each experiment, a sanger sequencing reaction was loaded to map 5'-termini of the investigated transcripts, which are indicated with a *black* arrowhead, and calculated distances from the 5'-end of the primer used in the reaction are indicated on the left of each panel.

34

A processed *as1043* transcript was identified also in the G27 strain, but no cleaved fragments were detected for *hp1043* (Fig 13 D, rightmost panel). This could be explained by an inefficient labelling reaction of the oligonucleotides used to detect the *hp1043* and *as1043* processed transcripts in this experiment, since the primer preparations obtained after purification from non-incorporated [γ-^{32}P]ATP exhibited significantly lower radioactivity levels in comparison with other primers prepared concomitantly. In support of this, the second primer proved capable of detecting *as1043* cleaved transcripts in a previous experiment (Fig 13 D, leftmost panel). A repetition of this experiment is required to prove this supposition. Moreover, in case our hypothesis on the biogenesis of the processed transcript is valid, another explanation could be that after the first cleavage operated by the dsRNase on the duplex, the produced downstream *hp1043* fragment is rapidly degraded and therefore undetectable.

Our results show that a transcript antisense to the *hp1043* gene is conserved among the two investigated *H. pylori* strains, as well as its processing, which is located nearby a processing site on the *hp1043* mRNA in the 26695 strain. Nonetheless, additional research is necessary to prove the proposed interaction between *hp1043* and *as1043* transcripts, as well as the involvement of RNases in the processing of the duplex and the biological output of the putative cleavage.

4. DISCUSSION

The increasing emergence of bacterial strains resistant to the conventional therapies based on antibiotics induced the World Health Organization (WHO) to publish for the first time a list of 12 pathogens that currently pose a substantial threat to the human population to stimulate the research for novel antimicrobial strategies (Tacconelli et al., 2018). Among these, *Helicobacter pylori* infections represent a major health-care burden because of the high prevalence and worldwide dissemination, as well as the life-long persistence and correlation with gastric malignancies.

The first part of the presented study fits in the search for new bioactive compounds able to inhibit the growth of *H. pylori* and proposes the transcriptional regulator HP1043 as a new molecular target. HP1043 proved to be essential for the viability of the bacterium (McDaniel et al., 2001) and a master regulator of multiple key cellular processes (Pelliciari et al., 2017), therefore it is conceivable that blocking its function would hamper *H. pylori* survival. With this aim, a molecular characterization of the HP1043 amino-acid determinants required for dimerization and DNA binding was carried out. In particular, hydrophobic interactions resulted non-fundamental for HP1043 dimerization, whereas mutation of residues involved in ionic interactions and H-bonds abolished dimer formation, confirming a previous hypothesis that ionic interactions might serve as a driving force for initial dimerization (Hong et al., 2007). Mutational analysis of the DNA-binding domain led to the identification of a few charged residues (R192, K197, T206, and R208) that are fundamental for interaction with DNA. These experimental data, in combination with the available HP1043 NMR structure and the DNA consensus binding sequence information, were used as restraints to guide an *in-silico* protein-DNA docking. The generated model shows an HP1043 dimer interacting in a head-to-head conformation with both the major and minor groove of a target DNA sequence. Although this configuration differs from the canonical head-to-tail conformation of other regulators of the OmpR/PhoB family, site-directed mutagenesis and *in vitro* binding assays validated the model, which was then used to conduct a virtual screening of small molecule libraries to identify compounds possibly able to interfere with the HP1043 regulatory activity. Among the identified 50 putative ligands, 8 compounds exhibited interactions with multiple domains of the dimeric HP1043 in the modeled conformation and represent therefore an interesting selection to investigate further.

The OmpR/PhoB family is the widest family of bacterial response regulators proteins, therefore the identification of ligands that interact selectively with a specific regulator is mandatory to avoid off-targeting effects and alterations to the microbiota. As an example, a recently published research proposed four natural flavonoids as HP1043-specific binders, due to inhibition of the *in vitro* DNA binding activity of the regulator and additive or synergic bactericidal activity in combination with the currently used antibiotics (González et al., 2019). Nonetheless, the best binding poses predicted by molecular docking analysis for these ligands show that interactions with HP1043 occur mainly with amino acid residues of the helix-turn-helix (HTH) DNA binding motif, which is the most conserved element among members of OmpR/PhoB family due to its importance for their activity as transcriptional regulators (Fig. 5). Additionally, flavonoids are natural compounds synthesized by plants in response to microbial infection, and this class of compounds was proven to have a broad-spectrum antimicrobial effect, supporting a non-specific action against *H. pylori* (Górniak et al., 2019). Interestingly, most of the ligands from our selection are predicted to interact with regions of the protein that show lower conservation scores in comparison with orthologues of the OmpR/PhoB family, such as the 4-stranded antiparallel β-sheet (β6-β9) of the DNA binding domain, at the interface with the N-terminal dimerization domain, suggesting that they might target specifically the *H. pylori* regulator. An *in vitro* and *in vivo* validation of the antimicrobial properties of the identified ligands will be carried out, as well as an analysis of their specificity for *H. pylori*. Moreover, the bacterial two hybrid system that was exploited to investigate the dimerization of HP1043 could be used as a reporter assay to evaluate *in vivo* ligand-dependent conformational changes of HP1043 that might interfere with its function.

The second line of research regarded the role of HP1043 as a transcriptional activator and the regulation of *hp1043* expression. A protein dot blot assay confirmed that HP1043 interacts with the RNA polymerase, particularly with the housekeeping σ^{80} factor and coherently with its putative role as a class II transcriptional activator. Additionally, to investigate how *hp1043* expression is regulated, we produced an *H. pylori* recombinant strain harboring an additional copy of the *hp1043* gene altered in nucleotide sequence but encoding a wild-type HP1043 protein (*hp1043S*) under the control of the iron inducible P$_{pfr}$ promoter and ribosome binding site. This set up allowed us, for the first time, to modulate intracellular HP1043 protein levels. However, we failed to demonstrate an upregulation of target genes *in vivo* following overexpression of the protein. Nonetheless, the experiment pinpoints the existence of multiple overlapping mechanisms

of regulation. In particular, since the HP1043 protein amount but not the *hp1043* mRNA level was downregulated after induction of *hp1043S*, we could speculate that a protease is involved in the regulation of HP1043 cellular concentration, as previously proposed but not yet investigated (Müller et al., 2007; Olekhnovich et al., 2014). Moreover, the fact that increased protein levels did not result in increased transcription of target genes suggests that HP1043 might undergo a yet uncharacterized post-translational modification, specific for *H. pylori* and different from phosphorylation, that is necessary for the activation of the protein or increases DNA binding efficiency. Another explanation could be that HP1043 requires a partner to operate its regulatory function or a chaperone to assist its folding and that the concentration of this partner/chaperone was out-titrated in our experiment. These hypotheses are further supported by the observation that an *H. pylori* cell extract fraction enriched in HP1043 (*hp*fHP1043) showed a higher affinity for DNA than the *E. coli* purified recombinant protein (rHP1043), and that a combination of *hp*fHP1043 and rHP1043 led to an enhancement in DNA binding that was greater than what expected from an additive interaction (Olekhnovich et al., 2014). This is also congruent with experimental results of *in vivo* reporter assays set up in the heterologous host *E. coli* in which the transcription of a *lacZ* gene put under the control of HP1043 target promoters failed to be activated after induction of HP1043 (data not shown). Lastly, the fact that *hp1043* mRNA levels were maintained steady during the experiment indicates the presence of a further transcriptional or post-transcriptional layer of regulation. A mechanism of post-transcriptional regulation of *hp1043* was postulated by multiple authors but never demonstrated (Delany et al., 2002; Müller et al., 2007). In particular, Müller and colleagues proposed that the ribosomal binding site in the *hp1043* mRNA could be rendered inaccessible by the formation of a stem-loop structure in its short untranslated leader sequence, due to the presence of an imperfect inverted repeat sequence. However, both in their work and in the previous one from Delany and colleagues, the full region was not used to construct the recombinant *H. pylori* merodiploid strains harboring the *hp1043* gene under the control of different regulatable promoters. Therefore, the proposed mechanism does not explain why increased protein levels were not obtained in the presence of a higher transcript amount (Delany et al., 2002; Müller et al., 2007). Conversely, our results show that if a post-transcriptional regulation exists, it probably involves the coding region or the 3' untranslated region of the *hp1043* mRNA, since the overexpression of a transcript with altered nucleotide sequence in the *hp1043*

open reading frame fused to the *pfr* leader sequence and ribosome binding site proved effective for obtaining increased HP1043 protein amounts.

We propose that a mechanism of post-transcriptional regulation could rely on an antisense transcript (asRNA) which overlaps the last third of the *hp1043* gene (HPnc5480, here indicated as *as1043*). Our results show that the expression of *as1043* is conserved among two *H. pylori* strains (26695 and G27), as well as its processing, favoring a functional relevance of this asRNA over spurious transcription. Moreover, the validated processing site was located in close proximity (~5 bp) to another putative cleavage site found on the *hp1043* mRNA in the 26695 strain. We speculate that the two processing derive from a single cleavage event of the *hp1043-as1043* mRNA duplex operated by a double-strand-specific endoribonuclease (ds-endoRNase). As an example, RNase III is a ds-endoRNase widely conserved among bacteria and eukaryotes that cleaves within RNA secondary structures or duplexes, generating fragments with a distinct 3′ overhang of two nucleotides. RNase III is mainly involved in the maturation of ribosomal RNAs (Iost et al., 2019), but it was also recently proven to regulate toxin-antitoxin duplexes in *H. pylori* (Arnion et al., 2017). Therefore it is conceivable that RNase III processes also other overlapping sense/antisense transcripts, as demonstrated in other bacteria (Lasa et al., 2011). Also, the accumulation of additional stable fragments downstream the cleavage site suggests a subsequent degradation by a 5'-to-3' exoribonuclease (5'-3'exoRNase) of the cleaved RNA ends. RNase J is the only known bacterial exoribonuclease with a 5'-3' directionality. Interestingly, it was recently reported that that the abundance of over 80% of mRNAs and antisense transcripts increased more than 2-fold upon RNase J depletion in *H. pylori*, hinting at an involvement of the enzyme in their degradation (Redko et al., 2016). It remains to be determined whether the reported cleavage in the *hp1043* mRNA represents an initialing signal for mRNA decay, or this processing stabilizes the upstream *hp1043* mRNA fragment harboring the coding sequence of the gene, for example by preventing degradation of the transcript operated by other single-stranded-RNA-specific RNases. Nonetheless, the cleavage site in the *hp1043* mRNA was not retrieved in the G27 strain in our pilot experiment. This could be due to a technical problem in the primer extension reaction, as well as rapid degradation of the processed transcripts. Another explanation could be that *as1043* acts *in trans* to regulate different targets, and the reported cleavage in *hp1043* is a spurious event specific for the 26695 strain. To rule out different hypothesis, additional research is necessary to validate the proposed direct interaction between the *hp1043* and *as1043* transcripts, as well as the

involvement of RNases in the processing of the duplex and the biological output of the putative cleavage.

Taken together, the results of the project shed light on the regulatory mechanism of the essential regulator HP1043 of *H. pylori* and represented a further step towards the design of novel molecules able to block *H. pylori* infection.

5. MATHERIALS AND METHODS

5.1   Bacterial strains and culture conditions.

Bacterial strains used in this study are listed in Table 1. *H. pylori* G27 strains were revived from -80 °C glycerol stocks and propagated at 37 °C on Brucella Broth agar (Oxoid) plates containing 5% complement inactivated fetal calf serum (FCS) and supplemented with Dent's antibiotic supplement. Bacteria were grown at 37°C in a water-jacketed thermal incubator (9% CO_2, 91% air atmosphere, 95% humidity) or in jars in the presence of CampyGenTM (Oxoid) gas-packs. Liquid cultures were grown in Brucella Broth (Sigma-Aldrich) supplemented with 5% complement inactivated FCS. When required, chloramphenicol was added to the medium to achieve a final concentration of 30 µg/mL. *E. coli* DH5α and BTH101 strains were grown in Luria-Bertani (LB) agar or in LB broth (Sigma-Aldrich). When required, media were supplemented with ampicillin (100 µg/mL), kanamycin (25 µg/mL) or chloramphenicol (30 µg/mL).

5.2   DNA manipulations.

DNA manipulations were performed as described previously (Sambrook et al., 1989). All restriction and modification enzymes were used according to the manufacturers' instructions (New England Biolabs). Plasmid DNA preparations were carried out with the NucleoBond Xtra Mini/Midi plasmid purification kits (Macherey-Nagel). DNA fragments for cloning purposes were extracted and purified from agarose gel using the NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel). PCRs were carried out in a MJ Research or Applied Biosystems thermal cycler using PCRBIO Classic Taq or HiFi polymerase (PCR Biosystems).

5.3   Overexpression and purification of recombinant His6-HP1043.

Recombinant N-terminal His-tagged HP1043 protein was overexpressed in *E. coli* DH5α cells transformed with plasmid ptrc::1043. For DNase I footprinting assays, HP1043 was affinity-purified as previously described (Pelliciari et al., 2017) and dialyzed against two changes of 1X 1043 Footprinting Buffer (1X 1043 FPB: 10 mM Tris-HCl pH 7.5; 50 mM NaCl; 10 mM $MgCl_2$; 1 mM DTT; 0.01% Igepal CA-630; 10% glycerol). For EMSA assays, HP1043 was purified as in (González et al., 2019) and dialyzed against the store buffer (50 mM Tris-HCl pH 8, 300 mM NaCl, 10% glycerol). Protein purity was assessed by SDS-PAGE and concentration of protein preparations was estimated by Bradford colorimetric assay (Bio-Rad). HP1043 mutants were obtained by all around PCR performed on plasmid ptrc::1043 with divergent primers listed in Table 3.

5.4   DNase I footprinting.

Plasmid pGEMt-P_{1227} WT harboring the HP1043 binding site on the P_{hp1227} promoter was linearized by enzymatic restriction and 5′-end labeled by T4 polynucleotide kinase in the presence

of $[\gamma^{32}P]$-ATP. DNA probes were obtained by second digestion with the appropriate enzyme and gel purified. Approximately 15 fmol of labeled probe were used per footprinting reaction. DNase I footprinting experiments were performed as previously described (Pelliciari et al., 2017). Briefly, approximately 15 fmol of labelled probes were incubated with increasing amounts of purified recombinant HP1043 in 1X 1043 FPB for 15 min at room temperature, in 50 µL reactions containing 300 ng of sonicated salmon sperm DNA as non-specific competitor. Partial digestion of the labelled probe was obtained by addition to the reaction mixture of 0.066 U of DNase I (Novagen) freshly diluted in 1X 1043 FPB containing 5 mM $CaCl_2$. The digestion was carried out for 90 s at room temperature and arrested with 140 µL of STOP buffer (192 mM NaOAc pH 5.2; 32 mM EDTA pH 8.0; 0.14% SDS; 64 µg/µL sonicated salmon sperm DNA). Samples were phenol-chloroform extracted, ethanol precipitated and resuspended in 10 µL of Formamide Loading Buffer (FLB: 95% formamide, 10 mM EDTA, 0.1% Xylene Cyanol FF, 0.1% Bromophenol Blue). Samples were then denatured at 100 °C for 5 min, chilled on ice, fractionated on a 6% polyacrylamide–8 M urea sequencing gel in TBE buffer, and autoradiographed. A modified G+A sequencing ladder protocol (Liu and Hong, 1998) was employed to map the binding sites.

5.5 Electrophoretic mobility shift assays.

The DNA binding activity of the recombinant HP1043 in the presence of DNA groove binders or IC-substituted DNA probes was assessed by electrophoretic mobility shift assays (EMSAs). A 100-bp promoter region of *hp1227* (26695 annotation; P_{hp1227}) encompassing HP1043 binding site was amplified by PCR with oligos Php1227_EMSA_F and Php1227_EMSA_R and used as target sequence. IC-box substituted P_{hp1227} probe (IC-P_{hp1227}) was obtained by annealing complementary oligonucleotides (Php1227_ICbox_F and Php1227_ICbox_R) to form a double stranded DNA. Recombinant HP1043 protein was mixed with approximately 10 ng of target promoter probe in a 20 µL reaction volume containing 10 mM Tris pH 7.5, 40 mM KCl, 100 mg/L BSA, 1 mM DTT, and 5% glycerol. For binding interference assays, DNA groove binders were added to final concentrations of 0.001, 0.01, 0.1, 1, 10, 100, 500 µM to the reaction mixtures. To ensure the specificity of EMSA assays, a 61-bp DNA fragment of the 16S rRNA gene was included as non-specific competitor DNA. Reactions were incubated at room temperature for 20 min and subsequently separated on a 6% non-denaturing polyacrylamide gel in 1X running buffer (25 mM Tris, 190 mM glycine) at 90 V. The gel was stained with 1X ethidium bromide and visualized with a Gel Doc XR+ image analyzer (Bio-Rad).

5.6 Generation of a HP1043-overexpressing strain of *H. pylori*.

To generate a plasmid for the overexpression of HP1043 under the control of an iron inducible promoter, the P_{pfr} promoter was cloned with oligos Ppfr_BamHI_F and Ppfr_R while the *hp1043S* synthetic gene (obtained by using alternative codons of the wild type gene and coding for the wild type HP1043 protein) was amplified with oligos hp1043S_F and hp1043S_BglII_F. The Ppfr_*hp1043S* fragment was obtained by a PCR-based fusion technique and cloned upstream the

CAT cassette in the pVAC::CAT plasmid after restriction with the appropriate enzymes to obtain pVAC::CAT-P*pfr*hp1043S*. The *H. pylori* strain overexpressing the recombinant HP1043 protein was generated by transformation of naturally competent wild-type *H. pylori* G27 cells with the suicide vector pVAC::CAT-P*pfr*hp1043S*. Overnight cultures were spotted onto plates and grown for a further 5 h. Bacteria were then mixed with 10 to 15 μg of plasmid DNA resuspended in 4 μL of brucella broth and incubated overnight at 37 °C. Transformants were selected on plates added of chloramphenicol (30 μg/mL), and single colonies were selected for further analysis. Insertion of the desired fragment in the *vacA* genomic locus was verified by PCR and sequencing. Overexpression of HP1043 was analyzed by Western blotting with a custom a-HP1043 polyclonal antibody (Biotem).

5.7 RNA isolation.

H. pylori cultures were liquid-grown at 37 °C with gentle shaking (140 rpm) until the appropriate phase of growth (optical density at 600nm [OD600] ~0.6), then bacterial cells were harvested, and total RNA was extracted with TRI-reagent (Sigma-Aldrich/Thermo Fisher Scientific) according to manufacturer's protocol. To measure the HP1043 transcriptional response, cultures of the wild-type and mutant strains grown to mid-log phase were treated with 1 mM $(NH_4)_2Fe(SO_4)_2$ (Sigma-Aldrich) for 1 or 2 hours prior to RNA extraction.

5.8 qRT-PCR analysis.

Synthesis of cDNA and qRT-PCR analysis were carried out as previously described. Briefly, after removal of contaminating genomic DNA through DNase I digestion (Novagen), 1 μg of DNA-free total RNA was mixed with 50 ng of random hexamers (Invitrogen), 1 mM dNTPs, AMV-Reverse Transcriptase (Promega) and incubated for 1 h at 37 °C for cDNA synthesis. Alternatively, RapidOut DNA Removal kit and RevertAid First Strand cDNA Synthesis kit (Thermo Fisher Scientific) were used according to the manufacturer's protocol. qRT-PCR analyses were performed using the qPCRBIO SyGreen Mix Lo-ROX kit (PCR Biosystems) following the manufacturer's protocol. Data were analysed applying the ΔΔCt method, and the 16S rRNA gene was used as internal reference. 16S rRNA amplification curves of qRT-PCR from samples untreated or treated with iron gave overlapping amplification curves, indicating that the amount of 16S rRNA was not changed during the time-course experiment.

5.9 Protein–DNA docking.

The HP1043_DBD-DNA interaction model was obtained using the data-driven docking program HADDOCK 2.2, adopting a two-stage protein–DNA docking approach. The first docking round was performed between a library of HP1043 DNA binding domains derived from the available NMR structure of the whole protein (PDB: 2HQR, residues 120-223) and a linear B-form DNA model of HP1043 binding site on P*hp1227* (TTT<u>TTTAAG</u>CAAAA<u>CTTAAA</u>CTTGTA, hexamers of consensus are underlined) generated through the 3DNA software implemented in the 3D-DART server. Amino acid residues of HP1043_DBD (R192, K197, Y206, K208) and nucleotides in

P*hp1227* (tttTTTAAGcaaaacTTAAacttgta, in uppercase) experimentally validated to be most relevant for HP1043-DNA interaction were used as restraints to guide the modeling procedure. Results of the first docking round were manually screened and used to generate a library of bent B-DNA structures that was used in the second and final docking round. To rebuild the whole HP1043 dimer, the dimeric N-terminal domain of HP1043 (PDB: 2PLN) was modelled with RosettaDock and linkers with MODELLER, as previously described (Musiani et al., 2011).

5.10 Virtual screening of small molecule libraries.

The molecular model of dimeric HP1043 bound to DNA and ligand dataset obtained from ZINC DB (http://zinc15.docking.org/) (Sterling and Irwin, 2015) were employed to conduct a virtual screening for putative HP1043 inhibitors. The about 18300 molecules included in the screening derived from 5 different datasets: substances with transcription factors as main target (11025 molecules), Drug Bank approved molecules (2868), Food and Drug Administration approved drugs (1585), Therapeutic Target Database approved drugs (2164), and NIH clinical collection (669). An in-house Perl script pipeline was employed for ligands preparation, parameter file for docking simulation, slurm job-scheduling input file and docking results analysis. The protein structure was prepared with the ADT interface, by utilizing a docking box of 126 x 126 x 126 points spaced 0.375 Å centered on the protein-DNA complex, including the whole complex with the exception of the DNA underside. The affinity maps for all the atom types available in AutoGrid were pre-calculated. Docking simulations were performed treating the protein as rigid and the ligands as flexible, and 50 runs per each simulation of Lamarkian genetic algorithm were performed with the AutoDock 4.2.6 suite (Morris et al., 2009). Selected molecules were submitted to a deep re-docking screening of 200 runs per simulation, maintaining the other conditions. About 1500 h/cpu of docking simulations were performed on the ITB LINUX cluster. The pipeline parses the docking results, in order to identify, for each simulation, the most representative conformation, that is the conformation with the best energy, and all the conformations with a docking energy within 1 kcal/mol from the first ranked solution. Docking energy, cluster population, estimated Ki value and atomic coordinates of each selected solution are extracted. The pipeline also reconstructs the coordinate files of protein–ligand complexes. The final molecule selection was based on Ki values (in the pM order or lower), and in the case of the re-docking results also on cluster numerosity (greater than 20 units).

5.11 Bacterial adenylate cyclase based two hybrid (BACTH) system.

The DNA region coding for the dimerization domain of HP1043 (HP1043N, residues 1-114) was PCR amplified with the oligos BATCH_HP1043N_BamHI_F and BACTH_HP1043N_KpnI_R, restricted with the appropriate enzymes and cloned in frame to the region encoding the N-terminus of fragments T18 and T25 of *Bordetella pertussis* adenylate cyclase in the plasmids pUT18 and pKNT25, respectively. Mutations in HP1043N were introduced by performing all around PCRs with divergent primers (Table 4) on plasmids pUT18_HP1043N and pKNT25_HP1043N. Plasmids were co-transformed in *cya E. coli* cells (BTH101 strain) and grown until mid-log phase

(OD$_{600}$ ~0.45) when expression of fusion proteins was induced with 1 mM IPTG at 37 °C for 75 min. To assess dimerization, 200 µL of each bacterial culture were transferred to a clear flat bottom 96 well plate (Falcon) to measure optical density at 600 nm (Abs$_{600}$). Then 20 µL of bacterial culture were added to 80 µL of permeabilization buffer (100 mM Na$_2$HPO$_4$, 20 mM KCl, 2 mM MgSO$_4$, 0.8 mg/mL hexadecyltrimethylammonium bromide [CTAB], 0.4 mg/mL sodium deoxycholate, 5.4 µL/mL b-mercaptoethanol) and incubated at 30 °C for 20-30 min. The enzymatic reaction was performed by incubating 14.4 µL of the mixture with 85.6 µL of prewarmed substrate solution (60 mM Na$_2$HPO$_4$, 40 mM NaH$_2$PO$_4$, 1 mg/mL o-nitrophenyl-β-D-Galactoside [ONPG], 2.7 µL/mL β-mercaptoethanol) until appearance of a pale-yellow coloration in the wild-type HP1043N dimerization control. Reactions were stopped by adding 100 µL of 1 M Na$_2$CO$_3$. Absorbance was measured at 420 nm (Abs$_{420}$) for ONPG and 550 nm (Abs$_{550}$) for cell debris, and Miller Units were calculated as follows:

$$1 \text{ Miller Unit (MU)} = 1000 \cdot \frac{\text{Abs}420 - (1.75 \cdot \text{Abs}550)}{t \cdot v \cdot \text{Abs}600}$$

In the formula, t is the reaction time in minutes and v is the volume of culture assayed in milliliters. All absorbance measures were performed with an EnSpire Multimode Plate Reader (Perkin Elmer).

5.12 Thermal shift assay.

Purified HP1043 wild-type and point mutants were diluted to a final concentration of 40 µM in 1X FPB and incubated with 4X SYPRO Orange dye (ThermoFisher) in 25 µL reactions. Protein stability was assessed as thermal unfolding by heating the protein solutions from 25 to 98 °C, at a speed of 1 °C per minute. The melting temperature (T_m) was determined by plotting the first negative derivative of the fluorescence emission at 569 nm as a function of temperature. Thermal denaturation was performed in a 96-well plate, using a CFX96 real time-PCR instrument (Bio-Rad).

5.13 Purification of H. pylori RNA polymerase subunits.

The DNA sequences encoding the full-length σ^{80} factor and σ80 region 4 (σ$^{80.4}$), and the isolated N-terminal (αNTD) and C-terminal (αCTD) domains of the α subunit of the *H. pylori* RNA polymerase (RNAP) were PCR amplified from the *H. pylori* G27 genomic DNA with oligos listed in table 4 and cloned in the pET15b after digestion with the appropriate enzymes. Overexpression of proteins was performed in the *E. coli* BL21 host strain and obtained by induction with 0.6 mM IPTG at 30 °C for 6 h. For protein purification, bacterial pellets were resuspended in 1X RNAP Lysis Buffer (1X RLB: 25 mM Tris pH 8.0, 0.5 M NaCl, 5% glycerol) with 5 mM imidazole and incubated in agitation with 1 mg/mL lysozyme for 1 h at 4 °C. Cells were lysed by sonication and the N-terminal 6xHis-tagged proteins were affinity-purified with the nickel-nitrilotriacetic acid agarose resin (Qiagen) pre-equilibrated in 1X RLB. The resin was washed twice with 1X RNAP Wash Buffer (1X RLB with 20 mM imidazole) and the bound protein was eluted in three steps with 50 µL of 1X RNAP Elution buffer (1X RLB containing 300 mM imidazole). Protein preparations were dialyzed against two changes of 1X RLB and stored in aliquotes at -80°C.

Protein purity was assessed by SDS-PAGE, while protein concentration was estimated by Bradford colorimetric assay (Bio-Rad).

5.14 Protein dot-blot assay.

To investigate HP1043-RNAP interactions, 8 μg of the purified *H. pylori* σ^{80} factor and equivalent molar amounts of αNTD and αCTD were spotted on a nitrocellulose membrane. The membrane was air-dried prior to 1 h of blocking with 5% skim milk in 1X PBST (1X PBS, 0.05% Tween). After a brief wash in 10 mM Tris, 40 mM KCl, 0.01% skim milk, the membrane was incubated with 1X PBST solution containing 5% skim milk and 0.08 mg of *E. coli* purified recombinant HP1043 for 45 min at room temperature (RT). To reveal HP1043-RNAP interactions, the membranes were incubated 1:5000 with a custom a-HP1043 polyclonal rabbit antibody (Biotem) and 1:10000 with an HRP conjugated anti-rabbit secondary antibody. In between incubations, membranes were washed three times in 1X PBST to remove non-specific interactions. To detect interactions, the membrane was incubated with a developing solution (6.8 mM p-Coumaric acid, 1.25 mM luminol, 0.009% H_2O_2) and visualized with a ChemiDoc MP Imaging system (Bio-Rad).

5.15 Primer extension analysis.

Primer extension analyses were performed as previously described (Delany et al., 2002). Briefly, oligonucleotides were 5'-end labeled by T4 polynucleotide kinase in the presence of [γ-^{32}P]ATP for 30 min at 37 °C. Unincorporated [γ-^{32}P]ATP was removed by passing the labeling reaction mixture through a MicroSpin G-25 Column (GE Healthcare). For the primer extension reactions, 12-15 μg of total *H. pylori* RNA were precipitated and resuspended in 10 μL reaction containing 0.1 pmol of radiolabeled primer, 2 mM dNTPs, in 1X AMV buffer (Roche). After a 2 min denaturation at 95 °C, reactions were added of 1 U of AMV reverse transcriptase (Roche) and further incubated at 42 °C for 30 min. Samples were then incubated with 1 μL of RNase A (10 mg/ml) to remove the initial RNA. Samples were phenol-chloroform extracted, ethanol precipitated and resuspended in 8 μL of Formamide Loading Buffer (FLB: 95% formamide, 10 mM EDTA, 0.1% Xylene Cyanol FF, 0.1% Bromophenol Blue). After denaturation for 5 min at 100 °C for 5 min, samples were fractionated on a denaturing 6% polyacrylamide gel, dried and autoradiographed. To ensure correct mapping of the investigated RNA 5' ends, a sequencing reaction was completed by using one of the primers used in primer extension reactions and the corresponding cloned promoter region.

TABLE 2: bacterial strains used in this study

Strain	Genotype	Source/reference
H. pylori **strains**		
G27	Clinical isolate; wild-type parental strain	(Xiang *et al.*,1995)
G27vacA::P$_{pfr}$1043S_CAT	G27 derivative obtained by double homologous recombination with plasmid pVAC::CAT-P$_{pfr}$*hp1043S*; Cm^R	This study
26695	Clinical isolate; wild-type parental strain	(Tomb et al., 1997)
E. coli **strains**		
DH5a	*supE44 DlacU169 (f80 lacZDM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1b*	(Hanahan, 1983)
BTH101	*F$^-$, cya-99, araD139, galE15, galK16, rpsL1 (Str^r), hsdR2, mcrA1, mcrB1*	Euromedex
BL21(DE3)	*F$^-$ ompT hsdS$_B$ (r$_B^-$, m$_B^-$) gal dcm (DE3)*	Invitrogen

TABLE 3: plasmids used in this study

Plasmid	Description	Source/reference
pGEMt-Easy	Cloning vector; Amp^R	Promega
pGEMt-P_{1227} WT	pGEMt-Easy derivative, containing a 61 bp probe corresponding to the region from 1.290.664 to 1.290.725 of *H. pylori* G27 genome. The region corresponds to the promoter of *HPG27_RS06145* gene (*hp1227* according to *H. pylori* 26695 strain annotation).	(Pelliciari, 2018)
pTrcHisA	Expression vector for N-terminal 6xHis-tag cloning; Amp^R	Invitrogen, Inc.
pTrc::1043	Derivative of pTrcHisA expressing the HP1043 response regulator; Amp^R	(Delany et al., 2002)
pVAC::CAT	Cloning vector; Amp^R, Cp^R	(Pepe et al., 2018)
pVAC::CAT-P$_{pfr}$*hp1043S*	Derivative of pVAC::CAT containing upstream of the CAT cassette the synthetic gene *hp1043S* under the control of the P$_{pfr}$ promoter and ribosomal binding site; Amp^R, Cp^R	This study
pKNT25	Cloning vector for BACTH assay; Km^R	Euromedex
pKNT25_HP1043N	Derivative of pKNT25 in which the DNA region coding for the dimerization domain of HP1043 (residues 1-115) was cloned in frame to the region coding for the N-terminus of the T25 fragment of *Bordetella pertussis* adenylate cyclase; Km^R	This study
pUT18	Cloning vector for BACTH assay; Amp^R	Euromedex
pUT18_HP1043N	Derivative of pUT18 in which the DNA region coding for the dimerization domain of HP1043 (residues 1-115) was cloned in frame to the region coding for the N-terminus of the T18 fragment of *Bordetella pertussis* adenylate cyclase; Amp^R	This study
pET15b	Cloning vector for protein expression; Amp^R	Novagen
pET15b_ *rpoD*F	Derivative of pET15b in which the full *H. pylori rpoD* gene encoding the σ^{80} factor was cloned in frame with the N-terminal 6xHis-tag; Amp^R	This study
pET15b_*rpoD*4	Derivative of pET15b in which the subunit 4 of the *H. pylori* σ^{80} factor (residues 599-683) was cloned in frame with the N-terminal 6xHis-tag; Amp^R	This study

pET15b *rpoA*N	Derivative of pET15b in which the segment of the *H. pylori* *rpoA* gene encoding the N-terminal domain (1-238) of the RNA polymerase α subunit was cloned in frame with the N-terminal 6xHis-tag; Amp^R	This study	
pET15b *rpoA*C	Derivative of pET15b in which the segment of the *H. pylori* *rpoA* gene encoding the C-terminal domain (249-337) of the RNA polymerase α subunit was cloned in frame with the N-terminal 6xHis-tag; Amp^R	This study	

TABLE 4: oligonucleotides used for cloning and mutagenesis

Primer	Sequence	*	Source / reference
HP1043_I184A_F	GGCTGAAGTGGCTATCAATCAAATCCG	M	This study
HP1043_E185A_F	GATTGCAGTGGCTATCAATCAAATCCG	M	This study
HP1043_I188A_F	GATTGAAGTGGCTGCCAATCAAATCCGCC	M	This study
HP1043_N189A_F	GATTGAAGTGGCTATCGCTCAAATCCGCC	M	This study
HP1043_N189A_R	ACATTAGGGGTAACCATTTCAGG	M	This study
HP1043_R192A_F	CGCCCAAAAAATGGATAAACCCTTGG	M	This study
HP1043_R192A_R	ATTTGATTGATAGCCACTTCAATC	M	This study
HP1043_Q193A_F	CCGCGCAAAAATGGATAAACCCTTGGGG	M	This study
HP1043_D196A_F	GCTAAACCCTTGGGGATTTCCACG	M	This study
HP1043_D196A_R	CATTTTTTGGCGGATTTGATTGATAG	M	This study
HP1043_K197A_F	GATGCACCCTTGGGGATTTCCAC	M	This study
HP1043_Y206A_F	GGTTGAAGCTGTAAGGCGCAGAGG	M	This study
HP1043_Y206A_R	GTGGAAATCCCCAAGGGTTTATC	M	This study
HP1043_R208A_F	GGTTGAAACCGTAGCGCGCAGAGGCTATC	M	This study
HP1043_EVKtoAAA_F	GCTGCAGCAGGGAAGCCTTTTGAAGTGCTTAC	M	This study
HP1043_EVKtoEAA_F	GAAGCTGCAGGGAAGCCTTTTGAAGTGCTTAC	M	This study
HP1043_EVKtoAAE_F	GCTGCTGAAGGGAAGCCTTTTGAAGTGCTTAC	M	This study
HP1043_EVK_R	AACTTCACGCCCCTTGTAAATAATC	M	This study
HP1043_Q190A_F	AATCCGCCAAAAAATGGATAAAC	M	This study
HP1043_Q190A_R	GCATTGATAGCCACTTCAATCAC	M	This study
HP1043_K194A_F	GCAATGGATAAACCCTTGGGG	M	This study
HP1043_K194A_R	TTGGCGGATTTGATTGATAGC	M	This study
HP1043_R209A_F	GCCAGAGGCTATCGTTTTTGC	M	This study
HP1043_R209A_R	CCTTACGGTTTCAACCGTG	M	This study
BATCH_HP1043N_BamHI_F	ATATGGATCCCATGCGCGTTCTACTGATTG	C	This study
BACTH_HP1043N_KpnI_R	ATATGGTACCCTCAAACGAGCCTCAATC	C	This study
HP1043_D93A_F	TTACATCGCTAAGCCTTACCGCAGC	M	This study
HP1043_D93A_R	GCGTCCGCACCCTGCTCAAAC	M	This study
HP1043_Y94A_F	CCTTACCGCAGCATTAAAGC	M	This study
HP1043_Y94A_R	CTTAGCGATAGCATCGTCCG	M	This study
HP1043_R100A_F	CAGCATTAAAGCTTTAGTCGCAAGG	M	This study
HP1043_R100A_R	GCGTAAGGCTTAGCGATGTAATCGTCC	M	This study
HP1043_A107K_F	AAAAGGATTGAGGCTCGTTTG	M	This study
HP1043_A107K_R	GACTAAAGCCTTAATGCTGCG	M	This study

48

HP1043_R108A_F	GATTGAGGCTCGTTTGAGGTTTTGG	M	This study
HP1043_R108A_R	GCTGCGACTAAAGCTTTAATGCTGCG	M	This study
HP1043_A111K_F	CGTTTGAGGGTACCGAGC	M	This study
HP1043_A111K_R	CTTCTCAATCCTTGCGACTAAAGC	M	This study
Ppfr_BamHI_F	ATAT<u>GGATCC</u>CTAGTTGATACCAACCCC	C	This study
Ppfr_R	AGTATCTCCTTTGTGTTGGTTTAAG	C	This study
hp1043S_F	CTTAAACCAACACAAAGGAGATACTATGCGTGTGCTTCTTATCGAG	C	This study
hp1043S_BglII_F	ATAT<u>AGATCT</u>TCACTCCTCGCAAGCAGG	C	This study
sigma80F_Hp_NdeI_F	ATAT<u>CATATG</u>AAAAAGAAAGCTAACGAAGAAAAAG	C	This study
sigma80_Hp_XhoI_R	ATAT<u>CTCGAG</u>TCAAATGCGCAAATAATTTCTTAAAATG	C	This study
sigma80.4_Hp_NdeI_F	ATAT<u>CATATG</u>TCCATTGATCACATCATGCGAG	C	This study
aNTD_Hp_NdeI_F	ATAT<u>CATATG</u>AAAGTTATCAAAACAGCACCTTTG	C	This study
aNTD_Hp_XhoI_R	ATAT<u>CTCGAG</u>TCAGGGTCTTTCGCCAAAAACG	C	This study
aCTD_Hp_NdeI_F	ATAT<u>CATATG</u>GCTCAAAGAGATGACGCTAAAG	C	This study
aCTD_Hp_XhoI_R	ATAT<u>CTCGAG</u>CTAATTTTCTAATCTTTTCTTTAAACTTTCTCTT	C	This study

* C = cloning, M = mutagenesis. Mutations/restriction sites are underlined.

TABLE 5: oligonucleotides used in Real-time PCR, primer extension and *in vitro* binding assays

Primer	Sequence	**	Source/reference
Php1227_EMSA_F	GCCAAAACGCCTAAAGCC	E	This study
Php1227_EMSA_R	TTGAAAGCGCAATAACCGC	E	This study
Php1227_ICbox_F	GTATAAAATCATATTTATTTCTCTTATTTTCACTTCATTTTTT<u>II</u>GCAAAACT<u>CAI</u>ACTTGTAATTGTATCATTTTAAGATCATTTTGATAAGTAGAGGA	E	This study
Php1227_ICbox_R	TCCTCTACTTATCAAAATGATCTTAAAATGATACAATTACAAGT<u>CTI</u>AGTTTTGC<u>CC</u>AAAAAATGAAGTGAAAATAAGAGAAATAAATATGATTTTATAC	E	This study
1043C_PE	AAAGCCTTTAACATTTAAGCC	PE	(Delany et al., 2002)
hp1043_PE_cleav_Sanger	CATGCCCCTAGCTAAACACC	PE	This study
as1043_PE_TSS	GAATGATGGTGTGGAGCGGG	PE	This study
as1043_PE_cleav	TTTTGCTACCCCAAACCGGC	PE	This study
hp1043-26695_PE_TSS	CAATCAGTAGAACGCGCATGG	PE	This study
hp1043-26695_PE_cleav	AACACAAATATGATATCATTGCTACCG	PE	This study
as1043-26695_PE_TSS	ATGAGGGTGTGGAGCGGG	PE	This study
as1043-26695_PE_cleav	GCTACCCAAAACCGGCGTG	PE	This study
1043RTF	CAGGGTGCGGACGATTACAT	RT	(Pelliciari et al., 2017)
1043RTR	AGCCCCAAAACCTCAAACGA	RT	(Pelliciari et al., 2017)
1043synth_RTF	GTGAGGAAGAGGTTCACGCT	RT	This study
1043synth_RTR	CGCAACCAACGCCTTGATAC	RT	This study

1227RTF	CGCTTGCCATGGGGGTTAAGT	RT	(Pelliciari et al., 2017)
1227RTR	CTTGTTGGCACCGCTTTTGA	RT	(Pelliciari et al., 2017)
1296RTF	TGTCAAAAGAAATGCCTACCGC	RT	This study
1296RTR	ATGGCAAGGATTGCTGGTGT	RT	This study
16SRTF	GGAGTACGGTCGCAAGATTAAA	RT	(Pelliciari et al., 2017)
16SRTR	CTAGCGGATTCTCTCAATGTCAA	RT	(Pelliciari et al., 2017)

** RT = Real-time PCR, PE = primer extension, E = EMSA. Mutations are underlined.

TABLE 6: synthetic genes

Primer	Sequence	Source/ reference
hp1043S	Gene encoding HP1043 with a modified nucleotide sequence: ATGCGTGTGCTTCTTATCGAGAAGAACTCCGTGTTAGGGGGCGAGATCGAGAA GGGGTTGAACGTGAAGGGGTTCATGGCGGACGTGACTGAAAGCTTGGAAGAC GGCGAGTACCTCATGGACATCAGAAACTATGATTTGGTGATGGTGAGCGACAA GAATGCGTTGAGCTTCGTGTCCAGGATTAAAGAGAAACACTCATCTATCGTGG TGTTGGTGTCATCTGACAATCCCACCAGTGAGGAAGAGGTTCACGCTTTCGAA CAAGGCGCTGATGACTATATTGCGAAACCCTATCGTAGTATCAAGGCGTTGGT TGCGAGAATCGAAGCGCGCTTAAGATTCTGGGGGTCAAACGTTATCGAGATCG GCGACTTAACTATCAGTCCCGACGAGGAGAAAATCATCTATAAAGGCCGCGA GGTGGAGGTTAAGGGCAAACCCTTCGAGGTTCTCACTCACCTCGCTAGGCACA GAGACCAAATCGTTTCTAAGGAGCAACTCTTGGATGCGATCTGGGAGGAACCC GAGATGGTGACTCCCAACGTTATCGAGGTTGCGATTAACCAGATTCGTCAGAA GATGGACAAGCCTTTAGGCATCTCTACTGTGGAGACTGTGAGACGTAGGGGGT ACCGCTTCTGTTATCCTAAGCCTGCTTGCGAGGAGTGA	Genewiz

6. SUPPLEMENTARY MATHERIAL

TABLE 7: full list of putative HP1043 ligands.

Ligand	Binding Energy	Run	Cluster numerosity	Ki	
ZINC000022443609	**-23.84**	**38**	**54**	**3.37**	**aM**
ZINC000013492579	-16.79	117	21	494.39	fM
ZINC000003938681	-16.41	66	62	937.24	fM
ZINC000003938681	-15.76	72	34	2.79	pM
ZINC000019363537	**-15.57**	**81**	**34**	**3.86**	**pM**
ZINC000001566899	-15.41	98	40	5.06	pM
ZINC000004214772	-15.31	147	51	5.96	pM
ZINC000100088802	-14.99	156	152	10.26	pM
ZINC000150338708 *	**-14.39**	**11**	**27**	**28.4**	**pM**
ZINC000013492553	-14.3	190	20	33.17	pM
ZINC000028823338	-14.14	124	36	43.44	pM
ZINC000004097448	-14.05	43	21	50.02	pM
ZINC000004097404	-13.99	79	58	55.55	pM
ZINC000222508879	-13.99	75	24	55.58	pM
ZINC000028823530	-13.93	60	29	61.88	pM
ZINC000013494070	-13.92	42	45	63.08	pM
ZINC000150338708 *	**-13.89**	**44**	**27**	**65.55**	**pM**
ZINC000019364225	**-13.84**	**70**	**51**	**71.87**	**pM**
ZINC000150338708	-13.81	87	76	75.8	pM
ZINC000013494068	-13.73	46	44	86.02	pM
ZINC000043206370	-13.56	176	53	115.23	pM
ZINC000001612996	-13.32	199	20	171.12	pM
ZINC000043206370	-13.2	129	35	210.22	pM
ZINC000013494069	**-13.1**	**10**	**31**	**250.98**	**pM**
ZINC000036701290	-13.08	103	26	256.35	pM
ZINC000013494068	-13.07	97	21	264.89	pM
ZINC000019364225	-13.04	169	34	274.57	pM
ZINC000028823491	-12.92	139	25	339.8	pM
ZINC000087666889	-12.92	22	32	335.64	pM
ZINC000098023177	-12.89	92	22	356.1	pM
ZINC000150598297	-12.82	121	22	401.34	pM

ZINC000004097383	-12.81	40	75	406.52	pM
ZINC000028134101	-12.79	153	23	421.91	pM
ZINC000028823512	-12.79	70	26	420.99	pM
ZINC000043206370	-12.79	122	26	422.27	pM
ZINC000087666882	-12.79	140	55	420.91	pM
ZINC000072316335	**-12.7**	**128**	**25**	**487.81**	**pM**
ZINC000022448983	-12.66	167	27	526.94	pM
ZINC000013494069	-12.65	92	50	534.17	pM
ZINC000028823490	-12.56	47	20	616.87	pM
ZINC000028823510	**-12.53**	**9**	**20**	**652.73**	**pM**
ZINC000003978083	-12.46	143	51	737.86	pM
ZINC000030690433	-12.46	46	38	740.65	pM
ZINC000052955754	-12.46	31	26	736.06	pM
ZINC000012503187	-12.42	134	35	791.48	pM
ZINC000014880002	**-12.39**	**15**	**52**	**827.22**	**pM**
ZINC000030690433	-12.38	91	23	841.79	pM
ZINC000087666886	-12.38	137	47	837.75	pM
ZINC000222508821	-12.32	63	45	936.79	pM
ZINC000028823531	-12.29	149	21	985.83	pM

7. BIBLIOGRAPHY

Alfizah, H., Norazah, A., Hamizah, R., and Ramelah, M. (2014). Resistotype of Helicobacter pylori isolates: the impact on eradication outcome. J. Med. Microbiol. *63*, 703–709.

Ansari, S., and Yamaoka, Y. (2019). Helicobacter pylori Virulence Factors Exploiting Gastric Colonization and its Pathogenicity. Toxins *11*.

Arnion, H., Korkut, D.N., Masachis Gelo, S., Chabas, S., Reignier, J., Iost, I., and Darfeuille, F. (2017). Mechanistic insights into type I toxin antitoxin systems in Helicobacter pylori: the importance of mRNA folding in controlling toxin expression. Nucleic Acids Res. *45*, 4782–4795.

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., and Ben-Tal, N. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res. *44*, W344–W350.

Bauer, S., Endres, M., Lange, M., Schmidt, T., Schumbrutzki, C., Sickmann, A., and Beier, D. (2013). Novel function assignment to a member of the essential HP1043 response regulator family of epsilon-proteobacteria. Microbiol. Read. Engl. *159*, 880–889.

Beier, D., and Frank, R. (2000). Molecular characterization of two-component systems of Helicobacter pylori. J. Bacteriol. *182*, 2068–2076.

Bik, E.M., Eckburg, P.B., Gill, S.R., Nelson, K.E., Purdom, E.A., Francois, F., Perez-Perez, G., Blaser, M.J., and Relman, D.A. (2006). Molecular analysis of the bacterial microbiota in the human stomach. Proc. Natl. Acad. Sci. U. S. A. *103*, 732–737.

Bischler, T., Tan, H.S., Nieselt, K., and Sharma, C.M. (2015). Differential RNA-seq (dRNA-seq) for annotation of transcriptional start sites and small RNAs in Helicobacter pylori. Methods San Diego Calif *86*, 89–101.

Blanco, A.G., Sola, M., Gomis-Rüth, F.X., and Coll, M. (2002). Tandem DNA Recognition by PhoB, a Two-Component Signal Transduction Transcriptional Activator. Structure *10*, 701–713.

Blanco, A.G., Canals, A., Bernués, J., Solà, M., and Coll, M. (2011). The structure of a transcription activation subcomplex reveals how σ(70) is recruited to PhoB promoters. EMBO J. *30*, 3776–3785.

Browning, D.F., and Busby, S.J.W. (2016). Local and global regulation of transcription initiation in bacteria. Nat. Rev. Microbiol. *14*, 638–650.

Buckler, D.R., Zhou, Y., and Stock, A.M. (2002). Evidence of Intradomain and Interdomain Flexibility in an OmpR/PhoB Homolog from Thermotoga maritima. Structure *10*, 153–164.

Canals, A., Blanco, A.G., and Coll, M. (2012). σ70 and PhoB activator. Transcription *3*, 160–164.

Croxen, M.A., Ernst, P.B., and Hoffman, P.S. (2007). Antisense RNA Modulation of Alkyl Hydroperoxide Reductase Levels in Helicobacter pylori Correlates with Organic Peroxide Toxicity but Not Infectivity. J. Bacteriol. *189*, 3359–3368.

Danielli, A., and Scarlato, V. (2010). Regulatory circuits in Helicobacter pylori: network motifs and regulators involved in metal-dependent responses. FEMS Microbiol. Rev. *34*, 738–752.

De la Cruz, M.A., Ares, M.A., von Bargen, K., Panunzi, L.G., Martínez-Cruz, J., Valdez-Salazar, H.A., Jiménez-Galicia, C., and Torres, J. (2017). Gene Expression Profiling of Transcription Factors of Helicobacter pylori under Different Environmental Conditions. Front. Microbiol. *8*.

Delany, I., Spohn, G., Rappuoli, R., and Scarlato, V. (2002). Growth Phase-Dependent Regulation of Target Gene Promoters for Binding of the Essential Orphan Response Regulator HP1043 of Helicobacter pylori. J. Bacteriol. *184*, 4800–4810.

Donczew, R., Makowski, Ł., Jaworski, P., Bezulska, M., Nowaczyk, M., Zakrzewska-Czerwińska, J., and Zawilak-Pawlik, A. (2015). The atypical response regulator HP1021 controls formation of the Helicobacter pylori replication initiation complex. Mol. Microbiol. *95*, 297–312.

Fallone, C.A., Chiba, N., Zanten, S.V. van, Fischbach, L., Gisbert, J.P., Hunt, R.H., Jones, N.L., Render, C., Leontiadis, G.I., Moayyedi, P., et al. (2016). The Toronto Consensus for the Treatment of Helicobacter pylori Infection in Adults. Gastroenterology *151*, 51-69.e14.

Foynes, S., Dorrell, N., Ward, S.J., Stabler, R.A., McColm, A.A., Rycroft, A.N., and Wren, B.W. (2000). Helicobacter pylori possesses two CheY response regulators and a histidine kinase sensor, CheA, which are essential for chemotaxis and colonization of the gastric mucosa. Infect. Immun. *68*, 2016–2023.

Friedland, N., Mack, T.R., Yu, M., Hung, L.-W., Terwilliger, T.C., Waldo, G.S., and Stock, A.M. (2007). Domain Orientation in the Inactive Response Regulator Mycobacterium tuberculosis MtrA Provides a Barrier to Activation. Biochemistry *46*, 6733–6743.

González, A., Salillas, S., Velázquez-Campoy, A., Espinosa Angarica, V., Fillat, M.F., Sancho, J., and Lanas, Á. (2019). Identifying potential novel drugs against Helicobacter pylori by targeting the essential response regulator HsrA. Sci. Rep. *9*, 11294.

Górniak, I., Bartoszewski, R., and Króliczewski, J. (2019). Comprehensive review of antimicrobial activities of plant flavonoids. Phytochem. Rev. *18*, 241–272.

Haley, K.P., and Gaddy, J.A. (2015). Helicobacter pylori: Genomic Insight into the Host-Pathogen Interaction (Hindawi).

Hong, E., Jung, J.-W., Shin, J., Kim, J.H., Jeon, Y.-H., Yamazaki, T., and Lee, W. (2004). Backbone 1H, 13C and 15N resonance assignments of the response regulator HP1043 from Helicobacter pylori. J. Biomol. NMR *28*, 85–86.

Hong, E., Lee, H.M., Ko, H., Kim, D.-U., Jeon, B.-Y., Jung, J., Shin, J., Lee, S.-A., Kim, Y., Jeon, Y.H., et al. (2007). Structure of an Atypical Orphan Response Regulator Protein Supports a New Phosphorylation-independent Regulatory Mechanism. J. Biol. Chem. *282*, 20667–20675.

Hooi, J.K.Y., Lai, W.Y., Ng, W.K., Suen, M.M.Y., Underwood, F.E., Tanyingoh, D., Malfertheiner, P., Graham, D.Y., Wong, V.W.S., Wu, J.C.Y., et al. (2017). Global Prevalence of Helicobacter pylori Infection: Systematic Review and Meta-Analysis. Gastroenterology *153*, 420–429.

Iost, I., Chabas, S., and Darfeuille, F. (2019). Maturation of atypical ribosomal RNA precursors in Helicobacter pylori. Nucleic Acids Res. *47*, 5906–5921.

Jeong, K.-W., Ko, H., Lee, S.-A., Hong, E., Ko, S., Cho, H.-S., Lee, W., and Kim, Y. (2013). Backbone Dynamics of an Atypical Orphan Response Regulator Protein, Helicobacter pylori 1043. Mol. Cells *35*, 158–165.

Lasa, I., Toledo-Arana, A., Dobin, A., Villanueva, M., de los Mozos, I.R., Vergara-Irigaray, M., Segura, V., Fagegaltier, D., Penadés, J.R., Valle, J., et al. (2011). Genome-wide antisense transcription drives mRNA processing in bacteria. Proc. Natl. Acad. Sci. U. S. A. *108*, 20172–20177.

Li, Y.-C., Chang, C., Chang, C.-F., Cheng, Y.-H., Fang, P.-J., Yu, T., Chen, S.-C., Li, Y.-C., Hsiao, C.-D., and Huang, T. (2014). Structural dynamics of the two-component response regulator RstA in recognition of promoter DNA element. Nucleic Acids Res. *42*, 8777–8788.

Lou, Y.-C., Wang, I., Rajasekaran, M., Kao, Y.-F., Ho, M.-R., Hsu, S.-T.D., Chou, S.-H., Wu, S.-H., and Chen, C. (2014). Solution structure and tandem DNA recognition of the C-terminal effector domain of PmrA from Klebsiella pneumoniae. Nucleic Acids Res. *42*, 4080–4093.

Lou, Y.-C., Weng, T.-H., Li, Y.-C., Kao, Y.-F., Lin, W.-F., Peng, H.-L., Chou, S.-H., Hsiao, C.-D., and Chen, C. (2015). Structure and dynamics of polymyxin-resistance-associated response regulator PmrA in complex with promoter DNA. Nat. Commun. *6*, 8838.

Mailhe, M., Ricaboni, D., Vitton, V., Gonzalez, J.-M., Bachar, D., Dubourg, G., Cadoret, F., Robert, C., Delerce, J., Levasseur, A., et al. (2018). Repertoire of the gut microbiota from stomach to colon using culturomics and next-generation sequencing. BMC Microbiol. *18*, 157.

Makino, K., Amemura, M., Kawamoto, T., Kimura, S., Shinagawa, H., Nakata, A., and Suzuki, M. (1996). DNA binding of PhoB and its interaction with RNA polymerase. J. Mol. Biol. *259*, 15–26.

Martínez-Hackert, E., and Stock, A.M. (1997). Structural relationships in the OmpR family of winged-helix transcription factors11Edited by M. Gottesman. J. Mol. Biol. *269*, 301–312.

McDaniel, T.K., Dewalt, K.C., Salama, N.R., and Falkow, S. (2001). New approaches for validation of lethal phenotypes and genetic reversion in Helicobacter pylori. Helicobacter *6*, 15–23.

Mizuno, T., and Tanaka, I. (1997). Structure of the DNA-binding domain of the OmpR family of response regulators. Mol. Microbiol. *24*, 665–667.

Morris, G.M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S., and Olson, A.J. (2009). AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. J. Comput. Chem. *30*, 2785–2791.

Müller, S., Pflock, M., Schär, J., Kennard, S., and Beier, D. (2007). Regulation of expression of atypical orphan response regulators of Helicobacter pylori. Microbiol. Res. *162*, 1–14.

Musiani, F., Ciurli, S., and Dikiy, A. (2011). Interaction of selenoprotein W with 14-3-3 proteins: a computational approach. J. Proteome Res. *10*, 968–976.

Nguyen, M.-P., Yoon, J.-M., Cho, M.-H., and Lee, S.-W. (2015). Prokaryotic 2-component systems and the OmpR/PhoB superfamily. Can. J. Microbiol. *61*, 799–810.

Niehus, E., Gressmann, H., Ye, F., Schlapbach, R., Dehio, M., Dehio, C., Stack, A., Meyer, T.F., Suerbaum, S., and Josenhans, C. (2004). Genome-wide analysis of transcriptional hierarchy and feedback regulation in the flagellar system of Helicobacter pylori. Mol. Microbiol. *52*, 947–961.

Nowak, E., Panjikar, S., Konarev, P., Svergun, D.I., and Tucker, P.A. (2006). The structural basis of signal transduction for the response regulator PrrA from Mycobacterium tuberculosis. J. Biol. Chem. *281*, 9659–9666.

Olekhnovich, I.N., Vitko, S., Chertihin, O., Hontecillas, R., Viladomiu, M., Bassaganya-Riera, J., and Hoffman, P.S. (2013). Mutations to essential orphan response regulator HP1043 of Helicobacter pylori result in growth-stage regulatory defects. Infect. Immun. *81*, 1439–1449.

Olekhnovich, I.N., Vitko, S., Valliere, M., and Hoffman, P.S. (2014). Response to Metronidazole and Oxidative Stress Is Mediated through Homeostatic Regulator HsrA (HP1043) in Helicobacter pylori. J. Bacteriol. *196*, 729–739.

Pelliciari, S. (2018). Biochemical and functional characterization of the HP1043 orphan response regulator of the human pathogen Helicobacter pylori. Tesi di dottorato.

Pelliciari, S., Pinatel, E., Vannini, A., Peano, C., Puccio, S., De Bellis, G., Danielli, A., Scarlato, V., and Roncarati, D. (2017). Insight into the essential role of the Helicobacter pylori HP1043 orphan response regulator: genome-wide identification and characterization of the DNA-binding sites. Sci. Rep. *7*.

Pepe, S., Pinatel, E., Fiore, E., Puccio, S., Peano, C., Brignoli, T., Vannini, A., Danielli, A., Scarlato, V., and Roncarati, D. (2018). The Helicobacter pylori Heat-Shock Repressor HspR: Definition of Its Direct Regulon and Characterization of the Cooperative DNA-Binding Mechanism on Its Own Promoter. Front. Microbiol. *9*.

Pflock, M., Finsterer, N., Joseph, B., Mollenkopf, H., Meyer, T.F., and Beier, D. (2006). Characterization of the ArsRS regulon of Helicobacter pylori, involved in acid adaptation. J. Bacteriol. *188*, 3449–3462.

Pflock, M., Müller, S., and Beier, D. (2007). The CrdRS (HP1365-HP1364) two-component system is not involved in ph-responsive gene regulation in the Helicobacter pylori Strains 26695 and G27. Curr. Microbiol. *54*, 320–324.

Redko, Y., Galtier, E., Arnion, H., Darfeuille, F., Sismeiro, O., Coppée, J.-Y., Médigue, C., Weiman, M., Cruveiller, S., and De Reuse, H. (2016). RNase J depletion leads to massive changes in mRNA abundance in Helicobacter pylori. RNA Biol. *13*, 243–253.

de Reuse, H., and Bereswill, S. (2007). Ten years after the first Helicobacter pylori genome: comparative and functional genomics provide new insights in the variability and adaptability of a persistent pathogen. FEMS Immunol. Med. Microbiol. *50*, 165–176.

Rhee, J.E., Sheng, W., Morgan, L.K., Nolet, R., Liao, X., and Kenney, L.J. (2008). Amino Acids Important for DNA Recognition by the Response Regulator OmpR. J. Biol. Chem. *283*, 8664–8677.

Robinson, V.L., Wu, T., and Stock, A.M. (2003). Structural Analysis of the Domain Interface in DrrB, a Response Regulator of the OmpR/PhoB Subfamily. J. Bacteriol. *185*, 4186–4194.

Roncarati, D., and Scarlato, V. (2018). The Interplay between Two Transcriptional Repressors and Chaperones Orchestrates Helicobacter pylori Heat-Shock Response. Int. J. Mol. Sci. *19*.

Salama, N.R., Hartung, M.L., and Müller, A. (2013). Life in the human stomach: persistence strategies of the bacterial pathogen Helicobacter pylori. Nat. Rev. Microbiol. *11*, 385–399.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular cloning: a laboratory manual (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory).

Scarlato, V., Delany, I., Spohn, G., and Beier, D. (2001). Regulation of transcription in Helicobacter pylori: simple systems or complex circuits? Int. J. Med. Microbiol. *291*, 107–117.

Schär, J., Sickmann, A., and Beier, D. (2005). Phosphorylation-independent activity of atypical response regulators of Helicobacter pylori. J. Bacteriol. *187*, 3100–3109.

Sharma, C.M., Hoffmann, S., Darfeuille, F., Reignier, J., Findeiß, S., Sittka, A., Chabas, S., Reiche, K., Hackermüller, J., Reinhardt, R., et al. (2010). The primary transcriptome of the major human pathogen Helicobacter pylori. Nature *464*, 250–255.

Slauch, J.M., Russo, F.D., and Silhavy, T.J. (1991). Suppressor mutations in rpoA suggest that OmpR controls transcription by direct interaction with the alpha subunit of RNA polymerase. J. Bacteriol. *173*, 7501–7510.

Sterling, T., and Irwin, J.J. (2015). ZINC 15 – Ligand Discovery for Everyone. J. Chem. Inf. Model. *55*, 2324–2337.

Sutton, P., and Boag, J.M. (2019). Status of vaccine research and development for Helicobacter pylori. Vaccine *37*, 7295–7299.

Svennerholm, A.-M., and Lundgren, A. (2007). Progress in vaccine development against Helicobacter pylori. FEMS Immunol. Med. Microbiol. *50*, 146–156.

Tacconelli, E., Carrara, E., Savoldi, A., Harbarth, S., Mendelson, M., Monnet, D.L., Pulcini, C., Kahlmeter, G., Kluytmans, J., Carmeli, Y., et al. (2018). Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. Lancet Infect. Dis. *18*, 318–327.

Vianna, J.S., Ramis, I.B., Ramos, D.F., VON Groll, A., and Silva, P.E.A. da (2016). DRUG RESISTANCE IN HELICOBACTER PYLORI. Arq. Gastroenterol. *53*, 215–223.

Visweswariah, S.S., and Busby, S.J.W. (2015). Evolution of bacterial transcription factors: how proteins take on new tasks, but do not always stop doing the old ones. Trends Microbiol. *23*, 463–467.

Walthers, D., Tran, V.K., and Kenney, L.J. (2003). Interdomain Linkers of Homologous Response Regulators Determine Their Mechanism of Action. J. Bacteriol. *185*, 317–324.

Yang, I., Nell, S., and Suerbaum, S. (2013). Survival in hostile territory: the microbiota of the stomach. FEMS Microbiol. Rev. *37*, 736–761.