Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

Biologia Cellulare e Molecolare

Ciclo XXX

Settore Concorsuale: 05/E2

Settore Scientifico Disciplinare: BIO/11

Biochemical and functional characterization of the HP1043 orphan response regulator of the human pathogen *Helicobacter pylori*

Presentata da: Pelliciari Simone

Coordinatore Dottorato	**Supervisore**
Giovanni Capranico	Vincenzo Scarlato
	Co-supervisore
	Davide Roncarati

Esame finale anno 2018

- **Annex 1:** Pelliciari S & Pinatel E, Vannini A, Peano C, Puccio S, De Bellis G, Danielli A, Scarlato V, 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,* 23;7:4106.

1. Aim of the project

The HP1043 protein of *Helicobacter pylori* is an essential transcriptional regulator and is expected to play a role in the regulation of crucial cellular processes. Even though some detailed structural information are available for HP1043, a deep understanding of its function and the definition of its target genes have been hampered so far by the fact that the *hp1043* gene cannot be deleted, nor the amount of protein modulated.

The first aim of the project was to elucidate the HP1043 regulon through genome-wide identification of the binding sites using Chromatin Immunoprecipitation-sequencing (ChIP-seq), one of the most powerful approaches to characterize protein-DNA interactions *in vivo*. To validate ChiP-seq results, *in vitro* DNA binding assays have been performed, enabling precise mapping of the HP1043 binding sites on a subset of identified targets. Their analysis revealed the presence of a conserved nucleotide sequence motif, that has been analysed through scanning mutagenesis to identify the most important nucleotide sequence determinants for protein binding.

A second aim was to elucidate the functional role of HP1043 in the control of gene transcription. Indeed, to establish whether the protein acts as an activator or as a repressor of transcription. However, none of the applied approaches was successful. To investigate the mechanism of action of the transcriptional regulator, experiment of *in vitro* transcription have been performed. This assay allowed to evaluate the effect of HP1043 transcription, and thus to establish a direct positive role of the regulatory protein bound to a promoter to stimulate transcription, acting as an activator of transcription

2. Introduction

2.1 Origin and features of bacterial cells

Prokaryotic cells are smaller than eukaryotic ones, with dimensions that vary between 0,1 and 10 µm, and show a various number of shape as rounded (cocci as *Staphylococcus aureus*), rods (bacilli as *Escherichia coli*) but also spirilliforms (as *Helicobacter pylori*), which present flagella in one or both extremities, and spirochetes (*Trepanoma pallidum*) that are helicoidal. In spite of all these different forms and shapes, bacteria present a similar cellular organization among species, with, from outside to inside, an external capsule, a cell wall (more or less thick, depending if the bacteria belonging to GRAM – or GRAM +, respectively), a plasmatic membrane, a cytoplasm and a circular nucleoid (Maresca et al. 2016).

Capsule

The majority of prokaryotes, and almost all pathogenic bacteria, has a capsule, a mucilaginous layer composed by monosaccharides principally involved in adhesion on other cells and biofilm formation. Some bacteria can strictly regulate the capsule formation, to generate this saccharides layer only in specific growth stages or during infection. In fact, it has been observed that in several pathogenic bacteria as *Neisseria meningitidis* and *S. aureus*, capsule defective strains have an increased biofilm formation and adherence to host cells, but less immune evasion and decreased virulence. For these reasons, capsule is often associated to a quiescent state of the bacterium, in which this structure guarantee protection against defensive mechanism of the host and diverse environmental stresses (Maresca et al. 2016).

Cell wall

Bacteria are commonly classified in two groups, depending on thickness of cell wall. This nomenclature derives from a staining technique developed by Hans Christian Gram and based on the crystal violet stain. Specifically, GRAM positive bacteria have a thick peptidoglycan layer (20 – 80 nm) outside the cell membrane and retain the dye, while GRAM negatives are characterized by a thinner peptidoglycan layer (2-7 nm), positioned between the membrane that enclose cytoplasm, called inner membrane, and another phospholipidic bilayer external from the peptidoglycan, called external membrane. The space between the cell wall and both membrane is called periplasm. Because of this peculiar physical characteristic, Gram-negative bacteria do not retain the crystal violet stain (Maresca et al. 2016).

In Gram-positive bacteria, peptidoglycan chains represent ~80% of the cell wall weight (5/10% in Gram-negatives), while lipids and proteins are almost absent. Moreover, they show an high concentration of teichoic acid, an antigenic compound formed by polyphosphate and D-alanine side-chains.

Gram-negative bacteria are characterized by a more complex cell wall, schematized in Fig. 1, with a periplasmic space that constitutes about 40% of the cell volume and a high number of proteins in the peptidoglycan. In addition, the external membrane has an internal layer made by phospholipids and an external layer with an exclusive presence of lipopolysaccharide (LPS), which contain sugars with antigenic proprieties.

LPS is made up of three components:

- *Lipid A*, presents two glucosamine molecules esterified with three fatty acids chains that anchor the compound to the membrane.

- *Polysaccharide core*, which has a variable sugar structure with at least one residue of 3 deoxy-D-manno-oct-2-ulosonic acid (KDO) and of L-glycero-D manno heptose (heptose).

- *O antigen*, the distal portion of the LPS; it is extremely variable in its composition of sugar repeating units (from one to six), which is the cause of the antigen variability among different pathogenic strains; in addition, in many of these species O antigen can be absent as strategy to evade immune system, as in *N. meningitidis* or *Bordetella pertussis.*
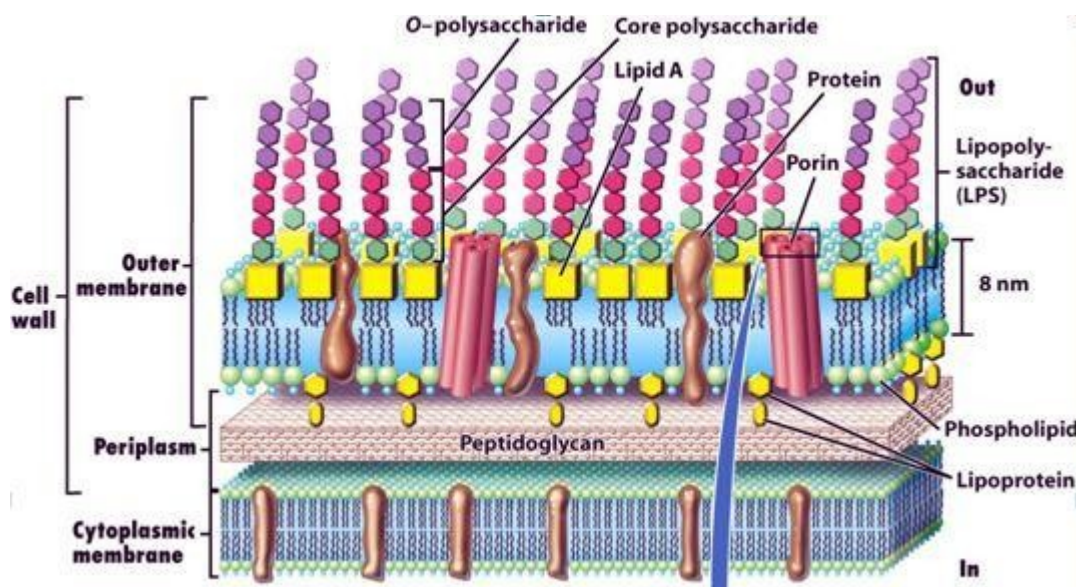
Fig. 1 Schematic representation of a GRAM – outer membranes: in the outer membrane is reported also the LPS, in which the different sugars are reported as hexagons of different colors and lipid A as a yellow box (Figure 4-35a Brock Biology of Microorganism 11/e 2006 © Pearson Prentice Hall, Inc.).

Bacterial membrane

Prokaryotic membranes are composed by a double layer of phospholipids, which are constituted by a glycerol molecule, esterified with two fatty acid chains (apolar tails of the molecules, hydrophobic) and a phosphate group (polar head, hydrophilic). Phospholipids can be *saturated*, in which fatty acids contain only single carbon bonds (C-C), and *unsaturated*, that contains also double bond (C=C) and are subdivided in *cis* isomers (when carbons are on the same side of the double bond) and *trans* isomers (when the carbons are on the opposite side): the few differences between the two classes of lipids have a great influence on the chemical and physical proprieties of the membrane. In fact, the double bond between two carbon atoms induce a kink in the phospholipid molecule, that generate a sort of disorder into the membrane. Membranes made up of saturated phospholipids appear very rigid, and have an high fusion point. On the contrary, when also unsaturated phospholipids are present, the membrane become more fluid, with a lower fusion point. For this reason, extremophile bacteria, such as *Cyanobacteria*, can modify the unsaturated/saturated phospholipid ratio in order to survive to extreme temperatures (Maresca et al. 2016).

Bacterial membranes have also a role as scaffold for outer membrane proteins (OMP), that are extremely important for essential functions of the cell. OMPs can be divided in *trans-membrane proteins*, that pass through the membrane with a part of their aminoacidic sequence (trans-membrane helix), and *peripheral proteins*, that are linked to the membrane through a partner trans-membrane protein or via interaction with the polar heads of the phospholipids (Maresca et al. 2016).

Cytoplasm

The cytoplasm is an heterogeneous mixture of different elements that represents the vital core of the bacterial cell. It is the location in which metabolic processes, protein synthesis, and all other pathway that allow bacteria to survive and proliferate take place.

The molecules inside cytoplasm are distributed and compartmentalized depending on their functions. This is possible due to the presence of cytoskeleton inside the cell, which is also fundamental to confer the characteristic shape to every bacteria (Maresca et al. 2016).

The cytoskeleton is organized in three components, that are *microfilaments*, which form a complex 3D structure immediately under the membrane and are composed by single monomers of actin (a globular protein of 55 kDa), *intermediate filaments* and *microtubules*, that are empty pipes-like structure made by tubulin monomers. These two lasr classes of filaments form the scaffold of the cell, due to their organization as a tridimensional web which make possible the transport and compartmentalization of cytoplasmic proteins and molecules (Maresca et al. 2016).

Above all, the most important molecule present in the cytoplasm is the bacterial chromosome, a long circular double strand DNA filament that encodes for the genetic information. The prokaryote genome is aploid, which means that contain only one copy of each gene, and is often anchored to the inner membrane at one pole of the cell.

2.2 Genomes Organization

The DNA is the source of the genetic information of all living cells, and is inherited through generations by daughter cells from the parental ones. This double-helix molecule is composed by 4 nitrogenous bases, that are adenine (A), guanine (G), cytosine (C) and thymine (T); the different sequence of these 4 nucleotides can encode all the information needed from cells for their maintenance and spreading.

In prokaryotic world, genomes have different dimensions, due to heterogeneity of bacterial cells. For example, among the sequenced genomes to date, the smallest one belongs to the intracellular symbiont *Carsonella ruddi*, with 180 Kb DNA, while the biggest belongs to *Sorangium cellulosum*, that is constituted of 13 Mb of DNA. Another important feature of bacterial genomes is the number of encoded genes. The minimal number of genes for bacterial survival is estimated between 300 and 700. By contrast there are genomes containing more than 8,000 genes. However, this parameter does not appear to be directly correlated with genomes size, due to the presence of intergenic

regions and non-coding portions whose size varies among different organisms (Maresca et al. 2016).

Besides dimensions or genes content, the genome has three main functions to fulfill, which are preservation, replication and expression. In addition, there are many characteristics that are common to any type of cells:

- The DNA is organized in a duplex that is capable of self-replication in a semi-conservative manner;
- The coding genetic material is allocated in specific regions called genes, spatially separated from each other;
- The genomes have non-coding DNA (promoter regions, 5' UTR and 3'UTR, etc…);
- Major changes in a given genomic DNA derive from internal rearrangements (for example gene duplication) or acquisition of foreign DNA (HGT, horizontal gene transfer).

Bacterial genomes are constituted by a single chromosome (excluding some exception as *Vibrio* or *Brucella*). However, prokaryotic cells can acquire and maintain circular plasmids that encode for extra-chromosomal genetic material, carrying their own replication origin. Plasmids do not often carry essential genes but appear to be useful for bacterial environmental adaptation, virulence, or stress response (Yamaichi et al, 1999, Jumas-Bilak et al, 1998)

The chromosome in prokaryotic cells looks "naked", with no proteic organization structures condensing DNA like that exerted by histones in eukaryotic cells, but organized in wire loops held together by a proteinaceous core (Kavenoff et al, 1976). These features make the prokaryotic genome easier to replicate and transcribe but less protected against external insults. However, the prokaryotic DNA double helix is continuously complexed with thousands of proteins as transcriptional factors, RNA polymerase (RNAp) molecules and nucleoid associated proteins (NAPs). The binding of all these proteins to the chromosome is likely to exert multiple roles. First of all, NAPs organize the bacterial nucleoid by packaging and condensing the DNA molecule in the cell. This likely contribute to protection of DNA from environmental stresses and insults, and in linking distal genome regions. Instead, transcriptional factors (that can act also include NAPs) and RNAp are part of the machinery transcribing DNA information in RNA molecules expressing the genetic information (Daubin et al, 2013, Rocha et al, 2003, Rocha et al, 2004, Visweswariah et al, 2015, Madan Babu et al, 2007).
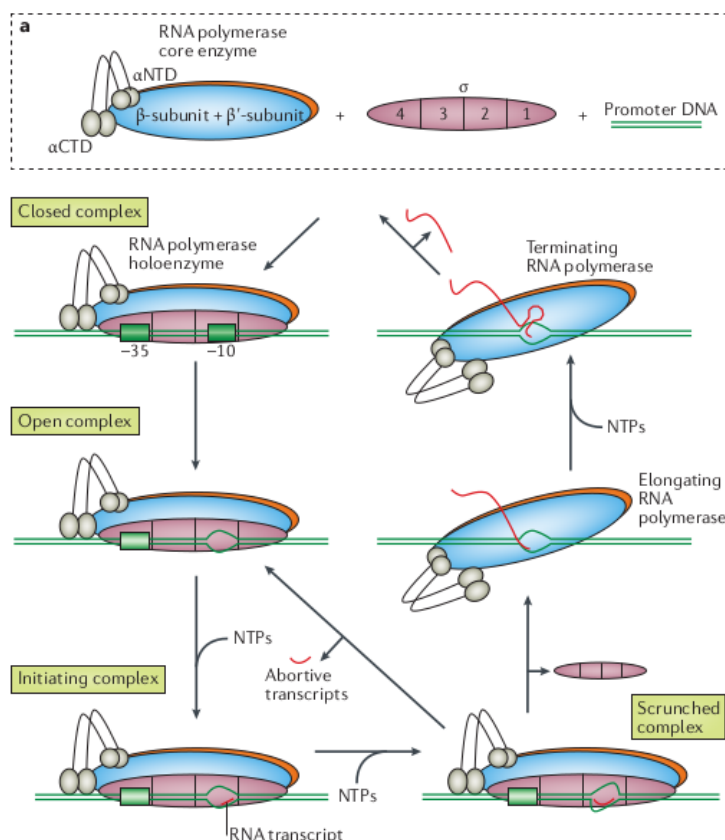
2.3 Transcription process

Transcription is the mechanism that transfers the genetic information from DNA to a single stranded RNA molecule. RNA is composed by three out of four nitrogenous bases that compose the DNA (A, C and G), as the thymine is substituted by uracil (U); moreover, the scaffold sugar of these bases is ribose instead of deoxyribose.

RNA synthesis is directed by the enzyme RNA polymerase (RNAp) a large protein complex composed of a *core* enzyme that includes the large β- and β′-subunits, two α-subunits and the small ω-subunit. To start RNA synthesis, the core enzyme requires binding of an accessory protein called sigma factor to form the RNAp *holoenzyme*.

The sigma factor allows RNAp to recognize, specific DNA regions at the beginning of genes, called promoters, in order to start RNA transcription. These regions show a -10 conserved sequence, TATAAT (TATAbox), and a -35 sequence, TTAGACA initially identified in *E.coli* and shown to be widely used among bacteria. An additional DNA element contacted by the C-terminal domain of the α-subunit of the RNAp holoenzyme, called *UP-element*, is located between -38 and -59 bp upstream of the transcriptional start site (TSS). An UP-element shows a degenerate consensus sequence AAAWWTWTTTTnnnAAAnn that facilitate RNAp binding and subsequent step(s) to enhance efficiency of transcription. These sequence elements reside in the *coding* DNA strand (Estrem et al, 1998).

Once the RNAp *holoenzyme* binds to a promoter region it forms a closed complex and then the transcription mechanism starts with a series of intermediate steps: first of all the protein unwinds the double-stranded DNA region of the TSS. Positions +1 and +2 of the unwound template strand (corresponding to the 5′ end of the RNA transcript) are then able to enter the active site to form the transcriptionally competent open complex. Addition of triphosphate ribonucleotides (NTPs) induce the formation of initiation complex which, in the first phase, pull the unwind DNA into the active site of the *holoenzyme* (a process known as scrunching) with the generation of small RNA fragments of 4/6 nucleotides called abortive transcripts. In this step of transcription process there is an unstable equilibrium between closed complex, open complex and scrunching complex, which vary depending on the promoter strength, as reported in Fig. 2 (Ruff et al, 2015). Variations of these promoter features usually correlate to the degree of similarity of the -10, the -35 and the UP-element sequences to their consensus and in turn to the promoter strength. Higher is the strength of a promoter, more favorite is the transition from closed complex to initiation complex. When RNAp is forming the scrunching initiation complex, it can either revert to closed complex and detaches

from the promoter or start the elongation process, in which the *holoenzyme* might releases the sigma factor and start synthesizing the RNA molecule. Once the RNAp reaches the end of the gene locus, it encounters a terminator sequence where the enzyme stops transcription and detaches from the DNA (Browning et al, 2016).

Fig. 2 RNAp transcriptional initiation phases; after the formation of a closed complex, the DNA is unwinded and, depending on promoter strength, the RNAp can escape promoter entering elongation phase; otherwise, the holoenzyme is stuck on the promoter (Browning D. F, Busby S. J, *Nat. Rev. Microbiol*, 2016, Figure courtesy by Nature publishing group).

2.4 Transcriptional control

The first step of transcriptional control is evidenced by the use of alternative sigma factors. An alternative sigma factor is a sigma factor that associates to the RNAp core enzyme to recognize different promoter DNA elements.

Sigma polypeptides are composed by 4 aminoacid domains which interact with different DNA elements of a promoter region: domains 1 and 2 are involved in the open complex formation, binding -10 sequence and the region between -10 and the transcriptional start site. Docking of the protein in this region is particularly important to prevent transcription start when sigma factor is still

bound to core enzyme. Domain 3 binds an extended TG of the -10 sequence, which improve the strength of the promoter element, whereas domain 4 is linked to recognition of the -35 sequence (Browning et al, 2016). The fact that each domain of sigma factors interacts with a specific promoter sequence, implies that different factors can discriminate among them to guide the RNAp only on specific promoters.

Regulation of transcription might be mediated by a class of proteins called regulators. These proteins exert their action by directly targeting specific stretches of nucleotide sequences usually neighboring promoter regions. Most of the regulatory proteins binds to specific DNA sequences by mean of a structural helix–turn–helix motif. Usually, this motif recognizes targets of 4–5 base pairs and affinity of binding is often increased by protein homodimerization (or higher order multimerization) of the transcriptional regulator. The role of transcription factors in controlling transcription is often exerted in response to environmental changes or stresses. In some cases, the activity of the regulator is mediated by the binding of a small molecule, ligand, or a covalent modification, in some other cases it depends on the availability of the regulator, which can be modulated by synthesis, turnover or sequestration. Transcription factors might also generate a complex regulatory networks (Browning et al, 2016).

Moreover, as the output of the action of a transcriptional factor on transcription could be either negative or positive, they have been divided in two major groups, repressors and activators of transcription, respectively.

Repressor factors

The simplest mechanism of transcription repression relies on the use of a dimeric protein that binds to DNA between -10 and -35 promoter regions, thus preventing RNAp docking by steric hindrance. Repression of gene transcription of several promoters is controlled in this way. In some instances, efficiency of repression is increased by the binding of additional repressor dimers to the operator that shows multiple adjacent binding sites. In other cases, the operators map outside the promoter region and efficient repression of transcription is achieved by a looping mechanism of the DNA molecule that prevents RNAp binding, as occurs on the galactose operon of *E. coli*. In this system, GalR represses transcription by binding to operators mapping upstream and downstream of the transcription start site. Moreover, in some promoters other proteins might act as anti-activator (CytR in *E.coli*), removing or preventing binding of an activator protein (Browning et al, 2016).

Activator factors

Activator factors exert their action to increase transcription, often from a low basal levels, by one out of three known mechanisms: class I activation, class II activation, and activation by a conformational change as schematized in Fig. 3. In class I activation, the activator binds upstream of the promoter elements to recruit the RNA polymerase on the promoter. Such a mechanism is often applied to promoters displaying a suboptimal consensus sequence in the -10 or the -35 region. In class II activation, the activator binds close or overlapping the −35 element of the promoter and activation of transcription is achieved by a direct interaction of the regulatory protein either with domain 4 of the sigma factor, or by the interaction with the N-terminal domain of the α-subunit of RNAp. The third mechanism of transcriptional activation relies on conformational changes of the promoter region when -35 and -10 sequence are incorrectly spaced. The binding of a regulatory protein can alter the DNA conformation to optimize the required distance between the -10 and -35 elements, allowing promoter recognition by the RNAp (Browning et al, 2016).
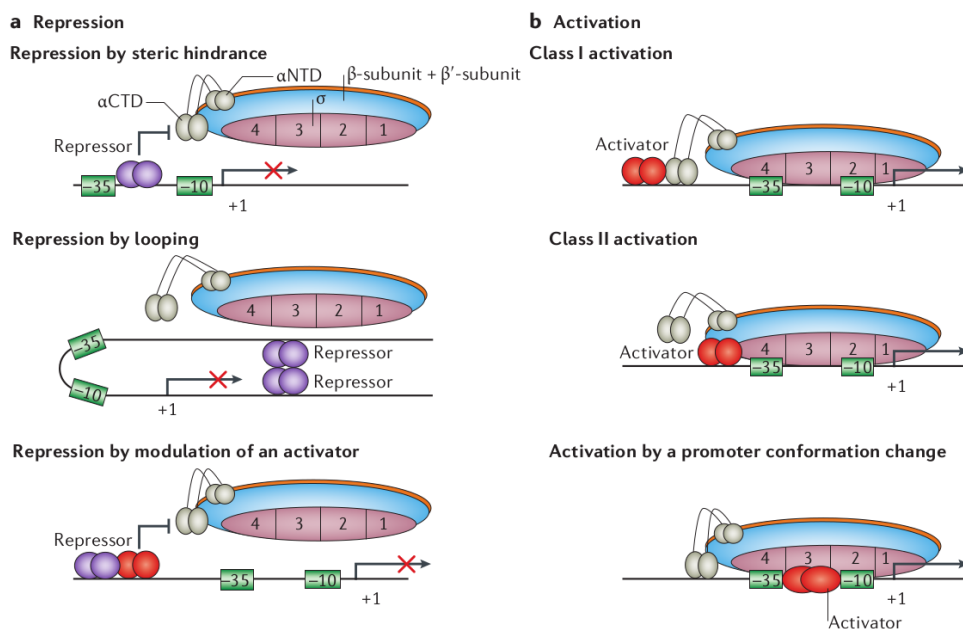
Fig. 3 Action mechanism of different regulators on core promoters and relative interaction with RNAp. In violet are reported the repressive transcriptional factors, in red the activators; the RNAp is represented as in Fig 2 (Browning D. F, Busby S. J, *Nat. Rev. Microbiol*, 2016, Figure courtesy by Nature publishing group).

2.5 Transcriptional regulation in *Helicobacter pylori*

H.pylori is a GRAM negative microaerophilic flagellated bacteria that colonize gastric mucus layer in human stomach; it is the principal causative agent of chronic gastritis, gastric cancer, duodenal ulcers and MALT lymphoma, and is associated with the development of adenocarcinoma. Comprehend its transcriptional regulation can be very important to open new perspective in *H. pylori* infection treatment, due to the emerging insurgence of antibiotic resistant strains.

The restricted gastric habitat of *H. pylori* has been associated with reduced functional redundancy of its small genome ($\sim$1.6 Mb), characterized by a constrained number of regulators strictly interconnected between them. The transcriptional machinery of *H. pylori* is similar to the one observed in many other organism, such as *E. coli*, but with some substantial differences. In fact, in the RNA polymerase holoenzyme of H. pylori, unlike in most species, the β and β′ subunits are found as a single fused gene product and this fusion likely confers a selective advantage. Also, the *H. pylori* σ^{80} specificity subunit of the RNA polymerase has several differences from the σ^{70} subunit of *E. coli* (32% identity, 51% similarity) and from other bacteria (Borin BN et al. 2014). In spite of the low promoters sequence homology between *E. coli* and *H. pylori*, the second ones can be activated in *E. coli* and *in vitro* by the *E. coli* RNA polymerase. Thus, functional similarities between the *E. coli* and *H. pylori* RNA polymerases have been postulated (Beier et al, 1998). The *H. pylori* promoters, in fact, do not show all the classical core elements found in bacteria: it is still possible the identification of a well conserved TATA box (-10 sequence, TATAAT), sometimes with an extended TG, but they lack the canonical -35 sequence, which seems to be substituted by an AT rich stretch region (Sharma et al, 2010).

Survival of *H. pylori* in the gastric niche depends on the concerted expression of virulence factors and housekeeping genes. To withstand the continuous stresses imposed by the harsh acidic environment and to counteract the host's responses, *H. pylori* needs fast transcriptional response through different two components systems with interconnected pathways, as reported in Fig. 4. For example, to respond to pH stress, *H. pylori* has the two system component ArsS-ArsR, which regulate genes still to be identified because of the impossibility to delete ArsR transcriptional regulator. However, it seem to be fundamental both for bacterial survival and infection (Pflock et al, 2005, Pflock et al, 2006).

During infection, *H. pylori* has to face off with metal starvation and these are fundamental enzyme cofactors triggering also the expression of virulence proteins. For this reason, the bacterium has to constantly compete with the host for these essential nutrients. On the other hand, metal ions are

toxic if present intracellularly in high amounts. Therefore, their homeostasis is tightly controlled. Three systems are dedicated to this fundamental task in *H. pylori*: the CrdRS two-component system, the ferric uptake regulator (Fur) involved in iron homeostasis, and a homolog of the Ni-responsive NikR regulator of *E. coli*.

Fur is a small (17kDa) regulatory protein that regulates iron homeostasis. This regulation typically occurs when Fur, which is bound by its ferrous iron cofactor (*holo* Fur), binds to specific DNA sequences (Fur boxes) in the promoters of genes involved in iron uptake and represses their expression; these Fur boxes tend to map close to the core promoter elements. *Holo* Fur repression is the best-characterized in *H. pylori* and its regulon includes hundreds of repressed genes. This regulatory protein can control transcription also in the absence of its iron cofactor (*apo* Fur) by repressing a set of different genes (Delany et al, 2001).

Another important metal transcriptional regulator is NikR, a pleiotropic regulator, acting as either a repressor or an activator of a variety of genes. This protein binds DNA only in his *holo* form, when bound by the nickel ion. It regulates transcription upon binding to the promoters of genes that encode factors involved in nickel import (*nixA, fecA, frpB4, exbB*), nickel storage (*hpn*), enzymatic activity (*ureAB*), autoregulation (*nikR*), as well as genes involved in iron homeostasis (*fur, pfr*) and possibly other cellular processes.

The last known metal ion traffic control system is the two components system CrdRS, primary involved in copper trafficking and resistance in *H. pylori*. This regulative pathway exerts its effect on the *crdA* gene, that is responsible for the detoxification of copper ions accumulated in bacterium cytoplasm (Waidner et al, 2005).

Another fundamental survival process that *H. pylori* has to face is the heat shock stress. This pathway is well studied and includes three multicistronic operons that are negatively controlled by two transcriptional repressors, HrcA and HspR. Only HspR is needed to repress its own gene promoter, while both regulators are necessary to repress the *groESL* and *hrcA* operons including also the *dnaK* gene (Roncarati et al, 2014). The others two components systems present in *H. pylori*, as well as the known alternative sigma factors, are all involved in chemiotaxis and flagella formation and regulation.

***H. pylori* transcription factors and regulators.**

TF	Gene	Description	Function
σ^{80}	*rpoD (HP0088)*	Vegetative sigma factor	Housekeeping
σ^{54}	*rpoN (HP0714)*	Alternative sigma factor	Flagellar regulation; *class II* genes; basal body, hook
σ^{28}	*fliA (HP1032)*	Alternative sigma factor	Flagellar regulation; *class III* genes; late structures
FlgR	*flgR (HP0703)*	NtrC-like response regulator	Flagellar regulation; RpoN-dependent regulation
FlgS	*flgS (atoS) (HP0244)*	NtrB-like cytoplasmic histidine kinase	Flagellar regulation; acid acclimation
CheA	*cheA (HP0392)*	Histidine kinase	Chemotaxis
CheY	*cheY (HP1067)*	Response regulator	Chemotaxis
CheY2	*(HP0392)*	CheY-like reciever domain fused to CheA	Chemotaxis; P~sink
ArsR	*arsR (HP0166)*	ompR-like response regulator	Acid acclimation
ArsS	*arsS (HP0165)*	Transmembrane histidine kinase	Acid acclimation; periplasmic acid sensor
HP1021	*(HP1021)*	Atypical orphan response regulator	Acetone metabolism
HP1043	*(HP1043)*	Atypical orphan response regulator	Growth
CrdR	*crdR (HP1364)*	Response regulator	Copper resistance
CrdS	*crdS (HP1365)*	Histidine kinase	Copper resistance
Fur	*Fur (HP1027)*	Ferric uptake regulator	Pleiotropic; metal ion homeostasis
NikR	*nikR (HP1338)*	*E. coli* NikR homologue	Pleiotropic; nickel ion homeostasis
HrcA	*hrcA (HP0111)*	*B. subtilis hrcA* repressor homologue; membrane associated	Heat shock; stress response
HspR	*hspR (HP1025)*	*Streptomyces* spp. *hspR* repressor homologue	Heat shock; stress response

Tab. 1 Summary table of *H. pylori* transcriptional factors with related description and function (putative for some of them). The two unknown regulators are enlightened with a red box (Danielli et al. *Plos Path.* 2010, article under Creative Commons CC-BY license).

Among the less studied transcriptional regulators of *H. pylori* there is HP1043. This because the *hp1043* gene appears to be essential for bacterial viability as suggested by the impossibility to generate a knock out mutant nor to modulate its expression *in vivo* (Delany et al, 2002). Our interest in the last few years has been focused on the identification of the genes regulated by this protein and to understand its biological function within the cell.

2.6 HP1043 codes for an essential orphan response regulator in *H. pylori*

HP1043 is an orphan response regulator of *H. pylori* that, probably, exerts its function also in the absence of phosphorylation. The term orphan is used because the protein is supposed to be part of a two components system but it lacks the kinase sensor. The structure of this transcriptional regulator was assessed some years ago through X-ray and NMR techniques and reveals that its molecular topology resembles that of the OmpR/PhoB subfamily. The N-terminal regulatory domain form a compact dimer and C-terminal transactivation domain is linked to it with a short flexible linker (Hong et al, 2007). The two alpha helixes α-7 and α-8 form a helix-turn-helix DNA binding motif and an electrostatic surface consisting of two regions with opposite charge distribution and the loop

between α-7 and α-8 is responsible for transcription activation and interaction with the σ^{80} subunit of the RNA polymerase (Hong et al, 2007). The crystal structure shows that a 2-fold symmetric dimeric interface is stabilized by the α-4/β-5/α-5 interface that resembles the phosphorylated and active form of the protein PhoB.

Despite the information on the HP1043 structure, its regulon is still unknown. In fact only three targets bound *in vitro* by the transcriptional regulator have been identified: the *hp1043* promoter itself, that can suggest an autoregulation of the protein, the *tlpB* promoter, that codes for a protein involved in chemotaxis (Delany et al, 2002), and the sRNA *cncr1* promoter, a small non-coding RNA involved in bacterial switch between adhesion mechanism on adenocarcinoma gastric cells (AGS) and bacterial motility (Vannini et al, 2017, Vannini et al, 2016). Another suggested potential target was the *porG* promoter, assessed with EMSA supershift assays (Olekhnovich et al, 2013). Moreover, the same authors hypothesized that HP1043 can have a role in response to oxidative stress, as treatment of growing cells with Metronidazole (MTZ) determined a decrease in protein levels as well as of the *hp1043* and *tlpB* transcripts, likely controlled by HP1043 (Olekhnovich et al, 2013).

It has been reported that attempts to modulate the HP1043 protein level was not successful in *H. pylori* cells with an insertion of an extra copy of the gene (Delany et al, 2002). In the past years, a few independent research groups attempted to delete *hp1043* gene after the insertion in a different locus of a second copy of the gene under the control of an inducible promoter, with no success (Bauer et al, 2013, Müller et al, 2007). The first attempt was performed by integrating a second copy of the gene under the control of the *pfr* promoter (inducible by addition of iron ions). Results from this strain submitted to iron showed an increase of *hp1043* mRNA levels with no changes in the amount of the HP1043 protein levels (Delany et al, 2002). In another attempt, an extra copy of the *hp1043* gene was inserted in the *H. pylori* genome under the control of *fecA1* promoter (repressed by addition of iron) to allow the deletion of endogenous copy of the *hp1043* gene (Schär et al, 2005). In this strain, the amount of HP1043 in the presence or absence of iron showed only slight variations, thus confirming that modulation of the amount of HP1043 in the cell is difficult to obtain.

The aim of this work is to expand and unravel the HP1043 regulon *in vivo* with Chromatin Immunoprecipitation (ChIP) assay, followed by deep sequencing. This approach allowed us to identify several new targets for HP1043. Analysis of the new binding sites reveal that the protein recognizes a specific consensus sequence on the DNA, identified by bioinformatics analysis and validated by scanning mutagenesis and binding properties assessed by DNaseI and hydroxyl-radical

footprints. Interestingly, the analysis of the HP1043 regulon revealed that a proportion of the regulated genes are involved in translational processes. Accordingly, block of translation and RNA-seq analysis confirmed that the majority of HP1043 target genes are induced.

Bibliography

- Bauer S, Endres M, Lange M, Schmidt T, Schumbrutzki C, Sickmann A, Beier D, (2013) Novel function assignment to a member of the essential HP1043 response regulator family of epsilon-proteobacteria, *Microbiology*, 159(Pt 5):880-9.

- Beier, D., Spohn, G., Rappuoli, R. and Scarlato, V. (1998), Functional analysis of the *Helicobacter pylori* principal sigma subunit of RNA polymerase reveals that the spacer region is important for efficient transcription. Molecular Microbiology, 30: 121–134.

- Borin BN, Tang W, Krezel AM, (2014) Helicobacter pylori RNA polymerase α-subunit C-terminal domain shows features unique to ε-proteobacteria and binds NikR/DNA complexes, *Protein Sci*, 23(4):454-63.

- Browning D. F, Busby S. J, (2016) Local and global regulation of transcription initiation in bacteria. *Nat Rev Microbiol*, 638-50.

- Danielli A, Amore G, Scarlato V, (2010) Built Shallow to Maintain Homeostasis and Persistent Infection: Insight into the Transcriptional Regulatory Network of the Gastric Human Pathogen *Helicobacter pylori*. *Plos. Pathogens*, 6(6):e1000938.

- Daubin V, Perriere G, (2003) G+C3 structuring along the genome: a common feature in prokaryotes. *Mol Biol Evol*, 20:471-483.

- Delany I, Spohn G, Rappuoli R, Scarlato V, (2001) The Fur repressor controls transcription of iron-activated and -repressed genes in *Helicobacter pylori*, *Mol. Microbiol*, 42(5):1297-309.

- Delany I, Spohn G, Rappuoli R & 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.

- Estrem S. T, Gaal T, Ross W, and Gourse R. L, (1998) Identification of an UP element consensus sequence for bacterial promoters. *PNAS*, 95(17): 9761–9766.

- Jones MD, Ademi I, Yin X, Gong Y, Zamble DB, (2015) Nickel-responsive regulation of two novel Helicobacter pylori NikR-targeted genes, *Metallomics*, (4):662-73.

- Jumas-Bilak E, Michaux-Charachon S, Bourg G, O'Callaghan D, Ramuz M, (1998) Differences in chromosome number and genome rearrangements in the genus *Brucella*. *Mol. Microbiol*, 27:99 –106.

- Kavenoff R, Bowen BC, (1976) Electron microscopy of membrane-free folded chromosomes from *Escherichia coli*. *Chromosoma*, 59:89 –101.

- Hong, E. et al, (2007) Structure of an atypical orphan response regulator protein supports a new phosphorylation-independent regulatory mechanism. *J. Biol. Chem*. 282, 20667–20675.

- Madan Babu M, Balaji S, Aravind L, (2007) General trends in the evolution of prokaryotic transcriptional regulatory networks. *Genome Dyn*, 3:66-80.

- Maresca B, Castellano S, Fortino V, Morello S, Porta A, (2016) Microbiologia molecolare e cellulare, *McGraw-Hill*.

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

- Olekhnovich I. N, Vitko S, Valliere M, Hoffman P, (2013) Response to Metronidazole and Oxidative Stress Is Mediated through Homeostatic Regulator HsrA (HP1043) in *Helicobacter pylori*. *J. Bacteriol,* 196(4):729-39.

- Pflock M, Kennard S, Delany I, Scarlato V, Beier D, (2005) Acid-induced activation of the urease promoters is mediated directly by the ArsRS two-component system of *Helicobacter pylori*. *Infect Immun,* 73(10):6437-45.

- Pflock M, Kennard S, Finsterer N, Beier D, (2006) Acid-responsive gene regulation in the human pathogen *Helicobacter pylori, J. Biotechnol,* 126(1):52-60.

- Rocha EP, Danchin A, (2003) Essentiality, not expressiveness, drives gene-strand bias in bacteria. *Nat Genet*, 34:377-378.

- Rocha EP, (2004) The replication-related organization of bacterial genomes. *Microbiology,* 150:1609-1627.

- Roncarati D, Danielli A, Scarlato V, (2014) The HrcA repressor is the thermosensor of the heat-shock regulatory circuit in the human pathogen Helicobacter pylori, *Mol. Microbiol*, 92(5):910-20.

- Ruff E. F, Record T. M, and Artsimovitch I, (2015) Initial Events in Bacterial Transcription Initiation. *Biomolecules,* 27;5(2):1035-62.

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

- Sharma CM, Hoffmann S, Darfeuille F, Reignier J, Findeiss S, Sittka A, Chabas S, Reiche K, Hackermüller J, Reinhardt R, Stadler PF, Vogel J, (2010) The primary transcriptome of the major human pathogen *Helicobacter pylori, Nature*, 464(7286):250-5.

- Vannini A, Roncarati D, Danielli A, (2016) The cag-pathogenicity island encoded CncR1 sRNA oppositely modulates Helicobacter pylori motility and adhesion to host cells, *Cell Mol Life Sci*, 73(16):3151-68.

- Visweswariah S. S, Busby S. J, (2015) Evolution of bacterial transcription factors: how proteins take on new tasks, but do not always stop doing the old ones. *Trends. Microbiol*, 23(8):463-7.
- Waidner B, Melchers K, Stähler FN, Kist M, Bereswill S, (2005) The *Helicobacter pylori* CrdRS two-component regulation system (HP1364/HP1365) is required for copper-mediated induction of the copper resistance determinant CrdA, *J. Bacteriol*, 187(13):4683-8.

- Yamaichi Y, Iida T, Park KS, Yamamoto K, Honda T, (1999) Physical and genetic map of the genome of *Vibrio parahaemolyticus*: presence of two chromosomes in *Vibrio* species. *Mol. Microbiol*, 31:1513–1521.

3. HP1043 regulon characterization

This study is focused on the characterization of the HP1043 orphan response regulator of *H. pylori*. For clarity of presentation it has been divided in two main parts. In the first part will be discussed the *in vivo* identification of the HP1043 transcriptional regulator target gene promoters, and of a putative consensus binding sequence. Notably, several of the identified genes are involved in translational functions. In the second part will be discussed the *in vitro* functional characterization of HP1043 binding to the targeted sequences and their mutants to exerts its regulatory action.

Most of the first part of this study has been recently published (Pelliciari et al 2017). Here is reported a brief summary of the Chomatin-ImmunoPrecipitation (ChIP) results for *H. pylori* HP1043 that allowed the identification of *in vivo* DNA-binding sites of the protein. The experiment has been performed on liquid culture of exponentially growing bacteria treated with formaldehyde to cross-link the interactions between the DNA-binding domain of the protein with bound nucleotides. Cells were then lysed and sonicated to obtain genome fragments of about 400 base pairs long.

The protein-DNA fraction was incubated with a polyclonal antibody against HP1043 and subsequently with IgG beads to recognize the constant portion of antibodies. This procedure allows separation of DNA fragment bound by HP1043, captured by the beads, from unbound genome regions. After several washes, the DNA-protein complexes were eluted from the matrix, the cross-link reverted and DNA fragments purified and subjected to nucleotide sequence determination.

ChIP assays data analysis allowed to identify genome regions in which the number of reads of immunoprecipitated sample, compared to the control experiment performed in absence of HP1043 antibody, were enriched at least 2-fold. From two biological replicates analyses it was possible to identify 37 highly reproducible peaks, with 89% of them mapping close to the 5'-end of annotated genes, positions that might be overlapping or associated to promoter regions. Analysis of the genes associated to the ChIP peaks was then carried out using the Clusters of Orthologous Groups of proteins (COGs) database. tRNA and rRNA genes were included in the translation category. Results of this analysis revealed that about one third of the peaks (37%) are linked to genes coding for proteins involved in translation and others involved in crucial cellular function such as energy production and transcription.

The validation of the binding sites identified in ChIP-seq experiments was carried out performing DNase I footprinting assays on a selection of targets among the list of 37 highly reproducible peaks. In this analysis was also included a probe mapping in a group of less reproducible peaks. Probes were radioactively labeled with ^{32}P and incubated *in vitro* with increasing amounts of recombinant

purified HP1043, then treated with DNase I to identify regions specifically bound by the protein. In all DNA probes appeared a clear region of protection from enzymatic digestion, representing the region of protein binding. These data confirmed the reliability of the highly reproducible peaks identified in ChIP-seq experiments and possibly expand the list to additional genes associated to poorly reproducible peaks *in vivo*.

Alignment of the nucleotide sequences of the promoter regions bound by HP1043 revealed that in all of them the -35 box shows an AT-rich motif region and a conserved -10 region, TATAAT. Alignments of the nucleotide sequences protected by HP1043 in DNase I footprinting assays identified an highly conserved motif constituted by two direct repeats (TTTAAG), separated by a 5 bp spacer, in which the second hemisite appears less conserved than the first one.

To further verify the correlation between this putative consensus sequence and HP1043 protein binding, hydroxyl-radical footprinting assays were performed on the same targets previously probed with DNase I footprinting. In all promoters tested, the protection pattern of HP1043 binding showed a triplet of 3/4 nucleotide spaced by two non-protected regions of 5/6 nucleotides. Notably, the central protected region overlaps the spacer between the hemisites, and other two protected regions map immediately upstream and downstream of the motifs. As hydroxyl-radical footprinting protections reflect minor accessibility of radical ions to the minor groove of the DNA, the protected nucleotides might mirror the effect of HP1043 binding to the major groove of the DNA to the hemisites that would narrow adjacent regions and become protected by hydroxyl-radical footprinting treatment.

These results validated the ChIP-seq experiments and confirmed the involvement of the conserved nucleotide motif in HP1043 binding. Subsequently, the study has been focused on the biological role of HP1043 on transcriptional regulation in *H. pylori*. As one third of the HP1043 regulon is linked to the translational process, we blocked protein synthesis by treating exponentially growing bacteria with a sub-lethal concentration of tetracycline, and analyzed transcripts levels by RNA-seq. Deep sequencing RNA revealed that almost all genes of the HP1043 regulon appeared upregulated in response to tetracycline treatment. Moreover, the percentage of genes with high transcript levels is twice (68%) the one randomly expected. RNA-seq results have been validated by qRT PCR on a selection of targets. Taken together, these data show that HP1043 is involved in the bacterial translational stress response. However, the exacts molecular mechanisms exerted by the HP1043 regulatory protein to control transcription remains to be elucidated. Specifically, whether the

transcriptional factor acts as a repressor, releasing promoters by tetracycline treatment, or as an activator, promoting transcription in response to the same stimulus.

The above summarized results have been included in the publication article included in this thesis as Annex 1 (Pelliciari et al 2017).

- Pelliciari S & Pinatel E, Vannini A, Peano C, Puccio S, De Bellis G, Danielli A, Scarlato V, 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,* 23;7:4106. (article under Creative Commons CC-BY license)

4. Molecular characterization of the HP1043 consensus sequence

Nucleotide sequence alignments of the HP1043 protected regions in DNase I footprinting experiments highlighted a putative consensus composed of two conserved hemisites (see Fig1, Panel C). The first hemisite appears to be more conserved than the second one and both share similarities to the TTTAAG sequence. The working hypothesis was that this motif is fundamental for HP1043 binding through the readout of the DNA in the major groove of the two conserved hemisites, likely on the same face of the DNA double helix.

Subsequently, the study has been focused on the further characterization of HP1043 binding to DNA and its effects on transcription. To these aims, mutants of the proposed consensus sequence were generated and assayed for HP1043 binding by footprinting assays, and transcriptional outputs. Results indicated that the conserved nucleotides of the identified consensus sequence play a crucial role in HP1043 recognition and binding. Moreover, *in vitro* transcription assays in the presence of HP1043 and template promoters harboring wild type or mutant consensus sequences indicated that the transcriptional regulator has a positive effect on transcription, and this activity correlates with HP1043 binding affinity on DNA.

Material and methods

DNA manipulations. General DNA manipulations were performed as described by Sambrook et al. (1989). All restriction and modification enzymes were used according to the manufacturers' instructions (New England Biolabs). Preparations of plasmid DNA were carried out with Nucleospin plasmid purification kit (Macherey-Nagel).

Overexpression and purification of recombinant HP1043 protein. Recombinant N-terminal His tagged HP1043 protein was overexpressed and purified from *DH5α E. coli* cells according to Pelliciari et al. (2017), using Ni-NTA resin (Sigma Aldrich). After elution from the resin, the protein was dialyzed against 200 ml of 1X 1043 FPB (10 mM Tris-Cl pH 7.5; 50 mM NaCl; 10 mM $MgCl_2$; 1 mM DTT; 0.01% Igepal CA-630, 10% glycerol) or 1X FPB_0 (10 mM Tris-Cl pH 7.5; 50 mM NaCl; 10 mM $MgCl_2$; 1 mM DTT; 0.01% Igepal CA-630) for 20 h, aliquoted and stored at -80° until utilization.

hp1227 **promoter mutants generation**. Overlapping oligonucleotides listed in Table 2 were used to generate P_{hp1227} promoter point mutants. About 250 pmol of each of them were phosphorylated with T4 polynucleotide kinase (New England Biolabs) in presence of 1mM ATP at 37° for one hour. After kinase inactivation at 65° for 20', 100 pmol of each oligonucleotide was annealed with its inverted repeat in 100 mM NaCl, heating the mixture for 5' at 100°C, then slowly cooled down to room temperature. Each annealed reaction was clone into the pBlueScript KSII (+) plasmid (Agilent) previously digested with EcoRV and dephosphorylated with Calf Intestinal Phosphatase (New England Biolabs). To generate plasmids with a luciferase gene controlled by *hp1227* promoter mutants (pLux-p1227, enlisted in Table 1), each of them was amplified from pBluescript plasmids with oligonucleotides 1227E FW and 1227E RV. PCR products were gel-purified and digested with EcoRI, then cloned in a pLux plasmid previously digested with the same enzyme and dephosphoylated.

DNase I footprinting. DNA fragments, obtained by digestion of pBlusescript P_{hp1227} mutant promoters plasmids listed in Table 1 with the appropriate restriction enzymes, were 5'-end labelled with [γ-^{32}P]-ATP and T4 polynucleotide kinase and gel purified. About 15 fmol of probe were used for each enzymatic cut reaction in the presence of purified HP1043 protein, which was performed in

1X 1043 FPB containing 300 ng of sonicated salmon sperm as non-specific competitor DNA, according to Pelliciari et al. (2017). DNA obtained were phenol-chloroform extracted, ethanol precipitated then resuspended in 10 µl of Formamide Loading Buffer (FLB: 95% formamide; 10 mM EDTA; 0.02% bromophenol blue; 0.02% xylene cyanol), denatured at 100°C for 3 min, and separated on a 8 M urea-6% polyacrylamide sequencing gel for the next auto-radiography.

Hydroxyl-radical footprinting. The probes were obtained as described above for DNase I footprinting. Hydroxyl-radical footprinting assays were performed as previously described (Pelliciari et al 2017). Briefly, the labeled probes were incubated with increasing amounts of purified HP1043 in 1X 1043 FPB_0, in a final volume of 30 µl for 20 min at room temperature. Digestion with radical ions was performed by the simultaneous addition of 2 µl each of the following solutions: 2 µl of Fe:EDTA solution (125 mM Fe $(NH_4)_2$ $(SO_4)_2$; 250 mM EDTA pH 8.0), 2 µl of 0.1 M DTT and 2 µl 1% H_2O_2. After a 2-min incubation, cutting reaction was stopped with 25 µl of STOP solution (600 mM NaOAc 3M pH 5.2; 100 ng/µl sonicated salmon sperm DNA; 4% glycerol), phenol-chloroform extracted and ethanol precipitated. Samples were resuspended in 6 µl of FLB, denatured at 100°C for 3 min, separated on 8 M urea-8.4% polyacrylamide sequencing gel and auto-radiographed.

In vitro **transcription.** About 6nM of each pLux plasmid were premixed in 1X *E. coli* RNA Polymerase Reaction Buffer (New England Biolabs) in a volume of 14 µl; 2 µl of 0,1 U/µl *E. coli* RNA polymerase (diluted in 1X Storage Buffer, New England Biolabs) were added and the mixes were incubated for 5' at 37°C. After that, different concentrations of purified HP1043 diluted in 1X 1043 FPB were added and incubated at 37° for 10'. To start transcription 2 µl of 5 mM NTPs, were added at 37°C. After 10' incubation, the reactions were stopped with 20 µl of 1M NaCl, phenol-chloroform extracted and ethanol precipitation.

Nucleic acids were resuspended in 12 µl of mqH_2O, and 6 µl of were treated with 10 U of DNase I (Merck Millipore) to digest plasmid templates, in 1X DNase Buffer (Merck Millipore) at 37° for one hour, Enzyme was inactivated at 65°C for 10' and the RNA was retrotranscribed as previously described (Pelliciari et al. 2017). As internal control, in cDNA generation reactions were added 1 ng of *in vitro* transcribed RNA of *cncr1*, performed with MEGAscript T7 Transcription kit (Thermo Fisher Scientific) following manufacturer's instructions.

qRT-PCR analysis. For qRT-PCR analyses, 2 µL of the diluted (1:10) cDNA samples were mixed with 5 µL of 2X Power Up SYBR Green master Mix (ThermoFisher Scientific) and oligonucleotides specific for the reporter genes Lux (LuxRT FW/LuxRT RV, Table 2) at 400 nM concentration in a final volume of 10 µl. Real time PCR was performed using the following cycling protocol: 95°C for 2 min, then 40 cycles consisting of a denaturation for 5s at 95°C followed by 30s at 60°C (annealing and extension steps). For each real time experiment, the specificity of the reaction was checked by including a melting profile at the end of the run. Data were analysed using the ΔΔCt method. The levels of cDNAs of interest were normalized against the measured level of the cDNA of the *cncr1* gene.

Table 1. Bacterial strains and plasmids

Bacterial strains /plasmids	Description	Source/Reference
Strain		
E. coli DH-5α	*supE44 ΔlacU169 (ϕ80 lacZΔM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1*	Hanahan, 1983
Plasmid		
pBlueScript KSII (+)	Cloning vector for PCR products; Amp^r.	Agilent
PBSK-*p1227* WT	pBlueScript KSII (+) derivative, containing a 61 bp probe corresponding to the region from 1.290.664 to 1.290.725 of *H. pylori* G27 genome obtained by oligos p1227WT FW/ p1227WT RV annealing. This region corresponds to the promoter of *HPG27_RS06145* gene (*hp1227* according to 26695 annotation). This probe was used for DNase I and hydroxyl-radical footprintings.	This work
PBSK-*p1227* T1F	pBlueScript KSII (+) derivative, containing a 61 bp probe corresponding to the mutated region from 1.290.664 to 1.290.725 of *H. pylori* G27 genome obtained by oligos p1227T1F FW/ p1227T1F RV annealing. This region corresponds to the mutated promoter of *HPG27_RS06145* gene (*hp1227* according to 26695 annotation). This probe was used for DNase I and hydroxyl-radical footprintings.	This work
PBSK-*p1227* T2F	pBlueScript KSII (+) derivative, containing a 61 bp probe corresponding to the mutated region from 1.290.664 to 1.290.725 of *H. pylori* G27 genome obtained by oligos p1227T2F FW/ p1227T2F RV annealing. This region corresponds to the mutated promoter of *HPG27_RS06145* gene (*hp1227* according to 26695 annotation). This probe was used for DNase I and hydroxyl-radical footprintings.	This work

PBSK-*p1227* T3F	pBlueScript KSII (+) derivative, containing a 61 bp probe corresponding to the mutated region from 1.290.664 to 1.290.725 of *H. pylori* G27 genome obtained by oligos p1227T3F FW/ p1227T3F RV annealing. This region corresponds to the mutated promoter of *HPG27_RS06145* gene (*hp1227* according to 26695 annotation). This probe was used for DNase I and hydroxyl-radical footprintings.	This work
PBSK-*p1227* A1F	pBlueScript KSII (+) derivative, containing a 61 bp probe corresponding to the mutated region from 1.290.664 to 1.290.725 of *H. pylori* G27 genome obtained by oligos p1227A1F FW/ p1227A1F RV annealing. This region corresponds to the mutated promoter of *HPG27_RS06145* gene (*hp1227* according to 26695 annotation). This probe was used for DNase I and hydroxyl-radical footprintings.	This work
PBSK-*p1227* A2F	pBlueScript KSII (+) derivative, containing a 61 bp probe corresponding to the mutated region from 1.290.664 to 1.290.725 of *H. pylori* G27 genome obtained by oligos p1227A2F FW/ p1227A2F RV annealing. This region corresponds to the mutated promoter of *HPG27_RS06145* gene (*hp1227* according to 26695 annotation). This probe was used for DNase I and hydroxyl-radical footprintings.	This work
PBSK-*p1227* GF	pBlueScript KSII (+) derivative, containing a 61 bp probe corresponding to the mutated region from 1.290.664 to 1.290.725 of *H. pylori* G27 genome obtained by oligos p1227GF FW/ p1227GF RV annealing. This region corresponds to the mutated promoter of *HPG27_RS06145* gene (*hp1227* according to 26695 annotation). This probe was used for DNase I and hydroxyl-radical footprintings.	This work
PBSK-*p1227* Δ1	pBlueScript KSII (+) derivative, containing a 61 bp probe corresponding to the mutated region from 1.290.664 to 1.290.725 of *H. pylori* G27 genome obtained by oligos p1227Δ1F FW/ p1227Δ1F RV annealing. This region corresponds to the mutated promoter of *HPG27_RS06145* gene (*hp1227* according to 26695 annotation). This probe was used for DNase I and hydroxyl-radical footprintings.	This work
PBSK-*p1227* T1S	pBlueScript KSII (+) derivative, containing a 61 bp probe corresponding to the mutated region from 1.290.664 to 1.290.725 of *H. pylori* G27 genome obtained by oligos p1227T1S FW/ p1227T1S RV annealing. This region corresponds to the mutated promoter of *HPG27_RS06145* gene (*hp1227* according to 26695 annotation). This probe was used for DNase I and hydroxyl-radical footprintings.	This work

PBSK-*p1227* T2S	pBlueScript KSII (+) derivative, containing a 61 bp probe corresponding to the mutated region from 1.290.664 to 1.290.725 of *H. pylori* G27 genome obtained by oligos p1227T2S FW/ p1227T2S RV annealing. This region corresponds to the mutated promoter of *HPG27_RS06145* gene (*hp1227* according to 26695 annotation). This probe was used for DNase I and hydroxyl-radical footprintings.	This work
PBSK-*p1227* T3S	pBlueScript KSII (+) derivative, containing a 61 bp probe corresponding to the mutated region from 1.290.664 to 1.290.725 of *H. pylori* G27 genome obtained by oligos p1227T3S FW/ p1227T3S RV annealing. This region corresponds to the mutated promoter of *HPG27_RS06145* gene (*hp1227* according to 26695 annotation). This probe was used for DNase I and hydroxyl-radical footprintings.	This work
PBSK-*p1227* A1S	pBlueScript KSII (+) derivative, containing a 61 bp probe corresponding to the mutated region from 1.290.664 to 1.290.725 of *H. pylori* G27 genome obtained by oligos p1227A1S FW/ p1227A1S RV annealing. This region corresponds to the mutated promoter of *HPG27_RS06145* gene (*hp1227* according to 26695 annotation). This probe was used for DNase I and hydroxyl-radical footprintings.	This work
PBSK-*p1227* A2S	pBlueScript KSII (+) derivative, containing a 61 bp probe corresponding to the mutated region from 1.290.664 to 1.290.725 of *H. pylori* G27 genome obtained by oligos p1227A2S FW/ p1227A2S RV annealing. This region corresponds to the mutated promoter of *HPG27_RS06145* gene (*hp1227* according to 26695 annotation). This probe was used for DNase I and hydroxyl-radical footprintings.	This work
PBSK-*p1227* GS	pBlueScript KSII (+) derivative, containing a 61 bp probe corresponding to the mutated region from 1.290.664 to 1.290.725 of *H. pylori* G27 genome obtained by oligos p1227GS FW/ p1227GS RV annealing. This region corresponds to the mutated promoter of *HPG27_RS06145* gene (*hp1227* according to 26695 annotation). This probe was used for DNase I and hydroxyl-radical footprintings.	This work
PBSK-*p1227* Δ2	pBlueScript KSII (+) derivative, containing a 61 bp probe corresponding to the mutated region from 1.290.664 to 1.290.725 of *H. pylori* G27 genome obtained by oligos p1227Δ2 FW/ p1227Δ2 RV annealing. This region corresponds to the mutated promoter of *HPG27_RS06145* gene (*hp1227* according to 26695 annotation). This probe was used for DNase I and hydroxyl-radical footprintings.	This work

pSB1075	Plasmid vector containing the 5.8-kb *Photorhabdus luminescens luxCDABE* operon cassette; Ap^r	Winson MK, et al. 1998.
pLux	PSB1075 derivative in which promoter *lasRI'* was deleted to generate a promoterless *lux* operon; Ap^r	This work
pLux-*p1227* WT	pLux derivative containing a DNA fragment of 61 bp amplified by PCR with oligos p1227Eco FW/p1227Eco RV from PBSK-*p1227* WT upstream *luxCDABE* operon.	This work
pLux-*p1227* T1F	pLux derivative containing a DNA fragment of 61 bp amplified by PCR with oligos p1227Eco FW/p1227Eco RV from PBSK-*p1227* T1F upstream *luxCDABE* operon.	This work
pLux-*p1227* T2F	pLux derivative containing a DNA fragment of 61 bp amplified by PCR with oligos p1227Eco FW/p1227Eco RV from PBSK-*p1227* T2F upstream *luxCDABE* operon.	This work
pLux-*p1227* T3F	pLux derivative containing a DNA fragment of 61 bp amplified by PCR with oligos p1227Eco FW/p1227Eco RV from PBSK-*p1227* T3F upstream *luxCDABE* operon.	This work
pLux-*p1227* A1F	pLux derivative containing a DNA fragment of 61 bp amplified by PCR with oligos p1227Eco FW/p1227Eco RV from PBSK-*p1227* A1F upstream *luxCDABE* operon.	This work
pLux-*p1227* A2F	pLux derivative containing a DNA fragment of 61 bp amplified by PCR with oligos p1227Eco FW/p1227Eco RV from PBSK-*p1227* A2F upstream *luxCDABE* operon.	This work
pLux-*p1227* GF	pLux derivative containing a DNA fragment of 61 bp amplified by PCR with oligos p1227Eco FW/p1227Eco RV from PBSK-*p1227* GF upstream *luxCDABE* operon.	This work
pLux-*p1227* Δ1	pLux derivative containing a DNA fragment of 61 bp amplified by PCR with oligos p1227Eco FW/p1227Eco RV from PBSK-*p1227* Δ1 upstream *luxCDABE* operon.	This work
pLux-*p1227* T1S	pLux derivative containing a DNA fragment of 61 bp amplified by PCR with oligos p1227Eco FW/p1227Eco RV from PBSK-*p1227* T1S upstream *luxCDABE* operon.	This work
pLux-*p1227* T2S	pLux derivative containing a DNA fragment of 61 bp amplified by PCR with oligos p1227Eco FW/p1227Eco RV from PBSK-*p1227* T2S upstream *luxCDABE* operon.	This work

pLux-*p1227* T3S	pLux derivative containing a DNA fragment of 61 bp amplified by PCR with oligos p1227Eco FW/p1227Eco RV from PBSK-*p1227* T3S upstream *luxCDABE* operon.	This work
pLux-*p1227* A1S	pLux derivative containing a DNA fragment of 61 bp amplified by PCR with oligos p1227Eco FW/p1227Eco RV from PBSK-*p1227* A1S upstream *luxCDABE* operon.	This work
pLux-*p1227* A2S	pLux derivative containing a DNA fragment of 61 bp amplified by PCR with oligos p1227Eco FW/p1227Eco RV from PBSK-*p1227* A2S upstream *luxCDABE* operon.	This work
pLux-*p1227* GS	pLux derivative containing a DNA fragment of 61 bp amplified by PCR with oligos p1227Eco FW/p1227Eco RV from PBSK-*p1227* GS upstream *luxCDABE* operon.	This work
pLux-*p1227* Δ2	pLux derivative containing a DNA fragment of 61 bp amplified by PCR with oligos p1227Eco FW/p1227Eco RV from PBSK-*p1227* Δ2 upstream *luxCDABE* operon.	This work

Table 2. Oligos

Oligonucleotides name	Sequence (5' to 3')
p1227WT FW	TTCTCTTATTTTCACTTCATTTTTTAAGCAAAACTTAAACTTGTAATTGTATCATTTTAAG
p1227WT RV	CTTAAAATGATACAATTACAAGTTTAAGTTTTGCTTAAAAAATGAAGTGAAAATAAGAGAA
p1227T1F FW	TTCTCTTATTTTCACTTCATTTATTAAGCAAAACTTAAACTTGTAATTGTATCATTTTAAG
p1227T1F RV	CTTAAAATGATACAATTACAAGTTTAAGTTTTGCTTAATAAATGAAGTGAAAATAAGAGAA
p1227T2F FW	TTCTCTTATTTTCACTTCATTTTATAAGCAAAACTTAAACTTGTAATTGTATCATTTTAAG
p1227T2F RV	CTTAAAATGATACAATTACAAGTTTAAGTTTTGCTTATAAAATGAAGTGAAAATAAGAGAA
p1227T3F FW	TTCTCTTATTTTCACTTCATTTTTAAAGCAAAACTTAAACTTGTAATTGTATCATTTTAAG
p1227T3F RV	CTTAAAATGATACAATTACAAGTTTAAGTTTTGCTTTAAAAATGAAGTGAAAATAAGAGAA
p1227A1F FW	TTCTCTTATTTTCACTTCATTTTTTTAGCAAAACTTAAACTTGTAATTGTATCATTTTAAG
p1227A1F RV	CTTAAAATGATACAATTACAAGTTTAAGTTTTGCTAAAAAAATGAAGTGAAAATAAGAGAA

p1227A2F FW	TTCTCTTATTTTCACTTCATTTTTTATGCAAAACTTAAACTTGTAATTGTATCATTTTAAG
p1227A2F RV	CTTAAAATGATACAATTACAAGTTTAAGTTTTGCATAAAAATGAAGTGAAAATAAGAGAA
p1227GF FW	TTCTCTTATTTTCACTTCATTTTTTAACCAAAACTTAAACTTGTAATTGTATCATTTTAAG
p1227GF RV	CTTAAAATGATACAATTACAAGTTTAAGTTTTGGTTAAAAAATGAAGTGAAAATAAGAGAA
p1227Δ1 FW	TTCTCTTATTTTCACTTCATTTGGATCCCAAAACTTAAACTTGTAATTGTATCATTTTAAG
p1227Δ1 RV	CTTAAAATGATACAATTACAAGTTTAAGTTTTGGGATCCAAATGAAGTGAAAATAAGAGAA
p1227T1S FW	TTCTCTTATTTTCACTTCATTTTTTAAGCAAAAGTTAAACTTGTAATTGTATCATTTTAAG
p1227T1S RV	CTTAAAATGATACAATTACAAGTTTAACTTTTGCTTAAAAAATGAAGTGAAAATAAGAGAA
p1227T2S FW	TTCTCTTATTTTCACTTCATTTTTTAAGCAAAACATAAACTTGTAATTGTATCATTTTAAG
p1227T2S RV	CTTAAAATGATACAATTACAAGTTTATGTTTTGCTTAAAAAATGAAGTGAAAATAAGAGAA
p1227T3S FW	TTCTCTTATTTTCACTTCATTTTTTAAGCAAAACTAAAACTTGTAATTGTATCATTTTAAG
p1227T3S RV	CTTAAAATGATACAATTACAAGTTTTAGTTTTGCTTAAAAAATGAAGTGAAAATAAGAGAA
p1227A1S FW	TTCTCTTATTTTCACTTCATTTTTTAAGCAAAACTTTAACTTGTAATTGTATCATTTTAAG
p1227A1S RV	CTTAAAATGATACAATTACAAGTTAAAGTTTTGCTTAAAAAATGAAGTGAAAATAAGAGAA
p1227A2S FW	TTCTCTTATTTTCACTTCATTTTTTAAGCAAAACTTATACTTGTAATTGTATCATTTTAAG
p1227A2S RV	CTTAAAATGATACAATTACAAGTATAAGTTTTGCTTAAAAAATGAAGTGAAAATAAGAGAA
p1227GS FW	TTCTCTTATTTTCACTTCATTTTTTAAGCAAAACTTAATCTTGTAATTGTATCATTTTAAG
p1227GS RV	CTTAAAATGATACAATTACAAGATTAAGTTTTGCTTAAAAAATGAAGTGAAAATAAGAGAA
p1227Δ2 FW	TTCTCTTATTTTCACTTCATTTTTTAAGCAAAAGGATCCCTTGTAATTGTATCATTTTAAG
p1227Δ2 RV	CTTAAAATGATACAATTACAAGGGATCCTTTTGCTTAAAAAATGAAGTGAAAATAAGAGAA
LuxRT FW	TTGGCAGATGTGTGTACCTTC
LuxRT RV	TGATGACTCCCAAGGAAAAATAG

536RT FW	AGGGCGGTGAACTTACAAAG
536RT RV	GGGGTATTCTTTGAGATTTGGAG
1227E FW	TTCTCTTATTTTCACTTCATTT
1227E RV	ATATGAATTCCTTAAAATGATACAATTACAAG

Results

Molecular validation of the first hemisite of the HP1043 consensus sequence. The putative consensus sequence identified by bioinformatic analysis, aligning HP1043 protected region from DNase I footprinting, is composed by two repeated hexamers with TTTAAG sequences separated by a spacer of 5 bp (Pelliciari et al, 2017). The two regions have the same nucleotide composition with different degree of conservation, since the first hemisite is more conserved than the second one. Our interest was to investigate the importance of the consensus on the HP1043 binding mechanism, and the relative involvement of each base. To fulfil this objective, we selected and analysed one HP1043 target, the *hp1227* promoter. This choice was based on the apparent high affinity of HP1043 binding shown in ChIP-seq and in DNase I footprinting experiments, and on the high degree of conservation of the HP1043 consensus sequence, TTTAAG – 5bp – CTTAAA.

The different degree of nucleotides conservation between the two hexamers suggested that the first repeated sequence might play a primary role in HP1043 binding. Accordingly, scanning mutagenesis of this site was carried out and mutants assayed in protein binding experiments. We substituted each nucleotide with the matching pair (A→T; C→G) to not alter the GC content of the region. We also substituted the TTTAAG sequence with a non-related one, to verify its involvement in HP1043 binding to the region. A summary of the generated mutants along with the results of the assays are reported in Fig. 1. In panel A the DNase I footprintings performed on each mutant are shown. Addition of HP1043 protein to the wild type probe clearly shows a region of protection by DNase I digestion that represents the binding of the protein and is indicated with a black bar to the right side of the image. Under the same experimental condition, the use of T1F mutant probe shows a similar area of extended protection starting with an higher amount of the HP1043 protein (compare lanes 2 and 3 of footprints T1F and WT, respectively). Using the mutant DNA probes T2F and T3F, protection of HP1043 appears at higher protein concentration (compare lanes 2 and 3 of footprints T2F and T3F with same lanes of WT). By contrast, HP1043 on the A1F and A2F mutant probes show binding restricted to the region spanning over the second hemisite only and is clearly detectable at higher protein concentration (Fig. 1, lanes 3 and 4, protection marked with grey boxes on the right side of the footprints). A similar footprint was obtained with the mutant probe GF, carrying the mutation on the last nucleotide of the consensus. Notably, the protection is clearly detected at lower HP1043 concentration (compare lane 3 of footprint GF with lanes 4 of footprints
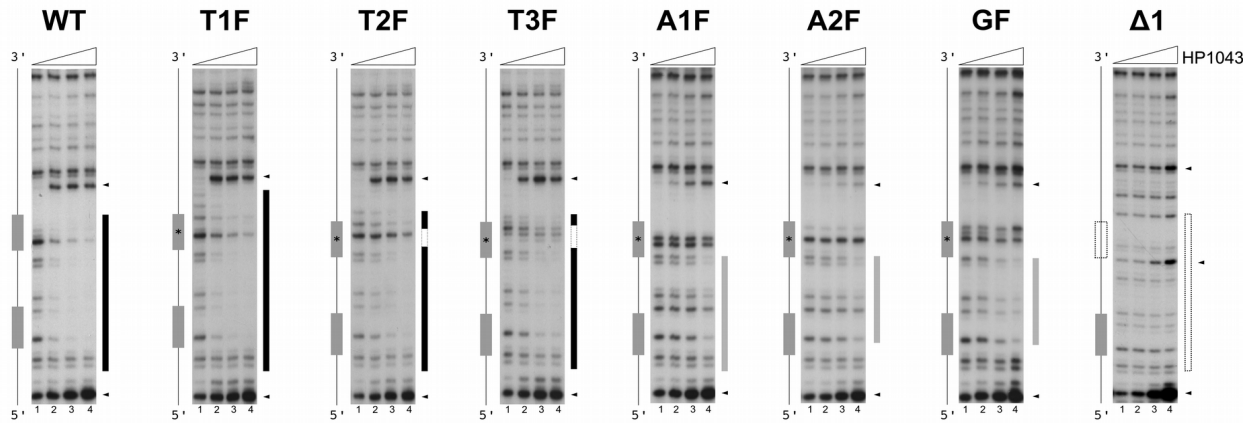
A1F and A2F). Finally, the probe in which the hemisite was substitute with a non-related nucleotide sequence, Δ1, show loss of protection. Noteworthy, the DNase I hypersensitive site mapping immediately upstream of the hemisite in the WT probe weakens on the A1F, A2F, and GF mutant probes. This observation is in agreement with the apparent weakened affinity of HP1043 for the latter mutant probes. Several DNase I hypersensitive sites are also detected at high HP1043 protein concentrations on the Δ1 probe probably representing non-specific contacts of HP1043 to this mutant probe.

To further study the binding of HP1043 to its targets, hydroxyl-radical footprinting assays were then performed. The experiments reported in Fig.1, panel B, show the hydroxyl-radical experiments on the same probes used for the DNase I footprints in panel A. The gel to the left side of the panel reports the experiment performed on the wild type sequence showing the appearance of three areas of protection, indicated with black boxes on the right side of the gel. These protections map on the spacer between the hemisites and in the regions immediately upstream and downstream of the consensus. As the protected region in hydroxyl-radical footprints represent a less accessibility of hydroxyl-ions attack to the DNA minor groove (Pelliciari et al. 2017), protections are likely arising from HP1043 binding to the major groove of the hemisites that narrow the minor groove in the adjacent regions, thus protecting them from radical ions cuts.
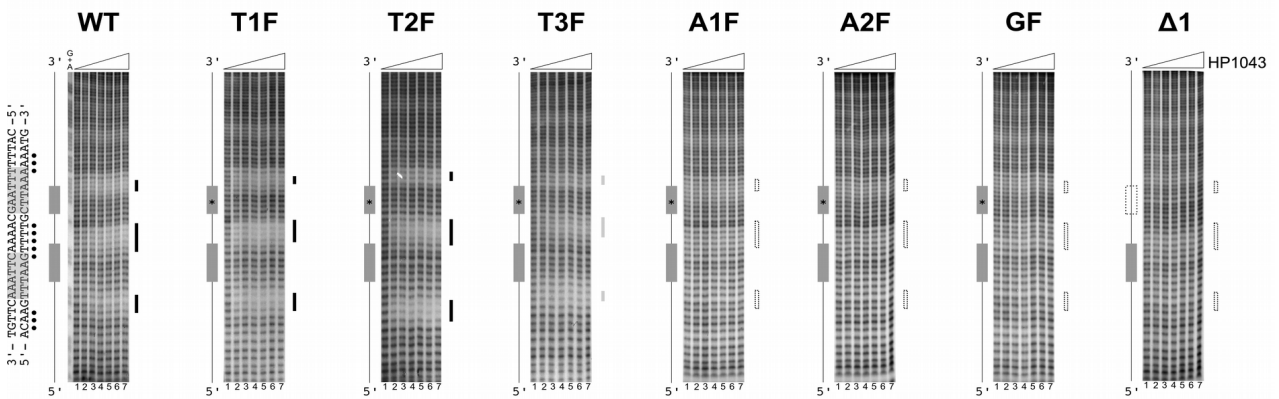
In the T1F and T2F mutants, protections appear at higher protein concentration than on the WT probe (compare lanes 5 and 4 of mutants and WT respectively Fig. 1B), likely representing a relative lower affinity of HP1043 for the mutants. Moreover, the protected region in T1F corresponding to the second hemisite of the consensus, appears at a lower HP1043 concentration than in the WT probe, suggesting an increased affinity of the protein for the hexamer closer to the transcription initiation start site. T2F exhibit the same behaviour of T1F, with an apparent lower affinity of HP1043 binding to the upper hemisite, which harbours the mutation. The T3F mutant shows a protection pattern similar to T2F, with a general lower affinity. The A1F, A2F, GF and Δ1 mutants show no protection. Barely area of protections or no protections were also observed in the DNase I footprints of panel.

From the above described results, it can be concluded that the first two nucleotides of the consensus of the first hemisite have minor influence on HP1043 binding, while the other four nucleotides of the hemisite ttTAAG (here indicated in capital letters) play a crucial role in protein binding and, possibly, in affinity of the protein for the target sites.
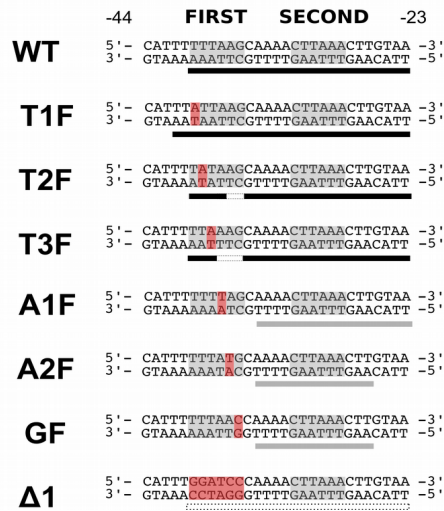
A

WT T1F T2F T3F A1F A2F GF Δ1

B

WT T1F T2F T3F A1F A2F GF Δ1

C

```
                    -44     FIRST   SECOND   -23

WT     5'- CATTTTTTAAGCAAAACTTAAACTTGTAA -3'
       3'- GTAAAAAATTCGTTTTGAATTTGAACATT -5'

T1F    5'- CATTTATTAAGCAAAACTTAAACTTGTAA -3'
       3'- GTAAATAATTCGTTTTGAATTTGAACATT -5'

T2F    5'- CATTTTATAAGCAAAACTTAAACTTGTAA -3'
       3'- GTAAAATATTCGTTTTGAATTTGAACATT -5'

T3F    5'- CATTTTTAAAGCAAAACTTAAACTTGTAA -3'
       3'- GTAAAAATTTCGTTTTGAATTTGAACATT -5'

A1F    5'- CATTTTTTTAGCAAAACTTAAACTTGTAA -3'
       3'- GTAAAAAATCGTTTTGAATTTGAACATT -5'

A2F    5'- CATTTTTTATGCAAAACTTAAACTTGTAA -3'
       3'- GTAAAAAATACGTTTTGAATTTGAACATT -5'

GF     5'- CATTTTTTAAGCCAAAACTTAAACTTGTAA -3'
       3'- GTAAAAAATTCGGTTTTGAATTTGAACATT -5'

Δ1     5'- CATTTGGATCCCAAAACTTAAACTTGTAA -3'
       3'- GTAAACCTAGGGTTTTGAATTTGAACATT -5'
```
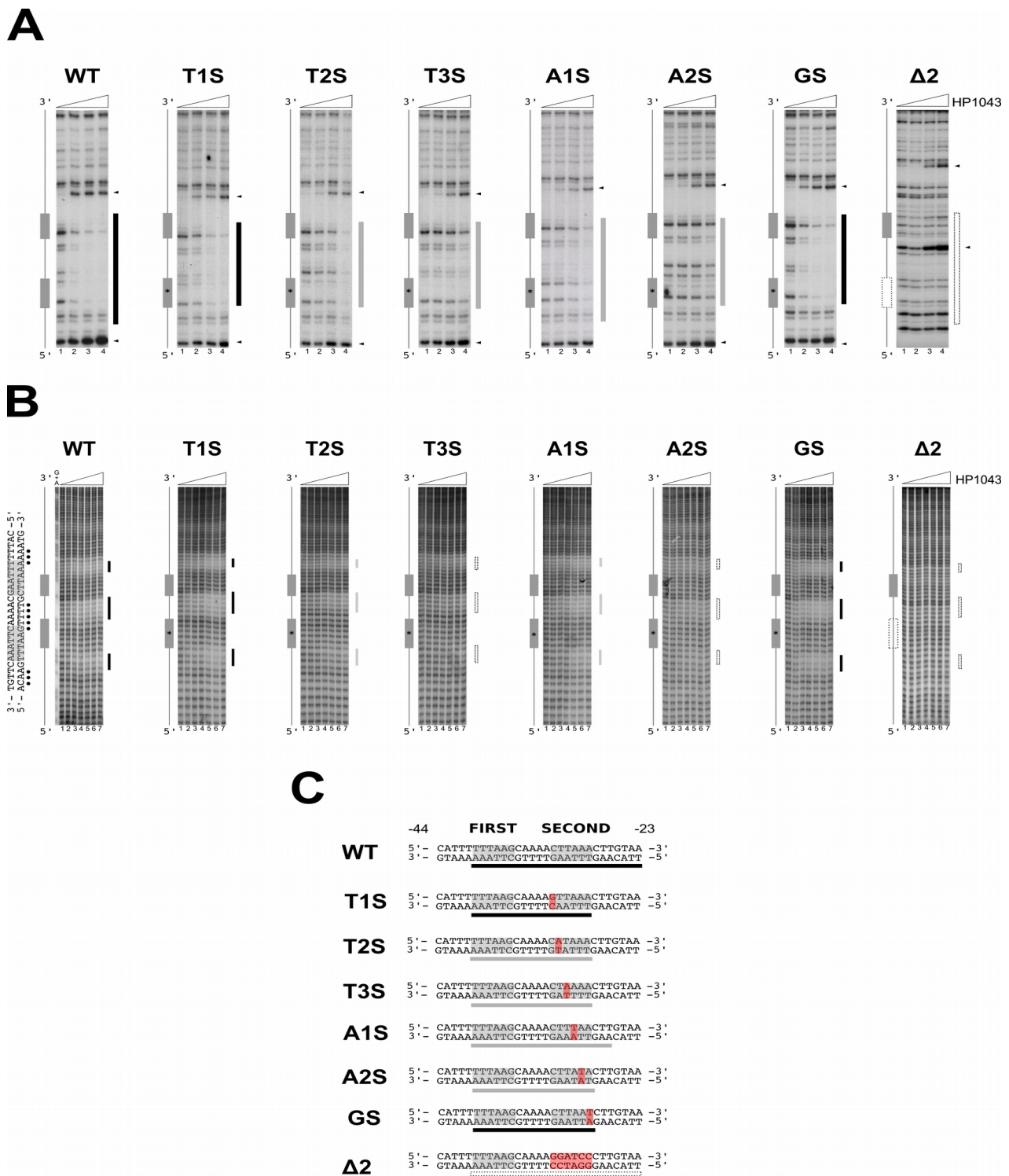
Fig.1 HP1043 protein-DNA interaction with WT and P_{hp1227} promoter mutant probes . DNase I footprinting experiments (Panel A) and hydroxyl-radical footprints (Panel B) were performed with increasing amounts of protein as depicted by the triangle on top of each gel. Lanes 1-4 of DNase I experiments contain 0, 1,7, 6,6 and 13,3 μM of dimeric HP1043 protein, respectively. Lanes 1-7 of hydroxyl-radical footprints contain 0, 0,17, 0,5, 1, 1,25, 1,7 and 2,5 μM of dimeric HP1043 protein, respectively. On the left of each experiment is reported a schematic representation of the probe, with grey boxes marking the positions of the hemisites of the HP1043 consensus sequence and an asterisk to indicate the mutated region. In the hydroxyl-radical experiments is reported also the nucleotide sequence in which black dots indicate the protected nucleotides. On the right of each gel ,black bars indicate protected regions, grey bar the weaker protected regions and empty bars the expected positions of binding. (Panel C) reports protected region in DNase I footprints; with red colour indicating mutations, grey boxes the hemisites, and bars under the sequences the strength and the extension of protein binding.

The second hemisite of the HP1043 consensus sequence is fundamental for protein binding.

Scanning mutagenesis of the first hemisite of the consensus sequence revealed that this element is important for DNA recognition by HP1043. The same approach was applied to the second hexamer of the consensus with results shown in Fig. 2.

In panel A are reported the DNase I footprintings carried out on mutants of the second hemisite, in panel B the hydroxyl-radical footprintings on the same DNA probes, and in panel C the wild type and mutant nucleotide sequences. Two mutants, T1S and GS, besides a slight difference in the appearance and intensity of a DNase I hypersensitive band, show a pattern of digestion similar to the wild type probe (WT). On the contrary, the other four mutants show a different behaviour, with an impairment of HP1043 binding that generated protected regions only at the higher protein concentration, for T2S and A1S, or no protection for T3S and A2S. Mutant Δ2, in which all the second hexamer was substitute with a different sequence, shows no protein binding. These evidences demonstrated that also the second hemisite, despite its lower bases conservation with respect to the first hemisite, is crucial for HP1043 binding. To better characterize and refine the protein binding analysis on mutant probes, hydroxyl-radical footprinting experiments were performed (Fig. 2, Panel B). Also in these assays, mutants T1S and GS show no relevant differences when compared to WT probes, while opposite was shown for the Δ2 mutant that exhibits no protected region, confirming no HP1043 binding when also the second hemisite is not present. The other four mutants showed loss of binding affinity; in fact, on mutants probes T2S and A1S the three areas of protection appear in the presence of high amounts of HP1043 protein in the reaction and mutants T3S and A2S show no area of protection.

Therefore, from these data collected, it appears clear that the most important nucleotides of the second hemisite for HP1043 binding are tTTAAg (indicated in capital letters), whereas the first and the last nucleotides of the hexamer seem to have no influence in specificity of protein /DNA interactions.

Fig. 2 HP1043 protein-DNA interaction with WT and P_{hp1227} promoter mutant probes . DNase I footprinting experiments (Panel A) and hydroxyl-radical footprints (Panel B) were performed with increasing amounts of protein as depicted by the triangle on top of each gel. Lanes 1-4 of DNase I experiments contain 0, 1,7, 6,6 and 13,3 μM of dimeric HP1043 protein, respectively. Lanes 1-7 of hydroxyl-radical footprints contain 0, 0,17, 0,5, 1, 1,25, 1,7 and 2,5 μM of dimeric HP1043 protein, respectively. On the left of each experiment is reported a schematic representation of the probe. Protection and hypersensitive bands are indicated as in Fig. 1.

Semi quantitative analysis of HP1043 binding to P_{hp1227} promoter mutants. To grasp information on the relative affinity of HP1043 for the mutants probes, the footprints of Fig. 1A and 2A were submitted to the FIJI software (Schindelin J. et al, 2012). This software performs a densitometric analysis and generates a peaks graph for a submitted gel lane, where the height of the spikes is proportional to the intensity of the corresponding band signal. The third lanes of each DNase I footprint experiment were selected for the analysis and mutants outputs were compared to the WT signals by overlying the obtained curves with results reported in Fig. 3. In Panels A and B are reported the results obtained for the first and second hexamer, respectively. Clearly, the region of protection shows probe-specific variations in the height of the peaks (coloured lines), and likely reflects changes in the affinity of HP1043 binding.

Quantification of each gel line under analysis was obtained by calculating the integral area subtended under the protected region of each curve (the region between dotted lines in panel A and B), normalized with respect to the first lane and then compared with that of WT. The obtained results are reported in Fig. 3, Panel C, in which the area calculated for the WT protected area was set to the maximum of protection (100%). Similar values of protection were obtained for the T1S and GS mutant probes. By contrast, mutants A1F, A2F, T3S and A2S show the most reduced values (less than 25% with respect to the WT value). Slight differences were detected in mutants T2S/A1S and T3S/A2SMutants T1F and T2F, shows approximatively a loss of 50% affinity and T3F more than 60%.

In conclusion, the above analysis highlights the contribution of each nucleotide of the HP1043 consensus sequence to the relative affinity of the protein for the binding site. Basically, the relevant nucleotides include the central part of the consensus, with major effect on the AAG out of TTTAAG for the first sequence and TTAA hexamer core for the second site. This finding well correlates with the footprints results.
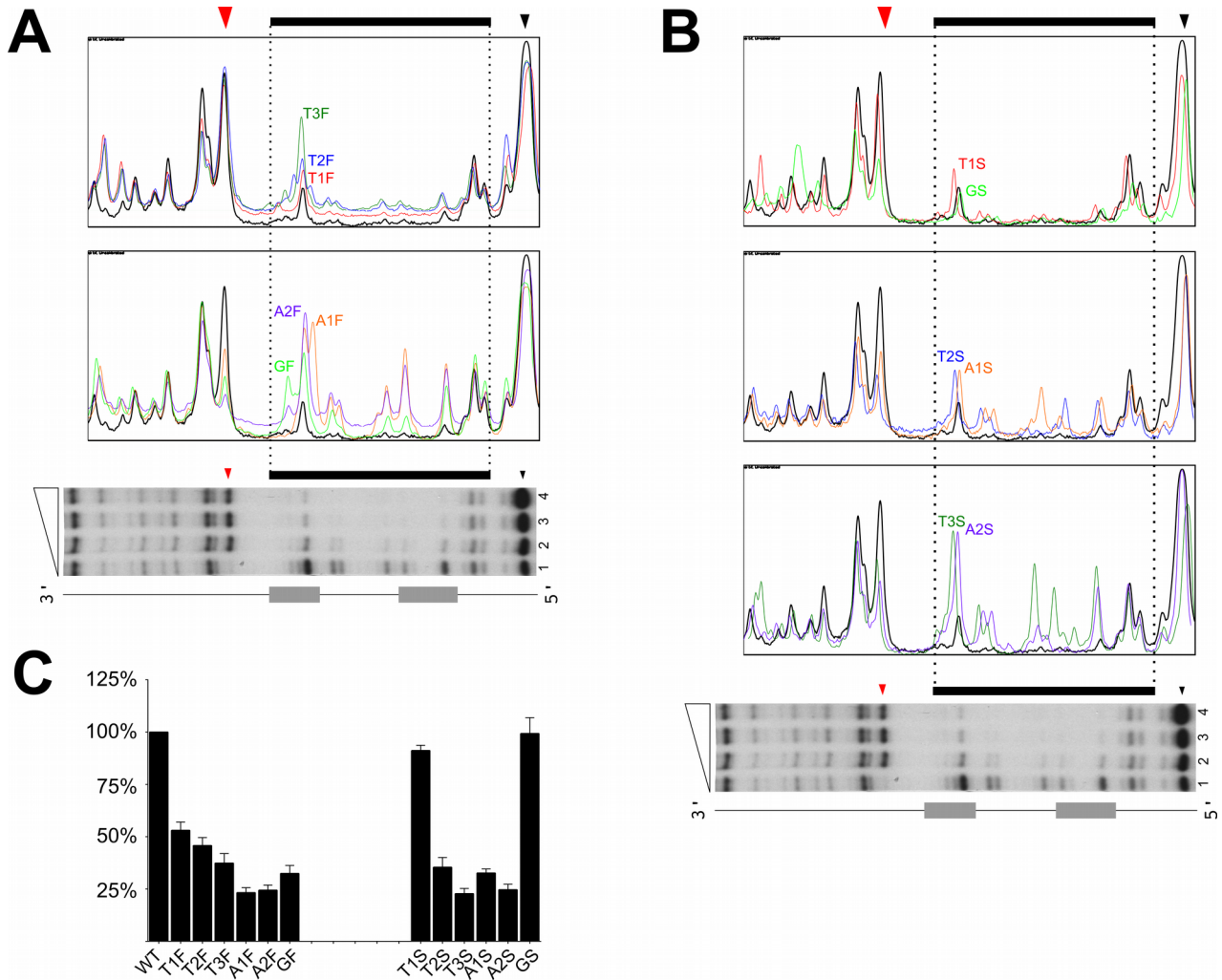
Fig. 3 Semi quantitative analysis of DNase I footprinting gels by densitometric quantification. The third lane of each footprint of Fig. 1 and Fig. 2 was analysed and plotted with the Fiji software. The analysis of the first hemisite is shown in Panel A, and that on the second hemisite in Panel B. Analysis of the WT curves is reported in in black. Black and red arrowheads mark the positions of two hypersensitive bands on the gel and the corresponding peaks in the graphs. The protected region is indicated with a black bar and its boundaries through the densitometric tracts by dotted lines. In panel C is reported a quantification of densitometric signal of the protected marked region, normalized with respect to first lanes value and then compared with WT signal. Standard deviations were calculated by quantification of two different experimental replicates.

HP1043 acts as an activator of transcription. In the first part of this study it has been reported that genes bound *in vivo* by HP1043 were upregulated in response to translational arrest (Pelliciari et al. 2017). These data suggested a direct regulation of HP1043, which might be achieved either by a mechanism of activation of transcription or by a mechanism of induction of transcription in response to translational arrest.

To further study the role of HP1043 on transcription, we performed *in vitro* transcription assays using the *E. coli* RNAp holoenzyme and a plasmid contain the *hp1227* promoter controlling the *lux*

reporter operon as template. Preliminary experiments were performed by incubating DNA with RNAp and then by adding purified HP1043. To a set of reactions, containing only template plasmid, were added a fixed quantity of RNAp and then increasing amounts of HP1043. In another set of reactions, the order of added components was inverted, that is, increasing amounts of the transcriptional regulator were added to the template, then the same amount of RNAp was added in each reaction. In the experiments with the RNAp added before HP1043, an increase in transcript levels was detected at lower protein concentration, then in the presence of high amounts of HP1043, the transcript levels decreased (Fig 4 panel A). A similar result was obtained in the parallel experiment in which the transcriptional factor was added before RNAp (*data not shown*). Subsequently, HP1043 was assayed on transcription using two additional regulated gene promoters, *cncr1* and *hp1043* itself as templates, with results reported on Fig.4 panel A. For all promoters tested it was observed an increase in mRNA levels up to DNA/protein ratio of 1:3, then a drop of mRNA to the baseline (transcription with no HP1043). The increase in transcript levels due to the presence of HP1043 in the transcriptional reaction on the tested promoters suggests that HP1043 is an activator of gene transcription. The observed reduction of the transcript levels in the presence of an excess amount of HP1043 is in agreement with observations reported from other systems (Tran CN et al. 2011) and it is likely due to an impairment of their function for protein aggregation or steric hindrance with the transcriptional machinery at high protein/DNA ratio. The different levels of mRNA detected upon addition of HP1043 in the *in vitro* transcription assays of the three promoters tested could reside in their different basal level of transcription To verify this hypothesis, *in vitro* basal activity of the promoters was analysed (Fig. 4 Panel B). Results show different levels of synthesized mRNA, with the lower amount obtained from the P_{hp1043} promoter and the highest amount from the P_{cncr1} promoter. However, upon addition of HP1043, the highest level of transcripts was obtained from the P_{hp1043} promoter, while comparable amounts of mRNA were observed from the P_{hp1227} and P_{cncr1}.

These results are in agreement with the hypothesis that the action of HP1043 on transcription is stronger on weak promoters than on strong promoters. According to the *in vitro* transcription data, HP1043 seems to act as a transcriptional activator, at least on the tested promoter. Moreover, the efficiency of transcriptional trans-activation exerted by HP1043 appears to be inversely linked to the basal transcriptional activity of the promoters.
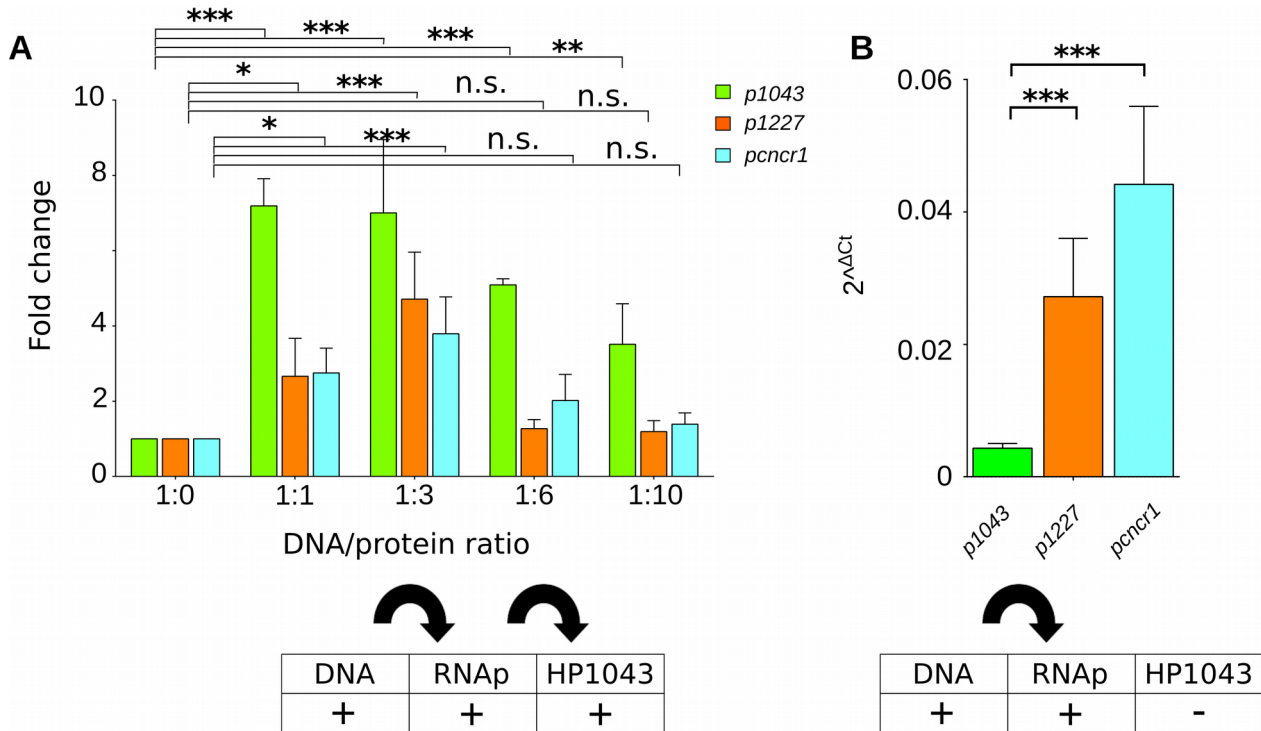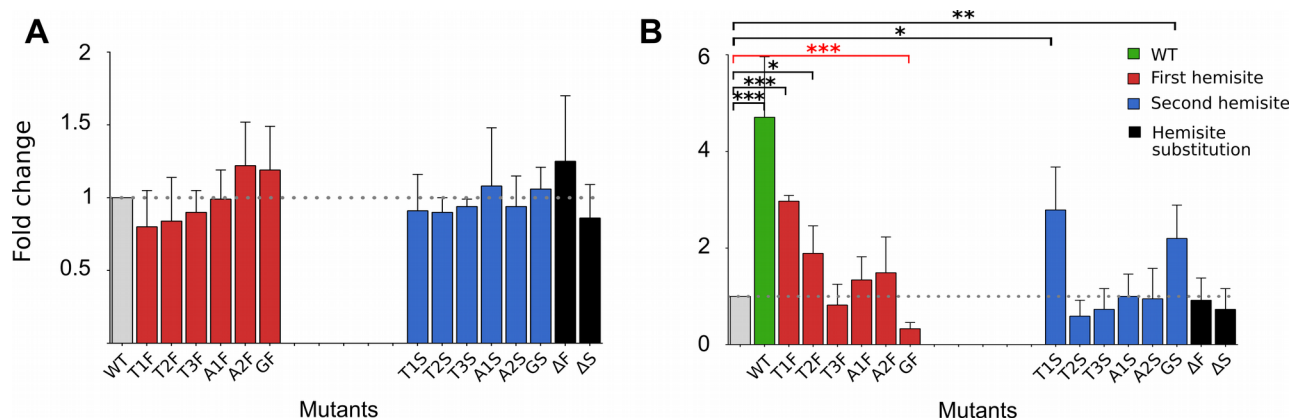
Fig. 4 *in vitro* transcription on different promoters with *E. coli* RNAp holoenzyme and HP1043 purified protein. In Panel B are reported the basal level of mRNA synthesized by incubating RNAp with plasmid containing different promoter that control *lux* operon, in Panel A the increases in transcription levels on the same constructs used before but with addition of HP1043 at different concentration. Under each graph is reported a schematic representation of the experiments, in which black arrows indicate the sequential addition of the components to the reaction. The asterisks indicate the statistical significance calculated by T student test, in which n.s. = non significative, * = $p < 0,01$, ** = $p < 0,005$, *** = $p < 0,001$.

***hp1227* promoter mutations impair HP1043 activity.** To verify if impairments of HP1043 binding to mutant binding sites correlate to the transcriptional activation of promoters, *in vitro* transcription experiments were performed as described above using all P_{hp1227} promoter mutants as templates. As the mutations affecting HP1043 binding map within promoter DNA regions it was first compared the level of transcripts synthesized from the WT promoter to the level of transcripts from the mutant templates by qRT PCR (Fig 5, Panel A). None of the tested plasmids show a statistical significant variation in the level of transcripts, indicating that mutations affecting HP1043 binding have no effects on promoter recognition and transcription by the RNA polymerase. Subsequently, transcription was carried out under the same experimental condition with the addition of the optimal concentration of HP1043 purified protein, established as the amount of protein required to obtain the maximum amount of transcripts.

Data obtained are reported in Fig. 5, Panel B, where each reaction was normalized to the control reaction with no HP1043 addition. While the amount of transcripts from the WT promoter increased

5-fold upon addition of HP1043 to the reaction, the amount of transcripts from all the mutants was lower (mutants T1F and T2F). The T3F, A1F, and A2F mutants showed a level of transcripts similar to the control experiments, indicating that the addition of the regulator had no effects on transcription. The GF mutant promoter showed an unexpected behaviour, as addition of HP1043 to the reaction caused a decrease in transcript levels. Mutatants in second hemisite, T1S and GS showed a level of transcripts similar to the amounts detected for the T1F and T2F mutants. The other four mutants, T2S, T3S, A1S, and A2S showed no variations with respect to their own controls with no HP1043 addition. The same result was obtained for the Δ1 and Δ2 mutants.

These findings correlate with the binding properties of HP1043 on WT and mutants assessed by both DNase I and hydroxyl-radical footprint, suggesting that a proper binding of HP1043 to its target exerts its action on the RNA polymerase by stimulating RNA transcription.

Fig. 5 *in vitro* transcription on *hp1227* mutant promoters upstream *lux* operon. In Panel A are reported the basal activity of each promoter cloned upstream *lux* operon; none of them show significant variation compared to WT. With grey bar is reported the WT promoter activity, in red the first hemisite mutants, in blue the second hemisite mutants and in black the two deletion mutants. In panel B are reported the transcriptional variation on the same promoter as before after addition of purified HP1043 in 3:1 ratio compared to RNAp. The bar follow the same colour scheme of Panel A, but in grey is reported the WT mRNA level in absence of HP1043 protein and in green the transcript level of the same construct with HP1043. The asterisks indicate the statistical significance calculated by T student test, n.s. = non significative, * = p< 0,01, ** = p<0,005, *** = p<0,001 . With black marks are indicated the promoters that conserve the increase in mRNA levels compared to WT without HP1043 addition, in red the promoter that shows a reduction in transcript levels.

Bibliography

- Hanahan D, (1983) Studies on transformation of Escherichia coli with plasmids. *J. Mol. Biol*, 166(4):557-80

- Pelliciari S & Peano C, Vannini A, Peano C, Puccio S, De Bellis G, Danielli A, Scarlato V, 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,* 23;7:4106.

- Schindelin, J.; Arganda-Carreras, I. & Frise, E. et al. (2012), "Fiji: an open-source platform for biological-image analysis", *Nature methods* 9(7): 676-682

- Tran CN, Giangrossi M, Prosseda G, Brandi A, Di Martino ML, Colonna B, Falconi M, (2011) A multifactor regulatory circuit involving H-NS, VirF and an antisense RNA modulates transcription of the virulence gene icsA of Shigella flexneri, *Nucleic Acids Res*, 39(18):8122-34.

- Winson MK, Swift S, Hill PJ, Sims CM, Griesmayr G, Bycroft BW, Williams P, Stewart GS, (1998) Engineering the luxCDABE genes from Photorhabdus luminescens to provide a bioluminescent reporter for constitutive and promoter probe plasmids and mini-Tn5 constructs, *FEMS Microbiol Lett*, 163(2):193-202.

5. Conclusions and perspectives

In this study, HP1043 was demonstrated to be one of the most important transcriptional regulator in the human pathogen *H. pylori*. With ChIP-seq approach several target promoters bound by HP1043 *in vivo* were identified. Intriguingly, they belong to crucial cellular pathways as translation, transcription and energy production and conversion. Moreover, the validation of *in vivo* binding, through DNase I and hydroxyl-radical footprinting experiments, suggests that HP1043 can recognize a specific nucleotide sequence formed by two conserved TTTAAG hexamers, spaced by a 5 bp non-conserved region. The importance of this putative consensus sequence was assessed, by single nucleotide mutagenesis performed on the P_{hp1227} promoter. The data collected showed that there might be different recognizing mechanism of the two hemisites, In fact, the corresponding nucleotides on the two hexamers appear to play a different effect on HP1043 binding. In the first hemisite the protein showed increasing impairment in DNA binding mutagenizing the nucleotides from the first T to the last G, with major effect on the AAG out of TTTAAG sequence. On the contrary, in the second hemisite, the first T and the last G seem to be irrelevant for protein binding, whereas the TTAA out of the TTTAAG hexamer appeared to play a key role, as their mutations completely abolished HP1043 binding to the DNA.

Also the functional role of the protein on its *in vivo* target was investigated. Due to the impossibility of modulating protein expression *in vivo, in vitro* transcription assays were performed to evaluate the effect of HP1043 promoter binding on the activity on the RNA polymerase. These experiments were performed on different promoters and in all of them there was an increment in mRNA levels after addition of the transcriptional regulator to the reactions. To correlate the binding of HP1043 to the increment in mRNA level generated by RNA polymerase, the same *in vitro* transcription experiments were performed on the P_{hp1227} promoter mutants. Results established that the increase in transcript levels is dependent on HP1043 promoter binding that in turns activates transcription.

Antibiotic resistance is increasing worldwide, and it has been regarded as the main factor reducing the efficacy of *H. pylori* therapy. The eradication rate of triple antibiotic therapy is currently less than 80% in most parts of the world. Antibiotic resistance is the main reason for treatment failure, therefore, the standard triple regimen is no longer suitable as a first-line treatment in most regions; For this reason, the discover of new strategy to counteract *H. pylori* infection is compelling. Novel alternatives based on microorganisms, peptides, polysaccharides, and intra-gastric violet light irradiation are developed, but most of them have not been effective in eradicating the bacteria and have been shown to maintain low bacterial levels. Moreover, all these approaches may lead

H. pylori to rapidly evolve in therapy resistant strains, because most of them were developed to interfere with enzymatic reactions.

Findings reported in this study can be the starting point to a search for new classes of antibiotic molecules and therapeutic treatments. In fact, until now, there are no studies reported in literature that aim to target *H. pylori* transcriptional regulator as new antibiotic approach for this bacterium. HP1043 has been demonstrated to be an essential protein for *H. pylori*, fine-tuned expressed and no responsive to modulation approaches of its expression. The possibility to identify a drug that might be able to impair binding of the HP1043 transcriptional regulator to its DNA target represents an appealing approach to specifically block infection and possibly, to eradicate *H. pylori* infection.

<u>Annex 1</u>

SCIENTIFIC REPORTS

Insight into the essential role of the *Helicobacter pylori* HP1043 orphan response regulator: genome-wide identification and characterization of the DNA-binding sites

Simone Pelliciari[1,*], Eva Pinatel[2,*], Andrea Vannini[1,†], Clelia Peano[2], Simone Puccio[2,3], Gianluca De Bellis[2], Alberto Danielli[1], Vincenzo Scarlato[1] & Davide Roncarati[1]

Many bacterial regulatory genes appear to be dispensable, as they can be deleted from the genome without loss of bacterial functionalities. In *Helicobacter pylori*, the hp1043 gene, also known as *hsrA*, is one of the transcriptional regulator that is essential for cell viability. This gene could not be deleted, nor the amount of protein modulated, supporting the hypothesis that HP1043 could be involved in the regulation of crucial cellular processes. Even though detailed structural data are available for the HP1043 protein, its targets are still ill-defined. Using Chromatin Immunoprecipitation-sequencing (ChIP-seq), one of the most powerful approaches to characterize protein-DNA interactions *in vivo*, we were able to identify genome-wide several new HP1043 binding sites. Moreover, *in vitro* DNA binding assays enabled precise mapping of the HP1043 binding sites on the new targets, revealing the presence of a conserved nucleotide sequence motif. Intriguingly, a significant fraction of the newly identified binding sites overlaps promoter regions controlling the expression of genes involved in translation. Accordingly, when protein translation was blocked, a significant induction of almost all HP1043 target genes was detected. These observations prompted us to propose HP1043 as a key regulator in *H. pylori*, likely involved in sensing and in coordinating the response to environmental conditions that provoke an arrest of protein synthesis. The essential role of HP1043 in coordinating central cellular processes is discussed.

Helicobacter pylori is a major pathogen, highly widespread among the human population, and it has been recognized as class I carcinogen by World Health Organization. It is considered the primary cause of severe gastrointestinal diseases such as peptic ulcer, gastric adenocarcinoma and MALT lymphoma[1,2]. The ability of this Gram-negative bacterium to colonize the harsh stomach niche and to establish a persistent infection depends on the coordinated expression of housekeeping genes as well as virulence factors that allow the pathogen to adapt to environmental conditions and to counteract host-defence mechanisms. A peculiar feature of the small-sized (1.66 Mb) *H. pylori* genome is the relative scarcity of genes encoding regulators of transcription. To date, only 17 transcriptional regulators have been identified and characterized to different extents[3]. Besides 3 σ factors (the housekeeping σ^{80} and the alternative RNA polymerase sigma subunits σ^{54}/σ^{28}, both involved in the transcription of flagellar genes) and 4 transcriptional repressors involved in metal homeostasis (Fur and NikR) or stress response (HrcA and HspR), *H. pylori* employs several two-component systems. Two-component systems are composed of a histidine-kinase and a response regulator. Upon signal perception, the sensor kinase catalyses its auto-phosphorylation and then transfers the phosphoryl group to a partner response regulator, a specific

[1]Department of Pharmacy and Biotechnology (FaBiT), University of Bologna, Italy. [2]Institute of Biomedical Technologies, National Research Council, Segrate, Milan, Italy. [3]Doctoral School of Molecular and Translational Medicine, University of Milan, Milan, Italy. [†]Present address: Department of Experimental, Diagnostic and Specialty Medicine, University of Bologna, Bologna, Italy. [*]These authors contributed equally to this work. Correspondence and requests for materials should be addressed to V.S. (email: vincenzo.scarlato@unibo.it) or D.R. (email: davide.roncarati@unibo.it)
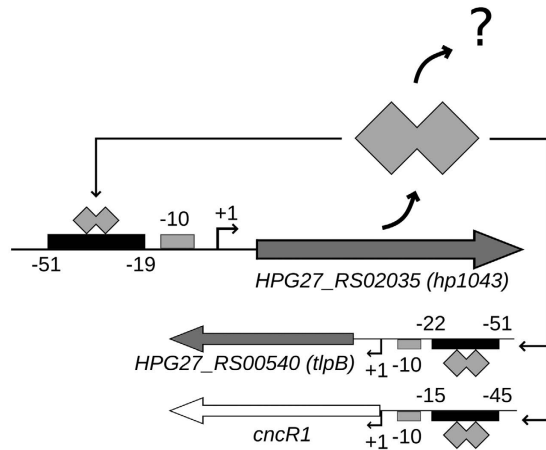
Figure 1. Schematic representation of the *hp1043* locus in *H. pylori* G27 along with the previously reported HP1043 binding sites. The binding sites of HP1043 are represented with black boxes, positions are relative to the transcription start sites. The transcriptional regulator, represented by a polyhedral shape, binds the promoter region of three target genes: *hp1043* and *tlpB*[11], and *cncR1*[12]. Protein coding genes are depicted with grey block arrows, the small non-coding *cncR1* RNA gene with a white block arrow. Transcription start sites are indicated with a bent arrow marked +1, the −10 region with grey boxes.

DNA binding protein that modulates transcription of target genes. Two-component systems are employed by *H. pylori* to coordinate gene expression in crucial cellular processes like chemotaxis (CheA/CheY)[4], copper resistance (CrdS/CrdR)[5], flagellar regulation (FlgS/FlgR)[6], and acid acclimation (ArsS/ArsR)[7]. Intriguingly, *H. pylori* genome harbours 2 genes, named *hp1043* (*HPG27_RS02035*) and *hp1021* (*HPG27_RS02145*), encoding for two so-called "orphan response regulators", as their partner sensor kinases are missing. HP1043, in particular, appears to be essential for cell viability. The *hp1043* gene, in fact, could not be deleted unless a second gene copy was integrated into the *H. pylori* chromosome[8]. The HP1043 regulator belongs to the OmpR family with a highly degenerate receiver sequence incapable of being phosphorylated. Different biochemical and structural studies suggest that HP1043 could exert its function in a phosphorylation-independent manner and it could be classified as belonging to a new response regulator family[8,9]. NMR-spectroscopy and X-ray crystallography suggested that HP1043 exists as a symmetric dimer, with two functional domains, an N-terminal regulatory domain and C-terminal DNA-binding domain. The dimer appears to be stable in solution in the un-phosphorylated state[10]. Even though detailed structural data are available for HP1043, the target genes bound and regulated by this regulator are still ill-defined. Specifically, to date only three HP1043 genomic binding sites have been characterized at the molecular level. *In vitro* DNA binding studies demonstrated that HP1043 binds its own promoter and the promoter region of *tlpB*, a gene encoding a methyl-accepting chemotaxis protein[11]. Moreover, in a recent study we demonstrated binding of HP1043 to the promoter region of *cncR1*, a small regulatory RNA of *H. pylori* G27 involved in the opposite modulation of motility and adhesion to host cells[12]. The impossibility of generating a knock-out mutant for *hp1043* gene, or even of modulating the amount of HP1043 protein in the cell, has hampered the detailed characterization of its regulatory function[11]. Initially proposed as a regulator of cell cycle-related functions[11], two recent studies attempted a link of HP1043 to homeostatic stress control (and named it HsrA for homeostatic stress regulator) of the bacterial cell and to a role in oxidative stress defence and nitrogen metabolism[13,14]. While gel mobility shift experiments support direct binding of HP1043 (HsrA) to the *porGDAB* promoter region, the regulation of other genes by HP1043 was inferred solely by the finding of a putative binding sequence[13,15].

In the present study we report the identification of HP1043 (HsrA) DNA targets *in vivo* and the characterization of selected binding sites *in vitro*. The combination of recent advances in deep sequencing coupled with well-established Chromatin Immunoprecipitation (ChIP) protocols for bacterial transcription factors, establishes ChIP-seq as one of the most powerful approaches to identify genome-wide all the *in vivo* binding sites of a regulator of interest. The ChIP-seq analysis here presented allowed us to identify several new *in vivo* HP1043 binding sites. Moreover, *in vitro* protein-DNA binding assays enabled precise mapping of the HP1043 binding sites on new targets, whose analysis revealed the presence of a conserved nucleotide sequence motif. Interestingly, a significant fraction of the newly identified motifs overlaps promoters associated to genes involved in the process of translation. Accordingly, a stress signal leading to the arrest of protein synthesis, resulted in a significant induction of almost all HP1043 target genes. These observations prompted us to propose HP1043 (HsrA) as a key regulator in *H. pylori*, likely involved in sensing and in coordinating the response to environmental conditions that provoke an arrest of protein synthesis.

Results

Genome-wide binding of HP1043 (HsrA). To date, the exact positions of only three HP1043 genomic binding targets, located on the promoter regions of *hp1043*, *tlpB* and *cncR1* genes, have been characterized in details and they are represented in Fig. 1. To identify genome-wide all the HP1043 targets and thus obtain an
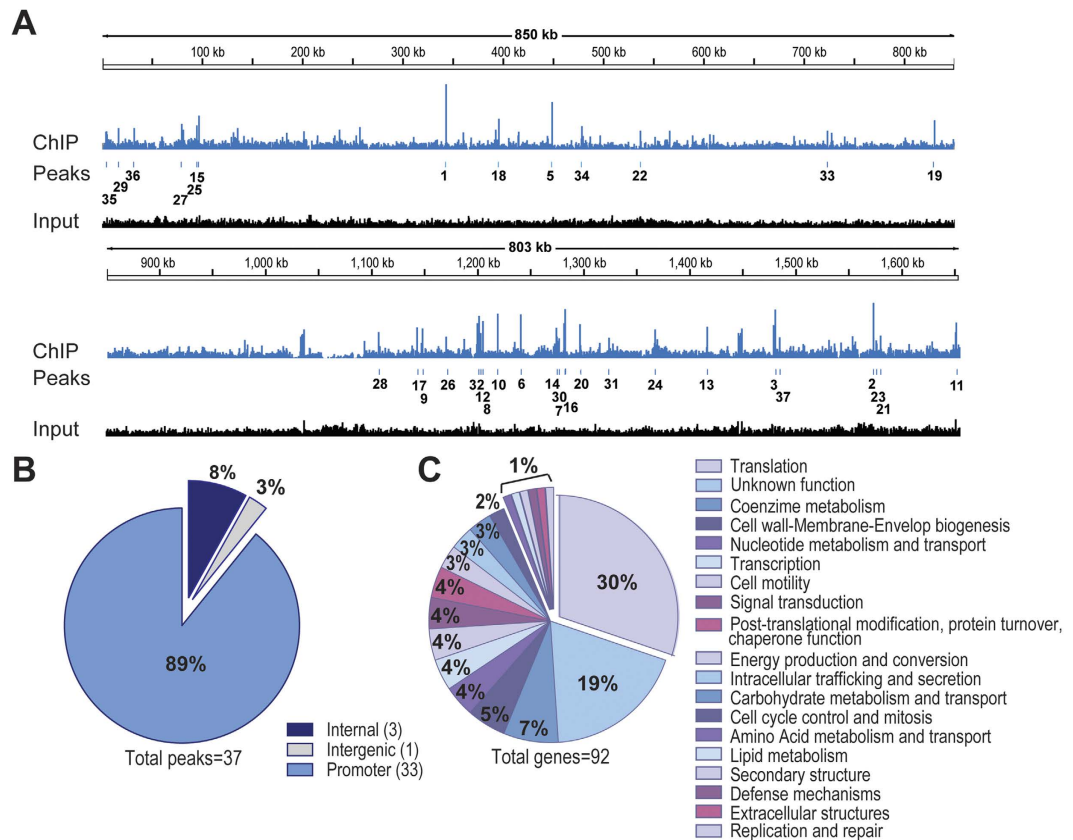
Figure 2. Genome-wide binding analysis of HP1043. (Panel A): genome-wide binding of HP1043 determined by ChIP-seq. The base count data, deriving from sequencing reads alignment on *H. pylori* G27 genome, of immunoprecipitated sample (labelled "ChIP", light blue track) and control sample (labelled "Input", black track) are represented; the position of the HP1043 binding sites, identified by peak-calling analysis, are highlighted by vertical bars and labelled according to the peak number. (Panel B): pie chart summarizing the positional analysis of the HP1043 binding sites relative to annotated genes. (Panel C): analysis of the functional categories of genes associated with novel HP1043 binding sites, identified by ChIP-seq.

unbiased overview of HP1043 regulatory role in *H. pylori,* we applied a ChIP-seq approach *in vivo*. In particular, liquid-grown *H. pylori* G27 wild type cells were cross-linked, sonicated and HP1043 protein-DNA complexes were immunoprecipitated (IP) with a specific polyclonal α-HP1043 antibody; at the same time a control sample (Input), deriving from sonicated but non-immunoprecipitated DNA, was prepared and used to estimate the background. A minimum of 2 million reads, with optimal mapping performances (>98.5%), for each sample and biological replicate were obtained (Supplementary Table S1) and used to generate the genome-wide binding profiles visualized in Fig. 2A. Combining Homer2 peak calling and Irreproducible Discovery Rate (IDR) procedure, we identified a set of 37 highly reproducible peaks (Fig. 2A and Table 1). These putative binding sites were the most enriched in the IP sample and highly reproducible among the two biological replicates (see Methods for a details). Afterwards, the peaks identified were annotated according to the *H. pylori* G27 RefSeq annotation (GCF_000021165.1), and cross-mapped to the transcription start sites and ncRNAs defined in strain 26695[16] (see Methods). Specifically, binding sites centred between position −150 and +50 with respect to a transcription start site were considered associated to a promoter region, while those centred outside of this positional range were classified as internal or intergenic on the basis of the peak central position (Table 1). According to this classification, 89% of the identified HP1043 binding sites (33/37) were associated to promoters, hence in a canonical position to exert a regulatory function, and about 1/3 of them were bidirectional promoters. Three out of 37 (8%) of the peaks mapped in internal regions, while only one single peak (3%) was classified as intergenic (Fig. 2B). Interestingly, several strongly enriched peaks were located in proximity of tRNAs genes (bolded in Table 1, Supplementary Table S2).

Among HP1043 target genes, we found the small RNA *cncR1*, previously demonstrated to be bound by this regulator[12], while the other three known targets (i.e. *hp1043* itself, *tlpB*[11] and *porG*[14]), were not present in the top scoring list. However, two regions mapping nearby the promoters of *hp1043* and *porG* were identified as enriched by the peak caller and are present in the full list of 107 peaks identified by Homer2 (Supplementary Table S3). These peaks were not included in the top list only because of the stringent criteria adopted to evaluate the peak calling reproducibility. This observation suggests that we may find some authentic and reliable HP1043 binding sites also among some of the less reproducible ChIP-seq peaks. For example, inspection of the latter peaks allowed to pinpoint an enriched region, present in both replicas, positioned on the *tlpB* promoter. Then, to obtain

Peak_name	start:end	Score	Class	Nearest genes
HsrA_Peak_1	342452:342626	33634,2	bidirectional promoter	nadE < **tRNA-Arg** > ilvC
HsrA_Peak_2	1570125:1570299	29290,7	promoter	> tRNA-Phe
HsrA_Peak_3	1476498:1476672	27064	promoter	16SrRNA <
HsrA_Peak_4$	1:126	26419	promoter	> HPG27_RS07980
HsrA_Peak_5	448262:448436	24526,6	promoter	**tRNA-Pro** <
HsrA_Peak_6	1212453:1212627	22077	promoter	fldA <
HsrA_Peak_7	1276504:1276678	21217,1	promoter	atpE <
HsrA_Peak_8	1194610:1194784	20592,6	promoter	16SrRNA <
HsrA_Peak_9	1140440:1140614	20533	bidirectional promoter	galE < **tRNA-Ile**
HsrA_Peak_10	1234366:1234540	20239,5	bidirectional promoter	nupC < **tRNA-Met** > HPG27_RS05875
HsrA_Peak_11	1648608:1648782	17967,3	promoter	> flgG2,flgG
HsrA_Peak_12	1198357:1198531	17780,6	promoter	rpS16,rpsP <
HsrA_Peak_13	1411804:1411978	17187,4	bidirectional promoter	asRNA_HPG27_RS06840 < > HPG27_RS06835 (pseudogene)
HsrA_Peak_14	1268342:1268516	16391,8	promoter	secE <
HsrA_Peak_15	96262:96436	16366,7	bidirectional promoter	fabD < **tRNA-Ser**
HsrA_Peak_16	1275920:1276094	15715,5	intergenic	tRNA-Leu « tRNA-Leu
HsrA_Peak_17	1135651:1135825	15701,4	bidirectional promoter	nrdB,nrdF < > tRNA-Leu
HsrA_Peak_18	395023:395197	15492,7	promoter	**tRNA-Ser** <
HsrA_Peak_19	829599:829773	14406,8	promoter	as_HPG27_RS03935 <
HsrA_Peak_20	1290665:1290839	14058,4	bidirectional promoter	hemN2,hemN < > cytc553
HsrA_Peak_21	1576835:1577009	13402,5	promoter	isoA <
HsrA_Peak_22	536813:536987	13358,6	promoter	cncR1_Hpnc2630 <
HsrA_Peak_23	1573020:1573194	13228,3	internal	dnaA <
HsrA_Peak_24	1362385:1362559	12858	promoter	rpl36,rpmJ <
HsrA_Peak_25	94349:94523	11431,6	promoter	rpoD <
HsrA_Peak_26	1163325:1163499	11422,2	promoter	> asRNA_HPG27_RS05530 (Hpnc3560)
HsrA_Peak_27	79140:79314	11103,6	promoter	**tRNA-Val** <
HsrA_Peak_28	1099169:1099343	10857,3	promoter	> rps1,rpsA
HsrA_Peak_29	16420:16594	10304,9	bidirectional promoter	putative_SRP_RNA < > HPG27_RS00110
HsrA_Peak_30	1270429:1270603	10176,2	promoter	> hetA
HsrA_Peak_31	1318842:1319016	9864	bidirectional promoter	bioC < > secG
HsrA_Peak_32	1196544:1196718	9212,8	promoter	rpl19,rplS <
HsrA_Peak_33	722928:723102	8986,8	internal	> flgR
HsrA_Peak_34	478010:478184	8089,2	promoter	> hofC
HsrA_Peak_35	4434:4608	7929,2	internal	**tRNA-Lys** <
HsrA_Peak_36	31591:31765	7905,6	bidirectional promoter	HPG27_RS00165 < > uspA
HsrA_Peak_37	1480392:1480566	7634,1	promoter	> rpl34,rpmH

Table 1. Location of putative HP1043 binding sites. $Located on G27 plasmid; Bold, feature included in the peak; >/<, target gene strand.

a general view of HP1043 regulatory function, each promotorial binding site was associated to the downstream operon. A functional enrichment analysis was then performed to determine if the genes included in the operons belonged to specific functional categories. For this purpose, protein coding genes were classified using Clusters of Orthologous Groups of proteins (COGs) database[17], while tRNAs and rRNAs were included to the translation category. According to this analysis, it turned out that a significant fraction (30%) of genes, associated to a HP1043 promotorial binding site and likely controlled by the regulator, is involved in the process of protein translation (Fig. 2C). Specifically, the promoters of several tRNA and rRNA genes, as well as of genes encoding ribosomal proteins (HPG27_RS05725/*rps16*, HPG27_RS06525/*rpl36*, HPG27_RS05215/*rps1*, HPG27_RS05705/*rpl19*, HPG27_RS07170/*rpl34*) or involved in rRNA and ribosome maturation and assembly (HPG27_RS05770/*ybeY* endoribonuclease and HPG27_RS06295/*frr* ribosome recycling factor) were targeted *in vivo* by HP1043. Strikingly, in the same analysis we identified HP1043 binding sites associated to genes involved in other fundamental cellular processes like energy production and conversion (flavodoxin, HPG27_RS05775/*fldA*; ATP synthase C chain, HPG27_RS06075/*atpE*; cytochrome c553, HPG27_RS06145/*cytc553*), as well as RNA transcription (HPG27_RS00460/*rpoD*, coding for the RNA polymerase housekeeping σ^{80} factor[18]). In this respect, we noticed that one more RNA polymerase gene, HPG27_RS06505/*rpoA*, encoding the RNA polymerase α subunit, is included in an operon likely controlled by HP1043. This last observation was manually curated because in the last version of *H. pylori* G27 genome annotation (GCF_000021165.1) this gene is mis-annotated to pseudogene.

Overall, from the above data we can infer that HP1043 may represents a key regulator in *H. pylori*, involved in the control of crucial cellular processes, such as transcription and translation.

Validation of novel HP1043 targets. To confirm HP1043 binding sites identified by ChIP-seq, a subset of them was selected and used as probes for *in vitro* binding to a purified recombinant HP1043 protein in DNase I footprinting experiments. Among the 37 putative highly confident HP1043 targets (Table 1) we selected 7 representative genomic regions: 6 binding sites mapping within promoters and 1 site falling inside a coding sequence. We also included a DNA probe covering the promoter of the HPG27_RS06315/*nuoA* gene. This putative HP1043 binding site was present in the total peak list and close to the ranking position of *hp1043*, but excluded from the top list. Similarly to some other peaks not included in the top list of Table 1, this binding site exhibits a broad region of enrichment encompassing the promoter, even if the peak centre appears mis-positioned with respect to the core promoter region. Thus, to assess if the peaks having this profile represent *bona-fide* HP1043 binding sites with a lower binding affinity rather than artifacts of the ChIP-seq experiment, we decided to validate further by DNase I footprinting one of them (Fig. 3). Consistent with ChIP-seq findings, all the 8 genomic regions selected were directly targeted *in vitro* by the purified recombinant HP1043 protein (all probes showed a clear area of DNase I protection), apparently with different relative affinities (some probes showed protection in the presence of higher amounts of HP1043 protein). Probes bound *in vitro* with higher relative affinity are reported in Fig. 3, panel A and B, while probes bound with lower relative affinity are shown in panel C and D. Increasing concentrations of HP1043 determined the appearance of a region of protection (black box) flanked at least on one side by a single DNase I hyper-sensitive band (black arrowhead). HP1043 binding to these 4 targets protected a DNA region of about 30 bp and the protection was saturated at $1.7\,\mu M$ protein concentration (Fig. 3, panel A, lanes 2, 7, 12 and panel B, lane 2); these were considered high affinity binding sites. HP1043 binding to the promoters of *isoA*, *rpoD*, *rpl36* and of *nuoA* genes (Fig. 3, panel C and D) appears at higher protein concentrations and with a slightly different binding pattern. Here the protection was saturated at a protein concentration of $13.3\,\mu M$, and again protected a region of about 30 bp flanked, only in some cases, by one or more bands of hyper-sensitivity to DNase I digestion. These were considered low affinity binding sites. Overall, the above *in vitro* footprinting assays confirm *in vivo* binding data. Notably, HP1043 binds *in vitro* exactly to the same positions of the enriched regions derived from ChIP-seq data analysis for all the top scoring peaks (Fig. 3A, B, C lower part of each panel). Moreover, the footprinting experiment reported in Fig. 3B (lanes 1 to 5) validates HP1043 intra-cistronic binding within the *flgR* coding sequence, further supporting ChIP-seq analysis and suggesting the existence of a minority of non-canonical binding sites, apparently not associated with regulatory functions.

In summary, we confirmed HP1043 binding to 8 new sites identified in *in vivo* ChIP-seq experiments and defined *in vitro* the nucleotide sequences bound by the protein.

Molecular characterization of HP1043 (HsrA) binding to DNA. The DNase I footprinting data indicated that the newly identified HP1043 binding sites overlap the core promoter region and map, in all cases, immediately upstream of the −10 box (Fig. 3, panels A, C and D). This observation parallels with the positions of HP1043 binding sites on the three previously characterized target sites (Fig. 1 [11,12]). To get a detailed picture of the promoter sequences encompassing HP1043 binding sites, the core promoter sequences of the newly and previously validated HP1043 targets were analysed using the GLAM2 software[19]. Results are shown in Fig. 4A. The HP1043 target promoters appear to be characterized by a −10 hexamer typical of the housekeeping σ^{80}-dependent transcription, preceded, in some cases, by an extended TG motif (Fig. 4A, upper part). Moreover, a sequence resembling the −35 box is lacking in these promoters, apparently replaced by two conserved AT-rich motifs separated by a spacer region (Fig. 4A, lower part). Intriguingly, the regions protected by HP1043 binding in DNase I footprinting assays (highlighted in Fig. 4A, upper part) encompass, in almost all cases (except in *rpoD*), these two conserved AT-rich motifs. To determine a possible consensus motif recognized by the HP1043 protein, we performed the same analysis using only the regions protected by HP1043 in the *in vitro* footprint experiments as input sequences. In this way GLAM2 outlined a highly conserved motif, reported in Fig. 4B. Specifically, the motif appears to be constituted by two direct repeats (TTTAAG) separated by a 5 bp-spacer, in which the first highly conserved hexamer is followed by a second less conserved one. Notably, the spacer length is conserved, positioning the centre of the two direct repeats at 11 bp distance that corresponds to 1 helical turn, suggesting that a dimer of HP1043 recognizes the DNA on the same face of the double helix. The conserved sequence motif (underlined and in bold in Fig. 4B, upper part) is either centred within the experimentally mapped binding sites (*secE*, *cyt553*, *cncR1* and *nuoA*), or shifted towards one side of the protected region (*isoA*, *rpoD*, *hp1043* and *tlpB*). Possibly, this is due to the intrinsic low level of resolution in mapping protein binding sites with DNase I footprinting assays. In fact, some regions of the DNA probes used for these assays naturally lack bands of digestion even in the sample without protein and this introduces a degree of uncertainty in defining the precise boundaries of the binding regions. Hence, to precisely map HP1043 (HsrA) binding sites and to further characterize the connection between the identified consensus motif and the intimate contacts of HP1043 with target DNA, hydroxyl-radical footprinting assays were carried out on selected targets, previously probed by DNase I footprintings (Fig. 3). Results are reported in Fig. 5. On all the promoter probes tested, HP1043 binding resulted in a periodic pattern of three short protected tracts of 3/4 nucleotides in length, separated by two non-protected regions of 5/6 nucleotides. Intriguingly, for all binding sites the central protection centers exactly within the spacer that separates the conserved direct repeats, while the other two protected DNA tracts fall immediately upstream and downstream of the direct repeats of the consensus binding motif (Fig. 5). It is worth mentioning that regions protected in hydroxyl-radical footprinting experiments reflect limited accessibility of radical ions to the DNA minor groove and, for this reason, these protected regions do not necessarily represent the portions of the probe directly contacted by the protein. Considering that HP1043 footprint regions surround the conserved direct repeats (Fig. 5), our data suggest that HP1043 could interact with the TTTAAG repeats in the DNA major groove narrowing
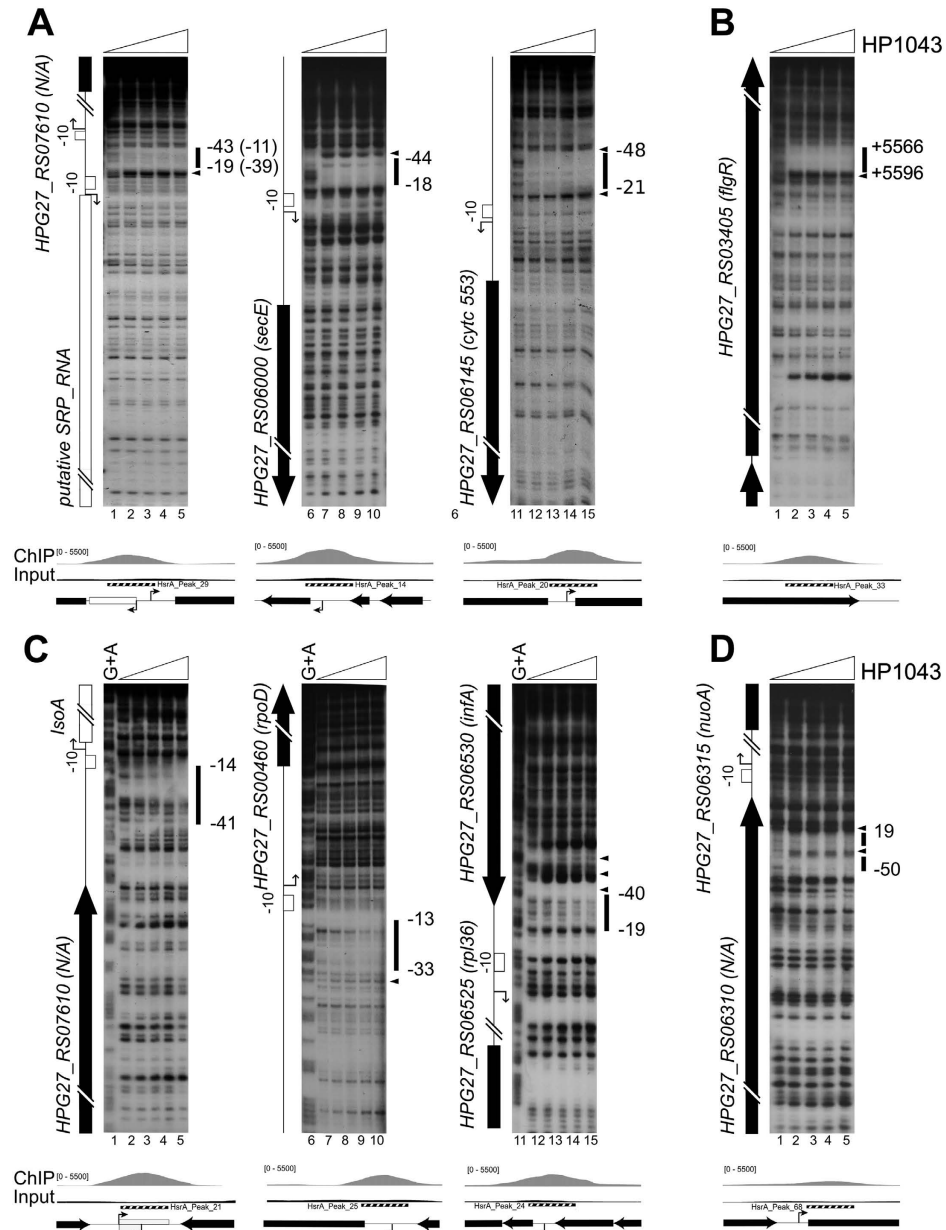
Figure 3. *In vitro* binding of purified recombinant HP1043 protein on selected genomic regions identified by ChIP-seq. In panel A high affinity promoter binding sites are shown, in panel B footprinting validation of an intragenic binding site is reported, in panel C low affinity promoter binding sites are represented and in panel D footprint validation of a binding site not included in the top scoring peak list. Radiolabelled DNA probes, harbouring the genomic regions of interest, were incubated with increasing concentrations of purified HP1043 (0, 1.7, 3.3, 6.6, 13.3 μM of dimeric HP1043 from left to right in each experiment) and subjected to DNase I digestion. Purified DNA fragments were separated on a polyacrylamide denaturing gel along with a G + A sequence reaction ladder (Panel C and data not shown) to map the binding sites. On the left of each autoradiograph, a schematic representation of the genomic region is drawn, with block arrows depicting coding sequences (black arrows or blocks) or putative sRNA (white arrow). Bent arrows represent the transcriptional start sites identified in *H. pylori* strain G27 by primer extension analyses (Supplementary Fig. 3S), white boxes the −10 promoter sequence. Notably, the initiation of RNA transcription at the analysed promoters is conserved between strain G27 and 26695. Black vertical lines on the right of each autoradiograph represent regions of protection from DNase I digestion, while black arrowheads highlight DNase I hyper-sensitive sites. Numbers refer to the positions with respect to the transcription start site (or with respect to the ATG translational start codon for the internal binding site on HPG27_RS03405). For the binding site mapped in the intergenic region between HPG27_RS00110 and putative SRP_RNA genes, HP1043 binding positions are reported with respect to the transcriptional start site of both genes (with no brackets for putative SRP_RNA, with brackets for HPG27_RS00110). Below each autoradiograph, a magnification of base-count data, deriving from ChIP-seq (as in Fig. 2A) of the genomic regions of interest, is reported together with a schematic representation of genes and features of the locus.
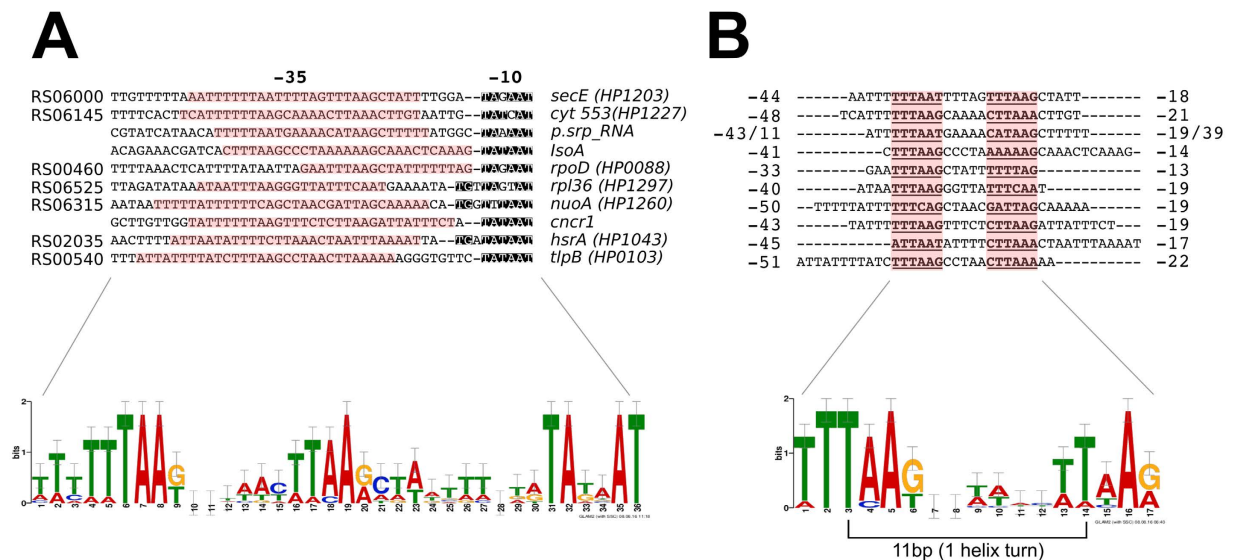
Figure 4. Characterization of HP1043 bound promoters (A) and definition of HP1043 consensus binding motif (B). Nucleotide sequence motifs were identified by using GLAM2 software[19], forcing strand specific alignment and maintaining all the default parameters to obtain the weblogos. Panel A, nucleotide sequence alignment (upper part) and weblogo (lower part) relative to 10 promoter regions bound by HP1043. Specifically, DNA sequences spanning from the transcriptional start site (+1) to position −70 of the 7 genes analysed by footprinting (Fig. 3A, C, D) were aligned together with the same DNA regions upstream *tlpB*, *cncR1* and *hp1043* genes. Nucleotides matching the consensus −10 box and the extended TG motif are highlighted in black, while DNA regions protected by HP1043 binding, as defined by DNase I footprintings (Fig. 3), are highlighted in red. Names reported on the left of each sequence refer to the new *H. pylori* G27 genome annotation, while names on the right of each sequence report the common gene name and, in brackets, the name referred to the *H. pylori* 26695 genome annotation. Panel B, nucleotide sequence alignment of the 10 protected regions in DNase I footprintings (Fig. 3), mapping in the promoters of the same genes reported in panel A. Aligned sequences and the resulting sequence logo are shown in the upper and lower part of panel B, respectively. The sequences of the two direct repeats are represented in bold and underlined; numbers flanking each sequence refer to the coordinates of the DNase footprinting protected regions (shown also in Fig. 3). In both sequence logos (lower part of both A and B panels) the height of each letter represents the relative conservation of each base.

the adjacent minor grooves which results in the protection observed *in vitro* by hydroxyl-radical footprinting. Interestingly, hydroxyl-radical footprinting analysis allowed the identification of a peculiar organization on *rpoD* promoter, which harbours multiple HP1043 binding sites. In fact, on this promoter, hydroxyl-radical footprinting assay (Fig. 5, lanes 26–30), besides the binding site already mapped with DNase I footprinting (Fig. 3C, lanes 6–10), revealed the existence of a distal HP1043 binding site, spanning positions −45 to −68 with respect to the initiation of RNA transcription. Moreover, the analysis of the DNA sequence of this distal binding site, revealed an HP1043 binding sequence highly similar to the consensus binding motif described above. The *rpoD* gene represents the first example of an HP1043 target with multiple binding sites, suggesting a complex mechanism of transcriptional regulation. In conclusion, hydroxyl-radical probing allowed to further refine the positions of HP1043 binding, providing the demonstration that the direct repeats of the proposed consensus motif are recognized by HP1043, likely through a major groove read-out mechanism.

Genes directly targeted by HP1043 are up-regulated upon translational arrest. ChIP-seq data indicated that HP1043 binds *in vivo* regulatory regions of several genes involved in translation (Fig. 2C). Hence, we hypothesized that the regulative function of HP1043 could be boosted by a signal affecting protein synthesis. To verify this hypothesis, in a preliminary survey, *H. pylori* G27 wild type cells were liquid-grown to mid-exponential phase, and HP1043 mRNA abundance was monitored at different time points after blocking protein synthesis by the addition of a sub-lethal concentration of tetracycline, which prevents aminoacyl tRNA attachment to ribosomal acceptor site[20]. Quantitative Real-Time PCR (qRT-PCR) assay revealed that the *hp1043* transcript amounts significantly increase upon tetracycline treatment, reaching a maximum induction upon 60 min of treatment (data not shown). Similar results were obtained upon treatment with chloramphenicol, which blocks ribosomes on mRNA during translation[21]. These preliminary results prompted us to further investigate the impact of translational stress on the HP1043 (HsrA) regulon. Total RNA collected from *H. pylori* wild-type cells exposed for 60 min to tetracycline and from a control culture were used to generate strand-specific cDNA libraries for RNA-sequencing analysis. Overall, differential gene expression outlined 1065 out of the 1592 total number of *H. pylori* genes to be deregulated upon antibiotic treatment. Of these, 543 genes were upregulated, representing 34% of the *H. pylori* genes (Fig. 6A, column X). However, focusing on the panel of genes belonging to all the operons controlled *in vivo* by HP1043, we observed that 52% of them were upregulated after tetracycline
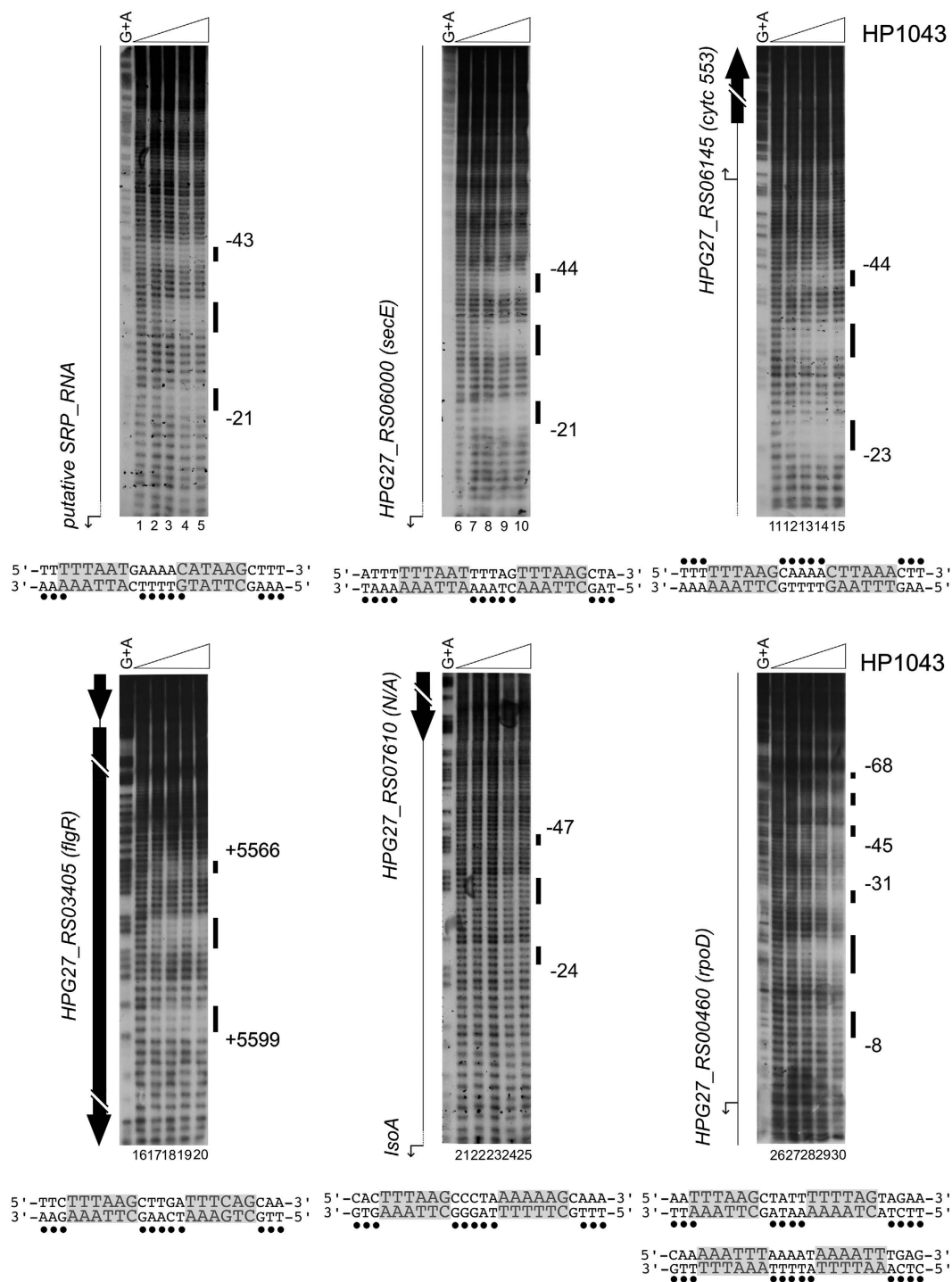
Figure 5. Hydroxyl-radical footprinting analysis of HP1043 binding to selected genomic regions. Radiolabelled DNA probes, harbouring the genomic regions of interest, were incubated with increasing concentrations of purified HP1043 dimer (0, 0.8, 1.7, 3.3 μM of dimeric HP1043 from left to right in each experiment) and subjected to hydroxyl-radical digestion. Purified DNA fragments were separated on a polyacrylamide denaturing gel along with a G + A sequence reaction ladder to map the binding sites. On the bottom of each autoradiograph, the nucleotide sequence of the HP1043 binding sites mapped is reported: grey-highlighted nucleotides depict the hexameric direct repeats of the binding motif, while the black dots indicate the nucleotides protected in hydroxyl-radical footprintings on the labelled DNA strands. Symbols are detailed in the legend to Fig. 3.
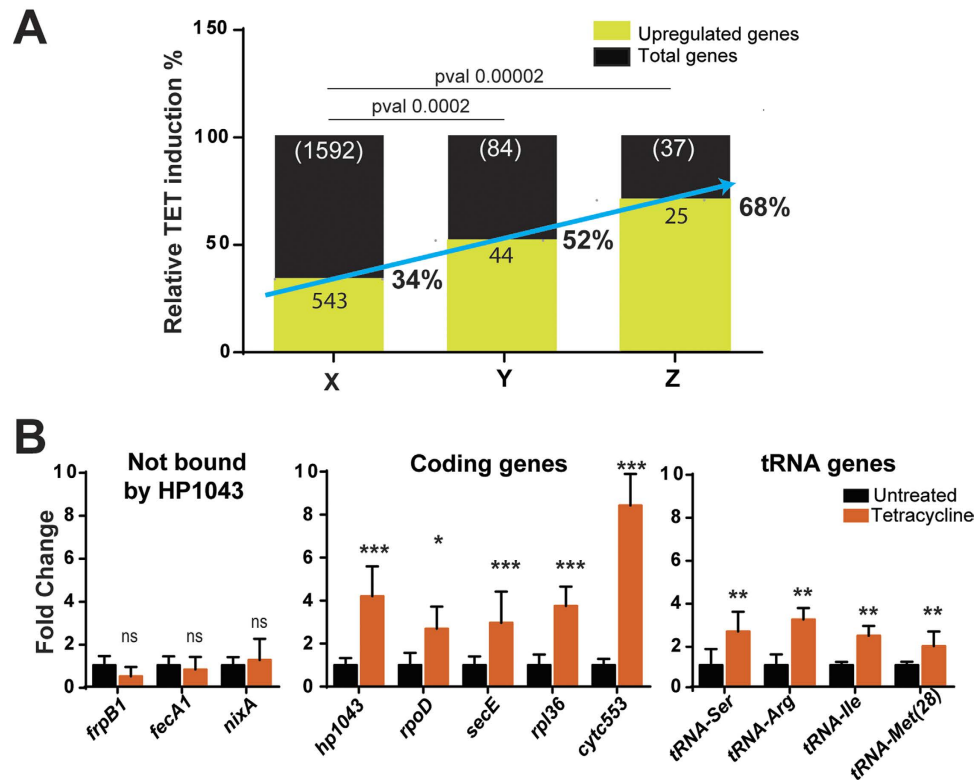
Figure 6. RNA-seq (A) and Real Time PCR (B) analyses of transcript level variations of novel HP1043 targets in response to translation arrest. Panel A): relative percentage of the genes upregulated after tetracycline treatment considering all annotated *H. pylori* G27 genes (column X, with exclusion of rRNA genes, depleted during sample preparation), the genes in operons under direct control of HP1043 (column Y with exclusion of 2 rRNA genes and of the genes absent in *H. pylori* annotation: 4 antisense RNA, and 2 small RNA), and the genes leading the operons under direct control of HP1043 (column Z with the exclusion of the genes indicated for column Y), respectively. The reported p-value was calculated using Fisher test. Panel B): effect of tetracycline treatment on transcript amounts of a selection of genes associated with an HP1043 binding site assessed by real time (qRT-PCR) analysis. Total RNA was extracted from cells treated or not treated with 0.5 μg/ml tetracycline for 60 min and reverse transcribed to cDNA. Transcript levels of a selection of genes not associated (panel B, left graph) or associated (panel B, central graph) with an HP1043 binding site were quantified by qRT-PCR, using the housekeeping 16 S rRNA gene as control. The same analysis was carried out on some tRNA genes (panel B, right graph) associated with HP1043 binding. Statistically significant differences were assessed by Student's t-test (Error bars indicate the standard deviation deriving from three independent biological samples, each analysed in duplicate technical replicates). Symbols: *p value < 0.05; **p value < 0.01; ***p value < 0.001; ns, p value > 0.05, not significant.

treatment (Fig. 6A, column Y), with a much more significant enrichment (68%) of upregulated HP1043 direct targets among the genes leading each operon (Fig. 6A, column Z). Several cellular pathways were affected by this treatment, supporting the notion that block of translation is a major challenge for bacterial cells, which respond with a wide transcriptional reprogramming of most key cellular processes. It is worth-noting that genes coding for rRNA and non-coding RNA were not included in the analysis because the first were depleted by the RNA sample preparation and the second are not annotated in strain G27.

To validate RNA-seq data, the differential mRNA levels of a subset of HP1043 targets were assayed upon tetracycline exposure by qRT-PCR analysis. Three negative controls (*frpB1*, *fecA1* and *nixA* i.e. three genes not targeted by HP1043) were included in the analysis and remained essentially unchanged upon antibiotic challenge (Fig. 6B, left histogram). Results reported in Fig. 6B (central histogram) show that transcript levels of selected HP1043 target genes (*hp1043, rpoD, secE, rpl36, cytc553*) increased upon tetracycline treatment, with fold variations ranging from 3- to 7-fold. Moreover, as shown in Fig. 6B (right histogram), a similar increasing trend upon translational arrest was observed for a selection of tRNA genes. These data suggest a possible involvement of the orphan response regulator HP1043 (HsrA) in the transcriptional response of *H. pylori* to environmental conditions or signals that promote arrest of protein synthesis.

Discussion

Two-component signal transduction systems typically regulate bacterial cellular functions in response to environmental conditions through a phosphorylation-dependent process. The human pathogen *H. pylori* relies on such regulatory systems to control important cellular functions such as motility, chemotaxis, acid acclimation

and copper resistance[22]. The orphan response regulator HP1043, also known as HsrA, is proven to be essential for cell growth and shows no requirement for the well-known phosphorelay scheme to be functional. In the present study, we have set up chromatin immunoprecipitation with α-HP1043 antibody followed by deep sequencing (ChIP-seq). This approach led to the identification of several new HP1043 genomic targets (Fig. 2A). To our knowledge, this is the first study that provides a genome-wide analysis of the HP1043 binding *in vivo*. Specifically, 37 genomic HP1043 binding sites were identified (Table 1), the majority of which are associated to a promoter region (Fig. 2B). A predominant fraction of genes associated to HP1043 binding encodes for proteins involved in crucial cellular functions, such as protein synthesis (tRNAs and ribosomal proteins coding genes), gene transcription (RNA polymerase subunits), and energy metabolism (operon containing genes coding for the NADH-ubiquinone oxidoreductase whole complex, as well as the ATP synthase C-chain gene promoter) (Fig. 2C, Table 1). Thus, it seems that HP1043 plays a key role for the fitness of the bacterium, which is a prerequisite for a successful infection. Moreover, our findings suggest that HP1043 regulator might represent a central regulatory switch mechanism that *H. pylori* exploits to modulate its metabolism and growth behaviour. In this respect, it is interesting to note the connection between HP1043 regulation and the arrest of translation (Fig. 6), a major challenge for bacterial cells.

Recent works by Olekhnovich and colleagues[14,15] suggested an involvement of HP1043 in directly regulating a list of about 70 genes encoding for proteins with disparate functions. However, it is worth mentioning that this proposed regulon was defined by searching against the *H. pylori* genome with a consensus binding motif defined by aligning two binding sites only[15]. In our study, we have defined the HP1043 regulon by identifying *in vivo* several new direct targets and noticed that many previously proposed binding sites appear not bound in our experimental conditions.

The list of 37 HP1043 genomic binding sites (Table 1) derives from a stringent peak-calling analysis, that takes into consideration the reproducible high fold enrichment of the immunoprecipitated DNA regions with respect to the input DNA in two independent biological replicates, thus representing high-confidence candidates. The identification of this list of *bona fide* HP1043 targets likely prevented from the inclusion of some false positives, but at the same time it may have determined the exclusion of several real binding sites characterized by low affinity binding levels and/or low reproducibility among replicates. In this respect, HP1043 binding on its own promoter, a known target previously characterized at the molecular level[11], was not included in the top list of highly significant binding sites. Moreover, we have shown that a putative HP1043 target, mapping upstream of the transcription start site of the HPG27_RS06315 gene, not included in the top list, is indeed an authentic HP1043 binding site (Fig. 3D). Hence, besides the new binding sites identified, it can be hypothesized that the HP1043 regulon may include additional members, not pinpointed by our analysis. Intriguingly, 8% of the newly identified genomic binding sites are not associated to promoters, mapping within protein coding sequences (Figs 2B and 3B). This observation poses some questions about the functional role of HP1043 binding to these internal sites. Even though we cannot exclude the existence of alternative and still ill-defined mechanisms of transcriptional regulation exerted by HP1043 bound inside coding sequences, possibly some HP1043 intracistronic binding sites could be associated with still-unknown internal promoters driving the transcription of intracistronic or anti-sense transcripts. Alternatively, these sites are not associated to regulation of transcription. The advent of the 'omics' revolution allowed the observation of binding sites not associated to regulation of several regulators, like *H. pylori* Fur repressor and *E. coli* CRP activator[23,24]. Even though it has recently been proposed that regulators with this behaviour may have evolved from nucleoid associated proteins[25], validation of this hypothesis needs further investigation.

The alignment of promoter regions harbouring HP1043 binding sites (Fig. 4A) revealed that genes controlled by this regulator are transcribed by a putative vegetative σ[80]-dependent promoter with a conserved −10 box. Moreover, the −35 hexamer, typical of housekeeping promoters appeared to be lacking, consistent with previous observations in the *H. pylori* 26695 strain[16]. Intriguingly, in the subset of promoters here analysed, the −35 motif is replaced by two conserved AT-rich motifs separated by a spacer region (Fig. 4A, lower panel), overlapping HP1043 binding site. The position of HP1043 binding sites, just upstream the −10 hexamer, is typical of activators of transcription[26,27]. For example the binding sites of *Bacillus subtilis* PhoP transcriptional activator, belonging to the large OmpR/PhoB subfamily of response regulators to which HP1043 has previously been associated[10], are typically centred between positions −17 to −66 of the activated promoters[28]. Accordingly, we speculate that HP1043 acts as an activator of transcription, boosting the activity of weak promoters controlling crucial genes involved in key cellular functions. Binding of HP1043 upstream of the −10 box would facilitate the contacts between the regulator and RNA polymerase, thereby stimulating initiation of transcription. This hypothesis might also be partly supported by the data summarized in Fig. 6B.

To identify sequence specific determinants for HP1043 binding, 10 nucleotide sequences protected in footprinting analyses were aligned using the GLAM2 computer program. A highly conserved motif (Fig. 4B) was identified and then proved to be bound by HP1043 through hydroxyl-radical footprintings (Fig. 5). The conserved binding motif is composed of two direct repeats (one highly conserved repeat followed by a second less conserved repeat) separated by a 5-bp spacer conserved in length. The HP1043 binding motif appears to be located on the coding strand in almost all promoters analysed. The only exception is represented by *hp1043* (Fig. 4B), in which the conserved direct repeat maps on the non-coding DNA strand. In a previous study, a portion of *hp1043* promoter was used to characterize HP1043 binding to DNA through electrophoretic mobility shift assay carried out on wild-type and mutated probes[10]. In particular, deletions of the second less conserved repeat, as well as single base mutations of highly conserved A and G of the first repeat, significantly impaired protein binding to DNA, supporting the identified HP1043 consensus binding motif.

In vitro characterization of HP1043 binding to selected promoters (Figs 3 and 5) revealed the existence of high affinity and low affinity binding sites. However, the comparison of the conserved direct repeats of these targets

did not suggest any evident sequence feature responsible for the discriminative HP1043 binding capacity. Further characterization of the HP1043 DNA binding motif is crucial to address this point.

The above mentioned conserved spacing between the direct repeats puts the centre of the two direct repeats at 11 bp distance that corresponds to 1 helical turn, suggesting that a dimer of HP1043 recognizes the DNA on the same face of the double helix. The structure of HP1043 determined by NMR and X-ray crystallography[10] supports this hypothesis, revealing a symmetrical dimer with two functional domains: the regulatory (dimer-ization) domain and the DNA-binding (transactivation) domain. In particular, the DNA-binding domains in the dimer appeared to be spaced by a distance compatible with one helix turn, allowing the interaction of each DNA-binding domain with one repeat of the conserved motif. Considering that the HP1043 regulatory domains form a symmetric dimer and that they are connected to their respective DNA-binding domains through a short 2-residues linker, it is conceivable that HP1043 forms a symmetric dimer in a head-to-head orientation with DNA. Consequently, HP1043 should be expected to contact a binding site made by an inverted repeat. Surprisingly, our sequence conservation analysis led to the identification of a binding motif characterized by two direct repeats (Fig. 4B). This observation can partially be explained considering the different conservation of the two hemi-sites. Specifically, the HP1043 binding motif appears to be constituted by a first highly conserved hex-amer followed by a second less conserved repeat. The different degree of conservation between the two hemi-sites could account for a different specificity of DNA recognition of the two HP1043 monomers. A DNA recognition mechanism like this has been proposed for HpNikR, an *H. pylori* transcriptional regulator of nickel homeosta-sis[29]. It has been shown that a dimer of symmetrical dimers of NikR, expected to contact inverted repeats, has more affinity for binding sites that deviate from a perfect inverted repeat architecture[29]. Accordingly, it has been proposed that during the interaction between NikR and DNA target sequence, the more conserved hemi-site acts as a recognition site, while the second less conserved repeat acts as a structural (stabilizing) binding site[29]. The DNA recognition mechanism of HP1043 could be similar, with the primary recognition event taking place only on the highly conserved repeat, thereby stabilized by a weaker interaction between the second repeat and the other DNA binding domain. Another possibility is that, upon DNA recognition, the HP1043 dimer undergoes a structural reorganization, allowing a prototypical interaction between a direct repeat motif and a proper oriented dimer. Further experiments will disentangle this apparent paradox in HP1043-DNA docking mechanism.

Methods

Bacterial strains and culture conditions.
Bacterial strains used in this study are listed in Table 2. *H. pylori* G27 wild type cells were revitalized from glycerol stocks on Brucella broth agar plates containing 5% fetal calf serum (FCS) in a 9% CO_2–91% air atmosphere at 37 °C and 95% humidity in a water-jacketed incubator (Thermo Forma Scientific). Liquid cultures were performed in Brucella broth medium supplemented with 5% FCS in glass flasks. *E. coli* cells were grown on Luria–Bertani (LB) agar plates or LB liquid broth; when required, ampicillin was added to the medium to achieve a final concentration of 100 μg/ml.

DNA manipulations.
DNA manipulations were performed as described by Sambrook *et al.*[30]. All restriction and modification enzymes were used according to the manufacturers' instructions (New England Biolabs). Preparations of plasmid DNA were carried out with NucleoBond Xtra Midi plasmid purification kit (Macherey-Nagel).

Overexpression and purification of recombinant HP1043 protein.
Recombinant N-terminal His-tagged HP1043 protein was overexpressed in *E. coli* and affinity-purified as previously described[11]. Purified HP1043 was dialysed against two changes of 1 × 1043 Footprinting Buffer (1 × 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 DNA-binding assays or against two changes of 1X PBS (137 mM NaCl; 2.7 mM KCl; 10 mM NaH_2PO_4; 1.8 mM KH_2PO_4; pH 7.4) for animal immunization. Protein concentration was determined by Bradford colorimetric assay (Bio-Rad) and the purity of the protein preparations was analysed by SDS-PAGE.

Chromatin Immunoprecipitation (ChIP) with α-HP1043 polyclonal antibody.
Available α-HP1043 polyclonal antibody from immunized rabbits[12] were purified by 3 sequential precipitations with 35% saturated $(NH_4)_2SO_4$ and subsequent resuspension in water. *H. pylori* G27 wild type cultures were liquid-grown to an OD_{600} of 1.0, harvested, crosslinked, sonicated and immunoprecipitated as previously described[12]. Briefly, protein-DNA complexes were chemically crosslinked with 1% formaldehyde, then DNA was shared by extensive sonication with Bioruptor (Diagenode). HP1043 covalently linked to target DNA fragments were immunopre-cipitated by incubating the crosslinked whole cell extracts with the polyclonal α–HP1043 antibody and then cap-turing the resulting complexes with Protein-G-conjugated sepharose beads. Cross-linking was reverted for 6 h at 65 °C, with occasional mixing. DNA was extracted once with phenol-chloroform and further extracted with chlo-roform. After the addition of 1% glycogen, DNA was ethanol precipitated and suspended in 50 μl double-distilled H_2O, as previously described[12].

ChIP-seq library preparation and sequencing.
Illumina libraries were prepared, for each biological rep-licate either from 5 ng of immunoprecipitated-DNA (IPs) or from 5 ng of the Input-DNA following the Illumina TruSeq ChIP-seq DNA sample preparation protocol; then each library was sequenced on a MiSeq Illumina sequencer and 51 bp single stranded reads were produced.

ChIP-seq data analysis.
Bowtie 2 (v2.2.6)[31] was used to align raw reads, produced from ChIP sequencing experiments, to *H. pylori* G27 genome. End-to-end mapping was performed and non-deterministic option was specified to force a single assignment of multi-mapping reads to the best scoring region (if present) or a ran-dom attribution in the case of regions with identical scores. High quality reads were then selected requiring: for

Bacterial strains/plasmids	Description	Source/Reference
Strain		
H. pylori G27 wild type	Clinical isolate, wild type	42
E. coli DH-5α	*supE44 ΔlacU169 (φ80 lacZΔM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1*	43
Plasmid		
pGEM-T-Easy	Cloning vector; Amp^r.	Promega
pGEM0703	pGEM-T-Easy derivative, containing a 437 bp DNA fragment corresponding to the region from 722.774 to 723.211 of *H. pylori* G27 genome amplified by PCR with oligonucleotides 0703FPF/0703FPR. This region corresponds to a portion of the coding sequence of HPG27_RS03405 (HP0703 according to 26695 annotation).	This work
pGEMp1203	pGEM-T-Easy derivative, containing a 302 bp DNA fragment corresponding to the region from 1.268.274 to 1.268.576 of *H. pylori* G27 genome amplified by PCR with oligonucleotides 1203FPF/1203FPR. This region encompasses the putative promoter region of HPG27_RS06000 (HP1203 according to 26695 annotation).	This work
pGEMp1227	pGEM-T-Easy derivative, containing a 308 bp DNA fragment corresponding to the region from 1.290.538 to 1.290.846 of *H. pylori* G27 genome amplified by PCR with oligonucleotides 1227FPF/1227FPR. This region encompasses the putative promoter region of HPG27_RS06145 (HP1227 according to 26695 annotation).	This work
pGEMsRNA17_18	pGEM-T-Easy derivative, containing a 426 bp DNA fragment corresponding to the region from 16.304 to 16.730 of *H. pylori* G27 genome amplified by PCR with oligonucleotides sRNA17_18FPF/sRNA17_18FPR. This region encompasses the intergenic region between putative SRP_RNA and HPG27_RS00110.	This work
pGEMA1.4	pGEM-T-Easy derivative, containing a 199 bp DNA fragment corresponding to the region from 1.576.846 to 1.577.045 of *H. pylori* G27 genome amplified by PCR with oligonucleotides A1.4FPF/A1.4FPR. This region encompasses the putative promoter region of *isoA* toxin/antitoxin system.	This work
pGEMp0088	pGEM-T-Easy derivative, containing a 275 bp DNA fragment corresponding to the region from 94.285 to 94.560 of *H. pylori* G27 genome amplified by PCR with oligonucleotides 0088FPF/0088FPR. This region encompasses the putative promoter region of HPG27_RS00460 (HP0088 according to 26695 annotation).	This work
pGEMp1296	pGEM-T-Easy derivative, containing a 477 bp DNA fragment corresponding to the region from 1.362.205 to 1.362.682 of *H. pylori* G27 genome amplified by PCR with oligonucleotides 1296FPF/1296FPR. This region encompasses the putative promoter region of HPG27_RS06525 (HP1297 according to 26695 annotation).	This work
pGEMp1260	pGEM-T-Easy derivative, containing a 242 bp DNA fragment corresponding to the region from 1.321.773 to 1.322.015 of *H. pylori* G27 genome amplified by PCR with oligonucleotides 1260FPF/1260FPR. This region encompasses the putative promoter region of HPG27_RS06315 (HP1260 according to 26695 annotation).	This work
pGEM0703HY	pGEM-T-Easy derivative, containing a 117 bp DNA fragment corresponding to the region from 722.983 to 723.100 of *H. pylori* G27 genome amplified by PCR with oligonucleotides 0703HYF/0703HYR. This region corresponds to a portion of the coding sequence of HPG27_RS03405 (HP0703 according to 26695 annotation).	This work
pGEMp1203HY	pGEM-T-Easy derivative, containing a 148 bp DNA fragment corresponding to the region from 1.268.391 to 1.268.539 of *H. pylori* G27 genome amplified by PCR with oligonucleotides 1203HYF/1203HYR. This region encompasses the putative promoter region of HPG27_RS06000 (HP1203 according to 26695 annotation).	This work
pGEMp1227HY	pGEM-T-Easy derivative, containing a 146 bp DNA fragment corresponding to the region from 1.290.664 to 1.290.810 of *H. pylori* G27 genome amplified by PCR with oligonucleotides 1227HYF/1227HYR. This region encompasses the putative promoter region of HPG27_RS06145 (HP1227 according to 26695 annotation)	This work
pGEMp0088HY	pGEM-T-Easy derivative, containing a 136 bp DNA fragment corresponding to the region from 94.431 to 94.567 of *H. pylori* G27 genome amplified by PCR with oligonucleotides 0088HYF/0088HYR. This region encompasses the putative promoter region of HPG27_RS00460 (HP0088 according to 26695 annotation).	This work
pGEMpsRNA17_18HY	pGEM-T-Easy derivative, containing a 129 bp DNA fragment corresponding to the region from 16.516 to 16.645 of *H. pylori* G27 genome amplified by PCR with oligonucleotides sRNA17_18HYF/sRNA17_18HYR. This region encompasses the intergenic region between putative SRP_RNA and HPG27_RS00110.	This work
pGEMpA1.4HY	pGEM-T-Easy derivative, containing a 189 bp DNA fragment corresponding to the region from 1.576.888 to 1.577.077 of *H. pylori* G27 genome amplified by PCR with oligonucleotides A1.4HYF/A1.4HYR. This region encompasses the putative promoter region of *isoA* toxin/antitoxin system.	This work

Table 2. Bacterial strains and plasmids.

uniquely mapping reads MAPQ (mapping quality) greater than 30 and alignment score greater than −10 while for multi-mapping reads alignment score was set equal or greater than −10. The quality of ChIP-Seq data was evaluated following ENCODE quality metrics (https://code.google.com/archive/p/phantompeakqualtools/) and the numerical values obtained are provided in Supplementary Table S1. The cross-correlation analysis resulted in good NSC and RSC values, obtained using the code from Dr. Kundaje at Stanford University (https://code.google.com/archive/p/phantompeakqualtools/) cited by ENCODE consortium (http://genome.ucsc.edu/ENCODE/encodeTools.html). Moreover, we obtained average PBC scores. Irreproducible Discovery Rate procedure (IDR v 2.0.2) following ENCODE guidelines[32], and using Homer (v4.7.2)[33] as peak caller, was performed to measure sample reproducibility and to identify consistent peaks. Homer parameters were set according to the authors' indication for IDR calculation (-P. 1 −LP. 1 −poisson. 1), -L was set to three. The "findPeaks score" column was

selected as ranking column for IDR calculations, as suggested by Homer authors. The pool of two independent input preparations was used as background for the analysis, as suggested by IDR procedure. Peaks were manually annotated to the genes having transcription start site within $-150/+50$ bp from the middle of the peak (promoter peaks), to the genes containing the middle of peak (intragenic) or to the two genes surrounding the intergenic regions containing the peak (intergenic), according to the current version of *H. pylori* G27 RefSeq annotation (GCF_000021165.1). Transcription start sites were derived by blasting the 50 bp before the initiation of transcription found by Sharma *et al.*[16] for 26695 and controlling their coherence with our RNA-seq signals. A transcriptional start site was automatically cross-mapped when fragments matched with at least 90% identity and at least with 45/51 nucleotides in length. To further define peaks annotation, for the regions where HP1043 was bound, we considered also the fragments matching with an identity of at least 80% and an overlap of 35/51 nucleotides, if showing an increase of transcription (according to our RNAseq data) in correspondence to the cross-mapped transcriptional start site. To obtain COG[34] annotation for the putative protein coding targets the protein accession number of those genes were submitted to CDD[35] online database and COG alphanumeric code was converted to function according to the official COG classification. Raw data are publicly available at Sequence Reads Archive under accession number BioProject PRJNA327549.

DNase I footprinting. Genomic regions harbouring the putative HP1043 binding sites identified by ChIP-seq analysis were PCR-amplified with oligonucleotides listed in Supplementary Table S4 and cloned in pGEM-T-Easy vector (Table 2). DNA fragments obtained by digestion with the appropriate restriction enzymes were 5′-end labelled with $[\gamma\text{-}^{32}P]$-ATP and T4 polynucleotide kinase and gel purified. Approximately, 15 fmol of labelled probe were used for each footprinting reaction. DNase I footprinting experiments were performed as previously described[36]. Briefly, the labelled probes were incubated with increasing amounts of purified recombinant HP1043 protein in 1×1043 FPB containing 300 ng of sonicated salmon sperm DNA as non-specific competitor in a final volume of 50 μl for 20 min at room temperature. The partial digestion of the labelled probe was carried out using 0.01U of DNase I (Novagen) diluted in 1×1043 FPB supplemented with 5 mM $CaCl_2$ and 1 mM DTT for 90 s at room temperature. Reactions were stopped 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), phenol-chloroform extracted and ethanol precipitated. Samples were resuspended in 10 μl of Formamide Loading Buffer (FLB: 95% formamide; 10 mM EDTA; 0.02% bromophenol blue; 0.02% xylene cyanol), denatured at 100 °C for 3 min, fractionated on a 8 M urea-6% polyacrylamide sequencing gel and auto-radiographed.

Hydroxyl-radical footprinting. Genomic regions of interest were 5′-end labelled and purified as described above for DNase I footprintings. Hydroxyl-radical footprinting assays were performed as previously described[37]. Briefly, the labelled probes were incubated with increasing amounts of HP1043 protein in 1×1043 FPB_0 (10 mM Tris-HCl pH 7.5; 50 mM NaCl; 10 mM $MgCl_2$; 1 mM DTT; 0.01% Igepal CA-630) containing 300 ng of sonicated salmon sperm DNA as non-specific competitor in a final volume of 30 μl for 20 min at room temperature. Partial digestion of the labelled probe was performed by the simultaneous addition of 2 μl each of the following solutions: 2 μl of Fe:EDTA solution (125 mM Fe $(NH_4)_2$ $(SO_4)_2$; 250 mM EDTA pH 8.0), 2 μl of 0.1 M DTT and 2 μl 1% H_2O_2. After a 2-min incubation, cutting reaction was stopped with the addition of 25 μl of STOP solution (600 mM NaOAc pH 5.2; 100 ng/μl sonicated salmon sperm DNA; 4% glycerol), phenol-chloroform extracted and ethanol precipitated. Samples were resuspended in 6 μl of FLB, denatured at 100 °C for 3 min, fractionated on 8 M urea-8.4% polyacrylamide sequencing gel and auto-radiographed.

RNA isolation. *H. pylori* cultures were liquid-grown until mid-exponential phase ($OD_{600} = 0.8$) then treated with 0.5 μg/ml of tetracycline for 60 min. Bacterial cells were harvested and total RNA was extracted with TRI-reagent (Sigma-Aldrich), according to manufacturer's protocol.

RNA-seq: library preparation, sequencing and analyses. Ribosomal RNAs were depleted starting from 1 μg of total RNA from each of the conditions analysed by using the RiboZero Gram negative kit (Epicentre, Illumina). Strand specific RNA-seq libraries were prepared by using the ScriptSeq^{TM} v2 RNAseq library preparation kit (Epicentre, Illumina) starting from 50 ng of previously rRNA depleted RNA from each biological replicate and for all the conditions analysed. Then each library was sequenced on a MiSeq Illumina sequencer and 76 bp reads were produced. Bowtie 2 (v2.2.6)[31] was used to align raw reads to *H. pylori* G27 genome with the same parameters used for ChIP-seq data. High quality reads were selected requiring: for uniquely mapping reads MAPQ (mapping quality) greater than 30 and alignment score greater than -15; for multi-mapping reads alignment score was set equal or greater than -15 rRNA depletion, strand specificity and gene coverage were evaluated using BEDTools (v2.20.1*)[38] and SAMtools (v0.1.19)[39] to verify the library preparation and sequencing performances (see Supplementary Table S1). Strand specific reads overlapping to the genes annotated in *H. pylori* G27 RefSeq annotation (GCF_000021165.1) for at least 50% of their length were considered to produce the raw-counts of each sample. Only rRNAs were excluded as they were depleted during the library preparation procedure. The R package DESeq2 (v1.4.5)[40] was then used to normalize the counts and to individuate differentially expressed features showing BH adjusted p-value lower than 0.01. DESeq2 uses one of the most common strategy to normalize data after sequencing, which is the median-of-ratio method. In brief for each non variable gene the ratio of the expression level between samples is calculated, then the median ratio across all expressed genes is used as the normalization scale.

Raw data are publicly available at Sequence Reads Archive under accession number BioProject PRJNA327549.

qRT-PCR analysis. Synthesis of cDNA and qRT-PCR analysis were carried out as previously described[41]. Briefly, after removal of contaminating genomic DNA through DNase I digestion, 1 μg of DNA-free total RNA

was mixed with 50 ng of random hexamers (Invitrogen), dNTPs (1 mM each), AMV-Reverse Transcriptase (Promega) and incubated for 1 h at 37 °C for cDNA synthesis. For qRT-PCR analyses, 2 μl of the diluted (1:10) cDNA samples were mixed with 5 μl of 2X Power Up SYBR Green master Mix (ThermoFisher Scientific) and oligonucleotides specific for the genes of interest (Supplementary Table S4) at 400 nM concentration in a final volume of 10 μl. qRT-PCR was performed using the following cycling protocol: 95 °C for 2 min, then 40 cycles consisting of a denaturation for 5 s at 95 °C followed by 30 s at 60 °C (annealing and extension steps). For each real time experiment, the specificity of the reaction was checked by including a melting profile at the end of the run. Data were analysed using the $\Delta\Delta$Ct method, using the 16 S rRNA gene as internal reference. qRT-PCR of 16 S rRNA on samples from cells untreated or treated with tetracycline gave overlapping amplification curves, indicating that the amount of 16 S rRNA was not changed during the time-course experiment.

References

1. Salama, N. R., Hartung, M. L. & Muller, A. Life in the human stomach: persistence strategies of the bacterial pathogen Helicobacter pylori. *Nat Rev Micro* **11,** 385–399 (2013).
2. Gisbert, J. P. & Calvet, X. Review article: Common misconceptions in the management of Helicobacter pylori-associated gastric MALT-lymphoma. *Aliment. Pharmacol. Ther*. **34,** 1047–1062 (2011).
3. Scarlato, V., Delany, I., Spohn, G. & Beier, D. Regulation of transcription in Helicobacter pylori: simple systems or complex circuits? *Int. J. Med. Microbiol*. **291,** 107–17 (2001).
4. Foynes, S. *et al.* 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 (2000).
5. Waidner, B., Melchers, K., Stähler, F. N., Kist, M. & Bereswill, S. The Helicobacter pylori CrdRS two-component regulation system (HP1364/HP1365) is required for copper-mediated induction of the copper resistance determinant CrdA. *J. Bacteriol*. **187,** 4683–4688 (2005).
6. Niehus, E. *et al.* Genome-wide analysis of transcriptional hierarchy and feedback regulation in the flagellar system of Helicobacter pylori. *Mol. Microbiol*. **52,** 947–961 (2004).
7. Pflock, M. *et al.* Characterization of the ArsRS regulon of Helicobacter pylori, involved in acid adaptation. **188,** 3449–3462 (2006).
8. Schär, J., Sickmann, A. & Beier, D. Phosphorylation-independent activity of atypical response regulators of Helicobacter pylori. *J. Bacteriol*. **187,** 3100–3109 (2005).
9. Müller, S., Pflock, M., Schär, J., Kennard, S. & Beier, D. Regulation of expression of atypical orphan response regulators of Helicobacter pylori. *Microbiol. Res*. **162,** 1–14 (2007).
10. Hong, E. *et al.* Structure of an atypical orphan response regulator protein supports a new phosphorylation-independent regulatory mechanism. *J. Biol. Chem*. **282,** 20667–20675 (2007).
11. Delany, I., Spohn, G., Rappuoli, R. & Scarlato, V. 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 (2002).
12. Vannini, A., Roncarati, D. & Danielli, A. The cag-pathogenicity island encoded CncR1 sRNA oppositely modulates Helicobacter pylori motility and adhesion to host cells. *Cell. Mol. Life Sci*. 1–18, doi: 10.1007/s00018-016-2151-z (2016).
13. Bauer, S. *et al.* Novel function assignment to a member of the essential HP1043 response regulator family of epsilon-proteobacteria. *Microbiol.* (*United Kingdom*) **159,** 880–889 (2013).
14. Olekhnovich, I. N., Vitko, S., Valliere, M. & Hoffmana, P. S. Response to metronidazole and oxidative stress is mediated through homeostatic regulator hsra (hp1043) in Helicobacter pylori. *J. Bacteriol*. **196,** 729–739 (2014).
15. Olekhnovich, I. N. *et al.* Mutations to essential orphan response regulator HP1043 of Helicobacter pylori result in growth-stage regulatory defects pathogenesis. *Infect. Immun*. **81,** 1439–1449 (2013).
16. Sharma, C. M. *et al.* The primary transcriptome of the major human pathogen Helicobacter pylori. *Nature* **464,** 250–255 (2010).
17. Tatusov, R. L., Koonin, E. V. & Lipman, D. J. A Genomic Perspective on Protein Families. *Science (80-.)*. **278,** 631–637 (1997).
18. Beier, D., Spohn, G., Rappuoli, R. & Scarlato, V. Functional analysis of the Helicobacter pylori principal sigma subunit of RNA polymerase reveals that the spacer region is important for efficient transcription. *Mol. Microbiol*. **30,** 121–134 (1998).
19. Frith, M. C., Saunders, N. F. W., Kobe, B. & Bailey, T. L. Discovering sequence motifs with arbitrary insertions and deletions. *PLoS Comput. Biol*. **4** (2008).
20. Choprat, A. N. Tetracycline Analogs Whose Primary Target Is Not the Bacterial Ribosome. *Antimicrob Agents Chemother* **38,** 637–640 (1994).
21. Fisunov, G. Y., Evsyutina, D. V., Arzamasov, A. a., Butenko, I. O. & Govorun, V. M. Ribosomal profiling of Mycoplasma gallisepticum. *ActaNaturae* **7,** 107–112 (2015).
22. Danielli, A. & Scarlato, V. Regulatory circuits in Helicobacter pylori: Network motifs and regulators involved in metal-dependent responses. *FEMS Microbiology Reviews* **34,** 738–752 (2010).
23. Danielli, A. *et al. In vivo* dissection of the Helicobacter pylori fur regulatory circuit by genome-wide location analysis. *J. Bacteriol*. **188,** 4654–4662 (2006).
24. Grainger, D. C., Hurd, D., Harrison, M., Holdstock, J. & Busby, S. J. W. Studies of the distribution of Escherichia coli cAMP-receptor protein and RNA polymerase along the E. coli chromosome. *Proc. Natl. Acad. Sci. USA* **102,** 17693–17698 (2005).
25. Visweswariah, S. S. & Busby, S. J. W. 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 (2015).
26. Busby, S. & Ebright, R. H. Promoter structure, promoter recognition, and transcription activation in prokaryotes. *Cell* **79,** 743–6 (1994).
27. Browning, D. F. & Busby, S. J. W. Local and global regulation of transcription initiation in bacteria. *Nat. Rev. Microbiol*. **14,** 638–650 (2016).
28. Eder, S., Liu, W. & Hulett, F. M. Mutational analysis of the phoD promoter in Bacillus subtilis: Implications for PhoP binding and promoter activation of Pho regulon promoters. *J. Bacteriol*. **181,** 2017–2025 (1999).
29. Dosanjh, N. S., West, A. L. & Michel, S. L. J. Helicobacter pylori NikR's interaction with DNA: A two-tiered mode of recognition. *Biochemistry* **48,** 527–536 (2009).
30. Sambrook, J., Fritsch, E. & Maniatis, T. *Molecular Cloning: A Laboratory Manual.* (Cold Spring Harbor Laboratory Press, 1989).
31. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9,** 357–359 (2012).
32. Landt, S. *et al.* ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res*. **22,** 1813–1831 (2012).
33. Heinz, S. *et al.* Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol. Cell* **38,** 576–589 (2010).
34. Tatusov, R. L. *et al.* The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* **4,** 41 (2003).
35. Marchler-Bauer, A. *et al.* CDD: NCBI's conserved domain database. *Nucleic Acids Res*. **43,** D222–D226 (2015).
36. Roncarati, D., Danielli, A. & Scarlato, V. CbpA acts as a modulator of HspR repressor DNA binding activity in Helicobacter pylori. *J. Bacteriol*. **193,** 5629–5636 (2011).
37. Agriesti, F. *et al.* FeON-FeOFF: the Helicobacter pylori Fur regulator commutates iron-responsive transcription by discriminative readout of opposed DNA grooves. *Nucleic Acids Res*. **42,** 3138–3151 (2014).

38. Quinlan, A. R. & Hall, I. M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).
39. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
40. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** 550 (2014).
41. Pelliciari, S., Vannini, A., Roncarati, D. & Danielli, A. The allosteric behavior of Fur mediates oxidative stress signal transduction in Helicobacter pylori. *Front. Microbiol.* **6,** 1–10 (2015).
42. Xiang, Z. *et al.* Analysis of expression of CagA and VacA virulence factors in 43 strains of Helicobacter pylori reveals that clinical isolates can be divided into two major types and that CagA is not necessary for expression of the vacuolating cytotoxin. *Infect. Immun.* **63,** 94–98 (1995).
43. Hanahan, D. Studies on transformation of Escherichia coli with plasmids. *J Mol Biol* **166,** 557–580 (1983).

Acknowledgements

Author Contributions

The experiments were conceived and designed by A. Danielli, C. Peano, D. Roncarati, V. Scarlato and performed by S. Pelliciari, E. Pinatel, C. Peano, A. Vannini. Data analysis was carried out by S. Pelliciari, E. Pinatel, A. Vannini, S. Puccio. The paper was written by D. Roncarati and V. Scarlato with contributions by C. Peano, G. De Bellis,. E. Pinatel. All authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Pelliciari, S. *et al.* 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,** 41063; doi: 10.1038/srep41063 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.