UNIVERSITÀ DEGLI STUDI DI PARMA

Department of Pharmacy
Laboratories of Biochemistry and Molecular Biology
PhD Program in
Biochemistry and Molecular Biology

XXVIII cycle

Identification and characterization of periplasmic proteins implicated in the adaptation of *Helicobacter pylori* to the human gastric niche

Coordinator:
  Prof. Andrea Mozzarelli
Tutor:
  Prof. Riccardo Percudani

PhD student:
Giulia Mori

2013-2015

CONTENTS

Summary

Helicobacter pylori, a Gram-negative, microaerophilic bacterium that colonizes the stomach of half the world population, is the main cause of chronic gastritis, peptic ulceration, gastric lymphoma, and gastric adenocarcinoma. *H. pylori* has been coevolving with humans for at least 50,000 years, colonizing the stomach in childhood and persisting throughout life, in the absence of antibiotic treatment. This implies near perfect adaptation to the niche and the ability to evade the host immune response.

The aim of this thesis was to identify and characterize important factors for the colonization and adaptation of *H. pylori* to the gastric environment. In particular, I present the results of the study of two periplasmic proteins: the first one involved in the bacterium antioxidant defense (*Hp* catalase-like, HP0485), the second one with a role in nutrient uptake from the stomach milieu (*Hp* dipeptide binding protein, HP0298).

Chapter one contains a general introduction on the bacterium, chapters two and three describe the studies of the two proteins and both include an abstract, introduction, results, discussion, materials and methods.

H. pylori catalase-like (HP0485) is a periplasmic protein, with a monomeric structure, belonging to a family of enzymes structurally related to catalase, but of undefined function. Despite the conservation of the catalase fold and the heme cofactor, HP0485 is not capable of catalyzing the dismutation of H_2O_2 (*catalatic* reaction), but a broad spectrum *peroxidatic* reaction, by coupling the hydrogen peroxide reduction with the oxidation of various one-electron donor substrates. Similarly to catalase, *Hp* catalase-like works at high H_2O_2 concentrations, being

not saturated up to elevated H_2O_2 levels (200 mM). The reaction rate increases exponentially and the enzyme does not loose activity even at these high values of the peroxide, at variance with peroxidases that are usually inactivated by H_2O_2 concentrations higher than 10-50 mM. This catalytic versatility and robustness suggests that Hp catalase-like has a role in H_2O_2 scavenging, and probably another function linked to the oxidation of a reduced substrate in the cell periplasm. In addition to the characterization of enzymatic activity, we propose a model for the heme assembly mechanism in the periplasm that involves the formation of a disulfide bond, identified as a hallmark of secreted catalase-like proteins.

The dipeptide binding protein of *H. pylori* (*Hp* DppA, HP0298) is the periplasmic component of an ABC importer system that appears to be specific of gastric species. Periplasmic binding proteins (PBPs) exhibit ligand promiscuity, but a very well conserved binding domain, that is found in many mammalian and human receptors, including for example, metabotropic glutamate and GABA receptors. To investigate *Hp* DppA physiologic substrate, we set up a protocol of "ligand-fishing" coupled with mass spectrometry. The His-tagged purified protein acts as the bait to capture its specific ligand within the endogenous environment in which the binding occurs, that is *H. pylori* cell extract. The protein-ligand complex is then purified through affinity chromatography and analyzed by HPLC-MS. The compounds that potentially interact with DppA were pentapeptides, rich in hydrophobic amino acids and containing at least one negative charged residue. Since *H. pylori* is auxotrophic for some amino acids, mostly hydrophobic, and the human stomach is rich of peptides produced by food protein digestion, *Hp* DppA could have a role in the uptake of peptides with specific length and amino acid composition that are naturally present in the gastric niche.

Riassunto

Helicobacter pylori è un batterio Gram-negativo in grado di colonizzare la mucosa gastrica umana e persistere per l'intero arco della vita dell'ospite. E' associato a patologie gastrointestinali, quali gastrite cronica, ulcere gastriche e duodenali, adenocarcinomi e linfomi gastrici. Si tratta di uno dei patogeni più diffusi, presente in circa metà della popolazione mondiale, e il solo che si è adattato a vivere nell'ambiente ostile dello stomaco umano.

Molteplici sono i fattori di virulenza che permettono al batterio la colonizzazione della nicchia gastrica e contribuiscono, anche attraverso l' induzione di una risposta infiammatoria, a profonde modificazioni dell' omeostasi gastrica. Queste ultime si associano, ad esempio, all'iperproduzione di fattori proinfiammatori, ad alterazioni sia della regolazione della secrezione acida gastrica sia del ciclo cellulare e della morte cellulare programmata (apoptosi) delle cellule epiteliali gastriche, a disordini nel metabolismo del ferro e a carenze di elementi essenziali.

Studi sulla diversità genetica di *H. pylori* osservata in ceppi isolati da varie regioni del mondo, dimostrano che tale batterio ha avuto una coevoluzione col genere umano attraverso la storia, ed è verosimile che *H. pylori* sia stato un costituente del microbiota gastrico per almeno 50.000 anni.

Scopo della tesi è stato quello di identificare e caratterizzare proteine importanti per la colonizzazione e l'adattamento di *H. pylori* alla nicchia

gastrica. In particolare gli sforzi si sono concentrati su due proteine periplasmatiche, la prima coinvolta nella difesa antiossidante (l'enzima catalasi-*like*, HP0485), e la seconda nel trasporto di nutrienti presenti nell'ambiente dello stomaco all'interno della cellula (la componente solubile di un ABC *transporter*, HP0298).

La strategia utilizzata prevede un'analisi bioinformatica preliminare, l'ottenimento del gene per amplificazione, mediante PCR, dal genoma dell'organismo, la costruzione di un vettore per il clonaggio, l'espressione eterologa in *E. coli* e la successiva purificazione. La proteina così ottenuta viene caratterizzata mediante diverse tecniche, quali spettroscopia UV, dicroismo circolare, gel filtrazione analitica, spettrometria di massa.

Il capitolo 1 contiene un'introduzione generale sul batterio, il capitolo 2 e il capitolo 3 descrivono gli studi relativi alle due proteine e sono entrambi suddivisi in un abstract iniziale, un'introduzione, la presentazione dei risultati, la discussione di questi ultimi, i materiali e i metodi utilizzati.

La catalasi-like (HP0485) è una proteina periplasmatica con struttura monomerica, appartenente ad una famiglia di enzimi a funzione per la maggior parte sconosciuta, ma evolutivamente correlati alla ben nota catalasi, attore fondamentale nella difesa di *H. pylori*, grazie alla sua azione specifica di rimozione dell'acqua ossigenata.

HP0485, pur conservando il fold catalasico e il legame al cofattore eme, non può compiere la reazione di dismutazione dell'acqua ossigenata; possiede invece un'attività perossidasica ad ampio spettro, essendo in grado di accoppiare la riduzione del perossido di idrogeno all'ossidazione di diversi substrati. Come la catalasi, lavora ad alte concentrazioni di aqua ossigenata e non arriva a saturazione a concentrazioni molto elevate di

questo substrato (200 mM); la velocità di reazione catalizzata rimane lineare anche a questi valori, aspetto che la differenzia dalle perossidasi che vengono in genere inattivate da concentrazioni di perossido di idrogeno superiori a 10-50 mM. Queste caratteristiche di versatilità e robustezza suggeriscono che la catalasi-*like* abbia un ruolo di *scavenger* dell'acqua ossigenata e probabilmente anche un'altra funzione connessa al suo secondo substrato, ossia l'ossidazione di composti nello spazio periplasmatico cellulare. Oltre alla caratterizzazione dell'attività è descritta anche la presenza di un ponte disolfuro, conservato nelle catalasi-like periplasmatiche, con un ruolo nell'assemblaggio dell'eme per ottenere un enzima attivo e funzionale.

La proteina periplasmatica HP0298, componente di un sistema di trasporto ABC, è classificata come trasportatore di dipeptidi e appartiene a una famiglia di proteine in grado di legare diversi substrati, tra cui di- e oligopeptidi, nichel, eme, glutatione. Benchè tutte associate a trasportatori di membrana batterici, queste proteine presentano un dominio di legame al substrato che risulta essere conservato nei domini extracellulari di recettori specifici di mammifero e uomo. Un esempio sono i recettori ionotropici e metabotropici del sistema nervoso. Per caratterizzare questa proteina è stato messo a punto un protocollo di *ligand-fishing* accoppiato alla spettrometria di massa. La proteina purificata, avente un *tag* di istidine, è stata incubata con un estratto cellulare di *H. pylori* per poter interagire con il suo substrato specifico all'interno dell'ambiente naturale in cui avviene il legame. Il complesso proteina-ligando è stato poi purificato per cromatografia di affinità e analizzato mediante HPLC-MS. L'identificazione dei picchi differenziali tra campioni con la proteina e

campioni di controllo ha portato alla caratterizzazione di pentapeptidi particolarmente ricchi in aminoacidi idrofobici e con almeno un residuo carico negativamente. Considerando che *H. pylori* necessita di alcuni aminoacidi essenziali, per la maggior parte idrofobici, e che lo stomaco umano è particolarmente ricco di peptidi prodotti dalla digestione delle proteine introdotte con il cibo, il ruolo fisiologico di HP0298 potrebbe essere l'internalizzazione di peptidi, con caratteristiche specifiche di lunghezza e composizione, che sono naturalmente presenti nella nicchia gastrica.

Chapter 1

General Introduction

Helicobacter pylori, a Gram-negative bacterium, persistently colonizes half the world's population and is associated with a variety of upper gastrointestinal disorders, including chronic gastritis, stomach and duodenal ulcers, adenocarcinomas and stomach lymphomas[1,2].

H. pylori is the sole pathogen able to colonize the human stomach, an ecological niche characterized by very acidic pH, hostile for most microbes. Since its original description in 1984[3], *H. pylori* has been extensively studied and many related species have been found to colonize the gastrointestinal tract of humans and other animals[4].

What makes *Helicobacter pylori* a successful and life-lasting colonizer of the human stomach is undoubtedly its ability to adapt and coevolve with the host; its early acquisition by mankind, long before the migrations of modern humans out of Africa[5] is certainly a proof. Another one is the plethora of factors the bacterium is equipped with, addressing the different challenges presented by the harsh environment. Among these factors urease, an enzyme that hydrolyses gastric juice urea into ammonia and carbon dioxide, plays a central role. *H. pylori* expresses urease at a level higher than that of any known microrganism[6], and uses the products of the urease reaction to buffer its periplasm, allowing maintenance of the cytoplasmic membrane potential[7]. This is a unique acid-acclimation strategy by which the bacterium faces acidity that encounters within the gastric lumen upon ingestion and periodically during the course of infection. Another key aspect of colonization is motility. Because of its neutralophilic nature *H. pylori* needs to leave the lumen and swim with its polar-sheathed flagella towards its preferred destination represented by the thick mucus layer that covers the epithelial lining of the stomach mucosa[8]. This movement is

guided by the presence of chemical attractants, such as urea and bicarbonate ions. *H. pylori* then binds to Lewis antigens present on host gastric cells through several types of adhesins, and secretes factors that stimulate inflammatory cells, as well as the multifunctional toxin VacA[9]. Furthermore, the presence of the cag pathogenicity island, a 40-kb DNA that encodes a type IV secretion system, seems to be necessary for optimal fitness of the bacterium and the appearance of pathogenic traits[10].

Most *H. pylori* cells are free-living and highly motile in the mucus layer, and only a small proportion (approximately 1–5%) attach to epithelial cells. The existence of two populations (adherent and non-adherent) with different survival characteristics, and the likelihood that there is interaction between these two populations, indicate that the mechanisms and features of *H. pylori* infection are multifaceted and complex[11].

Probably the most astonishing and paradoxical aspect of *H. pylori* infection is its persistence in the gastric niche in the presence of a host response. But, being a well-adapted microbe, *H. pylori* has evolved ways not only to circumvent these mechanisms, but also to utilize host responses to its own advantage[12]. The depletion of nutrients from the host, for example, is interconnected with the adaptively induced inflammatory response in the human stomach. Bacteria elaborate pro-inflammatory effectors provoking host responses leading to tissue damage with consequent nutrient release. Inflammation, while advantageous to the host for microbes that can be eliminated, may be deleterious when infection cannot be eradicated, since it leads to impairment of tissue structure and function. Thus, the ability of hosts to curtail inflammatory responses during persistent infections may be adaptive[13]. In the same way long-term, uncontrolled inflammation may be deleterious for *H. pylori*, since its niche would be lost[14]; thus it may also be adaptive for *H. pylori* to down-regulate inflammation[15]. Consistent with this view is the observation that important *H. pylori* surface molecules, such as the lipopolysaccharide, have low pro-inflammatory activity[16].

These regulated interactions of *H. pylori* with the host, incorporating positive and negative feedback systems, provide an exquisite example of mutual pathogen-host adaptation.

To deeply understand the origin and emergence of *Helicobacter pylori* as a human pathogen, it is essential to analyse its genome content and architecture and use the potent tool of comparative genomics. *H. pylori* possesses a relatively small genome, consisting of one 1,6-Mb circular chromosome. Such a feature is shared by many other pathogenic and non-pathogenic microorganisms of the ε-proteobacteria lineage that are strictly host-adapted, as their genomes have undergone a process known as reductive evolution, leading to reduced genome sizes[17]. The recent sequencing of numerous *H. pylori* strains[18], different gastric and enterohepatic Helicobacter species[19], and related ε-proteobacteria genera[20], enables the reconstruction of the ε-proteobacteria phylogenesis. Such an evolutionary context provides useful information on gains and losses of genes and lead to the identification of both shared and species-specific genes, making possible to describe similarities and differences in the lifestyle of these organisms, as well as common and divergent mechanisms for host adaptation. The genes that are present in all species are likely to be essential for maintenance of the microorganism in a mammalian or vertebrate host, thereby providing a general strategy for survival and growth, as well as suggesting common mechanisms for spread and transmission into larger host populations. On the other hand the identification of the genes that are unique to a species is an excellent starting point for functional analysis to gain insight on a specific microbe-host niche interaction.

Considering species of which a highly resolved whole-genome phylogeny is available we built an ultrametric tree describing the evolution of ε-proteobacteria (Fig. 1)[21]. The resulting pattern highlights the deep divergence of

Campylobacterales[22], which infect the intestine of human and other animals, the subsequent emergence of Helicobacteriaceae that colonize also the stomach, and the recent adaptation to the human gastric niche. To note, the very recent divergence of *H. pylori* and *H. acinonychis* (large feline pathogen) which is thought to have occurred <0.4 My ago via host jump from early human populations[23].

In an effort to identify and select interesting candidates among *H. pylori* still uncharacterized proteins, we focused both on genes specifically associated with the adaptation to the gastric niche and on genes more widespread among ε-proteobacteria. Accordingly, the studies described in this thesis concentrate on two periplasmic proteins with different distributions among ε-proteobacteria genomes. The dipeptide binding protein (DppA), part of an ABC importer system, appears to be specific of gastric species. The catalase-like protein is found in many *Helicobacter* species and in some Campylobacterales (Fig. 1).
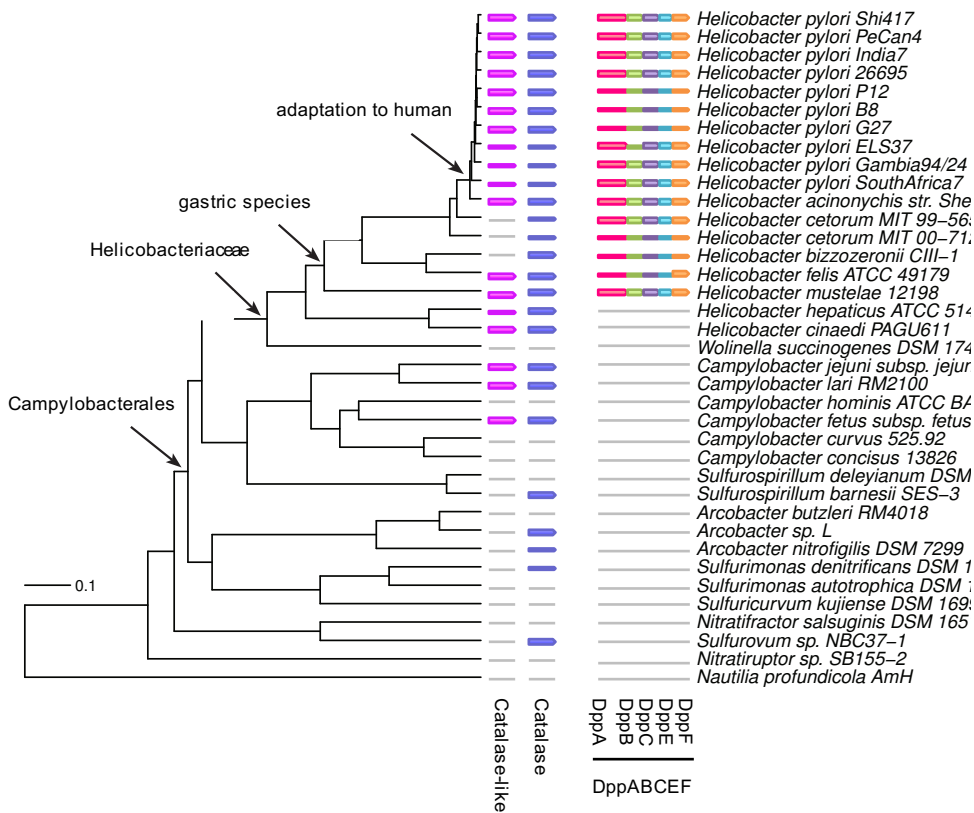
Figure 1: **Distribution of catalase-like and DppA with related genes in Epsilonproteobacteria genomes**. The ultrametric tree represents a whole-genome phylogeny of Epsilonproteobacteria. The presence or absence of genes is shown alongside the tree. A coloured pentagon indicates a gene identified in the bacterium genome; a gray line indicates a gene not identified in the bacterium genome. Catalase-like and DppA are the objects of study. DppA together with the whole ABC transporter are specific of gastric species. Catalase-like is more widespread in Epsilonproteobacteria, but to a lesser extent than catalase.

Aim of the study

The aim of this thesis was to identify and characterize proteins involved in the colonization and adaptation of *H. pylori* to the human gastric niche. Two aspects of the bacterium are highlighted: the potent antioxidant defence, which guarantees its survival in spite of a continuous inflammatory response activated by the host immune system, and the acquisition of nutrients from the stomach resulting in successful replication and persistence. I concentrated on two uncharacterized proteins, each of which is associated with one of the two aspects.

The first protein, *Hp* catalase-like, is a periplasmic enzyme linked to the antioxidant defence of the bacterium. To characterize its enzymatic activity, I compared it with *H. pylori* catalase and eukaryotic horseradish peroxidase, discovering some properties that place this *Hp* protein halfway between a peroxidase and a catalase. The mechanism of heme assembly, involving a conserved disulfide bond, was also investigated.

The second protein, *Hp* dipeptide binding protein (DppA), is the periplasmic component of an ABC importer system, implicated in the uptake of nutrient found in the gastric environment. To identify the specific ligand I used a strategy entailing ligand pull-down assay and mass spectrometry analysis. Such an approach could be used as initial interaction screening for the identification of putative ligands of other periplasmic binding proteins (PBPs) and related receptors bearing the PBP domain.

Chapter 2

Helicobacter-pylori catalase-like protein is a versatile and robust peroxidase

Abstract

In an environment in which hydrogen peroxide is both an internal and an external threat, bacterial pathogens require multiple H_2O_2 scavenging enzymes. *Helicobacter pylori*, an oxygen-sensitive microaerophilic bacterium, unique in its ability to colonize the human stomach, possesses many oxidative stress resistance factors to counteract the massive host inflammatory response, but only one, the ubiquitous catalase (KatA), is known to be specifically dedicated to H_2O_2 removal. The *H. pylori* (*Hp*) genome encodes a catalase-like protein (HP0485) that belongs to a family of proteins structurally related to catalase, but of undefined function. Despite its monomeric structure, and the weak sequence identity to catalase, the predicted three-dimensional structure reveals conservation of the catalase fold, and the heme active site. Nevertheless, our results show that *Hp* catalase-like is not capable of performing the *catalatic* reaction. The enzyme exhibits instead a broad-spectrum *peroxidatic* activity, rivaling that of typical peroxidases (i.e. horseradish peroxidase), with which the *Hp* protein shares no structural relation. At variance with peroxidases, but similarly to catalases, HP0485 shows a non-saturable interaction with H_2O_2 and can withstand high concentration of the peroxide without loosing activity. These features make *Hp* catalase-like a suitable H_2O_2 scavenger at elevated H_2O_2 concentration. The *Hp* catalase-like protein is secreted and assembled with the heme cofactor in the periplasm. The heme assembly mechanism involves the formation of a disulfide bond as a distinctive mark of catalase-like proteins. Our work unveils a role for *Hp* catalase-like in *H. pylori* antioxidant defence.

Introduction

Antioxidant systems are essential to bacterial pathogens. Their living at the expense of a host organism is an endless struggle against reactive oxygen species (ROS), generated not only intracellularly as by-products of aerobic metabolism, but also in the outside environment as weapons of the defence response activated in host tissue.

The human gastric pathogen *Helicobacter pylori*, because of its microaerophilic nature is highly vulnerable to oxygen toxicity, yet has the extraordinary ability to persists throughout the life of the host in spite of an aggressive inflammatory response[24]. A significant increase in the levels of ROS, such as superoxide anion ($O_2^{-\cdot}$), hydrogen peroxide (H_2O_2), and hydroxyl radical ($OH^\cdot$), has been measured in the stomach mucosa of *H. pylori* infected patients[25] as a result of the inflammatory reaction of gastric cells[26] and the oxidative burst of macrophages and neutrophils infiltrating the gastric epithelium[27]. This partially explains the inflammation-induced tissue damage (gastritis) and ulcer formation that is associated with persistent mucosal infection by *H. pylori*.

The ability of the bacterium to survive in the presence of high concentrations of oxygen radicals strongly depends on a battery of diverse antioxidant factors and enzymes[28]. Among these, hydrogen peroxide scavengers are key defences against oxidative stress[29]. Whereas superoxide (degraded by superoxide dismutase to hydrogen peroxide) and hydroxyl radical are highly reactive, and have thus a limited radius of action, hydrogen peroxide can easily penetrate cellular structures (cell walls, plasma membranes, etc.) and oxidize solvent-exposed iron-sulfur clusters and sulfhydryl groups, leading to inactivation of enzymes and other proteins[30]. Additionally, its decomposition product hydroxyl radical, formed in the Fenton reaction between H_2O_2 and metal ions, can react

with DNA, proteins, and lipids with potentially lethal consequences[31].

Rapid and efficient removal of H_2O_2 is therefore of essential importance for cell viability, and hydroperoxidases (catalases and peroxidases) are ubiquitous "housekeeping" enzymes well known to fulfil this need[32,33]. These oxidoreductases are capable of heterolytically disrupting the peroxidic bond of hydrogen peroxide (H-O-O-H), and also of some small organic peroxides (R-O-O-H).

Catalases convert two molecules of hydrogen peroxide to water and molecular oxygen, thus degrade hydrogen peroxide by dismutation (*catalatic* reaction):

$$2\ H_2O_2 \longrightarrow 2\ H_2O + O_2$$
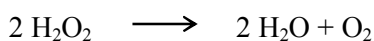
Peroxidases reduce H_2O_2 and use it for substrate oxidation (*peroxidatic* reaction):

$$H_2O_2 + 2\ AH_2 \longrightarrow 2\ H_2O + 2\ ^\bullet AH$$
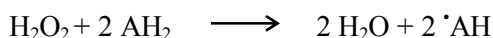
The first step of the enzymatic cycle of both heme catalases and peroxidases (Fig. 1) is in common: a molecule of H_2O_2 oxidizes the heme iron (Por Fe^{III}, where Por stands for the heme porphyrin) to an oxyferryl intermediate, named Compound I[34]. The oxygen–oxygen bond of H_2O_2 (H-O-O-H) is cleaved heterolytically, with one oxygen leaving as water and the other remaining at heme iron. Two electrons are removed from the heme in this process, one from the iron atom that becomes Fe^{IV}, and the second from the porphyrin, which acquires cation radical character ($^{\bullet+}$Por Fe^{IV}=O). However, the reactivity of this intermediate differs between catalases and peroxidases. In a catalatic cycle Compound I is reduced back to the ferric enzyme by a second molecule of hydrogen peroxide, with the axial oxygen atom becoming water and both oxygen atoms of the peroxide released as molecular oxygen. Thus, *catalatic* reaction occurs in two steps with one substrate, H_2O_2, acting both as oxidant and as reductant (both as electron acceptor and electron donor).

In contrast, peroxidases catalyse Compound I reduction by one-electron
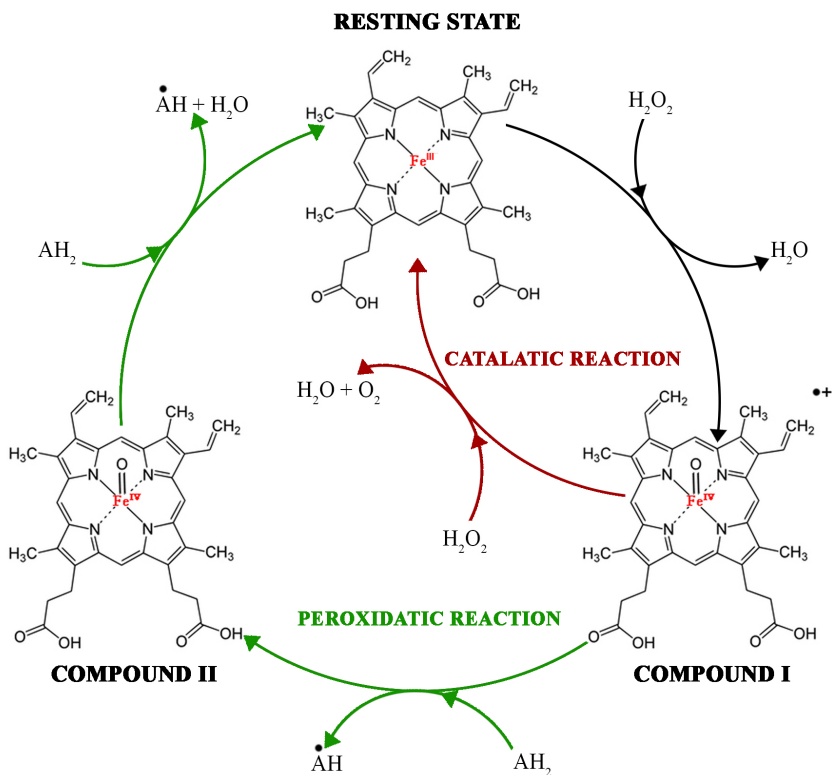
Figure 1. **Reactions catalyzed by peroxidases and catalases using hydrogen peroxide (H₂O₂)**. The catalytic cycle begins with Compound I formation for both oxidoreductase enzymes. In catalases, Compound I is reduced back to the resting state by a second molecule of H_2O_2 (*catalatic* reaction). In contrast, peroxidases preferentially recover the resting state via the *peroxidatic* reaction, in which Compound I sequentially reacts with two molecules of a one-electron donating substrate (AH_2), passing through another intermediate named Compound II.

donating substrates (AH$_2$), typically phenols or aromatic amines, to recover the native ferric state via another intermediate named Compound II. In this process, the porphyrin accepts one electron, thus loosing its radical character and becoming an oxoiron^{IV} or hydroxoiron^{IV} species (Por Fe^{IV}=O or Por Fe^{IV}−OH). The oxygen bound to iron is concomitantly protonated to give a molecule of water that dissociates from the heme to regenerate the resting ferric enzyme. Thus, a *peroxidatic* cycle includes three reactions (compound I formation, compound I reduction to compound II and compound II reduction back to the resting state) and two substrates, H$_2$O$_2$ acts only as oxidant (electron acceptor), whereas the electron-rich cosubstrate acts as reductant (electron donor). In the majority of the cases the porphyrin prosthetic group is heme *b* (with some exception of heme *d*) and its highly conserved amino acids surroundings, in particular the distal side, that differ between catalases and peroxidases, influence its reactivity and direct the reaction towards one of these pathways[35].

In the last thirty-five years, since when the first crystal structures of catalases were published[36,37], the mechanism of the *catalatic* reaction accomplished within such a strongly preserved active site has been extensively investigated. The comparison with *peroxidatic* reaction, the other side of the moon, has always been fascinating and helped to define the crucial residues directing towards one of the two pathways. Strong importance was attributed to the proximal heme ligands, tyrosine in catalases and histidine in peroxidases. Nevertheless, recent experimental and theoretical studies[38–40] have shown that the distinct reactivity of Compound I towards H$_2$O$_2$ in these enzymes cannot be explained in terms of differences in the nature of the proximal ligands, because of its little effect on the energy barrier of the oxidation of H$_2$O$_2$ (the second step of the *catalatic* reaction, Fig. 1). On the contrary, the O-O distance between Compound I and hydrogen peroxide affects significantly the reaction barrier, implying a major role for the distal side residues in determining the different behaviours. In catalases distal side properties and architecture allow to attain very short
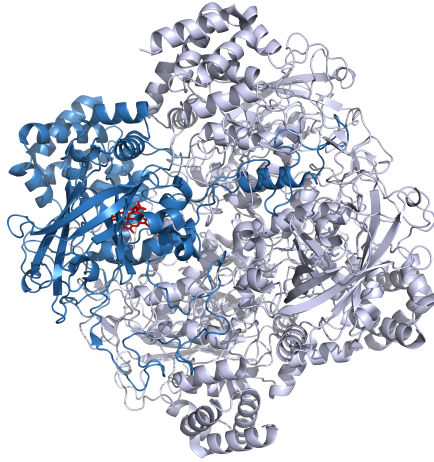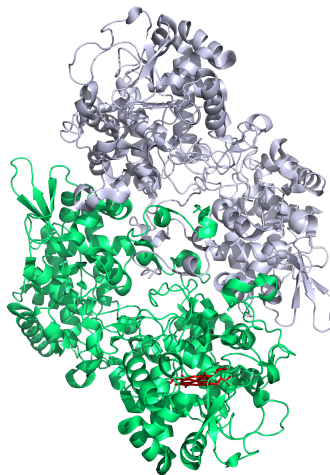
Figure 2. **Catalase and catalase-peroxidase structures**. (A) Tetrameric structure of catalase from *Helicobacter pylori* (PDB: 2A9E), coloured gray with a monomer evidenced in blue; heme is in red. (B) Dimeric structure of catalase-peroxidase from *Mycobacterium avium* (PDB: 1SJ2), coloured gray with a monomer evidenced in green; heme is in red.

distances to undergo the *catalatic* reaction, whereas in peroxidases a different distal environment is responsible for lengthier distances, and the reaction proceeds via the peroxidatic pathway.

Many bacteria rely on two unrelated heme catalases for H_2O_2 detoxification, a typical or "monofunctional" catalase and a "bifunctional" catalase–peroxidase (Fig. 2)[41]. They share high *catalatic* activity, but have significant differences in sequence, active site, tertiary, and quaternary structures, reflecting an independent origin. Catalase–peroxidases, besides their *catalatic* property, exhibit a *peroxidatic* activity similar to that of eukaryotic heme peroxidases, with which they share a common origin[42]. Catalases can also undergo a *peroxidatic* reaction involving organic reductants, similar to peroxidases; however, this activity is weak and usually restricted to small organic substrates, such as alcohols, nitrite and formate, because of limited accessibility to the active site[43].

Helicobacter pylori possesses only one catalase (KatA), a monofunctional catalase without peroxidase activity[44]. KatA differs from other catalases for some unique properties. It is very stable at high concentration of H_2O_2 (up to 500 mM), but is uniquely sensitive to organic hydroperoxides, due to perturbation of the heme environment and formation of a catalytically incompetent compound II species[45]. KatA is located both in the cytoplasm and in the periplasm[44]. *H. pylori* expresses another protein capable of degrading hydrogen peroxide, a cytosolic alkyl hydroperoxide reductase (AhpC)[46]. AhpC belongs to the peroxiredoxin family, which comprises enzymes lacking prosthetic groups with the ability to catalyse the reduction of hydrogen peroxide, peroxynitrite, and a wide range of organic hydroperoxides. Even if *H. pylori* AhpC has similar rate constant in reducing H_2O_2 and organic hydroperoxides (e.g. t-butyl hydroperoxide), its major role under physiological conditions is considered to be the detoxification of organic hydroperoxides, as demonstrated by the accumulation of lipid

hydroperoxides in ahpC mutant cells[45].

With any number of H_2O_2 scavenging enzymes in bacterial world[29], *H. pylori* possesses only one enzyme specifically dedicated to H_2O_2 removal, and even if its genome contains a periplasmic cytochrome *c* peroxidase[47], to date no peroxidatic activity has been demonstrated for any *H. pylori* proteins.

We found in *Hp* genome the presence of a catalase-like protein (HP0485), belonging to a poorly known family of proteins structurally related to catalases, of undefined function. Proteins in this family have a reduced size in respect to catalases, being no longer than 360 amino acids, and for this reason they are called "minicatalases". To our knowledge the catalase-related protein of *Mycobacterium avium* ssp. *paratuberculosis*, annotated as MAP-2744c, is the only member characterized for its function and crystal structure[48]. Even though these putative monomeric catalase-related proteins are found exclusively in prokaryotes, they do have a eukaryotic counterpart represented by the coral allene oxide synthase (cAOS). This heme peroxidase specialized for metabolism of a fatty acid hydroperoxide occurs in sea whip coral *Plexaura homomalla* as part of a fusion protein, the other component of which is a lipoxygenase that forms the hydroperoxide substrate[49]. cAOS and MAP-2744c are the only other reported structures presenting the catalase fold besides monofunctional catalases[48,50]. Their characterized catalytic activities reflect two different achievements within a catalase-related structure. *M. avium* enzyme, despite a predominant *peroxidatic* reaction with organic hydroperoxides, exhibits *catalatic* activity, whereas cAOS, being not capable of hydrogen peroxide dismutation, represents a complete switch from a H_2O_2-metabolizing catalase to a biosynthetic fatty acid hydroperoxidase.

To assess the potential role of the catalase-like in the bacterium antioxidant defence, we characterized the enzymatic activity of the *H. pylori* protein. We describe its peroxidatic activity and demonstrate that the assembly of fully active

form of this heme enzyme in the periplasm depends on a conserved disulfide bond.

Results

Helicobacter pylori encodes a protein with catalase fold

A catalase-like protein is present in the *Helicobacter pylori* 26695 genome and annotated as HP0485. Sequence alignment with *H. pylori* catalase (*Hp* KatA) reveals 22% of identity over ≈ 300 residues (Fig. 3). *Hp* catalase-like sequence contains 314 residues and that of *Hp* KatA contains 505 residues. The first amino acids of *Hp* catalase-like (up to His20) correspond to the signal peptide (see below) and were not included in the alignment. The region matching in the two proteins begins at residue 31 of *Hp* KatA and terminates at the end of HP0485 sequence. Despite of the low sequence identity between the two proteins, *Hp* catalase-like aligns perfectly with many essential, conserved residues of catalase heme active site, including the proximal heme-ligand, tyrosine (Tyr339 in KatA, Tyr306 in HP0485), and the distal catalytic histidine (His56 in KatA, His45 in HP0485)[34,35,51].

A 3D model was built for *Hp* catalase-like protein, using the structure of *Hp* KatA as template[52]. The QMEAN4[53] of the model is -6.48, probably due to the low level of identity between the two sequences. The predicted HP0485 structure was superimposed on a monomer of KatA revealing a striking similarity (Fig. 4). *Hp* catalase-like is essentially a "catalase core", in which the antiparallel eight-stranded β-barrel and the majority of the helical domain, both defining the heme binding pocket are maintained, whereas the first residues of the N-terminal arm and the extended wrapping loop, parts of the catalase structure important for tetramerization[54], are missing, along with the C-terminal α-helical domain.
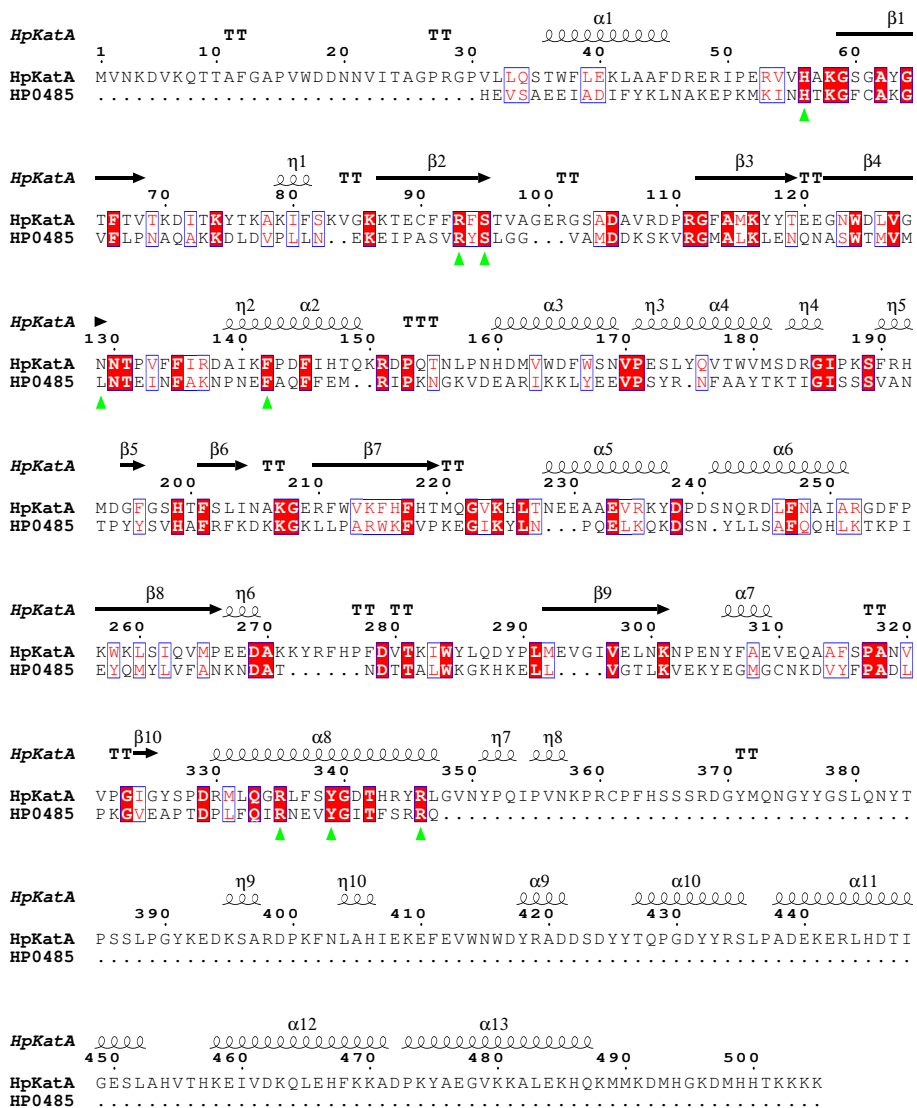
Figure 3. **Sequence alignment of *H. pylori* catalase and catalase-like**. Aligned sequence of *Hp* catalase (HpKatA) and *Hp* catalase-like (HP0485). Secondary structure elements inferred from the crystal structure of *Hp* KatA (PDB: 2A9E) are shown above the alignment. Upwards green triangles indicate the heme binding residues.
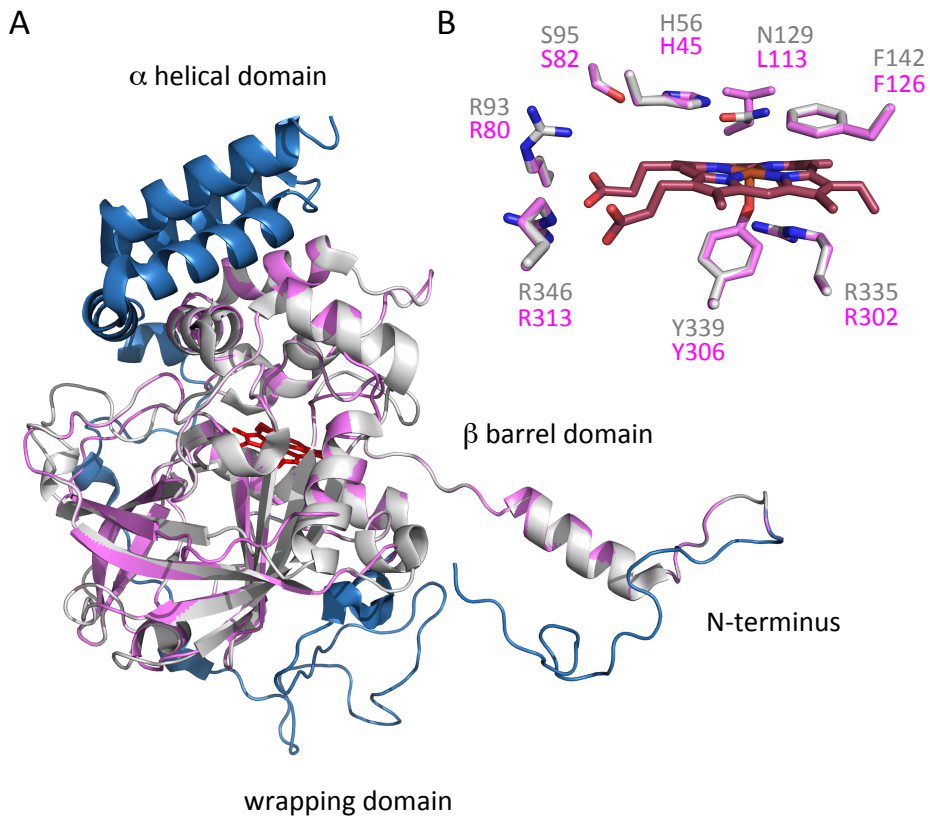
Figure 4. **Homology modelling of *H. pylori* catalase-like protein**. (A) Superimposition of HP0485 model (pink) with a monomer of *Hp* KatA (blue-gray; gray is the region overlapped by HP0485; heme is dark red). The two structures share the antiparallel eight-stranded β-barrel and the majority of the helical domain, whereas the first residues of the N-terminal arm, the extended wrapping loop and the C-terminal α-helical domain are absent in HP0485. (B) Superimposition of the heme environments of HP0485 model (pink carbons) and *Hp* KatA (gray carbons). The key residues are the same highlighted in the alignment in Figure 3.

Recombinant expression and periplasmic purification

In silico analysis with SignalP[55] predicted a signal peptide at the N-terminal of HP0485 sequence, with a cleavage site located between amino acid positions 20 and 21, which should result in a mature polypeptide of 35 kDa. This suggests that *Hp* catalase-like is a periplasmic protein exported into the periplasm via the Sec-pathway.

A predicted signal sequence is also present in *Hp* catalase-like homologous proteins of *Helicobacter* and *Campylobacter* species, so that multiple alignment of a selection of these sequences produced a sequence logo of the signal peptides typical of Gram-negative bacteria[56] (Fig. 5A and supplementary Fig. S1).

To recombinantly express *Hp* catalase-like protein in the periplasm, and avoid possible incompatibility with *H. pylori* and *E. coli* Sec-secretion systems, we inserted the gene encoding the mature form of HP0485 into the pET22 vector as a fusion with the PelB signal peptide[57]. The 6xHis-tag was fused at the C-terminal. The protein was expressed in *E. coli* BL21 (DE3), and both the precursor and the mature forms were detectable in the induced total cell extract on SDS-PAGE (Fig. 5B). After periplasmic extraction with a high-sucrose buffer, the mature form of 35 kDa became the predominant one and was purified by cobalt affinity chromatography. The binding of HP0485 turned the resin brown and the purified protein retained the brownish colour, suggesting the presence of bound heme.

Spectral properties

The absorption spectrum of *Hp* catalase-like protein showed a distinct Soret band at 406 nm, with a 280/406 absorbance ratio of 0.5, and three broad peaks at 508 nm, 540 nm and 620 nm, consistent with a ferric hemoprotein (Fig. 6A). Treatment with sodium dithionite caused a marked change in the spectrum. The

A

B
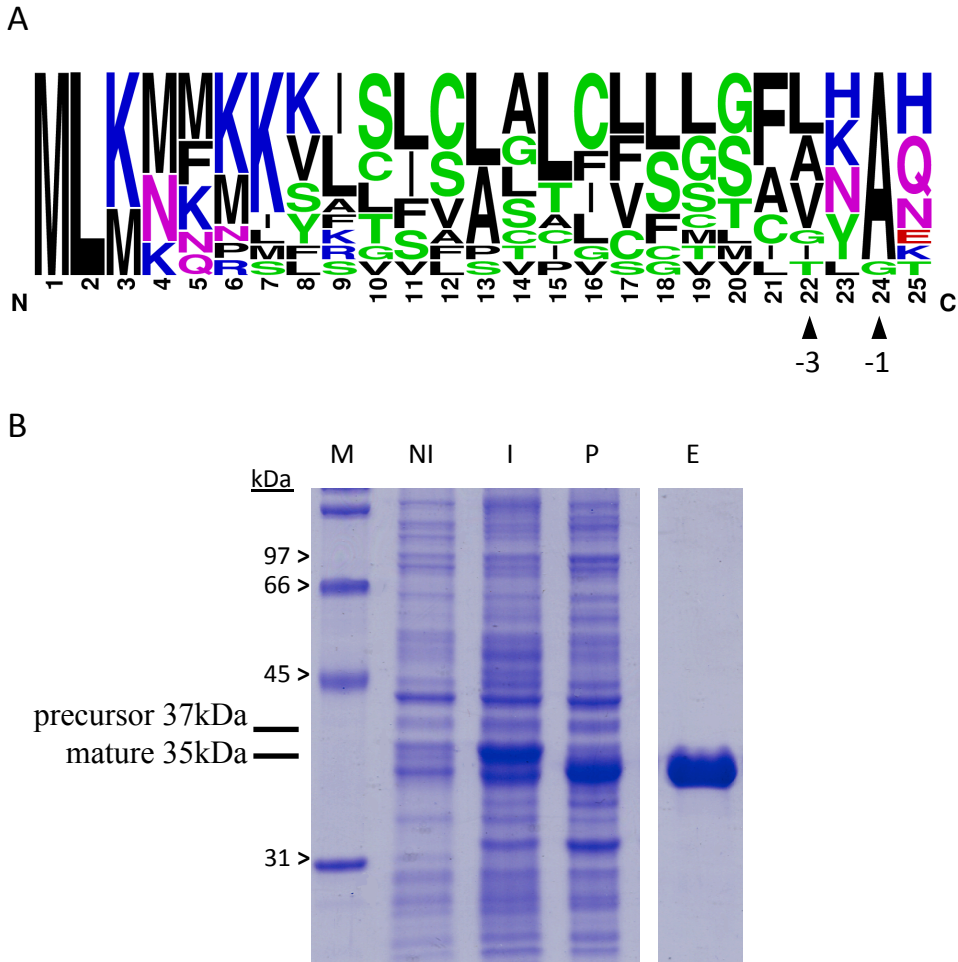
Figure 5. **Signal peptide and periplasmic purification**. (A) Sequence logos of signal peptides from HP0485 and homologous protein sequences, aligned by their cleavage sites (frequency plot representation). Conserved residues at -1 and -3 of the cleavage site are indicated with arrowheads. (B) SDS-PAGE with Coomassie staining of HP0485 purification from *E. coli* periplasm. Both the precursor and the mature forms are visible in the induced total cell extract, whereas only the mature form is found in the periplasm. M, molecular weight Marker, NI, Not Induced total cell extract, I, Induced total cell extract, P, post induction Periplasmic extract, E, affinity chromatography Elution with the purified protein.
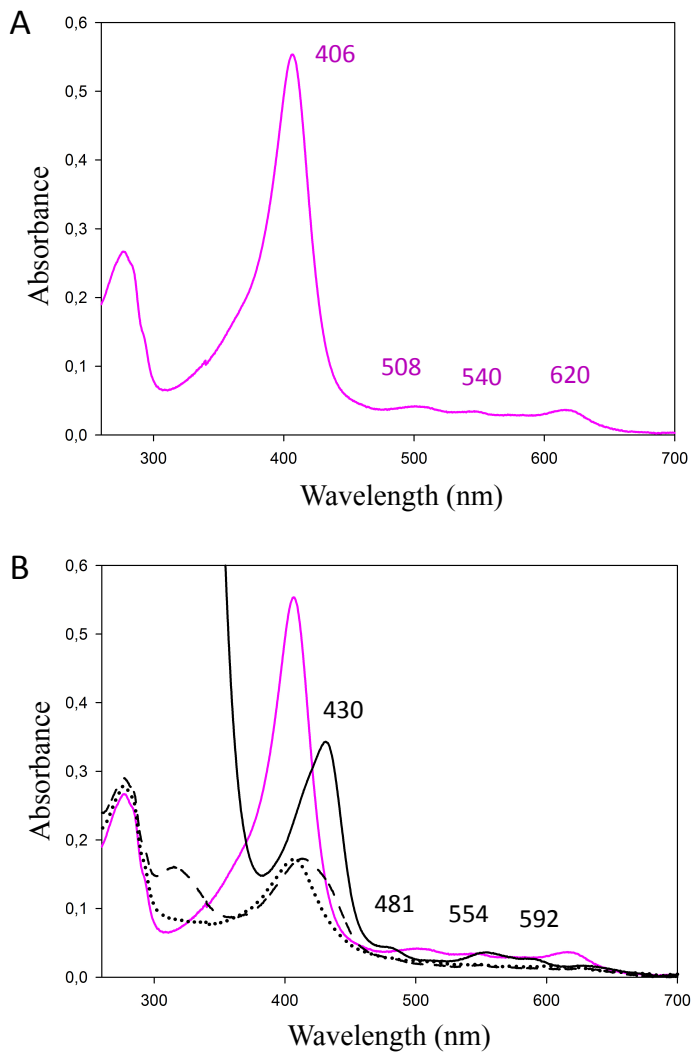
Figure 6. **UV-visible spectroscopy of *Hp* catalase-like.** The protein was in 50 mM sodium phosphate buffer (pH 7.4) containing 300 mM NaCl. (A) Native form (pink line); (B) Reduction with sodium dithionite. Black solid line, a few second after the addition of excess dithionite. Black dashed and dotted lines, 8 and 15 minutes after the addition of excess dithionite respectively.

Soret absorption decreased in intensity and shifted to 430 nm and the peaks at 508 nm, 540 nm and 620 nm were replaced with new peaks at 481 nm, 554 nm and 592 nm. These changes are characteristic features of a transition to a ferrous hemoprotein (Fig. 6B).

_Hp_ catalase-like has a monomeric structure

To observe the native oligomeric state of _Hp_ catalase-like protein, gel filtration chromatography and cross-linking experiments were performed.

Gel filtration chromatography revealed that _Hp_ catalase-like protein is a monomer of 35 kDa (Fig. 7A). A minor peak may represent a fraction of the purified protein organized as a dimer, owing to the high concentration of the sample.

Cross-linking experiments with glutaraldehyde were conducted on _Hp_ catalase-like and _Hp_ KatA in parallel (Fig. 7B). _Hp_ catalase-like migrated on SDS-PAGE as a monomer of 35 kDa, with and without glutaraldehyde treatment. Conversely, _Hp_ KatA appeared as a monomer of 58.6 kDa without glutaraldehyde treatment, and as a tetramer of 234.4 kDa upon the cross-linking reaction.

Enzyme activity

Given the structural similarity to catalase, the _catalatic_ activity of HP0485 was measured. Using the decrease in absorbance at 240 nm to follow disappearance of H_2O_2, the HP0485 showed no reaction, even at a concentration 100-fold higher than that of _Hp_ KatA (used as a positive control). Thus, _Hp_ catalase-like does not present _catalatic_ activity.

We next assayed the _peroxidatic_ activity, and found that the _Hp_ protein was able

Figure 7. **Monomeric structure characterization.** (A) Analysis of the purified HP0485 by gel filtration chromatography. The column was calibrated with molecular mass standards (arrowheads). (B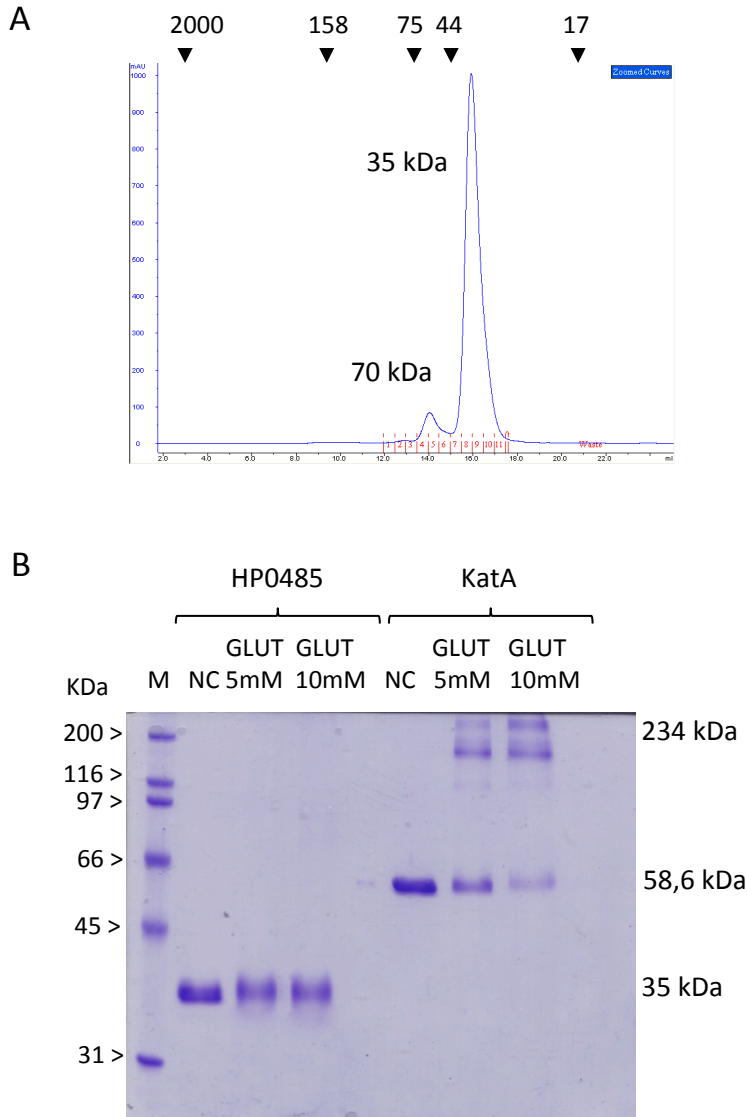) SDS-PAGE of the cross-linking experiments. HP0485 and KatA were treated with two concentrations of glutaraldehyde (5 mM and 10 mM). M, molecular weight Marker, NC, Not Cross-linked protein, GLUT, Glutaraldehyde treated protein.

to catalyse the oxidation of an organic electron-rich substrate, ABTS, by H_2O_2. Titration with increasing concentration of H_2O_2 (0.1 mM-200 mM), while ABTS concentration was kept constant (0.1 mM), showed a non-saturable kinetics and the predicted V_{max} could not be attained at H_2O_2 concentration up to 500 mM-1 M (Fig. 8A). This behaviour is atypical for a peroxidase, but it is common among catalases and represents a problem in defining K_m and k_{cat} values for these enzymes[41,58]. In the case of *Hp* catalase-like apparent values of 93.2 mM for K_m and 194 s^{-1} for k_{cat} were calculated. Moreover, the enzyme could withstand very high concentrations of H_2O_2 (molar ratio 1 : 10^7) without loosing catalytic efficiency.

To examine in more detail the *peroxidatic* activity, oxidation of different one-electron donors (pyrogallol, guaiacol, ABTS, ascorbate) was performed, using H_2O_2 as oxidant. Enzyme kinetics were measured at different concentrations of each electron-rich substrate (see Materials and Methods), while H_2O_2 concentration was kept constant (50 mM). The initial rates of oxidation by HP0485 showed a simple hyperbolic dependence on the concentration of the titrated substrates only in the case of pyrogallol. For the other tested substrates (guaiacol, ABTS, ascorbate) there was a reduction in the rate at high substrate concentration and the enzymatic behaviour was better described by a hyperbolic dependence with substrate inhibition at high concentration of the electron donors. This behaviour was reported for other enzymes with peroxidase activity and ascribed to the presence of an additional inhibitory binding site for these substrates with a regulatory function on the enzyme activity[59,60]. Figure 9 shows the Michaelis-Menten behaviour obtained from above experiments, and Table 1 summarizes K_m and k_{cat} values calculated for the oxidations. *Peroxidatic* activity of typical peroxidase of *A. rusticana* (horseradish peroxidase, HRP) was measured in parallel with the same substrates and exhibited hyperbolic dependence on the concentration of substrates (Table 1).

Hp catalase-like was able to oxidize all the tested compounds; in particular, it
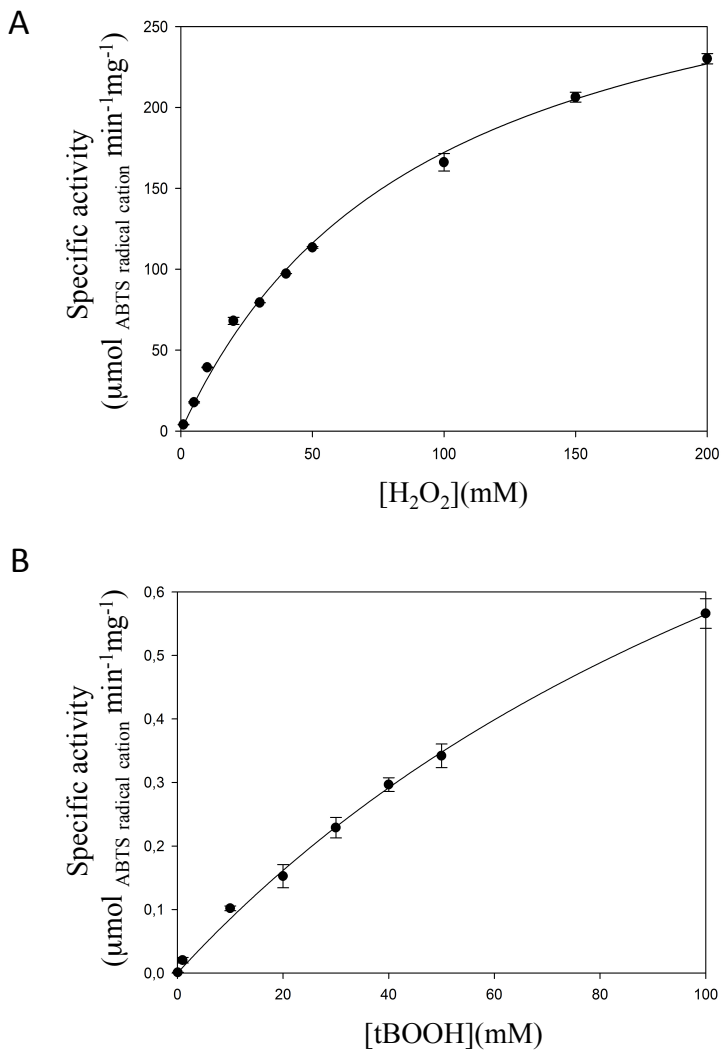
Figure 8. **Peroxidatic activity with ABTS.** (A) Kinetics with ABTS and H_2O_2. ABTS concentration was 0.1 mM and H_2O_2 concentration was varied (0.1-200 mM) when monitoring the formation of the ABTS radical cation at 420 nm. (B) Kinetics with ABTS and tBOOH. ABTS concentration was 0.1 mM and tBOOH concentration was varied (0.1-100 mM).
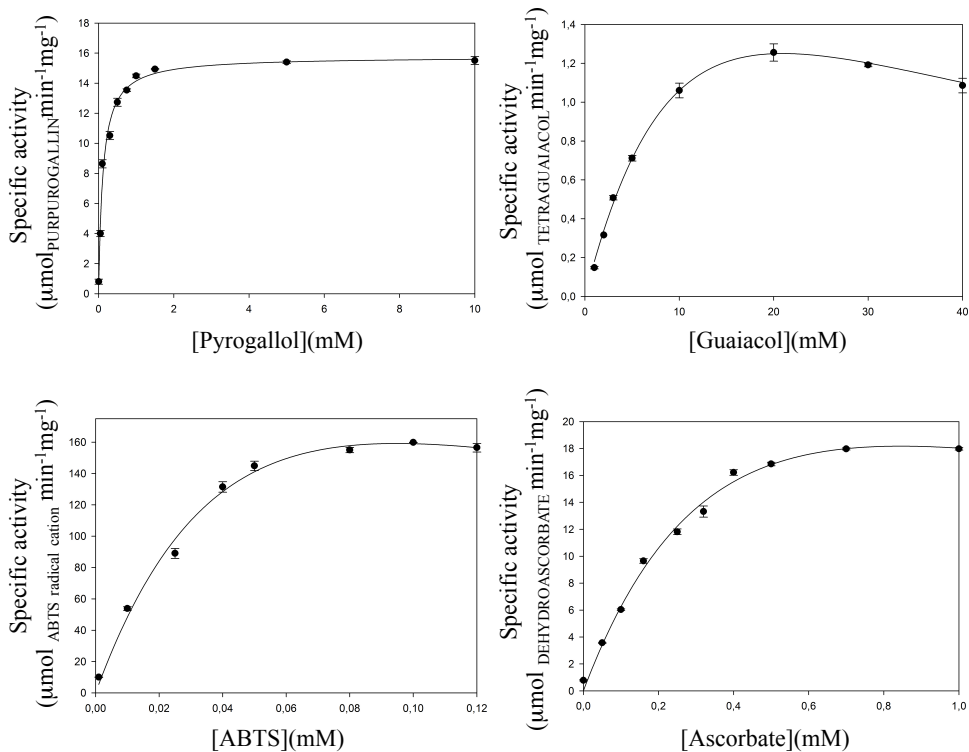
Figure 9. *Peroxidatic* **activity with H$_2$O$_2$ and different one-electron donors**. *Hp* catalase-like protein was assayed for *peroxidatic* activity using hydrogen peroxide and different one-electron donors (pyrogallol, guaiacol, ABTS, ascorbate). H$_2$O$_2$ concentration was 50 mM while the other substrate concentration was varied, when monitoring the formation of the final oxidative product at the corresponding λ (see Material and Methods and Table S1).

could oxidize ascorbate, at variance with HRP. Calculated K_m and k_{cat} values for ascorbate, respectively of 585 μM and 25 s^{-1}, were reported for some plant ascorbate peroxidases[61–63]. The other electron donors are typical synthetic substrates for HRP. Hp catalase-like exhibited a reaction rate ≈ 200-fold lower than that of HRP in oxidizing pyrogallol (k_{cat} values of 9.2 s^{-1} for HP0485 and 1580 s^{-1} for HRP), but with both guaiacol and ABTS the catalytic turnover was similar between the two enzymes (k_{cat} values for guaiacol of 1.96 s^{-1} for HP0485 and 12.2 s^{-1} for HRP; k_{cat} values for ABTS of 235.3 s^{-1} for HP0485 and 143.9 s^{-1} for HRP).

The ability of Hp catalase-like to utilise organic hydroperoxides was also evaluated (Fig. 8B and Table 2). Kinetics in presence of increasing concentration of t-butyl hydroperoxide (t-BOOH; 0.1 mM-100 mM) and constant ABTS (0.1 mM), showed a weaker peroxidatic activity (k_{cat}: 0.88 s^{-1}) compared to that with H_2O_2 (k_{cat}: ≈ 200 s^{-1}) and again non-saturable interaction with the peroxide substrate. We followed the reaction rate up to 100 mM t-BOOH (molar ratio HP0485 : t-BOOH = 1 : 10^5); at this value the increase was still exponential and the enzyme did not exhibit loss in activity. These results suggest that Hp catalase-like is able to catalyse *peroxidatic* reaction with both H_2O_2 and t-BOOH, being not saturated up to very high peroxide concentrations, but it is more efficient with hydrogen peroxide. For comparison, the *peroxidatic* activity of the catalase-like from *Mycobacterium avium* in equivalent assays, ABTS-H_2O_2 and ABTS-t-BOOH, was reported to have k_{cat} values of 13 s^{-1} and 320 s^{-1} respectively. Thus, curiously, the catalase-like of *M. avium* exhibited opposite behaviour with a preference for organic hydroperoxides over hydrogen peroxide.

HP0485

Substrate	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M (M^{-1} s^{-1})
PYROGALLOL	0.105 ± 0.014	9.2 ± 0.21	8.8×10^4
GUAIACOL	17.6 ± 2.7	1.96 ± 0.36	1.1×10^2
ABTS	0.073 ± 0.038	235.3 ± 90	3.2×10^6
ASCORBATE	0.585 ± 0.202	25.2 ± 6.4	4.3×10^4

HRP

Substrate	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M (M^{-1} s^{-1})
PYROGALLOL	2.2 ± 0.6	1580 ± 121	7.2×10^5
GUAIACOL	2.8 ± 0.8	12.2 ± 3.4	4.3×10^3
ABTS	4.6 ± 1.6	143.9 ± 19	3.1×10^4
ASCORBATE	Activity not detected		

Table 1. **Kinetics parameters for *Hp* catalase-like and horseradish peroxidase**.

HP0485

Substrate	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M (M^{-1} s^{-1})
H_2O_2	93.2 ± 7.7	194 ± 7.6	2.081×10^3
t-BOOH	165.6 ± 20.6	0.88 ± 0.08	5.3

MAP_2744c

Substrate	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M (M^{-1} s^{-1})
H_2O_2	30	13	4.33×10^2
t-BOOH	0.022	320	1.45×10^4

Table 2. **Kinetics parameters for *Hp* catalase-like and *M. avium* MAP_2744c**.
Values are from this work (HP0485) or from the literature (MAP_2744c; Pakhomova *et al*., 2009

Heme environment

Our next step was to investigate the origin of the disparity between *Hp* catalase and *Hp* catalase-like activity. To understand why two enzymes extremely similar in structure exhibit such a marked difference in function, the heme active site has to be considered. As can be seen from figures 4 and 10, there is a strictly conservation of key amino acids close to the heme. In addition to the proximal tyrosine (Tyr339 in KatA, Tyr306 in HP0485) and the distal histidine (His56 in KatA, His45 in HP0485), an arginine that is hydrogen bonded to Tyr in KatA (Arg335) is found at the same position in HP0485 (Arg302), and similarly, a serine that participates in hydrogen bond interaction with the catalytic histidine in KatA (Ser95) is present in HP0485 (Ser82). All these residues have been proposed to be important for Compound I formation that is the first common step of the *catalatic-peroxidatic* cycle (Fig. 1)[34,51,64]. But there are at least two significant differences in the neighbouring residues lining the heme distal side that determine a different occupancy and electrostatics in the active site and could contribute to the different reactivity of the two enzymes. In particular the distal side of the heme in KatA also includes an asparagine (Asn129) positioned over the heme near to distal histidine (His56), and is implicated along with the distal His in modulating or tuning catalysis[51]. In HP0485, the distal asparagine is not conserved and a hydrophobic residue, leucine (Leu113), is found at this position. Another notable difference is the substitution of the valine next to the catalytic His in KatA (Val55) with a polar residue, asparagine (Asn44) in HP0485 that may alter its availability for the reaction. It is intriguing to note that the characteristic of these residues flanking the catalytic His in the heme active site are interchanged between the two enzymes: hydrophobic-polar in catalase and polar-hydrophobic in catalase-like.
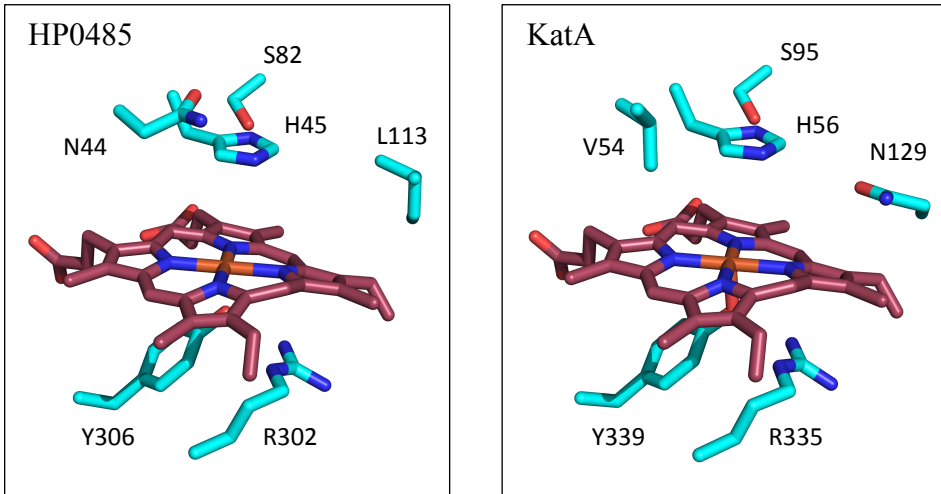
Figure 10. **Key residues in the heme environment in the active sites of *H. pylori* HP0485 and KatA**. The heme (dark red) has conserved His and Ser residues on the distal side. The proximal heme ligand, Tyr, is flanked by a conserved Arg. In HP0485 Leu and Asn residues substitute catalase residues Asn and Val respectively. KatA structure is from PDB file 2A9E (*Hp* KatA), HP0485 is a model obtained from SWISS-MODEL.
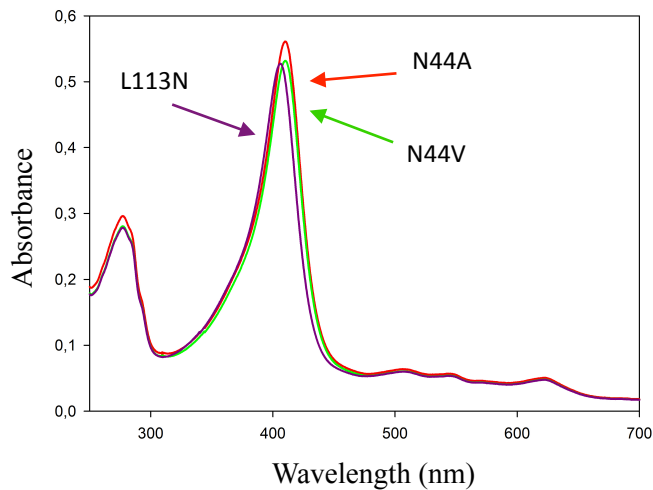
Figure 11. **Absorption spectra of *Hp* catalase-like mutants**. Purified N44V, N44A and L113N mutants in 50 mM sodium phosphate buffer (pH 7.4) with 300 mM NaCl. The Soret peak is at 407 nm in L113N (purple line), 409 nm in N44V (green line), 410 nm in N44A (red line).

Analysis of *HP* catalase-like mutants

In the light of these observations, *Hp* catalase-like single mutants were created to eliminate the differences at the heme distal side with *Hp* catalase. In the first and second mutant, Asn44, next to the catalytic His, was replaced with a hydrophobic residue, Val (N44V) and Ala (N44A) respectively. In the third mutant, Leu113 was mutated to Asn (L113N). Each substitution gave a protein with similar heme content, but a shifted Soret peak in respect to the wt (406 nm). As shown in figure 13, N44V and N44A mutants display a peak at 409 nm and 410 nm respectively, while L113N presents a 407 nm peak. The *catalatic* activities of the mutants were assayed by spectrophotometric measurement of H_2O_2 consumption but, as for the wt protein, no absorbance decrease at 240 nm was detected. These results suggest that a single mutation is not sufficient for the acquisition of the *catalatic* activity.

Assembly and disassembly of the catalase-like heme cofactor

To understand the process of *Hp* catalase-like assembly in the periplasm the apo form of the protein was generated through acid-butanone extraction of bound heme[65]. This method has been successfully used for other hemoproteins like cytochromes and globins, which retain their tertiary structure and can be readily reconstituted with free heme[66]. Upon phase-separation, the lower phase remained colourless, whereas the upper phase was coloured and contained free heme, as demonstrated by the absorption at 385nm. The final spectrum of the dialyzed lower phase containing the protein, lacked the peak at 406 nm and the minor peaks (Fig. 12A), still the protein conserved unaltered tertiary structure as shown by the CD spectra of the holo and the apo forms (Fig. 12B). This result clearly indicates that heme is non-covalently bound to *Hp* catalase-like protein,
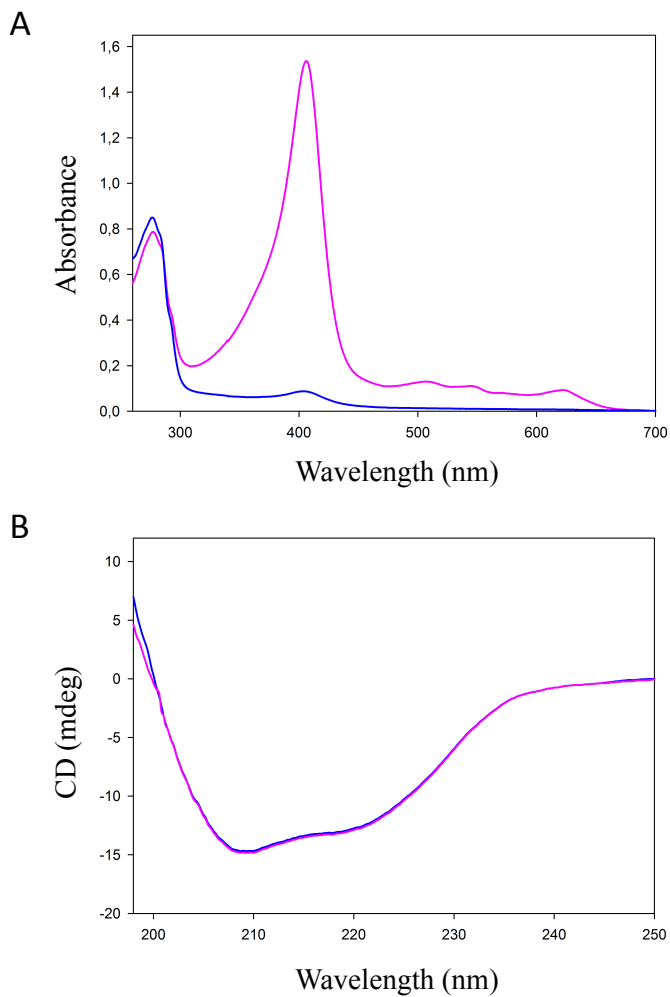
Figure 12. **Holo and apo forms of *Hp* catalase-like**. The two protein forms were both in 10 mM sodium phosphate buffer (pH 7.4); the spectra were collected at 25 °C. Holo form is pink; apo form is blue. (A) Absorption spectra; (B) Far UV CD spectra.
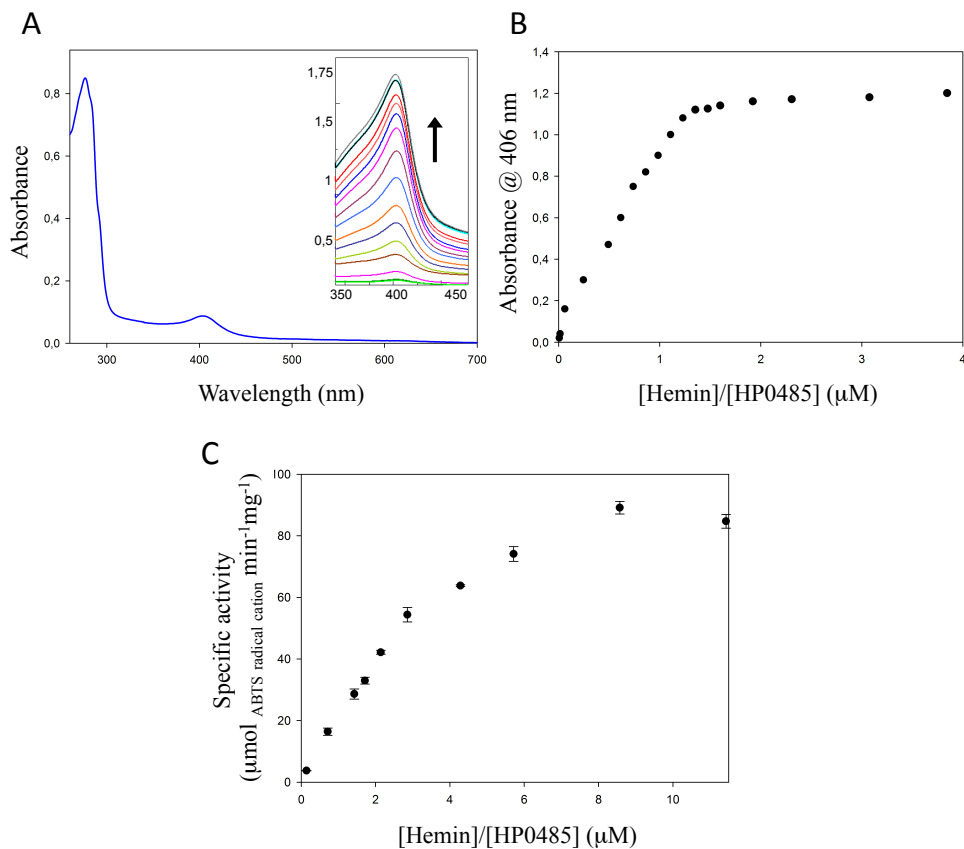
Figure 13. **Heme reconstitution of the apo *Hp* catalase-like**. (A) Hemin titration of the apo *Hp* catalase-like (12 μM) to re-form the holo protein, monitored by visible absorption spectroscopy immediately after mixing of the samples. (B) The change in A406 with increasing amounts of hemin (100 nM-20 μM) shows the essentially stoichiometric incorporation of heme. (C) *Peroxidatic* assay with H_2O_2 and ABTS coupled to the hemin titration of the apo protein. Aliquots of apo *Hp* catalase-like (7 mM) were incubated with hemin at concentrations ranging from 1 μM to 80 μM at 25 °C for 15 minutes. After this time the peroxidatic activity of 5 nM of the protein from the sample was determined.

as reported for catalases[35], and that heme binding is not necessary for the maintenance of the folding.

Titration of the apo catalase-like with free hemin at physiological conditions (25 °C, pH 7.4) was able to regenerate the holo form of the enzyme. The Soret absorption peak concomitantly increased and reached a maximum when HP0485 and hemin were present at approximately equimolar amounts (Fig. 13A-B). The *peroxidatic* activity of the reconstituted apo form was almost fully recovered ($\approx$ 80%) within fifteen minutes of incubation at room temperature after the addition of excess hemin. This was tested with an ABTS-H_2O_2 assay coupled to the hemin titration of the apo protein (Fig. 13C). Aliquots of apo *Hp* catalase-like (7 μM) were incubated with hemin at concentrations ranging from 1 μM to 80 μM at 25 °C for 15 minutes. After this time, the peroxidatic activity of 5 nM of the protein from the sample was determined. Control incubations with only apo catalase-like or hemin, yielded no reaction. The *peroxidatic* activity increased with the addition of hemin and reached a plateau at a 4-fold amount of hemin with respect to the protein.

A conserved disulfide bond is essential for heme binding

To gain insights on the mechanism of heme assembly within the catalase-like active site, the presence of a putative disulfide bond was investigated. In the *Hp* catalase-like model structure, in proximity of the heme cavity, two cysteines (Cys50 and Cys277) appear to be at a distance of 2.5 Å, consistent with a disulfide bond (Fig. 14A). These two cysteines are not present in *Hp* KatA, but are conserved in HP0485 homologous containing a putative signal peptide in these protein sequences. Additional evidence for the existence of a disulfide bond is given by the co-variance of these two cysteines in multiple sequence alignments that include catalase-like proteins lacking the signal peptide (e.g. from Gram-positive bacteria) (Fig. 14D).
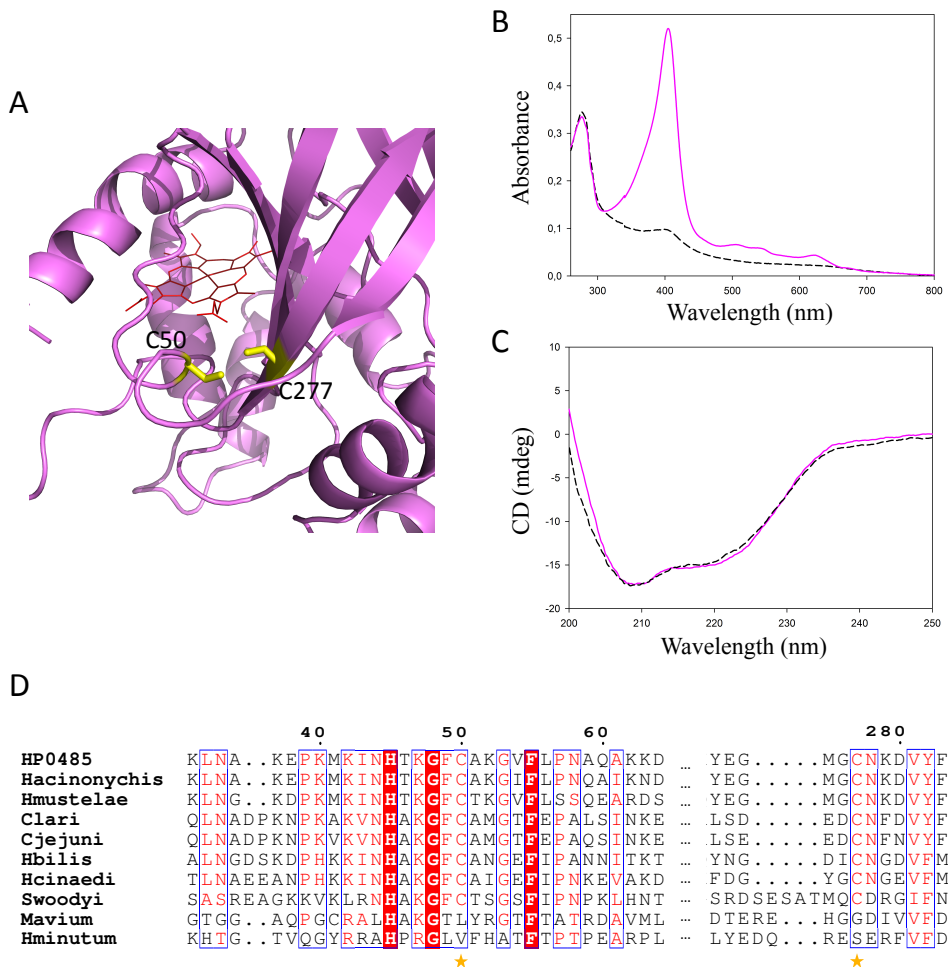
Figure 14. **Evidence for the presence of a disulphide bond in secreted catalase-like proteins implicated in the heme assembly**. (A) Cysteine residues (yellow sticks) in HP0485 model structure at a distance of 2.5 Å, consistent with a disulfide bond. (B) SS-bond reduction with 5 mM TCEP lead to the detachment of heme from the active site. Absorption spectroscopy of holo HP0485 untreated (pink line) and treated with TCEP (black dashed line). (C) Far UV CD spectra of holo HP0485 untreated (pink line) and treated with TCEP (black dashed line). (D) Multiple sequence alignment of HP0485 and related proteins displaying the covariance of the two Cys (yellow stars).

To provide experimental evidence, HP0485 was treated with 5 mM TCEP to reduce the disulfide bond; interestingly, the bound heme was lost, as indicated by a spectrum resembling that of the apo protein (Fig. 14B). Far UV CD measurement demonstrated that the protein still retained a folded conformation with no appreciable differences from the holo form (Fig. 14C).

The addition of free heme to this apo form, in presence of TCEP, resulted in a visible spectrum resembling that of free heme rather than that of a heme-protein complex, and after gel filtration chromatography the protein spectrum was still featureless above 300 nm. Thus, reduction of the disulfide bond impaired the heme-reconstitution of apo catalase-like. Overall these experiments demonstrate the essentiality of disulfide bond formation for the incorporation of heme into the active site.

Evolutionary relationship between catalase and catalase-like proteins

Based on the alignment of the entire amino acid sequences of various prokaryotic catalase and catalase-like proteins, a tree was constructed to gain an overview of the phylogenesis of *Hp* catalase-like and its related proteins, as well as explore the features that are either in common or different between the two groups of enzymes (Fig. 15).

The phylogenesis shows the distinct group of catalase-like that separated quite early from catalases, as there is no mixture of members between the two lineages. As well as catalases that are divided in "small" subunit and "large" subunit enzymes[43], catalase-like proteins display differences in sequence length and therefore in dimension. In agreement with the catalase-like phylogenesis presented by Pakhomova *et al.*[48], two major groups are found: catalase-like with up to 330 residues and catalase-like with a number of residues ranging from 330 up to 360. Whatever the exact dimension is, it does not reach that of catalases,
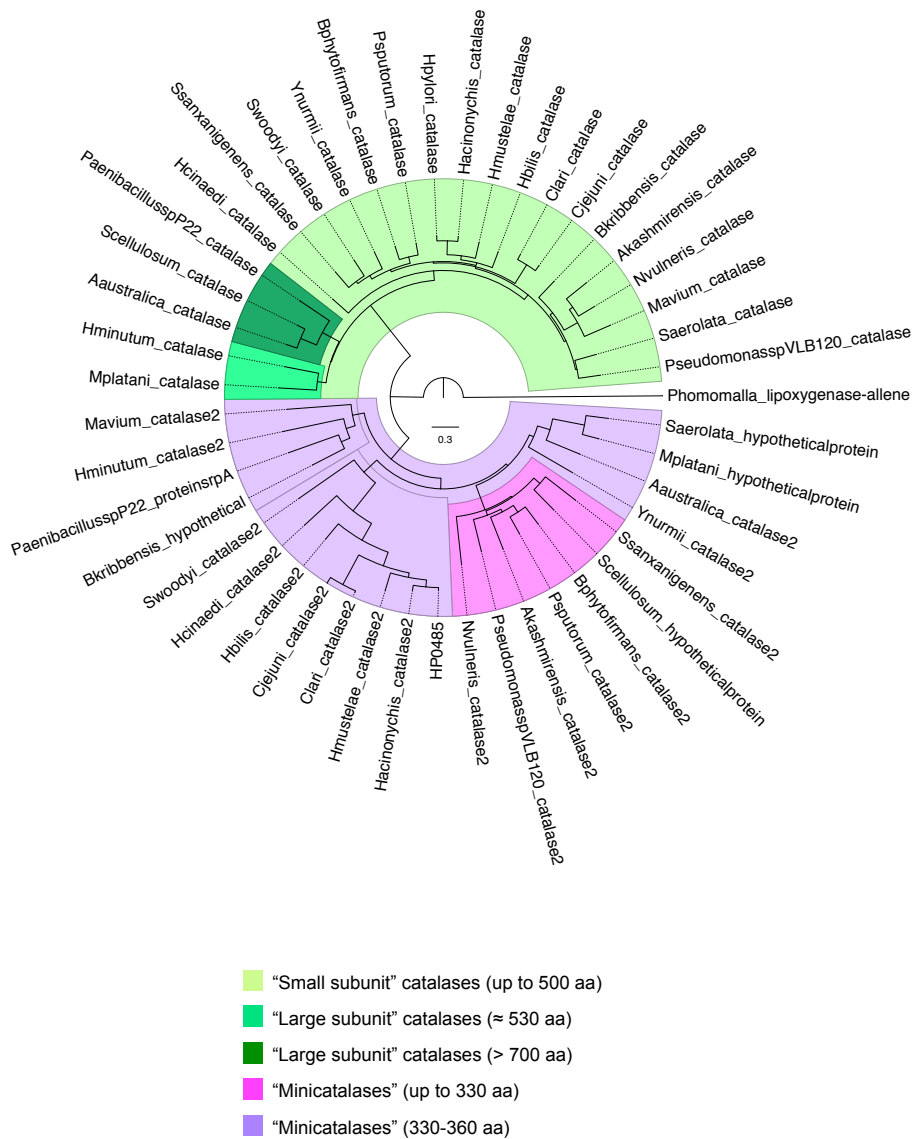
Figure 15. **Phylogenesis of HP0485 and related proteins**. Midpoint-rooted tree representing the evolutionary relations between catalase and catalase-related enzymes. HP0485 segregates with the subgroup with proteins < 330 amino acids from Gram negative bacteria. Groups are named according to Klotz and Loewen for catalases, and Pakhomova *et al*. for catalase-like. *P. Homomalla* AOS is equally distant from both groups.

owed to the absence of two domains (Fig. 3), and for this reason catalase-like proteins are also called "mini-sized" catalases or "minicatalases".

Selected portions of the multiple sequence alignments of catalase-like proteins and catalases (Fig. S2 and S3) show the conservation of key residues within the heme binding pocket. The proximal Tyr and the distal His are always conserved. In contrast, the other two residues, the Val preceding the His, and the vicinal Asn, that are present in almost all catalases, are variant positions in catalase-like enzymes. Overall, a lower degree of conservation in catalase-like compared to catalases is evident from the comparison of branch lengths in the two groups, hinting at the heterogeneous activity expected for these proteins.

Discussion

Helicobacter pylori has the extraordinary ability to survive within the human stomach although it induces an intense and persistent inflammation in the infected gastric mucosa, that culminates in the heavy generation of reactive oxygen metabolites. To overcome the severe oxidative stress the bacterium employs a great armoury of ROS detoxifying factors.

Being so powerful and specific for the degradation of hydrogen peroxide, catalase is one of *H. pylori* most prominent antioxidant enzymes, with a unique, strongly preserved three-dimensional organization.

Within the bacterium genome another protein is present that bears a catalase domain, and belongs to a family of enzymes structurally related to catalase, but of undefined function.

To uncover another potential actor in the antioxidant resistance of this highly successful human pathogen, we characterized *H.pylori* catalase-like protein, HP0485.

Hp catalase-like is located in the cell periplasm, as shown by the presence of a signal peptide at the N-terminal of the sequence. After recombinant expression and periplasmic purification, the protein appeared to be a monomer of 35 kDa with a ferric heme group. Despite the weak sequence identity to catalase (22%), the predicted tertiary structure resembles a catalase core and accurately conserves the heme environment. In particular, key catalytic residues such as the proximal iron-ligand tyrosine (Tyr306) and the distal histidine (His45), with their respective neighbours, arginine (Arg302) and serine (Ser82), are preserved. Quite unexpectedly, we found that *Hp* catalase-like protein does not present *catalatic* activity. The enzyme catalyses the H_2O_2-dependent oxidation of

different organic electron donors (i.e. pyrogallol, guaiacol, ABTS, ascorbate). Catalase-like interaction with hydrogen peroxide appears to be non-saturable at physiological levels, as the reaction rate constantly increases at H_2O_2 concentration up to 500 mM - 1 M, yet the *peroxidatic* activity in the presence of a suitable electron-rich substrate is efficient enough ($k_{cat}/K_m \approx 10^6 \, M^{-1} \, s^{-1}$ for the kinetics with ABTS) to suggest an important role for *Hp* catalase-like in cellular H_2O_2 metabolism. HP0485 can use also organic hydroperoxides, i.e. t-butyl hydroperoxide, as oxidant, but its involvement in organic hydroperoxide detoxification is less likely, since it shows a feeble catalytic activity.

A close look at the two structures with a catalase fold, the experimental KatA structure and the HP0485 model, pointed out the presence of two important differences in the heme active site. The valine preceding the catalytic histidine in KatA (Val55) is substituted with an asparagine (Asn44) in catalase-like, while the distal asparagine in KatA (Asn129) is replaced by a leucine in catalase-like (Leu113). Single mutations introduced in *Hp* catalase-like protein to alternately abolish these differences were not useful for the acquisition of *catalatic* activity, suggesting that the two newly generated active sites still retain significant divergences from that of catalase.

Extraction of the heme from the active site produced the apo catalase-like, which retained a folded conformation as a proof that heme is not required for the maintenance of the structure. Reconstitution of the apo protein with free heme led to the recovery of the 80% of the *peroxidatic* activity, but the stoichiometry of the heme titration calculated by monitoring the increase of the 406 nm peak and that obtained by monitoring the peroxidatic reaction did not match. Whether this discrepancy is due to non-optimal conditions (temperature, timing, pH, etc.) used for the *in vitro* assay, or to the in vivo uncoupling of the two processes, it is still an open question that needs further investigation.

A disulfide bond identified as a hallmark of secreted catalase-like proteins, was found to be implicated in the heme assembly. Since two cysteines (Cys50 and Cys277) were positioned at a distance compatible with the formation of a disulfide bond in HP0485 model structure and displayed a clear pattern of covariance in proteins multiple alignments, being both conserved only in the enzymes with a signal peptide, we hypothesized a link between the correct formation of a disulfide bond and the assembly of a functional hemoprotein in the periplasm. Accordingly, reduction of the two cysteines led to the loss of the bound heme, and reconstitution of the holo form was impaired in the presence of reducing agents.

The phylogenetic analysis of catalase-like proteins identifies this group of prokaryotic enzymes as a novel distinct branch of the catalase superfamily, equally distant from catalases and cAOS[48]. As suggested by the lower degree of conservation at sequence level, the bacterial catalase-related structure appears to be less constrained than the catalase one, and this likely reflects heterogeneous activity among these catalase-like enzymes. This idea is supported by the two family members characterized so far. As perfect example of promiscuous activity, *M. avium* catalase-like, exhibits weak *catalatic* activity, minor *peroxidatic* activity using H_2O_2, and strong *peroxidatic* activity using organic hydroperoxides[48]. The other family member, the *Hp* catalase-like characterized here, does not show *catalatic* activity, has minor *peroxidatic* activity using organic hydroperoxides, and strong *peroxidatic* activity using H_2O_2.

Hp catalase-like protein: a peroxidase in catalase clothing

Concealed behind a catalase appearance, *Hp* catalase-like turned out to be a versatile and robust peroxidase. The first attribute arises from its broad

peroxidatic activity, the second one from its ability to withstand high concentration of hydrogen peroxide.

The presence of *peroxidatic* activity instead of *catalatic* activity is a property typical of eukaryotic peroxidases. Indeed the comparison with one of the most famous peroxidases, HRP, seems appropriate. Compared to HRP in terms of catalytic rate, k_{cat}, HP0485 is inferior in the reaction with pyrogallol, whereas it is similarly effective in oxidizing ABTS and guaiacol (Table 1). Interestingly, however HP0485 is effective in oxidizing ascorbate, a substrate that is not attacked by HRP. Ascorbate is the physiological substrate of eukaryotic peroxidases from higher plants and green algae, constituting the family of ascorbate peroxidases. These enzymes represent a different evolutionary lineage from that of HRP and share a prokaryotic origin with bacterial catalase-peroxidases[42]. In prokaryotes ascorbate peroxidase activity was reported several times[67,68], but none of the putative enzymes were purified. An ascorbate-oxidizing activity was detected in *H. pylori*[69] and in agreement with this, the concentration of vitamin C was found to be decreased in gastric juice from *H. pylori* infected patients, and replaced by its oxidation product, dehydroascorbic acid[70]. However, Ødum and Andersen attributed the ascorbate oxidizing activity detectable in *H. pylori* lysate supernatant to a protein band of $\approx$ 14 kDa, consistent with a cytochrome *c*-like molecule[69]. In this work we showed that purified recombinant *Hp* catalase-like is able to catalyse the H_2O_2-dependent oxidation of ascorbate, with efficiency comparable to some ascorbate peroxidases, but at higher H_2O_2 concentrations (10-500 mM vs 0,1-10 mM)[60–62].

In the light of its broad specificity for electron donors, the complete inability of *Hp* catalase-like protein to use H_2O_2 for a *catalatic* reaction is surprising. Investigation of *Hp* catalase-like *peroxidatic* property revealed two features shared with catalase enzymes. First, HP0485 shows non-saturable interaction

with H_2O_2, as the V_{max} is not attained at H_2O_2 concentration up to 500 mM - 1 M. Second, the reaction rate constantly increases even at high excess of H_2O_2 (molar ratio $1:10^7$) without the loss of enzyme activity. Conversely, most peroxidases, including HRP, are inactivated at concentration of H_2O_2 higher than 10-50 mM[71].

HP0485 exhibits the same non-saturable behaviour with *t*-butyl hydroperoxide at physiological levels. The *Hp* enzyme can withstand very high concentrations of t-BOOH (molar ratio $1:10^5$) without loosing activity, even if the turnover number is over 200-fold lower compared to the reaction with hydrogen peroxide. This latter aspect in contrast with catalase that is extremely sensitive to *t*-butyl hydroperoxide, being totally inactivated by 1:800 molar ratio[45].

The substrate inhibition observed in HP0485 kinetics with three of the tested electron donors was described for other peroxidases. Interestingly, for the multifunctional enzyme, dehaloperoxidase-hemoglobin (DHP), it represents a strategy to solve the toxicity problem posed by some of its substrates. Indeed its substrate inhibition behaviour has been attributed to the presence of an internal binding site that plays a role in regulating its peroxidase function and preventing the collapse of the protein stability during the catalytic reaction at high substrate concentration[59]. In this light, the substrate inhibition behaviour of HP0485 could be related to its increased stability in respect to other peroxidase enzymes.

A common feature on the heme distal side in catalatic reaction

The residue considered unanimously of vital importance for catalase function on the heme distal side is the conserved histidine. The distal asparagine is also highly preserved in catalase (Fig. S3) and its role in catalysis, according with molecular dynamic studies, seems to be related mainly with the orientation and flow of substrates and products[51]. The importance of this residue was called into

question when several different mutants were obtained, which retained *catalatic* function and only exhibited overall lower activity[72,73]. This effect was marked in particular at high H_2O_2 concentration, with lower values of both V_{max} and K_m, suggesting a possible role for the Asn in binding the peroxide and thus increasing affinity for the substrate[72].

The valine flanking the distal histidine is well conserved in catalases, but to a lower extent compared to Asn, as also methionine or proline natural variants can be found (Fig. S2). This position certainly came to prominence with the structural characterization of coral allene oxide synthase[50]. The absence of *catalatic* activity in cAOS was attributed to the presence of a threonine replacing the valine, and the hypothesis was confirmed by the acquisition of *catalatic* function in Thr to Val mutant[74]. Molecular dynamic analysis showed that the Thr in cAOS forms a strong hydrogen-bond with H_2O_2 and, with the aid of the catalytic His, helps to pull H_2O_2 away from the heme iron center[75]. On the other hand, mutations of the valine in catalase decrease, but not abolish *catalatic* activity, suggesting a minor but still important role in helping optimizing catalysis for this residue[76].

To understand why HP0485 single mutations of asparagine (Asn44) and leucine (Leu113) to their counterparts in catalase, Val and Asn respectively, did not lead to the acquisition of catalatic activity, it can be helpful to consider these two residues together and not separately. Noteworthy, their positioning in *Hp* catalase-like creates an electrostatic pattern on the distal side of heme that is exactly opposite to that found in catalase: polar-hydrophobic *versus* hydrophobic-polar (Fig. 10). Keeping this in mind, it is interesting to reconsider the other known catalase-related enzymes. In *M. avium* protein leucine is found in place of the valine, and glutamine substitutes the asparagine, such as the catalase pattern hydrophobic-polar is maintained, resulting in a weak but detectable *catalatic* activity. A completely different pattern is present in cAOS,

with two polar residues, threonine and asparagine, at these positions, and the *catalatic* reaction is absent. Our results enrich this framework, as we characterized a catalase-related protein with a natural inverted pattern (polar-hydrophobic instead of hydrophobic-polar) on the heme distal side that is incapable of dismutating H_2O_2. We introduced single mutations, leading to disruption of the pattern and its conversion to polar-polar and hydrophobic-hydrophobic, without achieving any catalatic activity. All together these lines of evidence suggest the requirement of a hydrophobic-polar pattern on the heme distal side for the accomplishment of the *catalatic* reaction. It will be thus interesting to see if a double mutant would fulfil this active site requirement in *Hp* catalase-like.

A disulfide bond for heme assembly

It is not completely known how most hemoprotein, including catalase, are assembled from their constituents, heme and polypeptide, within the cell.

In catalase binding to the heme internalization is a complex mechanism that likely involves the oligomerization process, as the heme pocket is formed with residues from two different subunits[35]. A model hypothesizes the initial formation of holo monomers, followed by holo dimers, and finally by holo tetramers[54]. Dissociation of the tetramer that can be obtained at extremes of pH, leads in the majority of the cases to enzyme inactivation, as monomers are formed and usually followed by release of the heme[77]. Attempts to reconstitute active catalase by addition of heme to the apo protein prepared in this way have not been successful, not even with the aid of molecular chaperones[54]. In vitro assembly was reported only for an organism, *Enterococcus faecalis*, that cannot synthesize heme, but can acquire it from the environment to form a cytoplasmic catalase[78]. Apart from this exception, the problem of *in vitro* folding and assembly of catalases still remains unsolved[54].
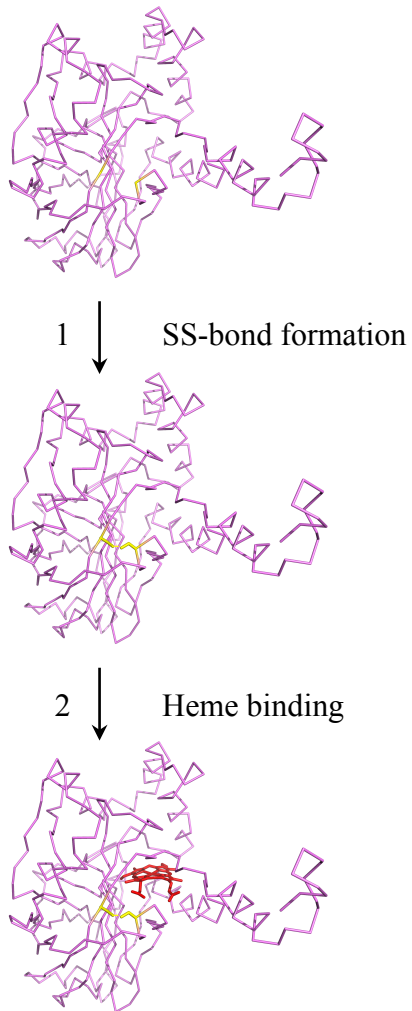
Figure 15. **Model for *Hp* catalase-like protein assembly in the periplasm**. The representation with HP0485 model (pink ribbon rendering, with two Cys in yellow and heme in red) shows the order of events during the protein assembly and heme incorporation to generate active *Hp* catalase-like, as suggested by our experimental results. Step 1, Disulphide bond formation within a folded or partially folded apo catalase-like. Step 2, insertion of heme in the active site.

In our work we successfully obtained the apo form of *Hp* catalase-like as well as its reconstitution at physiological conditions. We discovered that the recovery of a full active enzyme depends on the formation of a conserved disulfide bond. To illustrate our results we propose a model for the assembly of catalase-like (Fig. 16). The protein first acquires its folded conformation in the periplasm upon cleavage of the signal peptide (apo form); next disulfide bond formation permits the insertion of heme in the active site (holo form).

The successful *in vitro* assembly of catalase-like suggests that this process does not require additional cellular components, and can promptly occur inside the cell. The dependency on a disulfide bond can be a control mechanism that allows the insertion of the heme only when the protein is correctly folded to avoid formation of nonproductive enzymes. Another possible explanation is that heme binding is regulated to produce a holo peroxidase only when its activity is needed. The *peroxidatic* reaction, besides hydrogen peroxide degradation, produces oxidized compounds that can be damaging for the cell, if not correctly metabolised. Therefore this process is likely coupled to other co-regulated detoxifying systems. Furthermore, the presence of disulfide bond is a mark of rigidity and stability for a protein (e.g. free exposed cysteines are target of attack by hydrogen peroxide) and represents an important aspect when considering the prominent stability of *Hp* catalase-like capable to cope with extremely high doses of hydrogen peroxide.

What is the physiological role of *Hp* catalase-like?

Bacterial cells typically own multiple scavenging systems to keep intracellular hydrogen peroxide at low non-harmful levels. The various catalases and peroxidases can be distinguished by variation in substrate specificity, kinetic properties, enzyme stability, expression pattern, and compartmentalization. *E.*

coli, for example, expresses at least three authentic scavengers, two heme catalases, one cytoplasmic and one periplasmic, plus AhpC peroxidase. *H. pylori* has only one catalase located both in cytoplasm and in periplasm[79], and it possesses likewise AhpC in the cytoplasm. Yet, differently to *E. coli* AhpC, acting primarily in H_2O_2 degradation, *H. pylori* AhpC is mainly dedicated to detoxification of organic hydroperoxides, providing also a protecting role for catalase in preventing its organic hydroperoxides-dependent inactivation[45].

The evaluation of *Hp* catalase-like role as H_2O_2 scavenger inside the cell could start from considering its catalytic efficiency. It rivals that of eukaryotic peroxidases (i.e. HRP), but is certainly inferior to that of powerful catalase. The latter aspect might suggest that HP0485 function in H_2O_2 degradation becomes important when that of catalase is failing. This is typical of cell multiple scavenging systems, that cannot be entirely appreciated when they are all together present. *H. pylori* catalase-deficient mutants are viable when cultured *in vitro*, suggesting that endogenously generated H_2O_2 is efficiently removed by another scavenger[80,81]. AhpC could supply catalase lack in the cytoplasm but the periplasm would remain defenceless. However, we believe that the presence of HP0485 in the periplasm cannot be merely explained with the balance of H_2O_2 scavengers between cytoplasm and periplasm, as it is not self-evident that another peroxidase is needed.

An additional line of evidence comes from regulation, and thus from gene expression data. *Hp* catalase-like results significantly up-regulated under aerobic condition (20% O_2), and this connotes its role in *H. pylori* antioxidant defence[82]. Furthermore, its expression under normal growth conditions is induced with that of two other important antioxidant factors, superoxide dismutase (SOD) and neutrophil activating protein (NapA), after the late log-to-stationary phase transition, in which there is a major switch in gene expression, mainly involving

gene related to virulence[83]. Conversely expression of KatA is not enhanced by the cells entering stationary phase, as the highest activity occurs during exponential phase[44].

As a peroxidase, *Hp* catalase-like requires an electron donor for function. For most peroxidases the naturally occurring substrate is unknown; there are only few examples of a known specific substrate for *peroxidatic* activity. One of these is provided by the catalase-peroxidase from *Mycobacterium tuberculosis* that can activate the antituberculosis drug isoniazid[84]. However, the requirement of another substrate might imply another function in addition to hydrogen peroxide removal, for example the oxidation of toxic reductants and thus, their detoxification[29]. HP0485 ability to oxidize different compounds could reflect a versatile and flexible function within the cell, and this is certainly an intriguing aspect that needs further exploration. Our efforts are underway to clarify the precise function of *Hp* catalase-like protein and understand its contribution to *H. pylori* antioxidant defence.

Materials & Methods

Homology modelling

Homology modelling of *H. pylori* catalase-like was performed with SWISS-MODEL[85] using the structure of *H. pylori* KatA (PDB: 2A9E) as template. Images were created with the PyMOL Molecular Graphics System program.

Cloning and expression of recombinant *H. pylori* catalase-like

The HP0485 gene was amplified by PCR from genomic *H. pylori* G27[86], with primers 5'-ATATAT<u>CCATGG</u>GACATGAAGTGAGCGCTG-3' (forward, NcoI recognition site underlined) and 5'-ATATAT<u>CTCGAG</u>TTGCCTTCTGCTAAAAG-3' (reverse, XhoI recognition site underlined), using Phusion DNA Polymerase (Thermo Scientific). In this way the signal peptide at the N-terminal of the sequence predicted with SignalP[55] was removed. The amplified HP0485 gene was cloned into the pET22b(+) vector (Novagen) in frame with N-terminal PelB sequence and C-terminal His-tag. The integrity of the construct was confirmed by DNA sequencing.

20 ml of an O/N culture (40ml, 37°C) of *E.coli* BL21(DE3) cells harbouring the pET22-HP0298 plasmid were used to inoculate 1 litre of Luria-Bertani (LB) medium containing 100 µg ml^{-1} ampicillin. Cells were grown for 2.5 hours at 37 °C until OD_{600} of 0.6-0.8 was reached. HP0485 expression was induced with 0,1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and cells cultivation for another 20 hours at 30 °C. Culture was harvested at 4°C and stored at -80 °C.

Periplasmic purification

Hp catalase-like was isolated from the periplasmic fraction. To isolate the periplasm, after the cells from 1 litre culture had been spun down, the cell pellet was resuspended in 10 ml osmotic shock buffer (500 mM sucrose, 100 mM Tris, and 1 mM EDTA, pH 8) and incubated O/N at 4°C on a rotating shaker, followed by centrifugation at maximum speed for 45 min. The supernatant, containing the periplasmic extract was subjected to affinity chromatography with TALON metal affinity resin (Clontech). Briefly, the periplasmic fraction was mixed with 1-2 ml of resin and incubated for 45 min at 4°C on a rotating shaker. The sample was packed in a polypropylene tube and washed four or five times with 10 resin volumes of wash buffer (50 mM Tris, 10 mM imidazole, pH 8.0). The resin was transferred to a plastic column, washed with another resin volume of wash buffer, and finally the bound protein was eluted three times by the addition of 1 ml of elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, pH 7.4). The protein fractions were pooled and dialyzed at 4°C against 4000 volumes of pH 7.4 storage buffer consisting of 50 mM NaH_2PO_4, 300 mM NaCl. Final concentration and purity were estimated through UV-visible spectra ($\varepsilon_{280} = 37.485$ $M^{-1} cm^{-1}$) and SDS-PAGE respectively. The protein was stored at 4 °C.

Spectroscopic measurements

Spectroscopic analyses were performed at room temperature in pH 7.4 storage buffer containing 50 mM NaH_2PO_4, 300 mM NaCl in a quartz cell with a 1 cm path length (105.201-QS from Hellma), using a JASCO V-750 spectrophotometer. In some experiments, the enzyme was reduced with few grains of sodium dithionite (DTO) or 5 mM tris (2-carboxyethyl) phosphine (TCEP).

CD spectra were collected using JASCO J-715 spectropolarimeter equipped with a Peltier thermal controller. Measurements were performed in far-UV (wavelength range form 250-200 nm) in 10 mM NaH_2PO_4 pH 7.4 at room temperature, using a 1 mm path quartz cuvette (QS-110 from Hellma). Three runs were accumulated with 0.4 mg/ml protein, a bandwidth of 0.2 nm and a response time of 8 s at a scan speed of 50 nm/min. Each spectrum is subtracted of the buffer contribution.

Size exclusion chromatography

The native monomeric structure of *Hp* catalase-like was determined by gel filtration chromatography using a Superdex 200™ 10/300 GL (GE Healthcare), equilibrated with 50 mM NaH_2PO_4, 300 mM NaCl, pH 7.4. The His-tagged HP0485 eluted as a single peak, corresponding to a monomer of 35 kDa. The system was operated on an AKTA FPLC instrument (GE Healthcare). Prior to perform the experimental measures the system was calibrated using markers of different molecular weight: dextran (2000 kDa), aldolase (158 kDa), conalbumine (75 kDa), ovalbumin (44 kDa) and myoglobin (17 kDa).

Chemical cross-linking

Chemical cross-linking was performed to determine the oligomeric state of *Hp* catalase-like in comparison with *Hp* KatA. The reactions were performed by adding 5 mM and 10 mM of glutaraldehyde to protein solutions containing 7 μM of HP0485 or KatA, in 50 mM NaH_2PO_4, 300 mM NaCl, pH 7.4. Each reaction was incubated at room temperature for 2 min and then quenched by the addition of glycine to 300 mM. The samples were resolved by SDS-PAGE (12% gel) and the cross-linked products were visualized by Coomassie Brilliant Blue staining.

Peroxidatic activity assays

The kinetic assays were conducted in 50 mM NaH_2PO_4, pH 6 at room temperature, using a JASCO V-750 spectrophotometer.

Peroxidatic activity of *Hp* catalase-like was determined with assays containing H_2O_2 and one of the reducing substrates: pyrogallol, guaiacol, ABTS or ascorbate (all purchased from Sigma). The activity with pyrogallol and guaiacol was determined at 420 nm for pyrogallol and 470 nm for guaiacol, which corresponds to the absorbance peaks of purpurogallin and tetraguaicol products with extinction coefficients $\varepsilon_{420} = 2640$ M^{-1} cm^{-1} and $\varepsilon_{470} = 26600$ M^{-1} cm^{-1} respectively. The assays contained 50 mM H_2O_2 and 0.1-10 mM pyrogallol or 1-40 mM guaiacol. For ascorbate the activity was followed as the decrease in A_{290} due to the consumption of ascorbate in an assay containing 50 mM H_2O_2 and 0.05-1 mM ascorbate ($\varepsilon_{280} = 2800$ M^{-1} cm^{-1}) in 50 mM NaH_2PO_4, pH 6, containing 0.1 mM EDTA. The ABTS assay followed the formation of the ABTS radical cation at 420 nm ($\varepsilon_{420} = 36000$ M^{-1} cm^{-1}) in a solution containing 50 mM H_2O_2 and 0.01-0.12 mM ABTS. These assays were also performed for HRP (Sigma) at the same experimental conditions, with the exception of a 10 mM H_2O_2 constant concentration.

The ABTS assay was also performed to compare HP0485 ability to utilize different peroxides. For these kinetics ABTS concentration was 0.1 mM, and H_2O_2 or t-butyl hydroperoxide concentration was varied (0.1-200 mM for H_2O_2 and 0.1-100 mM for t-butyl hydroperoxide).

The total volume of each assay was 1 ml in a quartz cuvette with a 1 cm path length. All the assays were started by enzyme addition. The exact final concentration of enzyme added varied for each reducing substrate used. In particular, a concentration of 5 nM in ABTS oxidation assays, and of 0.2 μM in oxidation assays with the other one-electron donors, generated a satisfactory linear absorbance change over the 3±5 min assay period. Activity was calculated

from the gradient of the change in absorbance with time (ΔA) divided by the molar extinction coefficient (ε). The kinetic data were then fit to the Michaelis-Menten equation with substrate inhibition (for HP0485 with guaiacol, ABTS and ascorbate) or without substrate inhibition (for HP0485 with pyrogallol and HRP with all substrates) to estimate the kinetic parameters.

Site direct mutagenesis, cloning, expression and purification of *Hp* catalase-like mutants

The N44V, N44A and L113N HP0485 mutants were obtained using 30 ng of pET22-HP0485 construct as template. We employed site-direct mutagenesis by primer extension, in which flanking primers (complementary to the ends of the target sequence) and internal primers (containing the mutate bases) were required, as described by Steffan *et al.*[87]. Briefly, during the first round of PCR two reactions were assembled using one flanking and one internal primer. The products were then mixed for the second PCR reaction that was assembled with the flanking primers. The complementary ends of the products hybridized in this PCR reaction to create the final product, which contains the mutated internal sequence. For all mutants the flanking primers were those used for HP0485 amplification. For N44V the internal primers were 5'-CTAAAATGAAAATC**GTG**CACACTAAGGGG-3' (forward, mutation in bold) and 5'-CCCCTTAGTGTG**CAC**GATTTTCATTTTAG-3' (reverse, mutation in bold). For N44A the internal primers were 5'-CTAAAATGAAAATC**GCG**CACACTAAGGGG-3' (forward, mutation in bold) and 5'-CCCCTTAGTGTG**CGC**GATTTTCATTTTAG-3' (reverse, mutation in bold). For L113N the internal primers were 5'-CAATGGTGATG**AAC**AATACAGAAATC-3' (forward, mutation in bold) and 5'- GATTTCTGTATT**GTT**CATCACCATTG-3' (reverse, mutation in bold).

As for the HP0485 wild-type sequence, the mutant sequences were cloned into the pET22b(+) vector (Novagen) in frame with N-terminal PelB sequence and C-terminal His-tag. The integrity of the constructs were confirmed by DNA sequencing prior to transform *E.coli* BL21(DE3) cells. Strategy of expression and purification of *Hp* catalase-like mutants was in accordance with the procedure described for the wild-type catalase-like.

Catalase activity assay

Hp catalase-like and its mutants were assayed for catalase activity by measuring the decrease at 240 nm due to the consumption of H_2O_2 in 50 mM NaH_2PO_4, pH 6, at room temperature, using a JASCO V-750 spectrophotometer.

Preparation of apo *Hp* catalase-like

The heme was extracted from *Hp* catalase-like using 2-butanone/HCl, according to the method of Teale[65]. 300 µl of 150 µM purified HP0485 were adjusted with 0,1 M HCl to pH 1.7-2, followed by the addition of two volumes of 2-butanone. The mixture was shaken briefly and was allowed to stand at room temperature for 1-2 min. The aqueous fraction containing apo catalase-like was dialyzed first against 1000 volumes of Milli-Q water to remove any remaining ketone, then against 1000 volumes of 50 mM NaH_2PO_4, 300 mM NaCl, pH 7.4.

Reconstitution of apo protein

For heme reconstitution 130 µl of 12 µM apo protein in 50 mM NaH_2PO_4, 300 mM NaCl, pH 7.4, were titrated with 0.3 µl aliquots of a solution of hemin (Sigma) in 0.1 N NaOH to 100 nM-20 µM final concentration, at room

temperature. Samples were mixed by inversion and spectra were recorded immediately. For peroxidatic assay with H_2O_2 and ABTS coupled to the hemin titration of the apo protein, 1 µl aliquots of apo Hp catalase-like (7 µM) in 3 µl of storage buffer (50 mM NaH_2PO_4, 300 mM NaCl, pH 7.4), were incubated with 1 µl aliquots of hemin at concentrations ranging from 1 µM to 80 µM at room temperature for 15 minutes. After that the peroxidatic activity of 5 nM of the protein from the sample was determined.

Phylogenetic reconstruction

The midpoint-rooted circular phylogram representing a maximum-likelihood phylogeny of *H. pylori* catalase-like and related proteins was created with the ClustalW program[88] and visualized with FigTree. Multiple sequence alignments were built with ClustalW, analyzed with Genedoc, and the online resource Espript[89] was used to generate alignment figures.

Supplementary material

>WP_029647928.1 .
-----MKKISLSLCLVFSLGFLKAH
>WP_013023348.1 | catalase-like [Helicobacter mustelae].
---MKKISLSVCAVLFSCMGFAHAH
>WP_039629043.1 catalase-like [Campylobacter lari].
----MKKYISSCLAICCLSSAIYAN
>WP_034554113.1 | catalase-like [Helicobacter apodemus].
--MKNKKYLSLALACCLLGSIALAQ
>WP_027327717.1 | catalase-like [Helicobacter pametensis].
---MKKSLSLIVAALLCGSVAGYAQ
>WP_042960407.1 | catalase-like [Campylobacter fetus].
----MNKVFSFCAAACLFVSALNAE
>WP_040462952.1 | catalase-like [Helicobacter bilis].
------MSKLFVSTLFVSTMCVYGN
>WP_005022233.1 | catalase-like [Helicobacter pullorum].
MLKNFKKVLTICASTCLLGTFVNAQ
>WP_006655894.1 catalase-like [Helicobacter canadensis].
MLKNFKKVLTICASTCLLGTFVNAQ
>WP_013469182.1 | catalase-like [Helicobacter felis].
---MQRLKRVLFPLLGVFCLCAHAT
>WP_033788399.1| catalase-like [Helicobacter pylori].
-----MKKIGLSLGLIFSLGFLKAH
>WP_011577620.1| catalase-like [Helicobacter acinonychis].
-----MKKICLSLGLIFSLGFLKAH
>WP_053945167.1| catalase-like [Helicobacter ailurogastricus].
----MPKFACSLLLPLVLLSLTHAK

Figure S1: **Signal peptides from HP0485 and homologous protein sequences, aligned by their cleavage site**.
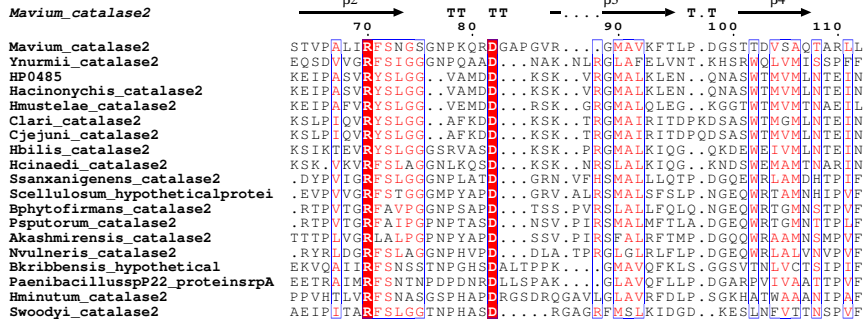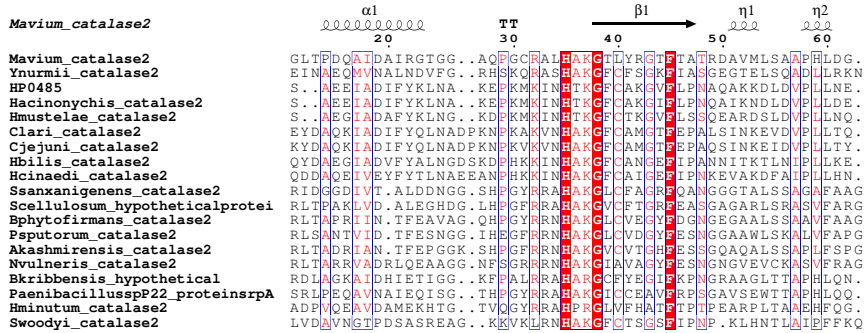
Figure S2: **Selected parts of the multiple sequence alignment of 19 catalase-like proteins**. Secondary structure elements inferred from the crystal structure of *M. avium* catalase-like (PDB: 3E4W) are shown above the alignment. (A) Distal side of the prosthetic heme group. (B) Proximal side of heme. Upwards pink triangles indicate the essential histidine and tyrosine residues; upwards cyan triangles indicates the other heme binding residues (same residues highlighted in figure S3).

A

Hpylori_catalase

ℓℓℓℓ   β1   η1   TT
          50   60   70   80

```
Hpylori_catalase             LAAFDRERIPERVVHAKGSGAYGTFTVTK.....DITKYTKAKIFSKVGK
Hacinonychis_catalase        LAAFDRERIPERVVHAKGSGAYGTFTVTK.....DITKYTKAKIFSKVGK
Hcinaedi_catalase            LAAFDRERIPERVVHARGSGAYGVFEATA.....DISKYSKAAVF.KKGT
Hmustelae_catalase           LAHFDRERIPERVVHAKGSGAYGTFTVTN.....DITKYTKAKIFSKVGK
Hbilis_catalase              LAHFDRERIPERVVHAKGSAAYGELTITN.....DITMYSKADIFSQVGK
Clari_catalase               LAHQNRERIPERTVHAKGSGAYGELKITK.....DISQYTKAKVL.QLGE
Cjejuni_catalase             LAHQNRERIPERTVHAKGSGAYGEIKITA.....DLSAYTKAKIF.QKGE
Saerolata_catalase           LAHFNRERIPERVVHAKGAGAYGVFRVTR.....DVSAWTTAAFLNKVGK
PseudomonasspVLB120_catalase LAHFNRENIPERVVHAKGAGAHGYFEVTN.....DITHYTHAKLFSEVGK
Bkribbensis_catalase         LAHFNRERVPERVVHAKGAGAHGYFEVTN.....DMSKYTKAKLFHSEGK
Akashmirensis_catalase       MAHFNREKVPERQPHAKGAFGVFETTE.....DVSKYTKAALF.QKAC
Nvulneris_catalase           LAQFNRERVPERIVHAKGAGAYGELVVTN.....DVSRYTKAKLF.QPGA
Mplatani_catalase            ITHFDRERIPERVVHARGAGAHGYFEAYGTIGGEPASKYTRARVLNETGV
Hminutum_catalase            LARFDRERIPERVVHARGVGAYGTFESYG.....DFSQLTKASLLTAKGK
Aaustralica_catalase         ITHFDHERIPERVVHARGSGAHGYFECTE.....DITALTRAAPFAKVGK
Scellulosum_catalase         ITHFDHERIPERVVHARGSGAHGYFECTE.....DITSLTKAAPFAKVGK
PaenibacillusspP22_catalase  MTHFDHERIPEIVHARGFCAHGYFQVVYK.....PMAELTEAGFLQDPSV
Ssanxanigenens_catalase      LAHFDREVIPERRMHAKGWGAYGSFTVTH.....DITRFTRAKIFSEVGK
Swoodyi_catalase             LAHFDREVIPERRMHAKGSGAFGFRFTVTH.....DISEFTKAKLFESVGK
Ynurmii_catalase             LAHFDREVIPERRMHAKGSGAYGTFTVTH.....DITKYTRAKIFSQIGK
Bphytofirmans_catalase       LAHFDREVIPERRMHAKGSGAFGTLKVTH.....DISRYTKAKVFAQVGT
Psputorum_catalase           LAHFDREVIPERRVHAKGSGAFGKLTITH.....DITKYTRARVFEAVGK
```

β2   TT   β3   TT   β4
          90   100   110   120   130

```
Hpylori_catalase             KTECFFRFSTVAGERGSADAVRDPRGFAMKYYTEEGNWDLVGNNTPVFFI
Hacinonychis_catalase        KTECFFRFSTVAGERGFADAVRDPRGFAMKYYTEEGNWDLVGNNTPIFFI
Hcinaedi_catalase            KTPMFIRFSTVAPELGSADAVRDPRGFAVKFYTTEGNWDLVGNNTPIFFI
Hmustelae_catalase           KTPCFFRFSTVAGERGAADAERDPRGFAMKYYTEEGNWDLVGNNTPIFFI
Hbilis_catalase              KTKAFLRFSTVAGERGAADAERDVRGFALKFYTNEGNWDLVGNNTPIFFI
Clari_catalase               NTPLFIRFSTVAGEAGAADAERDVRGFAIKFYTKEGNWDLVGNNTPIFFI
Cjejuni_catalase             VTPLFLRFSTVAGETGAADAERDVRGFAIKFYTKEGNWDLVGNNTPIFFI
Saerolata_catalase           ETETFLRFSTVGGEKGSADAERDPRGFALKFYTEEGNYDLVGNNTPVFFV
PseudomonasspVLB120_catalase QTETFLRFSTVGGERGSADTERDPRGFAVKFYTEEGNWDLVGNNTPVFFI
Bkribbensis_catalase         RTPMFIRFSTVAGELGSADTVRDPRGFAVKFYTEDGNYDLVGNNTPVFFI
Akashmirensis_catalase       KTEMLARFSTVAGESPDTWRDVRDSLKFYTDEGNYDLVGNNTPVFFV
Nvulneris_catalase           RTESLVRFSTVAGEQGSPDTWRDPRGFAVKFYTEDGNYDLVGNNTPIFFI
Mplatani_catalase            KTPMFVRFSTVAGAKESPETARDPRGFAVKFKTVDGNWDLVGNNLKVFFI
Hminutum_catalase            KTPMFVRFSTVIHPQGSPETLRDPRGFALKFYTDEGNWDLVGNNLPVFFI
Aaustralica_catalase         KTPVFVRFSTVAGERGSADTVRDVRGFAVKFYTDEGNWDLVGNNIPVFFI
Scellulosum_catalase         RTPVFVRFSTVAGERGSADTVRDVRGFAVKFYTEEGNWDLVGNNIPVFFI
PaenibacillusspP22_catalase  KTPVFVRFSTVAGSRGSADTVDVRGFAVKFYTEEGNFDLVGNNMPVFFI
Ssanxanigenens_catalase      TTDLFMRFSSVAGERGAADAERDIRGFAIKFYTEEGNWDLVGNNTPVFFF
Swoodyi_catalase             ETEVFVRFSTVAGERGAADAERDIRGFALKFYTEQGNWDLVGNNTPVFFL
Ynurmii_catalase             KPDMFVRFSTVAGERGAADAERDIRGFAMKFYTEEGNWDLVGNDTPVFYL
Bphytofirmans_catalase       ETPLFMRFSTVAGERGAADAERDVRGFSIKFYTEEGNWDLVGNNTPVFFI
Psputorum_catalase           ETPVFLRFSTVAGERGAADAERDVRGFSLKFYTDEGNWDLVGNNTPVFFI
```

B

Hpylori_catalase

ℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓ   η7   η8   TT
α9
          340   350   360   370

```
Hpylori_catalase             QGRLFSYGDTHRYRLG.VNYPQIPVN.KPRCP.FHSSSRDGYMQN.GYY.
Hacinonychis_catalase        QGRLFSYGDTHRYRLG.VNYAQIPVN.KPRCP.FHSSSRDGHMQN.GYY.
Hcinaedi_catalase            QGRLFAYGDTQRYRLG.INHHQIPIN.APKSA.LNSSVRDGYMTN.GAY.
Hmustelae_catalase           QGRLFAYGDTARYRLG.INHGQLPVN.APKVP.VHNTHRDGFMQN.GNY.
Hbilis_catalase              QGRLFSYGDTQRYRLG.INHSQLPVN.AAIAS.VSNTHRDGYMQR.GSF.
Clari_catalase               QARIFSYPDAHRYRIG.TNYHLLPVN.RAKSE.VNTYNVAGAMNFDTYK.
Cjejuni_catalase             QARIFSYPDAHRYRIG.TNYHLLPVN.RAKSE.VNTYNVAGAMNFDNYK.
Saerolata_catalase           QGRLFSYADAHRYRLG.GNYGLLPVN.KPHAAPVHSYQRDGAMRFDGNG.
PseudomonasspVLB120_catalase QGRVFAYADAHRYRVG.TNHQQIPVN.APR.CPMHSYQRDGSMAN.GSY.
Bkribbensis_catalase         QGRLFAYSDAHRYRVG.ANHNSLPIN.RPKSD.VNNYQRDGQMRFDTNG.
Akashmirensis_catalase       LGRAFAYNDAQRNRIG.TNFHQLPVN.RPKVP.VNTYMFDGQMAYEHS..
Nvulneris_catalase           LGRVFAYADAHRYRLG.TNYAQLPPN.LPRAAQVNSYSKEGAMRYFYND.
Mplatani_catalase            QGRTLSYSDTQRYRVG.PNYLQLPINAPAPGVKTATNQRDGQMTFYVDGV
Hminutum_catalase            QGRLFSYADTQRYRVG.ANYQSLPVNRSRATVN..NNNQGGAMNIAHT..
Aaustralica_catalase         QGRNFSYLDTQLITRLGGPNFHLLPIN.APVAQ.VHNNQRDGFHQHGIFR.
Scellulosum_catalase         QGRNFSYLDTQLTRLGGPNFHELPIN.APVAE.VHNNQRDGFHRQAINR.
PaenibacillusspP22_catalase  QGRLFSYTDTQLSRLGGPNFHELPIN.RPMCP.VHNNQRDGMHRMTINR.
Ssanxanigenens_catalase      QARLFSYPDAARYRLG.VNHDHIPVN.AARCP.VRSYHRDGQMRVDGNA.
Swoodyi_catalase             QGRLFSYGDSQRYRLG.VNHSHIPVN.APKCP.FHSYHRDGQMRTDGNA.
Ynurmii_catalase             QGRLFSYGDAHRYRLG.VNHHQIPVN.GAKCP.FHNYHRDGAMRVDGNS.
Bphytofirmans_catalase       QGRLFSYGDTQRYRLG.INHHQIPVN.AARVPSPHSFHRDGAMRTDGNL.
Psputorum_catalase           QGRLFSYGDTQRYRLG.VNHSAIPVN.APKCPFHHSFHRDGAMRVDGNL.
```
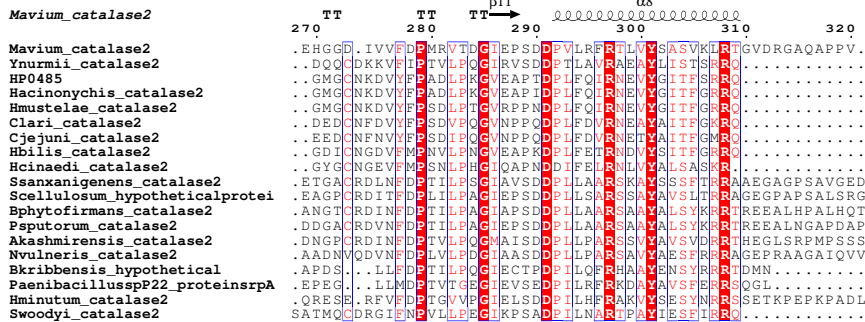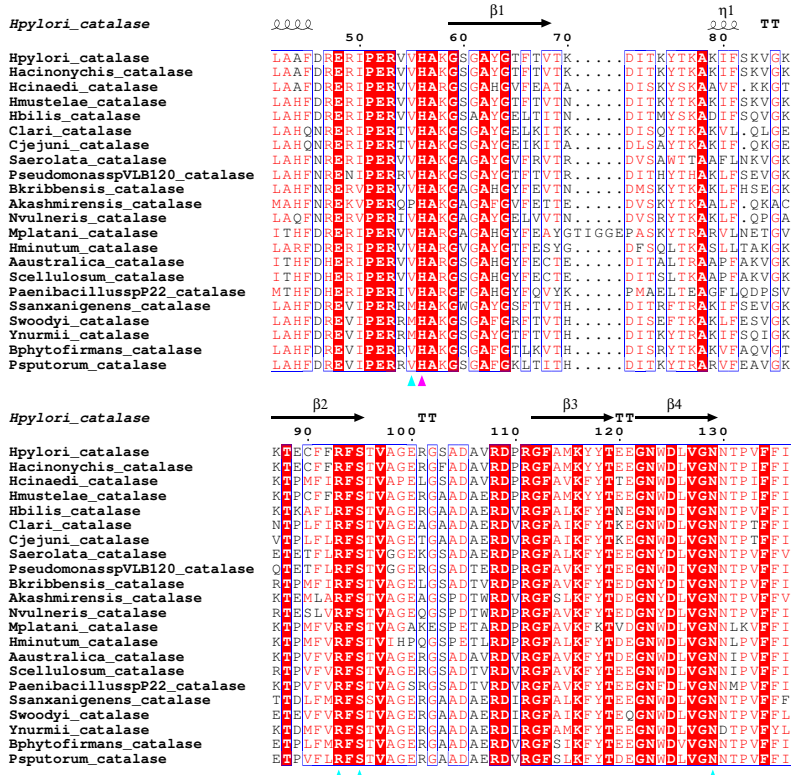
Figure S3: **Selected parts of the multiple sequence alignment of 22 monofunctional catalases**. Secondary structure elements inferred from the crystal structure of *Hp* KatA (PDB: 2A9E) are shown above the alignment. (A) Distal side of the prosthetic heme group. (B) Proximal side of heme. Upwards pink triangles indicate the essential histidine and tyrosine residues; upwards cyan triangles indicates the other heme binding residues (same residues highlighted in figure S2).
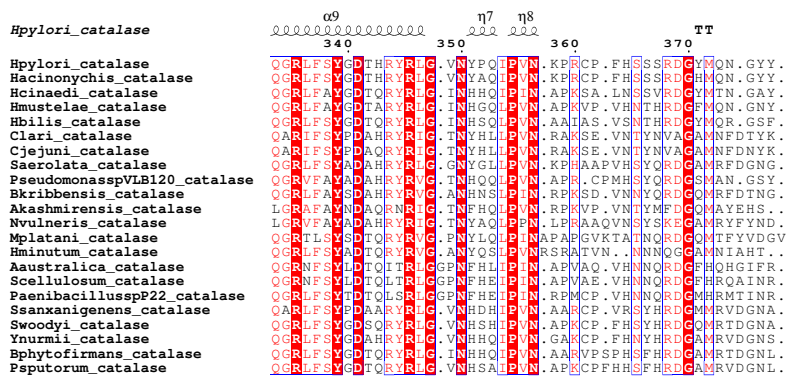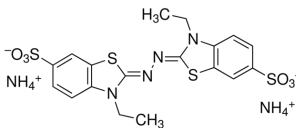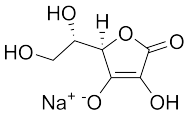
Substrate	Chemical structure	Kinetics λ (nm)	Product	ε (mM cm)$^{-1}$
Pyrogallol		420	Purpurogallin	2,64
Guaiacol		470	Tetraguaiacol	26,6
ABTS		420	ABTS radical cation	36
Sodium ascorbate		290	Dehydroascorbate	2,8

Table S1: **One-electron donors substrates assayed for *Hp* catalase-like *peroxidatic* activity**.

Chapter 3

Functional characterization of the gastric specific periplasmic binding protein HP0298

Abstract

As essential components of ABC importer systems in Gram-negative bacteria, periplasmic binding proteins (PBPs) have a crucial function in nutrient uptake and are often related to the environment wherein the microorganisms live. The highly adapted human pathogen, *Helicobacter pylori*, requires a number of amino acids for growth, many of which are hydrophobic. Its dipeptide binding protein, DppA (HP0298), part of the ABC dipeptide transport system (*dppABCDE*) was previously assessed for peptide uptake, but its function is not fully understood. The protein shares 39% identity with the well-characterized *E. coli* DppA and differs significantly from all the other PBPs of the same group. HP0298 is closely related to proteins of other *Helicobacter* species that are all gastric pathogens in different hosts. To gain insight on *H. pylori* (*Hp*) DppA physiological role we set up a protocol of "ligand-fishing" assay coupled with mass spectrometry aimed to purify the specific PBP ligand from the native environment in which the binding occurs. A recombinant His-tagged DppA was produced, and incubated with *H. pylori* cell extract; the in-solution formed His-tagged protein-ligand complex was trapped and purified by affinity chromatography, and then subjected to HPLC-MS analysis. The compounds potentially interacting with HP0298 were pentapeptides, rich in hydrophobic amino acids and containing at least one negative charged residue. Considering the nutrient requirement of *H. pylori* and the high peptide content of its unique niche, we conclude that *Hp* DppA could be involved in the uptake of gastric

peptides derived from food proteins, with specificity based on length and amino acid composition.

Because the PBP domain-containing proteins not only function as the initial receptors in metabolite transport in bacteria, but also they are employed in several signal transduction pathways in mammals (including humans), this experimental method could be extended to other still orphan receptors as initial interaction screening for ligand identification.

Introduction

Periplasmic binding proteins (PBPs) are a large bacterial family of proteins serving as primary receptors for a large number of metabolites. Located in the periplasm of Gram-negative bacteria, PBPs mediate substrate transport or induce chemotaxis towards the nutrient source[90]. In both cases, they are involved in the active transport of solutes inside the bacterial cell, as constituents of the multicomponent transport systems ATP-binding cassette (ABC transporters)[91,92]. In general, the ABC transport systems for solutes uptake consists of two transmembrane domains, which assist in ligand translocation across the inner membrane, two peripheral membrane ATP-binding domains, responsible for energizing the systems, and a periplasmic binding protein, which binds its ligand with high affinity and delivers it to the translocator[93].

PBPs bind a broad spectrum of diverse molecules, ranging from amino acids, carbohydrates, amines, metal ions, di- and oligo-peptides and vitamins. Besides the wide variation in ligand size and chemical functionality, PBPs have different sizes (25 to 70kDa) and little sequence similarity, but they share an extremely conserved three-dimensional structure (Fig. 1)[94]. They are monomers formed by two large polypeptide lobes connected by a flexible hinge region with a large cleft, wherein the ligand binds. Ligand binding induces a conformational change such that the two lobes twist and close to entrap the ligand and the protein undergoes the transition from the open conformation to the closed one. This model of binding has led to the term "Venus flytrap" mechanism[95,96]. In addition to ligand binding, hinge bending in the PBP fulfils a key role in signal transduction. It enables the formation of a ligand-stabilized closed structure which is presumably recognized, in preference to the open structure, by the appropriate membrane-bound components and, thus, initiates nutrient
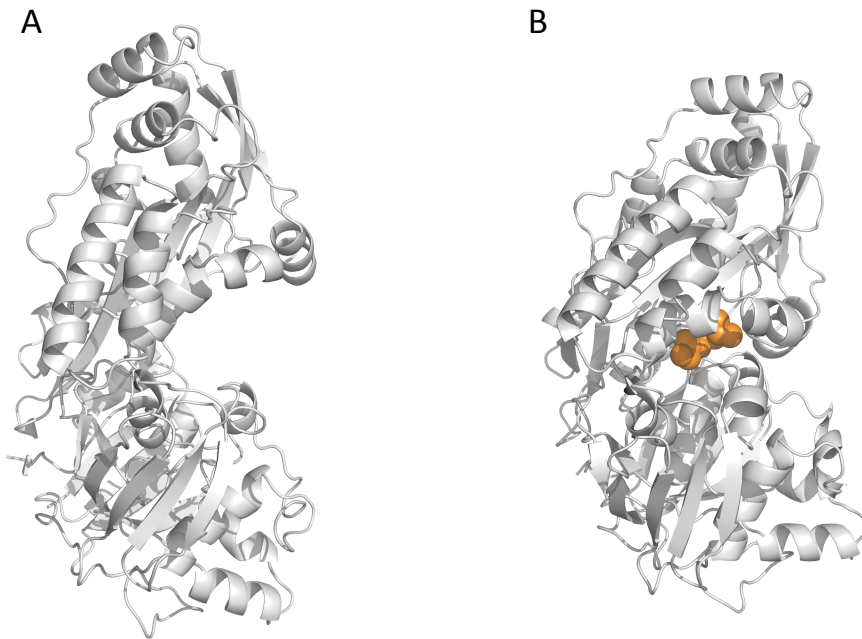
Figure 1. **Structure of *E. coli* DppA**. (A) Structure in open unliganded conformation (PDB: 1DPE) (B) Structure in closed liganded conformation with the bound Gly-Lys dipeptide (PDB: 1DPP). Structures in gray cartoon rendering with ligand in orange spheres.

translocation or chemotaxis[97]. The PBP fold is also used inside the cell as the ligand binding domain of many transcription factors, including the classical Lac repressor[98].

Even though the PBPs are restricted to Gram-negative bacteria, homologous proteins exist in Gram-positive bacteria as extracellularly bound lipoproteins[99], and in Eukaryotes as ligand binding domain of diverse receptors. These include glutamate/glycine-gated ion channels such as the NMDA receptor, G protein-coupled receptors, including metabotropic glutamate, GABA-B, calcium sensing, and pheromone receptors, and atrial natriuretic peptide-guanylate cyclase receptors[100]. All these receptors, arisen via gene fusion events between genes encoding the PBPs and genes for integral membrane proteins, are regulated by conformational changes of the protein extracellular domain in response to ligand binding. Thus, the PBP module is exceptionally widespread and thanks to the extensive information available about three-dimensional structure and mechanism of ligand binding, it can serve as a guide in drug design and development.

In bacteria, PBPs together with the other members of the ABC transporter are important virulence factors because they play roles in nutrient uptake. It has been demonstrated that transporters encoded in bacterial genomes correlate with the physiological niches the bacteria live in[101], thereby playing a role in the adaptation to a specific environment and its available sources.

Helicobacter pylori, a highly adapted human pathogen, possess diverse ABC transport systems for nutrient uptake, by which he depletes the host of essential elements. Studies focusing on the nutrients requirements of *H. pylori* for growth have shown that the bacterium is auxotrophic for several amino acids, including methionine, alanine, histidine, isoleucine, and valine, and consequently dependent on a steady supply of these growth factors from the environment[102]. Furthermore, despite its capability to grow on glucose[103], *H. pylori* does not have

the enzymatic machinery to break down more complex carbohydrates, and is thus limited in carbohydrates utilization[104]. *H. pylori* can actually survive employing amino acids as the basic nutrients and the sole carbon and nitrogen energy sources[105]. In addition to transporting amino acids, *H. pylori* may have the ability to transport the abundant peptides, which are found in the stomach.

Two ABC transporters have been classified for peptide uptake, one for dipeptides (dppABCDF) and the other for oligopeptides (oppABCD)[47]. Both of these systems are extremely conserved in *H. pylori* strains, indicating their essential nature[106].

A previous study showed that the dipeptide system mutants (in particular, *dppA*, *dppB*, *dppC* mutants) lacked the ability to use only certain dipeptides, hexapeptides, and nonapeptides, while the oligopeptide system mutants (*oppA*, *oppB*, *oppC* mutants) were not particularly affected, or only partially, in their peptide uptake capability[107]. Since the *dpp* gene cluster is homologous to the *nik* operon of *Escherichia coli*, *H. pylori* DppA was assessed for a role in nickel import, but the hypothesis has been ruled out, because the gene product showed no contribute to the synthesis of catalytically active urease[108].

Based on sequence homology analyses, the bacterial PBP superfamily has been classified into 8 clusters, and *H. pylori* DppA was assigned to cluster 5, comprising dipeptide (DppA family), oligopeptide (OppA family) and nickel specific PBP's (NikA family)[90]. Continuous family updates by the Transporter Classification Database has now led to a cluster 5 PBPs containing up to 42 different subfamilies that are associated with translocation cargos that can vary from di- and oligo-peptides, and nickel substrates to antimicrobial peptides, δ-aminolevulinic acid, heme, plant opines, carbohydrates, glutathione and others.

Investigation of PBP specific ligand has proven to be a challenge for several reasons. Although many PBP structures are available, it is almost impossible to make an accurate prediction about the exact substrates that are bound, owed to the versatile nature of the PBPs that usually can accommodate different ligands

within their binding pockets. A similar problem is encountered when performing ligand-binding studies. Many PBPs present comparable affinity for different substrates, and the discrimination between the physiological ligand and an unspecific one is challenging. Often, analysis of the genome context of an ABC transporter helps by finding linked genes that give clues to function, but this is not applicable in many cases. Besides, many PBP genes are misannotated with a wrong attribution in respect to the transported substrate. One example is provided by the glutathione-binding protein (GbpA) of *Haemophilus influenzae* that was initially classified as heme-binding protein[109,110].

Our interest in *H. pylori* DppA (HP0298) was corroborated by our findings that it represents an exclusive trait of pathogenic gastric species and the discovery of its specific substrate could broad the knowledge on an uptake system, with its linked intracellular pathways, that can be relevant for the human host.

This study describes the unbiased method that we employed to investigate its physiological substrate. This method, named "ligand-fishing" is a ligand pull-down assay coupled with mass spectrometry (MS). MS-based approaches are extensively used in interactomics studies for analyzing protein–protein and protein–ligand interactions, and also in the expanding field of metabolomics. MS provides high sensitivity, wide dynamic range and the possibility of distinguishing metabolites by their accurate m/z signals and fragmentation patterns. Chromatography can be easily coupled with MS to differentiate metabolites by their retention times providing an additional dimension for identification. In our protocol the purified recombinant His-tagged DppA is immobilized on cobalt affinity chromatography resin and acts as the bait to capture its specific ligand from its native environment, being in our case, *H. pylori* cell extract. The eluted samples are then subjected to HPLC-MS with Orbitrap mass spectrometer, and the untargeted metabolite profiling of extracts are inspected for mass spectral features that differ significantly between the

protein and control samples. The results could give clues on the specific ligand or a category of substrates that bind to the periplasmic binding protein.

Considering the involvement of PBP domain-containing proteins in several signal transduction mechanisms in mammals and humans, this experimental method could be extended to other still orphan receptors in search of their ligand.

Results

Phylogenesis of *H. pylori* DppA and related PBPs

To reconstruct the evolutionary relation of *H. pylori* DppA within the PBP superfamily a homology research was performed with BlastP. The sequences with significant E values were all PBPs of cluster 5 (DppA-OppA-NikA)[90]. The majority is annotated as generic solute binding proteins (Sbp), but some of them are classified as dipeptide binding proteins (DppA), oligopeptide binding proteins (OppA), heme binding proteins (HbpA), nickel binding proteins (NikA), or glutathione binding proteins (GbpA). 51 sequences were selected for multiple alignment and tree reconstruction (Fig. 2). Among the five main phylogenetic clusters, HP0298 grouped with proteins all belonging to *Helicobacter* species. These proteins appeared to be distantly related to DppA-OppA, NikA, HbpA and GbpA families. Only a few of these proteins have already been characterized, but many of them are putative proteins and often their annotation disagrees with their position in the phylogenesis. For example, many proteins classified as HbpA are found within the GbpA group.

This phylogenesis reflects the principle postulating that periplasmic binding proteins from evolutionarily divergent organisms, which are specific for the same solute are, in general, much more closely related to each other than are proteins specific for different solutes from the same organism[90]. Consistently, the *Helicobacter* species are found in two different groups. *H. canis*, *H. bilis* and *H. hepaticus* are found among NikA group, whereas all the other *Helicobacter* species, including *H. pylori*, form a separate cluster. Noteworthy, these divergent *Helicobacter* species are all gastric pathogens of different organisms, human (*H.
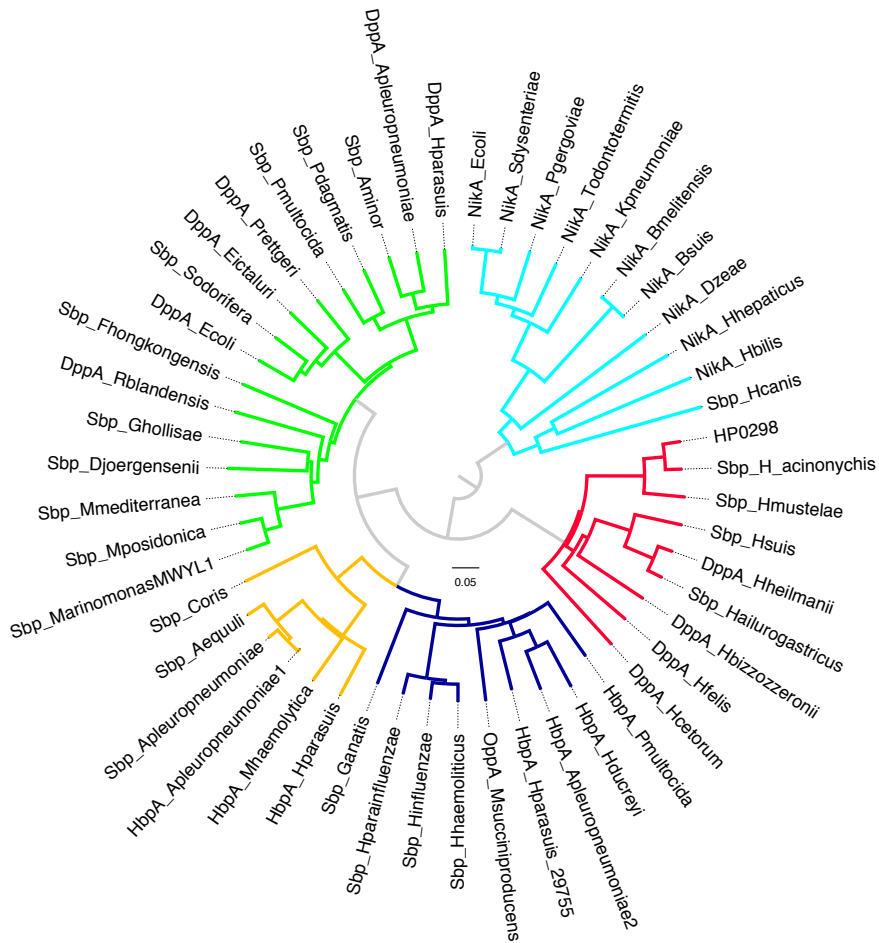
Figure 2. **Phylogenesis of HP0298 and related proteins**. Midpoint-rooted tree representing the evolutionary relations between 51 periplasmic binding proteins of cluster 5 (Tam and Saier, 1993). HP0298 segregates with other PBPs all belonging to gastric *Helicobacter* species (red), and it is distantly related to the other PBP families of dipeptide and oligopeptide binding proteins, DppA-OppA (cyan), nickel binding proteins, NikA (green), putative heme binding proteins, HbpA (yellow), glutathione binding proteins, GbpA (blue)

pylori) and other mammals. This led us to hypothesize that a specific ligand could exist for *Hp* DppA that is linked to the gastric environment.

Expression and periplasmic purification of recombinant *Hp* DppA

At the N-terminal of the HP0298 sequence a signal peptide was predicted (SignalP)[55], that directs the protein to the periplasmic space via the Sec-pathway. According to this analysis, the protein is cleaved at residue 23, leading to a mature form of 61 kDa in the periplasm. For the recombinant expression, the endogenous signal peptide was removed and substituted with a sequence (PelB) recognized as a signal peptide by *E. coli*[57]. The gene sequence was cloned into pET22 plasmid fused with the 6xHis tag at the C-terminal. After HP0298 overexpression, the periplasm was extracted with osmotic shock buffer and the protein purified by cobalt affinity chromatography. To evaluate the obtained preparation both Glycine–SDS-PAGE and Tricine–SDS-PAGE[111] were employed. The latter is the preferred electrophoretic system for the resolution of proteins smaller than 30 kDa and a substantial contamination due to the presence of small peptides was detected. To further remove the contaminants, size exclusion chromatography was performed. Fractions containing HP0298 eluted consistently with a monomer of 61 kDa and yielded a preparation that was over 95% pure (Fig. 3).

pH stability and denaturation studies

In the ligand pull-down assay the protein, in complex with the ligand, is eluted from the cobalt resin, with two different buffers. In particular, prior to elution the resin is split in two aliquots and the release of the protein with the bound ligand is attained through competition by the addition of imidazole, or through
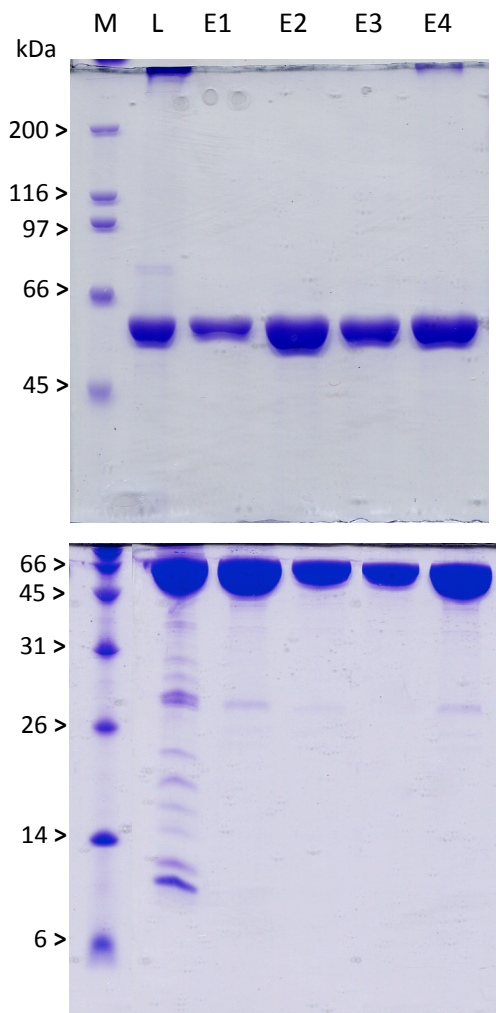
Figure 3. **Gel filtration chromatography**. The upper gel was obtained with Glycine SDS-PAGE, while the lower gel was obtained with Tricine SDS-PAGE. A higher amounts of protein contaminants is visible from the Tricine SDS-PAGE gel that was extensively removed after size-exclusion chromatography. M, molecular weight Marker, L, protein Loaded on column, E, protein Elutions.
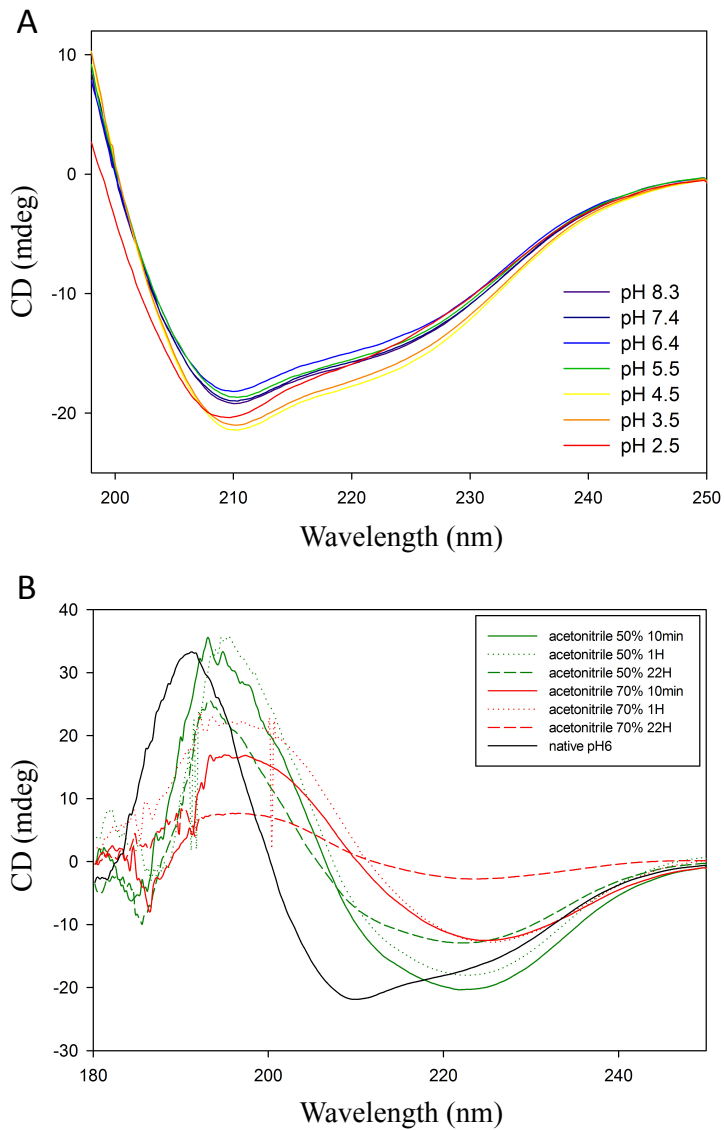
Figure 4. **CD measurements**. (A) Spectra of HP0298 in 10 mM NaH$_2$PO$_4$ ranging from pH 2 to 8.3 at 25 °C. (B) Spectra of HP0298 in 50% and 70% acetonitrile solutions.

the protonation of the tag histidines by a reduction in pH. Two different elution buffers are used in parallel to reduce the possible interference in the subsequent analysis. During the pull-down elution step in native conditions the ligand is still bound to the protein. In contrast, the elution from the HPLC column is performed in a gradient of acetonitrile and the ligand is released.

Before setting the protocol, denaturation studies were performed to evaluate (1) the stability of the protein in different conditions, to confirm that the acid pH elution did not affect the structure of the protein, (2) if the degree of denaturation in acetonitrile was marked enough to ensure the release of the ligand from the protein.

To assess the effect of changes in pH on the secondary structure, CD spectra of HP0298 were recorded from 250–200 nm in buffers ranging from pH 2 to 8.3 at 25 °C (Fig. 3A). CD spectrum appeared that of a mixed α-helix/β-sheet structure as described for the PBP fold. The spectra were collected immediately, after 2 hours and after 18 hours after the dilution in the proper buffer. The protein was stable in all the tested conditions up to 18 hours.

CD measurements were also carried out in denaturant conditions with acetonitrile at two different concentrations (50% and 70%) at 25 °C (Fig. 3B). The spectra were recorded after 10 minutes, after 1 hour and after 22 hours after the dilution in acetonitrile solution. All spectra resulted deeply affected. In 70% acetonitrile, after only 10 minutes, the protein lost the majority of the secondary structure elements.

Ligand pull-down assay

For the ligand pull-down assay a *H. pylori* G27 total cell extract from 50 ml total culture was prepared as described in Material and Methods. The suspension was split into two identical aliquots and mixed with purified His-tagged HP0298 (600 ng; 8 μM) or buffer for the control sample; both samples were further
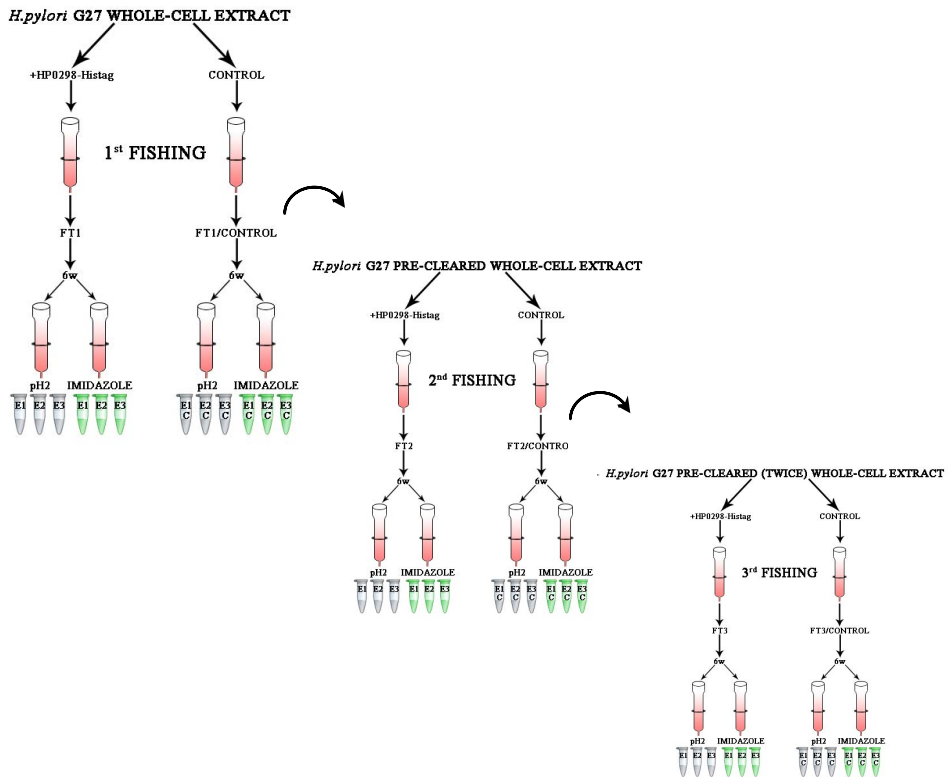
Figure 5. **Ligand pull-down assay**. *H. pylori* cell extract with the prey ligand is split in two aliquots; one is incubated with the protein, and then with the resin, while the other one is incubated only with the resin serving as a control. After recovery of the control from the resin, the obtained precleared extract is again employed for another fishing round. The overall procedure includes three fishing steps to minimize the number of false positives within a single experiment. Prior to elution the resin is split in two aliquots and the release of the protein with the bound ligand is attained through competition by the addition of imidazole, or through the protonation of the tag histidines by a reduction in pH (pH 2).

incubated with Co-resin (Fig. 4). After that, the control was passed through a column, and the cleared lysate was collected for the second round of the fishing experiment. The resin with the protein–ligand complex bound was washed several times and then split for elution with imidazole buffer (50 mM Tris pH 8, 250 mM imidazole) and pH 2 buffer (50 mM Tris pH2). The cleared lysate was split again in two aliquots and mixed with purified His-tagged HP0298 (300 ng; 8 μM) or buffer for the control sample, and the procedure was repeated as described for the first round. A new cleared (twice) lysate was collected and the third round of the fishing experiment was performed (His-tagged HP0298 150 ng; 8 μM). The overall procedure includes three fishing steps to minimize the number of false positives within a single experiment.

Characterization of peptides complexed with *Hp* DppA

To investigate the substrate preference of *Hp* DppA, the eluted samples were analysed by HPLC-MS. Untargeted metabolite profiling of extracts were explored to identify differential m/z features between protein and control samples. We employed XCMS software program to detect the species enriched in the samples with the protein (Table S1). Four m/z peaks with ≥ 5-fold differences between all protein and control samples in both the elution conditions were considered (Table 1). They correspond to m/z 516.3028, m/z 562.2725, m/z 578.2496, m/z 602.3402. When searched against the Metlin[112] metabolite database to make putative metabolite assignment, the m/z values in positive ion mode matched the mass of different tetrapeptides and pentapeptides with ≤ 5 ppm mass accuracy.

To confirm the identity of the compounds MS/MS analysis was performed, followed by *de novo* sequencing with PEAKS software[113]. For each of the m/z values one or two possible pentapeptides were identified with < 2 ppm mass accuracy (Table 2). Notably, all the peptides copurified with HP0298 contained

A

m/z med	RT med	Npeaks TOT	Npeaks Control	Npeaks +HP0298	1 Control	2 Control	3 Control	1 +HP0298	2 +HP0298	3 +HP0298	Log odds	fold	tstat	pvalue
602.340474	1354.82	3	0	3	0.00E+00	1.79E+03	0.00E+00	6.54E+06	3.37E+06	1.06E+06	8.72	6137.7120	2.30	0.15
578.253432	971.46	2	0	2	3.01E+04	7.45E+03	1.89E+03	5.34E+07	9.19E+05	6.32E+06	7.34	1536.2775	1.21	0.35
562.272288	807.36	2	0	2	1.38E+03	0.00E+00	0.00E+00	3.79E+06	1.88E+06	4.44E+05	8.40	4427.6263	2.10	0.17
516.303180	1203.54	2	0	2	1.13E+04	2.55E+03	1.50E+03	2.21E+07	1.13E+07	3.66E+06	7.79	2404.8004	2.31	0.15

B

m/z med	RT med	Npeaks TOT	Npeaks Control	Npeaks +HP0298	1 Control	2 Control	3 Control	1 +HP0298	2 +HP0298	3 +HP0298	Log odds	fold	tstat	pvalue
602.3404	1390.61	2	0	2	0.00E+00	3.54E+03	0.00E+00	9.50E+04	2.31E+06	1.44E+05	6.58	541.1106	1.14	0.34
578.2470	968.16	2	0	2	0.00E+00	0.00E+00	0.00E+00	1.31E+07	1.42E+07	1.86E+07	NA	Inf	9.08	0.01
562.2697	805.96	2	0	2	0.00E+00	0.00E+00	0.00E+00	1.07E+06	9.80E+05	1.33E+06	NA	Inf	10.71	0.01
516.3021	1190.06	2	0	2	0.00E+00	0.00E+00	1.19E+03	9.73E+06	1.17E+07	1.51E+07	1.03	30813.4329	7.71	0.02

Table 1. **XCMS report of the selected m/z features enriched in the samples with HP0298**. (A) Experiment with pH2 elution buffer. (B) Experiment with imidazole elution buffer. Log odds. obtained with the formule "ln[(Σ +HP0298 intensity values)/(Σ control intensity values)]; fold. mean fold change; tstat. Welch's two sample t-statistic; pvalue. p-value of t-statistic.

m/z	z	Mass	Peptide	ppm
516.303	1	515.2955	LGLDV	0.3
562.2725	1	561.2646	LTVDD	1.1
578.2496	1	577.2418	AMEDL	1
602.3402	1	601.3323	LELDL	1.1

Table 2. **Results of *de novo* sequencing of the selected m/z features enriched in the samples with HP0298 obtained with PEAKS software**.
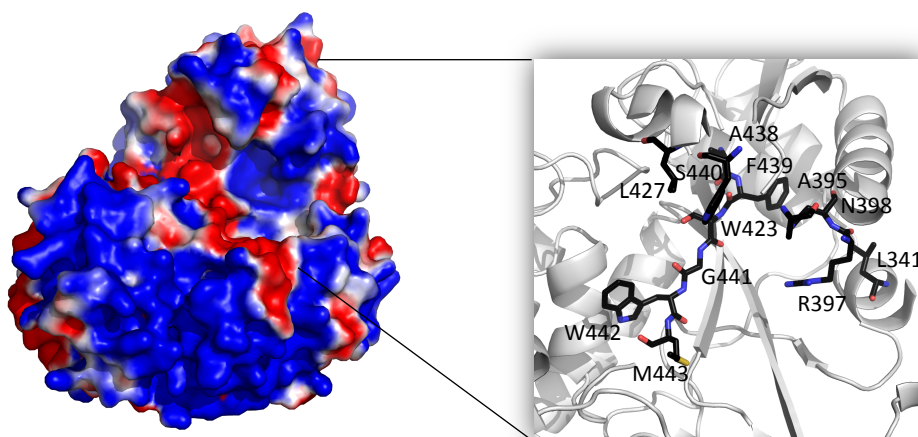

Figure 6. **Homology modelling of *H. pylori* DppA**. HP0485 model was created with SWISS-MODEL using the structure of E. coli DppA (PDB: 1DPE) as template. In the surface charge representation negative charges are shown in red, and positive charges are shown in blue. In the inset, detail of the binding pocket in gray cartoon rendering. Residues lining the cleft are labeled (black carbon sticks).

at least one negatively charged residue. The other amino acids were almost all hydrophobic with abundance of leucine. Since the masses of leucine and isoleucine are the same and the list contains leucine only, a switch between these two residues is possible. All the pentapeptide sequences were verified not to belong either to the HP0298 sequence or the PelB sequence. After that, a homology search of *H. pylori* G27 genome was performed and the pentapeptide sequences were found in some proteins. Since these peptides could be internalized from the external environment and *H. pylori* was cultured in a medium with bovine serum, a search against the Bos taurus proteome was carried out. Several proteins were found that contain the putative sequences: almost all these proteins have a homolog in human.

Surface charge distribution of the peptide-binding pocket of *Hp* DppA

To examine the hydrophobic and electrostatic properties of HP0298 binding pocket a model structure was generated by homology modelling using the structure of *E. coli* DppA as template (Fig. 6). A prevalence of hydrophobic residues is found lining the binding cleft between the two domains. The determined surface charge distribution shows a positively charged surface inside the pocket and on the secondary elements above that are enclosing the pocket upon ligand binding. According to this model, that of *Hp* DppA appears to be an attractive binding site for hydrophobic negative charged peptides.

Generation of *H. pylori dppA*-negative mutant

To further investigate the contribution of *Hp* DppA in the physiology of the cell, we constructed *Hp dppA* knock-out mutant. For this purpose, *H. pylori* strain G27 was transformed with the vector pBS::ΔHP0298, in which most of the

sequence encoding for HP0298 was replaced with a kanamycin resistance cassette. Mutants carrying the correct replacement of the wild type sequence with the kanamycin cassette were confirmed by PCR, and were viable on Brucella agar plates.

Discussion

Periplasmic binding proteins, coupled with inner membrane and ATP-binding proteins in a multicomponent ABC transport system, are important virulence factors for bacteria, as they serve roles in nutrient uptake and chemotaxis[90].

The gastric pathogen *Helicobacter pylori* requires several growth factors and has evolved different mechanisms to utilize the sources provided by it colonizing niche. In particular the bacterium possesses a dipeptide ABC transporter (dppABCDF) with a soluble periplasmic component, DppA (HP0298), homologous to PBPs of cluster 5 (DppA-OppA-NikA). Despite the efforts for the identification of the specific ligand, a definite resolution of the physiological role accomplished by *Hp* DppA and the entire ABC system has not been achieved[107,114], therefore we decided to investigate on the specific substrate of this PBP with a different approach.

To gain insight into the protein phylogeny we constructed a tree with a selection of cluster 5 PBPs and discovered that *H. pylori* DppA was distantly related to the other PBP families and grouped with other proteins all belonging to *Helicobacter* species sharing the same colonizing niche, the stomach, in different host organisms.

For substrate searching in the protein natural milieu we adopted a ligand pull-down strategy coupled with mass spectrometry analysis. The His-tagged recombinant DppA, after purification, was mixed with *H. pylori* extract in which the endogenous ligand should be present, and immobilized on the cobalt resin to be purified in complex with the bound ligand. The MS analysis helped to distinguish between compounds that significantly bound to the protein, and matrix adducts or contaminations. We observed four m/z features that differed significantly between all protein and control samples pairs. The accurate masses

were first searched in Metlin metabolite database and matched that of different tetrapeptides and pentapeptides. Next, *de novo* sequencing of the MS/MS data gave diverse possible pentapeptide sequences, all rich in hydrophobic and negative charged residues. The sequences matched with some the proteins of *H. pylori* G27. Furthermore, different proteins of *Bos taurus* (since *H. pylori* was cultured in a medium supplemented with bovine serum) and *Homo sapiens* were found to contain the putative sequences.

Hp dppA knock-out mutant was obtained and will be characterized and tested in appropriate *in vitro* and/or *in vivo* models, following the confirmation of the ligand-fishing assay results.

Homology modelling with *E. coli* DppA as template, allowed the creation of a three-dimensional model for HP0298. Interestingly, the binding pocket located in the cleft between the two domains has a distinctly positive bottom and is lined with hydrophobic residues that could nicely accommodate a hydrophobic negative charged peptide.

Many studies have been performed on the oligopeptide binding receptor, OppA, of different Gram-negative and Gram-positive bacteria and the protein was produced endogenously where natural peptides are available for binding[115–119]. These works pointed out the existence of a mechanism for peptide selection based on amino acid composition rather than sequence. For example, Klepsch *et al*. characterized *E. coli* OppA and discovered a preference for positively charged peptides[119]. In our work we showed that HP0298 preferentially binds pentapeptides that are rich in negative and hydrophobic amino acids (in particular aspartate and leucine or isoleucine). This result suggests that also for *H. pylori* DppA the peptide amino acid composition might be a determinant for the binding. Remarkably, the amino acids that are considered to be essential for the bacterium are mostly hydrophobic. Further binding studies with the purified protein and different peptides, with both similar and different characteristics

from those identified, will be carried out to confirm the evidence. These properties can be related to the natural niche of the bacterium. The human stomach is very rich in peptides that are produced during digestion by the action of proteases and peptidases. Pepsin, a well-known aspartic protease of the human gastric mucosa, is active at low pH (pH < 3.0) and quickly denatured when the pH exceeds 5.5, thus preventing continuous proteolytic action that may damage the stomach[120]. This enzyme seems to have a preference for hydrophobic amino acid residues on both sides of the sensitive peptide bond[121], producing peptides with at least one hydrophobic end. Since *H. pylori* DppA expression was shown to be dependent on pH, in particular increased with decreasing medium pH[122], it is tempting to speculate that this periplasmic binding protein functions as a receptor for the specific uptake of gastric peptides derived from food proteins. Future research will focus on the link between the internalization of specific peptides and the physiology of both the cell and its native environment. With clues about the substrate, *Hp dppA*-negative mutant will be helpful for further experiments.

We have developed a protocol that couples affinity purification with high performance liquid chromatographic tandem mass spectrometry (HPLC-MS/MS) for the identification of unknown ligand of periplasmic binding proteins. This procedure can be applied to virtually any other PBP-ligand couple of which the ligand is unknown. The PBP module is found in many mammalian membrane receptors containing an extracellular binding domain that can be recombinantly expressed in soluble form. Whatever the need, either to discover the physiological ligand that is not known, or to design a specific drug, this approach offers several advantages:

1. Being both soluble and stable proteins, production of recombinant His-tagged PBP baits and its purification by cobalt affinity chromatography is accessible, simple, and compatible with assessment by HPLC and MS.

2. Prey ligands are extracted from the endogenous environment of the bait protein.

3. Co-resin provides a negative control to identify non-specific binding proteins.

4. Since the ligand from the binding protein has to be released for mass-spectrometric detection, there is no limit of preserving intact complex of PBP-ligand as in other techniques (e.g. native MS, crystallization, etc.).

5. Gel-free MS workflow aids comprehensive detection of low-abundance peptides, low-affinity interactions, and overcomes the challenge of having a non-protein ligand.

6. Numerous tools are available for the analysis of MS data, including specialized metabolic databases that can aid metabolite identification; MS/MS experiments can be performed on the differential peaks for accurate metabolite assignment.

7. Comparison of dependent (three rounds of the same fishing experiment) and independent (different prey ligand containing extracts) replicates plus negative controls can minimize false positive identifications.

Additional independent binding experiments with bait protein and prey ligand are needed to verify the interaction and quantitatively describe it. An important benchmark for the proper description and characterization of a new ligand-receptor interaction is the measurement of the binding affinity. There are many potential approaches that can yield an estimate of the binding affinity in terms of association and dissociation constants, between a ligand and receptor; these include isothermal titration calorimetry[123], differential scanning fluorimetry[124], or microscale thermoforesis[125]. However, the ligand-fishing protocol that we

described can be an up-front and sensitive interaction screening, aimed to identify ligands for PBPs and related proteins.

Materials & Methods

Phylogenetic reconstruction

The midpoint-rooted circular phylogram representing a maximum-likelihood phylogeny of *H. pylori* DppA and related periplasmic binding proteins was created with the ClustalW program[88] and visualized with FigTree.

Multiple sequence alignments were built with ClustalW, analyzed with Genedoc, and the online resource Espript[89] was used to generate alignment figures. The list of residues involved in protein-ligand interactions was downloaded from the NCBI's Conserved Domain Database (CDD)[126].

Cloning and expression of recombinant *H. pylori* DppA

The HP0298 gene was amplified by PCR from genomic *H. pylori* G27[86], with primers 5'-ATATAT<u>CCATGG</u>GAAGCGAAAGCCCGAATGC-3' (forward, NcoI recognition site underlined) and 5'-ATATAT<u>CTCGAG</u>TCTCAGCGCTTTTTCCAG-3' (reverse, XhoI recognition site underlined), using Phusion DNA Polymerase (Thermo Scientific). In this way the signal peptide at the N-terminal of the sequence predicted with SignalP[55] was removed. The amplified HP0298 gene was cloned into the pET22b(+) vector (Novagen) in frame with N-terminal PelB sequence and C-terminal His-tag. The integrity of the construct was confirmed by DNA sequencing.

20 ml of an O/N culture (40ml, 37°C) of *E.coli* BL21(DE3) cells harbouring the pET22-HP0298 plasmid were used to inoculate 1 litre of Luria-Bertani (LB) medium containing 100 µg ml^{-1} ampicillin. Cells were grown for 2.5 hours at 37 °C until OD_{600} of 0.6-0.8 was reached. HP0298 expression was induced with

0.25 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and cells were incubated for another 20 hours at 20 °C. Culture was harvested at 4°C and stored at -80 °C.

Periplasmic purification

HP0298 was isolated from the periplasmic fraction. To isolate the periplasm, after the cells from 1 litre culture had been spun down, the cell pellet was resuspended in 10 ml osmotic shock buffer (500 mM sucrose, 100 mM Tris, and 1 mM EDTA, pH 8) and incubated for 2 hours at 4°C on a rotating shaker, followed by centrifugation at maximum speed for 45 min. The supernatant, containing the periplasmic extract was subjected to affinity chromatography with TALON metal affinity resin (Clontech). Briefly, the periplasmic fraction was mixed with 1-2 ml of resin and incubated for 45 min at 4°C on a rotating shaker. The sample was packed in a polypropylene tube and washed four or five times with 10 resin volumes of wash buffer (50 mM Tris 10 mM imidazole, pH 8.0). The resin was transferred to a plastic column, washed with another resin volume of wash buffer, and finally the bound protein was eluted three times by the addition of 1 ml of elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, pH 7.4). The protein fractions were pooled and dialyzed at 4°C against 4000 volumes of pH 7.4 storage buffer consisting of 50 mM NaH_2PO_4, 300 mM NaCl.

Purity of His-tagged DppA was assessed by Glycine SDS-PAGE and Tricine SDS-PAGE[111]; gels were stained with Coomassie brilliant blue. The protein was further purified by Sephacryl S-200 HR (GE Healthcare), equilibrated with 50 mM NaH_2PO_4, 300 mM NaCl, pH 7.4. The His-tagged HP0298 was eluted as a single peak, corresponding to a monomer of 61.4 kDa. Final concentration and purity were estimated through UV-visible spectra (ε_{280} = 104170 M^{-1} cm^{-1}) and Glycine/Tricine SDS-PAGE respectively. Prior to store at -80°C, HP0298 was added of glycerol at a final concentration of 15%.

CD measurements

CD measurements were performed using a Jasco J715 spectropolarimeter equipped with a Peltier thermal controller. CD spectra were recorded to evaluate tertiary structure stability for HP0298. Three runs were accumulated with 0.3 mg/ml protein in 10 mM NaH_2PO_4 (pH 2–8.3) and acetonitrile solution (50-70%) using a 1 mm path quartz cuvette (QS-110 from Hellma). Spectra were recorded for each sample from 250 to 200/180 nm at 25 °C using a bandwidth of 0.2 nm and a response time of 8 s at a scan speed of 50 nm/min. Each spectrum is subtracted of the buffer contribution.

Ligand pull-down assay

For the pull-down assay total extracts from *H. pylori* G27 cells were prepared. First, *H. pylori* G27 was revitalized from glycerol stocks on Brucella broth agar plates added with 5% fetal calf serum and Skirrow's antibiotic supplement under microaerophilic conditions in jars (Oxoid gas packs). After restreaking on fresh plates, bacteria were cultured in a 9% CO2–91% air atmosphere at 37°C and 95% humidity in a water-jacketed incubator (Thermo Forma Scientific). Liquid cultures were grown in modified Brucella broth medium supplemented with 5% fetal bovine serum. Total cell extract from *H. pylori* G27 for a single pull-down experiment was obtained from a 50 ml culture. In particular, cells were spun down and resuspended in 2 ml of lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, pH 7.4) containing 1 mg/ml lysozyme and 10 μg/ml (each) DNase and RNase, sonicated on ice, and centrifuged at maximum speed for 30 min; the supernatant was collected and stored at 4°C.

For the pull-down assay, the suspension was split into two 1 ml aliquots and mixed with 100 μl of purified His-tagged HP0298 (600 ng; 8 μM) or 100 μl of storage buffer (50 mM NaH_2PO_4, 300 mM NaCl, pH 7.4) for the control sample,

and incubated for 1 hour at 4 °C. Both samples were mixed with 100 μl of TALON resin and incubated for another hour at 4 °C on a rotating shaker. After that, the control was passed through a column, and the cleared lysate was collected for the second round of the fishing experiment. The resin with the protein–ligand complex bound was washed six times with 1 ml of 50 mM Tris buffer, pH 8, and then split for elution with imidazole buffer (50 mM Tris pH 8, 250 mM imidazole) or pH 2 buffer (50 mM Tris pH 2). In both cases, three successive elutions with 100 μl of buffer were carried-out. The cleared lysate (1 ml) was split again in two 500 μl aliquots and mixed with 50 μl of purified His-tagged HP0298 (300 ng; 8 μM) or 50 μl of storage buffer for the control sample, and the procedure was repeated as described for the first round, using 50 μl of TALON resin. Three elutions with 80 μl of buffer were carried-out in this second step. A new cleared (twice) lysate was collected and the third round of the fishing experiment was performed. 250 μl aliquots from the lysate were mixed with 25 μl of purified His-tagged HP0298 (150 ng; 8 μM) or 25 μl of storage buffer for the control sample, and further incubated with 25 μl of TALON resin. Three elutions with 60 μl of buffer were carried-out in this third step. All samples were stored at 4 °C prior to HPLC-MS analysis.

HPLC-MS and HPLC-MS/MS analysis

The analysis of the eluted samples with the protein-ligand complex and that of the control samples were performed using a HPLC DIONEX Ultimate 3000 coupled with LTQ Orbitrap XL (Thermo Fisher Scientific Inc., San Jose, CA) and fitted with a heated ESI probe (H-ESI-II, Thermo Fisher Scientific Inc.). Chromatographic separations were carried out by means of an Aeris Peptide 3.6u XB-C18 column (150x2.13mm; Phenomenex). The mobile phases consisted of 0.2% (v/v) formic acid in water (A) or in acetonitrile (B). A

gradient elution was applied as follows with a flow rate of 0.2 ml/min: 0-5 min, 2% B, 5-30 min, 40%B, 30-31 min, 50% B, 31-43 min, 80% B, 50-51 min, 2% B, plus reconditioning.

The MS worked in positive ionization mode, with a capillary temperature of 275 °C, while the source was maintained at 250 °C. The sheath gas flow was 40 units, while auxiliary and sweep gases were set to 10 units. The source voltage was 3.5 kV. The capillary voltage and tube lens were 12.76 and 70.13 V, respectively.

The MS was performed following two scan events: (1) FT-MS in positive ionization mode with a resolution of 7500 and MS scan range from m/z of 50 to 400; (2) FT-MS in positive ionization mode with a resolution of 15000 and MS scan range from m/z of 400 to 2000. Chromatograms and spectral data were acquired using XCalibur software 2.1. XCMS software[127] was employed to analyze the untargeted metabolomic datasets to identify metabolite features that differed significantly between the protein and the control samples.

The identification of potential DppA peptide ligands was performed with metabolite fragmentation in MS/MS, with parent masses at m/z of 516.30350, 562.27250, 578.24970, 602.34050 and MS scan range from m/z of 400 to 2000; isolation width of 1 m/z, activation type of CID, normalized collision energy of 35, activation Q frequency of 0.25, an activation time of 30 ms. *De novo* sequencing of peptides was carried out with PEAKS software[113].

Homology modelling

Homology modelling of *Hp* DppA was performed with SWISS-MODEL[85] using the structure of *E. coli* DppA (PDB: 1DPE) as template. Images were created with the PyMOL Molecular Graphics System program and the APBS plug-in in PyMOL[128]. Negative charges are shown in red, and positive charges are shown in blue.

Construction of *Hp dppA* knock-out mutant

H. pylori G27-derivative *dppA* knock-out mutant was obtained using a pBluescript KS II vector carrying DNA regions flanking the *dppA* gene on the *H. pylori* chromosome, as well as a *Campylobacter coli aphA-3* kanamycin resistance (Km^R) cassette. Primer forward 5'-ATAT<u>GCGGCCGC</u>AACCAAGCGTGGTTTTAGAAGAG-3' (NotI recognition site underlined) and primer reverse 5'-ATAT<u>GGATCC</u>TTAGATGGATTAAGAGCAGCGTTTGG-3' (BamHI recognition site underlined) were used to amplify and clone a 573 bp NotI-BamHI fragment encompassing the region upstream *Hp dppA*, which corresponds to 458 bp upstream the HP0298 open reading frame (ORF) and 101 bp of the 5' region of the HP0298 ORF. Primer forward 5'-ATAT<u>GGATCC</u>TAAAACGACCGGAGTGAGCGTG-3' (BamHI recognition site underlined) and primer reverse 5'-ATAT<u>GAATTC</u>ACAATGCTAGGCAAAATCAAGTGCTTG-3' (EcoRI recognition site underlined) were used to amplify and clone a 697 bp BamHI-EcoRI fragment encompassing the region downstream *Hp dppA*, corresponding to 55 bp of the 3' region of the HP0298 ORF and 630 bp downstream the HP0298 ORF. The kanamycin cassette, obtained as BamHI-BamHI fragment from pVac::Km^R vector[129], was inserted between these two fragments and the final construct was used to transform *H. pylori* G27. *H. pylori* transformants were obtained by double homologous recombination of the naturally competent G27 strain: freshly grown overnight cultures were spotted onto plates and grown for a further 5 hours, at which point 15 μg of plasmid DNA (pBS::ΔHP0298) was added onto the growing strain and incubated overnight. Positive clones were selected on Brucella agar plates supplemented with kanamycin (25 μg/ml), according to the resistance phenotype conferred by the Km^R cassette. The kanamycin-selected mutant strains were confirmed by PCR.
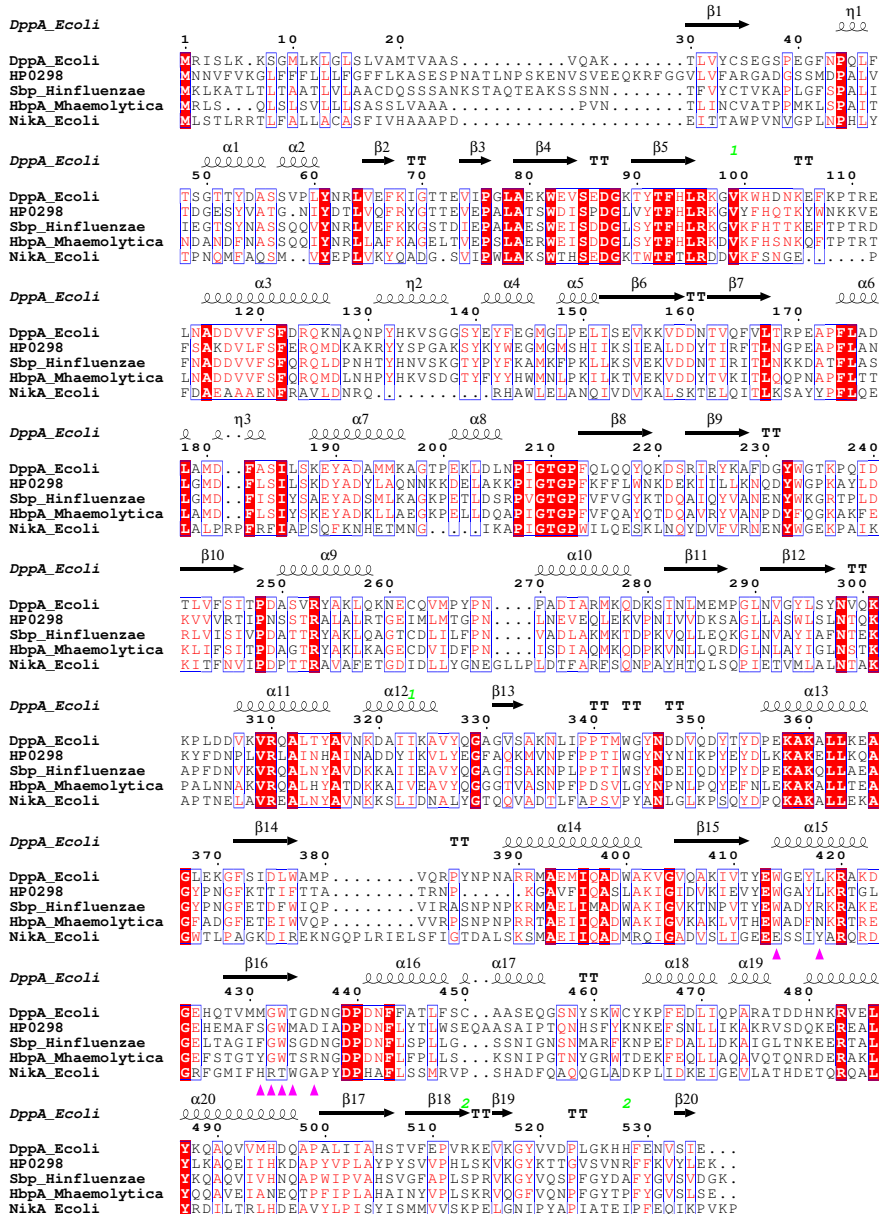
Supplementary material

Figure S1. **Multiple sequence alignments of 5 PBPs**. Each of these sequences represents one of the clusters of the tree in Figure 2. Pink upwards triangles indicate essential residues involved in protein-ligand interactions, according to CDD. Green numbers below the alignment indicate putative disulfide bonds in *E. coli* DppA.
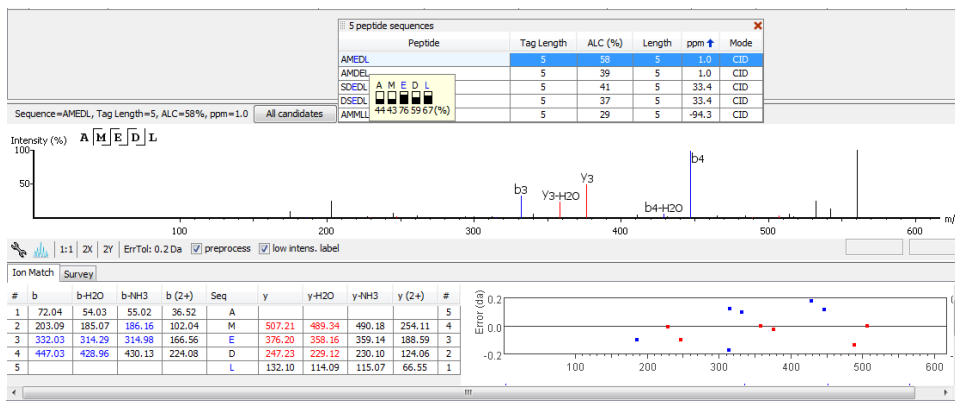
Figure S2. **Example of de novo sequencing of the m/z 578.2496 with PEAKS software**.

m/z med	RT med	Npeaks TOT	Npeaks Control	Npeaks +HP0298	1 Control	2 Control	3 Control	1 +HP0298	2 +HP0298	3 +HP0298	Log odds	fold	tstat	pvalue
602.3404736	1354.82	3	0	3	0.00E+00	1.79E+03	0.00E+00	6.54E+06	3.37E+06	1.06E+06	8.72	6137.712002	2.30	0.14816405
562.2722876	807.36	2	0	2	1.38E+03	0.00E+00	0.00E+00	3.79E+06	1.88E+06	4.44E+05	8.40	4427.626337	2.10	0.170278885
516.3031795	1203.54	2	0	2	1.13E+04	2.55E+03	1.50E+03	2.21E+07	1.13E+07	3.66E+06	7.79	2404.800367	2.31	0.147485834
578.2534318	971.46	2	0	2	3.01E+04	7.45E+03	1.89E+03	5.34E+07	9.19E+05	6.32E+06	7.34	1536.277525	1.21	0.349196039
621.2890143	1241.16	3	0	3	4.55E+03	0.00E+00	0.00E+00	3.15E+06	1.65E+06	4.67E+05	7.05	1156.286243	2.26	0.152453539
557.2614981	1198.95	2	0	2	1.56E+03	1.23E+03	0.00E+00	1.11E+06	5.40E+05	1.44E+05	6.47	641.6995829	2.13	0.166767863
445.264286	849.82	2	0	2	3.11E+03	1.31E+03	0.00E+00	7.32E+05	3.77E+05	1.22E+05	5.63	278.397491	2.31	0.146918716
423.1734354	843.00	2	0	2	8.23E+03	0.00E+00	0.00E+00	1.38E+05	5.74E+05	5.02E+05	4.99	147.3960553	2.98	0.096736129
1629.51507	1981.98	2	0	2	0.00E+00	0.00E+00	1.85E+05	7.02E+06	1.96E+06	2.59E+06	4.14	62.64797192	2.38	0.139539603
591.2811185	834.21	2	0	2	2.09E+04	1.79E+03	0.00E+00	8.19E+05	3.73E+05	1.13E+05	4.05	57.49626371	2.07	0.173824861
1635.337439	1978.67	3	0	3	2.65E+05	1.20E+05	6.21E+04	1.00E+07	7.41E+06	4.08E+06	3.87	48.0441196	4.09	0.054654042
1726.678825	1977.43	2	0	2	7.91E+04	1.92E+05	3.10E+04	5.27E+06	6.00E+06	7.74E+05	3.69	39.88131006	2.40	0.138692945
1938.859753	1977.43	2	0	2	0.00E+00	7.64E+04	1.81E+04	3.07E+06	5.24E+05	8.76E+03	3.64	38.14166283	1.24	0.341964154
1427.403955	1978.67	3	0	3	3.74E+05	1.65E+05	7.25E+04	9.77E+06	5.78E+06	3.97E+06	3.46	31.90927226	3.68	0.06612283
664.7609066	1980.92	2	0	2	0.00E+00	5.25E+04	1.11E+04	1.44E+06	5.15E+05	3.78E+04	3.44	31.27730661	1.56	0.258146863
1445.363007	1982.16	2	0	2	7.00E+05	3.39E+05	6.28E+04	1.70E+07	1.02E+07	6.38E+06	3.42	30.53536724	3.48	0.072713484
1502.032393	1976.99	3	0	3	4.55E+05	4.30E+04	5.41E+04	7.30E+06	6.36E+06	2.31E+06	3.36	28.92188928	3.34	0.077386029
1515.854938	1976.99	3	0	3	1.82E+05	5.98E+05	1.23E+05	1.51E+07	5.53E+06	4.81E+06	3.34	28.21025418	2.46	0.132678852
1576.746385	1977.51	2	0	2	4.63E+05	0.00E+00	4.38E+04	5.72E+06	6.04E+06	3.07E+06	3.15	23.28436592	1.92	0.193182281
1627.265362	1977.31	2	0	2	1.86E+05	1.80E+05	6.36E+04	6.26E+06	3.21E+06	2.31E+04	3.10	22.09224234	1.68	0.235239761
1381.14721	1980.35	3	0	3	1.03E+06	5.14E+05	9.40E+04	1.76E+07	1.17E+07	5.98E+06	3.07	21.53939558	3.34	0.07769346
1775.406602	1979.19	2	0	2	1.11E+05	4.67E+05	3.40E+04	8.84E+06	2.41E+06	1.48E+06	3.03	20.79507849	1.74	0.222419615

Table S1. **XCMS report of 23 selected m/z features enriched in the samples with HP0298 in the experiment with imidazole elution buffer.** Log odds. obtained with the formule "ln[(Σ +HP0298 intensity values)/(Σ control intensity values)]; fold. mean fold change; tstat. Welch's two sample t-statistic; pvalue. p-value of t-statistic.

m/z med	RT med	Npeaks TOT	Npeaks Control	Npeaks +HP0298	1 Control	2 Control	3 Control	1 +HP0298	2 +HP0298	3 +HP0298	Log odds	fold	tstat	pvalue
602.340451	1390.61	2	0	2	0.00E+00	3.54E+03	0.00E+00	9.50E+04	2.31E+06	1.44E+05	6.58	541.1106	1.14	0.336669855
578.247201	968.16	2	0	2	0.00E+00	0.00E+00	0.00E+00	1.31E+07	1.42E+07	1.86E+07	NA	Inf	9.08	0.011923641
562.272261	805.96	2	0	2	0.00E+00	0.00E+00	0.00E+00	1.07E+06	9.80E+05	1.33E+06	NA	Inf	10.71	0.008618375
516.303220	1190.06	2	0	2	0.00E+00	0.00E+00	1.19E+03	9.73E+06	1.17E+07	1.51E+07	1.03	30813.4329	7.71	0.016459273
512.2388003	2990.7	2	0	2	0.00E+00	0.00E+00	0.00E+00	3.91E+05	0.00E+00	3.06E+05	NA	Inf	1.96	0.189528382
689.4359673	1326.965	2	0	2	0.00E+00	4.32E+04	2.69E+04	3.16E+05	1.35E+05	1.09E+05	2.08	7.985541772	2.47	0.123777449
507.2661384	785.1005	2	0	2	3.35E+03	5.53E+04	4.27E+04	3.29E+05	1.52E+05	1.10E+05	1.76	5.825631866	2.37	0.128692865
652.4044005	1091.17	2	0	2	3.64E+04	1.91E+05	2.01E+05	1.34E+06	5.03E+05	4.49E+05	1.67	5.338187223	2.12	0.15968585
571.3097255	864.8905	2	0	2	7.67E+03	4.76E+04	6.38E+04	3.76E+05	1.24E+05	1.24E+05	1.66	5.242343228	1.96	0.178908867
612.2699873	871.5255	2	0	2	2.28E+04	6.91E+04	1.14E+05	6.53E+05	1.80E+05	1.82E+05	1.60	4.928497909	1.69	0.226571364
611.3487919	1558.69	3	1	2	4.08E+04	1.65E+05	2.09E+05	1.11E+06	3.10E+05	3.80E+05	1.47	4.343089611	1.77	0.209530186
543.1786237	94.50895	2	0	2	2.35E+04	1.26E+05	1.29E+05	1.99E+05	1.89E+05	1.50E+05	0.66	1.929999585	2.29	0.115332061
636.4177796	1463.59	2	0	2	1.97E+05	2.44E+05	3.98E+05	4.53E+05	5.47E+05	5.88E+05	0.64	1.894248048	3.44	0.033160558
551.1960613	94.6129	3	0	3	1.87E+04	2.05E+05	7.12E+04	1.99E+05	1.62E+05	1.49E+05	0.55	1.731497409	1.25	0.323138046
167.012874	81.623	3	1	2	1.86E+06	4.64E+06	2.86E+06	6.26E+06	4.15E+06	5.26E+06	0.51	1.673312911	2.07	0.11310941
413.2672635	2989.63	3	1	2	0.00E+00	5.33E+07	0.00E+00	4.67E+07	4.02E+07	0.00E+00	0.49	1.631420749	0.49	0.652231205
391.2850023	2990.96	3	1	2	0.00E+00	0.00E+00	3.37E+07	1.72E+07	1.93E+07	1.81E+07	0.48	1.618090503	0.62	0.599694202
986.8607273	94.6129	3	1	2	2.09E+04	9.46E+05	8.29E+05	1.03E+06	1.02E+06	8.55E+05	0.48	1.617061152	1.25	0.33108295
599.2255174	95.48555	2	0	2	3.07E+03	2.07E+05	1.41E+05	2.07E+05	2.27E+05	1.33E+05	0.48	1.610927433	1.07	0.364568741
831.576193	2989.7	4	2	2	5.86E+05	0.00E+00	1.35E+06	1.12E+06	9.44E+05	1.06E+06	0.48	1.610630581	0.99	0.419870127
440.1754209	94.6129	3	1	2	1.98E+04	5.26E+05	4.71E+05	5.98E+05	5.55E+05	4.66E+05	0.47	1.593563229	1.22	0.335706538
557.2554961	96.1403	3	0	3	7.86E+03	3.31E+05	3.38E+05	3.32E+05	3.86E+05	3.40E+05	0.45	1.563625395	1.15	0.363122526
255.1553998	83.9826	4	3	1	2.01E+06	1.32E+07	1.30E+07	1.81E+07	1.46E+07	1.15E+07	0.44	1.56003103	1.26	0.294867024

Table S2. **XCMS report of 23 selected m/z features enriched in the samples with HP0298 in the experiment with pH2 elution buffer.** Log odds. obtained with the formule "ln[(Σ +HP0298 intensity values)/(Σ control intensity values)]; fold. mean fold change; tstat. Welch's two sample t-statistic; pvalue. p-value of t-statistic.

Bibliography

1. Kusters, J. G., van Vliet, A. H. M. & Kuipers, E. J. Pathogenesis of Helicobacter pylori Infection. *Clin. Microbiol. Rev.* **19,** 449–490 (2006).

2. Ernst, P. B. & Gold, B. D. The disease spectrum of Helicobacter pylori: the immunopathogenesis of gastroduodenal ulcer and gastric cancer. *Annu. Rev. Microbiol.* **54,** 615–40 (2000).

3. Marshall, B. & Warren, J. R. Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. *Lancet* **323,** 1311–1315 (1984).

4. Solnick, J. V & Schauer, D. B. Emergence of diverse Helicobacter species in the pathogenesis of gastric and enterohepatic diseases. *Clin. Microbiol. Rev.* **14,** 59–97 (2001).

5. Falush, D. Traces of human migrations in Helicobacter pylori populations. *Science (80-.).* **299,** 1582–1585 (2003).

6. Bauerfeind, P., Garner, R., Dunn, B. E. & Mobley, H. L. Synthesis and activity of Helicobacter pylori urease and catalase at low pH. *Gut* **40,** 25–30 (1997).

7. Krulwich, T. A., Sachs, G. & Padan, E. Molecular aspects of bacterial pH sensing and homeostasis. *Nat. Rev. Microbiol.* **9,** 330–343 (2011).

8. Montecucco, C. & Rappuoli, R. Living dangerously: how Helicobacter pylori survives in the human stomach. *Nat. Rev. Mol. Cell Biol.* **2,** 457–466 (2001).

9. Palframan, S. L., Kwok, T. & Gabriel, K. Vacuolating cytotoxin A (VacA), a key toxin for Helicobacter pylori pathogenesis. *Front. Cell. Infect. Microbiol.* **2,** 1–9 (2012).

10. Censini, S. *et al.* cag, a pathogenicity island of Helicobacter pylori, encodes type I-specific and disease-associated virulence factors. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 14648–14653 (1996).

11. Kirschner, D. E. & Blaser, M. J. The dynamics of Helicobacter pylori

infection of the human stomach. *J. Theor. Biol.* **176,** 281–90 (1995).

12. Blaser, M. J. Ecology of Helicobacter pylori in the human stomach. *J. Clin. Invest.* **100,** 759–762 (1997).

13. Blaser, M. J. Malaria and the natural history of Helicobacter pylori infection. *Lancet (London, England)* **342,** 551 (1993).

14. Karnes, W. E. *et al.* Positive serum antibody and negative tissue staining for Helicobacter pylori in subjects with atrophic body gastritis. *Gastroenterology* **101,** 167–74 (1991).

15. Blaser, M. J. & Parsonnet, J. Parasitism by the 'slow' bacterium Helicobacter pylori leads to altered gastric homeostasis and neoplasia. *J. Clin. Invest.* **94,** 4–8 (1994).

16. Muotiala, A., Helander, I. M., Pyhälä, L., Kosunen, T. U. & Moran, A. P. Low biological activity of Helicobacter pylori lipopolysaccharide. *Infect. Immun.* **60,** 1714–6 (1992).

17. Andersson, J. O. & Andersson, S. G. E. Insights into the evolutionary process of genome degradation. *Curr. Opin. Genet. Dev.* **9,** 664–671 (1999).

18. Ahmed, N., Loke, M. F., Kumar, N. & Vadivelu, J. *Helicobacter pylori* in 2013: Multiplying Genomes, Emerging Insights. *Helicobacter* **18,** 1–4 (2013).

19. Flahou, B. *et al.* Gastric and enterohepatic non-Helicobacter pylori Helicobacters. *Helicobacter* **18,** 66–72 (2013).

20. Lefébure, T., Bitar, P. D. P., Suzuki, H. & Stanhope, M. J. Evolutionary dynamics of complete Campylobacter pan-genomes and the bacterial species concept. *Genome Biol. Evol.* **2,** 646–655 (2010).

21. Cravedi, P., Mori, G., Fischer, F. & Percudani, R. Evolution of the selenoproteome in Helicobacter pylori and Epsilonproteobacteria. *Genome Biol. Evol.* evv177 (2015). at <http://gbe.oxfordjournals.org/lookup/doi/10.1093/gbe/evv177>

22. Battistuzzi, F. U., Feijao, A. & Hedges, S. B. A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land. *BMC Evol. Biol.* **4,** 44 (2004).

23. Eppinger, M. *et al.* Who ate whom? Adaptive Helicobacter genomic changes that accompanied a host jump from early humans to large felines. *PLoS Genet.* **2,** 1097–1110 (2006).

24. Crabtree, J. E. Immune and inflammatory responses to Helicobacter pylori infection. *Scand. J. Gastroenterol.* **31,** 3–10 (1996).

25. Davies, G. R. *et al.* Helicobacter pylori stimulates antral mucosal reactive oxygen metabolite production in vivo. *Gut* **35,** 179–185 (1994).

26. Bagchi, D., Bhaitacharya, G. & Stohs, S. J. Production of reactive oxygen species by gastric cells in association with Helicobacter Pylori. *Free Radic. Res.* **24,** 439–450 (1996).

27. Ramarao, N., Gray-Owen, S. D. & Meyer, T. F. Helicobacter pylori induces but survives the extracellular release of oxygen radicals from professional phagocytes using its catalase activity. *Mol. Microbiol.* **38,** 103–13 (2000).

28. Wang, G., Alamuri, P. & Maier, R. J. The diverse antioxidant systems of Helicobacter pylori. *Mol. Microbiol.* **61,** 847–860 (2006).

29. Mishra, S. & Imlay, J. Why do bacteria use so many enzymes to scavenge hydrogen peroxide? *Arch. Biochem. Biophys.* **525,** 145–60 (2012).

30. Jang, S. & Imlay, J. A. Micromolar intracellular hydrogen peroxide disrupts metabolism by damaging iron-sulfur enzymes. **282,** 929–937 (2007).

31. Park, S., You, X. & Imlay, J. A. Substantial DNA damage from submicromolar intracellular hydrogen peroxide detected in Hpx- mutants of Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 9317–9322 (2005).

32. Loew, O. A new enzyme of general occurrence in organisms. *Science* **11,** 701–2 (1900).

33. Dunford, H. On the function and mechanism of action of peroxidases. *Coord. Chem. Rev.* **19,** 187–251 (1976).

34. Fita, I. & Rossmann, M. G. The active center of catalase. *J. Mol. Biol.* **185,** 21–37 (1985).

35. Díaz, A., Loewen, P. C., Fita, I. & Carpena, X. Thirty years of heme catalases structural biology. *Arch. Biochem. Biophys.* **525,** 102–10 (2012).

36. Murthy, M. R., Reid, T. J., Sicignano, A., Tanaka, N. & Rossmann, M. G. Structure of beef liver catalase. *J. Mol. Biol.* **152,** 465–99 (1981).

37. Vainshtein, B. K., Melik-Adamyan, W. R., Barynin, V. V., Vagin, A. A. & Grebenko, A. I. Three-dimensional structure of the enzyme catalase. *Nature* **293,** 411–412 (1981).

38. Takahashi, A., Kurahashi, T. & Fujii, H. Effect of imidazole and phenolate axial ligands on the electronic structure and reactivity of oxoiron(IV) porphyrin π-cation radical complexes: drastic increase in oxo-transfer and hydrogen abstraction reactivities. *Inorg. Chem.* **48,** 2614–25 (2009).

39. Robert, A., Loock, B., Momenteau, M. & Meunier, B. Catalase modeling with metalloporphyrin complexes having an oxygen ligand in a proximal position. Comparison with complexes containing a proximal nitrogen. *Inorg. Chem.* **30,** 706–711 (1991).

40. Vidossich, P., Alfonso-Prieto, M. & Rovira, C. Catalases versus peroxidases: DFT investigation of H_2O_2 oxidation in models systems and implications for heme protein engineering. *J. Inorg. Biochem.* **117,** 292–7 (2012).

41. Zámocký, M., Furtmüller, P. G. & Obinger, C. Evolution of catalases from bacteria to humans. **10,** 1527–1548 (2008).

42. Zámocký, M., Janecek, S. & Koller, F. Common phylogeny of catalase-peroxidases and ascorbate peroxidases. *Gene* **256,** 169–82 (2000).

43. Klotz, M. G. & Loewen, P. C. The molecular evolution of catalatic hydroperoxidases: evidence for multiple lateral transfer of genes between prokaryota and from bacteria into eukaryota. *Mol. Biol. Evol.* **20,** 1098–1112 (2003).

44. Hazell, S. L., Evans, D. J. & Graham, D. Y. Helicobacter-Pylori catalase. *J. Gen. Microbiol.* **137,** 57–61 (1991).

45. Wang, G. *et al.* Role of a bacterial organic hydroperoxide detoxification system in preventing catalase inactivation. **279,** 51908–51914 (2004).

46.   Baker, L. M. S., Raudonikiene, A., Hoffman, P. S. & Poole, L. B. Essential thioredoxin-dependent peroxiredoxin system from Helicobacter pylori : genetic and kinetic characterization. **183,** 1961–1973 (2001).

47.   Tomb, J. F. *et al.* The complete genome sequence of the gastric pathogen Helicobacter pylori. *Nature* **388,** 539–547 (1997).

48.   Pakhomova, S., Gao, B., Boeglin, W. E., Brash, A. R. & Newcomer, M. E. The structure and peroxidase activity of a 33-kDa catalase-related protein from Mycobacterium avium ssp. paratuberculosis. *Protein Sci.* **18,** 2559–68 (2009).

49.   Koljak, R., Boutaud, O., Shieh, B.-H., Samel, N. & Brash, A. R. Identification of a naturally occuring peroxidase-lipoxygenase fusion protein. *Science (80-.).* **277,** 1994–1996 (1997).

50.   Oldham, M. L., Brash, A. R. & Newcomer, M. E. The structure of coral allene oxide synthase reveals a catalase adapted for metabolism of a fatty acid hydroperoxide. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 297–302 (2005).

51.   Alfonso-Prieto, M., Biarnes, X., Vidossich, P. & Rovira, C. The molecular mechanism of the catalase reaction. *J. Am. Chem. Soc.* **131,** 11751–11761 (2009).

52.   Loewen, P. ., Carpena, X. & Fita, I. Helicobacter pylori catalase compound   I.   *To   be   Publ.*   at <http://www.rcsb.org/pdb/explore.do?structureId=2a9e>

53.   Benkert, P., Kunzli, M. & Schwede, T. QMEAN server for protein model quality estimation. *Nucleic Acids Res.* **37,** W510–W514 (2009).

54.   Zámocký, M. & Koller, F. Understanding the structure and function of catalases: clues from molecular evolution and in vitro mutagenesis. *Prog. Biophys. Mol. Biol.* **72,** 19–66 (1999).

55.   Nielsen, H., Engelbrecht, J., Brunak, S. & Heijne, G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **10,** 1–6 (1997).

56.   Schneider, T. D. & Stephens, R. M. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* **18 ,** 6097–6100 (1990).

57.   Lei, S. P., Lin, H. C., Wang, S. S., Callaway, J. & Wilcox, G. Characterization of the Erwinia carotovora pelB gene and its product

pectate lyase. *J. Bacteriol.* **169,** 4379–4383 (1987).

58.   Chelikani, P., Fita, I. & Loewen, P. C. Diversity of structures and properties among catalases. *Cell. Mol. Life Sci.* **61,** 192–208 (2004).

59.   Zhao, J., Lu, C. & Franzen, S. Distinct enzyme–substrate interactions revealed by two dimensional kinetic comparison between dehaloperoxidase-hemoglobin and horseradish peroxidase. *J. Phys. Chem. B* **119,** 12828–12837 (2015).

60.   Pozdnyakova, N., Makarov, O., Chernyshova, M., Turkovskaya, O. & Jarosz-Wilkolazka, A. Versatile peroxidase of Bjerkandera fumosa: substrate and inhibitor specificity. *Enzyme Microb. Technol.* **52,** 44–53 (2013).

61.   Dalton, D. A., del Castillo, L. D., Kahn, M. L., Joyner, S. L. & Chatfield, J. M. Heterologous expression and characterization of soybean cytosolic ascorbate peroxidase. *Arch. Biochem. Biophys.* **328,** 1–8 (1996).

62.   Dai, Y., Huang, C., Wen, L., Sheu, D. & Lin, C. Lemon ascorbate peroxidase : cDNA cloning and biochemical characterization. 9–17 (2012).

63.   Lu, H., Han, R.-L. & Jiang, X.-N. Heterologous expression and characterization of a proxidomal ascorbate peroxidase from Populus tomentosa. *Mol. Biol. Rep.* **36,** 21–7 (2009).

64.   Green, M. T. The Structure and spin coupling of catalase compound I: a study of noncovalent effects. *Journal of the American Chemical Society* 123 (37), 9218–9219 (2001). at <http://pubs.acs.org/doi/pdf/10.1021/ja010105h>

65.   Teale, F. W. J. Cleavage of the haem-protein link by acid methylethylketone. *Biochim. Biophys. Acta* **35,** 543 (1959).

66.   Feng, Y. & Sligar, S. G. Effect of heme binding on the structure and stability of Escherichia coli apocytochrome b562. *Biochemistry* **30,** 10150–10155 (1991).

67.   Mittler, R. & Tel-Or, E. Oxidative stress responses in the unicellular cyanobacterium Synechococcus PCC 7942. *Free Radic. Res. Commun.* **12-13 Pt 2,** 845–50 (1991).

68.   Miyake, C., Michihata, F. & Asada, K. Scavenging of hydrogen peroxide

in prokaryotic and eukaryotic algae: acquisition of ascorbate peroxidase during the evolution of Cyanobacteria. *Plant Cell Physiol.* **32,** 33–43 (1991).

69. Ødum, L. & Andersen, L. P. Investigation of Helicobacter pylori ascorbic acid oxidating activity. *FEMS Immunol. Med. Microbiol.* **10,** 289–294 (1995).

70. Sobala, G. M. *et al.* Acute Helicobacter pylori infection: clinical features, local and systemic immune response, gastric mucosal histology, and gastric juice ascorbic acid concentrations. *Gut* **32,** 1415–8 (1991).

71. Valderrama, B., Ayala, M. & Vazquez-Duhalt, R. Suicide inactivation of peroxidases and the challenge of engineering more robust enzymes. *Chem. Biol.* **9,** 555–565 (2002).

72. Gao, B., Boeglin, W. E. & Brash, A. R. Role of the conserved distal heme asparagine of coral allene oxide synthase (Asn137) and human catalase (Asn148): mutations affect the rate but not the essential chemistry of the enzymatic transformations. *Arch. Biochem. Biophys.* **477,** 285–90 (2008).

73. Melik-Adamyan, W. *et al.* Substrate flow in catalases deduced from the crystal structures of active site variants of HPII from Escherichia coli. *Proteins* **44,** 270–81 (2001).

74. Tosha, T., Uchida, T., Brash, A. R. & Kitagawa, T. On the relationship of coral allene oxide synthase to catalase: a single active site mutation that induces catalase activity in coral allene oxide synthase. *J. Biol. Chem.* **281,** 12610–12617 (2006).

75. De Luna, P., Bushnell, E. A. C. & Gauld, J. W. A molecular dynamics examination on mutation-induced catalase activity in coral allene oxide synthase. *J. Phys. Chem. B* **117,** 14635–41 (2013).

76. Mashhadi, Z., Boeglin, W. E. & Brash, A. R. Inhibitory effects of a novel Val to Thr mutation on the distal heme of human catalase. *Biochimie* **106,** 180–3 (2014).

77. Tanford, C. & Lovrien, R. Dissociation of catalase into subunits. *J. Am. Chem. Soc.* **84,** 1892–1896 (1962).

78. Baureder, M., Barane, E. & Hederstedt, L. In vitro assembly of catalase. *J. Biol. Chem.* **1,** M114.596148– (2014).

79. Harris, A. G. & Hazell, S. L. Localisation of Helicobacter pylori catalase in both the periplasm and cytoplasm, and its dependence on the twin-arginine target protein, KapA, for activity. *FEMS Microbiol. Lett.* **229,** 283–289 (2003).

80. Manos, J., Kolesnikow, T. & Hazell, S. L. An investigation of the molecular basis of the spontaneous occurrence of a catalase-negative phenotype in Helicobacter pylori. *Helicobacter* **3,** 28–38 (1998).

81. Odenbreit, S., Wieland, B. & Haas, R. Cloning and genetic characterization of Helicobacter pylori catalase and construction of a catalase-deficient mutant strain. *J. Bacteriol.* **178,** 6960–7 (1996).

82. Park, S. A. & Lee, N. G. Global regulation of gene expression in the human gastric pathogen Helicobacter pylori in response to aerobic oxygen tension under a high carbon dioxide level. *J. Microbiol. Biotechnol.* **23,** 451–458 (2013).

83. Thompson, L. J. *et al.* Gene expression profiling of Helicobacter pylori reveals a growth-phase-dependent switch in virulence gene expression. *Infect. Immun.* **71,** 2643–55 (2003).

84. Zhao, X. *et al.* Hydrogen peroxide-mediated isoniazid activation catalyzed by Mycobacterium tuberculosis catalase-peroxidase (KatG) and its S315T mutant. *Biochemistry* **45,** 4131–40 (2006).

85. Bordoli, L. *et al.* Protein structure homology modeling using SWISS-MODEL workspace. *Nat. Protoc.* **4,** 1–13 (2009).

86. 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. **63,** 94–98 (1995).

87. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K. & Pease, L. R. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77,** 51–59 (1989).

88. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23,** 2947–8 (2007).

89. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42,** W320–4 (2014).

90. Tam, R. & Saier, M. H. Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. *Microbiol. Rev.* **57,** 320–46 (1993).

91. Davidson, A. L. & Chen, J. ATP-binding cassette transporters in bacteria. *Annu. Rev. Biochem.* **73,** 241–268 (2004).

92. Rees, D. C., Johnson, E. & Lewinson, O. ABC transporters: the power to change. *Nat. Rev. Mol. Cell Biol.* **10,** 218–227 (2009).

93. Higgins, C. F., Hiles, I. D., Whalley, K. & Jamieson, D. J. Nucleotide binding by membrane components of bacterial periplasmic binding protein-dependent transport systems. *EMBO J.* **4,** 1033–9 (1985).

94. Quiocho, F. A. & Ledvina, P. S. Atomic structure and specificity of bacterial periplasmic receptors for active transport and chemotaxis: variation of common themes. *Mol. Microbiol.* **20,** 17–25 (1996).

95. Mccammonj, J. A. Hinge-bending in L-Arabinose- binding Protein. 1131–1134 (1982).

96. Flint, A., Sun, Y.-Q. & Stintzi, A. Cj1386 is an ankyrin-containing protein involved in heme trafficking to catalase in Campylobacter jejuni. *J. Bacteriol.* **194,** 334–45 (2012).

97. Spurlino, J. C., Lu, G. Y. & Quiocho, F. A. The 2.3-A resolution structure of the maltose- or maltodextrin-binding protein, a primary receptor of bacterial active transport and chemotaxis. *J. Biol. Chem.* **266,** 5202–19 (1991).

98. Lewis, M. *et al.* Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science* **271,** 1247–54 (1996).

99. Gilson, E. *et al.* Evidence for high affinity binding-protein dependent transport systems in gram-positive bacteria and in Mycoplasma. *EMBO J.* **7,** 3971–4 (1988).

100. Felder, C. B., Graul, R. C., Lee, a Y., Merkle, H. P. & Sadee, W. The Venus flytrap of periplasmic binding proteins: an ancient protein module present in multiple drug receptors. *AAPS PharmSci* **1,** E2 (1999).

101. Garmory, H. S. & Titball, R. W. ATP-binding cassette transporters are targets for the development of antibacterial vaccines and therapies. *Am. Soc. fot Microbiol.* **72,** 6757–6763 (2004).

102. Nedenskov, P. Nutritional requirements for growth of Helicobacter pylori. *Appl. Environ. Microbiol.* **60,** 3450–3453 (1994).

103. Mendz, G. L. & Hazell, S. L. Evidence for a pentose phosphate pathway in Helicobacter pylori. *FEMS Microbiol. Lett.* **84,** 331–336 (1991).

104. Doig, P. *et al.* Helicobacter pylori physiology predicted from genomic comparison of two strains. *Microbiol. Mol. Biol. Rev.* **63,** 675–707 (1999).

105. Mendz, G. L. & Hazell, S. L. Aminoacid utilization by Helicobacter pylori. *Int. J. Biochem. Cell Biol.* **27,** 1085–1093 (1995).

106. Salama, N. *et al.* A whole-genome microarray reveals genetic diversity among Helicobacter pylori strains. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 14668–73 (2000).

107. Weinberg, M. V. & Maier, R. J. Peptide transport in Helicobacter pylori: Roles of Dpp and Opp systems and evidence for additional peptide transporters. *J. Bacteriol.* **189,** 3392–3402 (2007).

108. Davis, G. S. & Mobley, H. L. T. Contribution of dppA to urease activity in Helicobacter pylori 26695. *Helicobacter* **10,** 416–23 (2005).

109. Vergauwen, B., Elegheert, J., Dansercoer, A., Devreese, B. & Savvides, S. N. Glutathione import in Haemophilus influenzae Rd is primed by the periplasmic heme-binding protein HbpA. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 13270–5 (2010).

110. Vergauwen, B., Van der Meeren, R., Dansercoer, A. & Savvides, S. N. Delineation of the Pasteurellaceae-specific GbpA-family of glutathione-binding proteins. *BMC Biochem.* **12,** 59 (2011).

111. Schägger, H. Tricine–SDS-PAGE. *Nat. Protoc.* **1,** 16–22 (2006).

112. Smith, C. A. *et al.* METLIN: a metabolite mass spectral database. *Ther Drug Monit* **27,** 747–751 (2005).

113. Ma, B. *et al.* PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **17,** 2337–42 (2003).

114. Davis, G. S. & Mobley, H. L. T. Contribution of dppA to urease activity in Helicobacter pylori 26695. *Helicobacter* **10,** 416–23 (2005).

115. Guyer, C. A., Morgan, D. G. & Staros, J. V. Binding specificity of the periplasmic oligopeptide-binding protein from Escherichia coli. *J. Bacteriol.* **168,** 775–9 (1986).

116. Tame, J. R. *et al.* The structural basis of sequence-independent peptide binding by OppA protein. *Science* **264,** 1578–81 (1994).

117. Sleigh, S. H., Seavers, P. R., Wilkinson, A. J., Ladbury, J. E. & Tame, J. R. Crystallographic and calorimetric analysis of peptide binding to OppA protein. *J. Mol. Biol.* **291,** 393–415 (1999).

118. Berntsson, R. P.-A. *et al.* The structural basis for peptide selection by the transport receptor OppA. *EMBO J.* **28,** 1332–1340 (2009).

119. Klepsch, M. M. *et al.* Escherichia coli peptide binding protein OppA has a preference for positively charged peptides. *J. Mol. Biol.* **414,** 75–85 (2011).

120. Fruton, J. S. The specificity and mechanism of pepsin action. *Adv. Enzymol. Relat. Areas Mol. Biol.* **33,** 401–43 (1970).

121. Tang, J. Specificity of pepsin and its dependence on a possible 'hydrophobic binding site'. *Nature* **199,** 1094–1095 (1963).

122. Wen, Y. *et al.* Acid-adaptive genes of Helicobacter pylori. *Infect. Immun.* **71,** 5921–5939 (2003).

123. Freyer, M. W. & Lewis, E. A. Isothermal titration calorimetry: experimental design, data analysis, and probing macromolecule/ligand binding and kinetic interactions. *Methods Cell Biol.* **84,** 79–113 (2008).

124. Pantoliano, M. W. *et al.* High-density miniaturized thermal shift assays as a general strategy for drug discovery. *J. Biomol. Screen.* **6,** 429–40 (2001).

125. Wienken, C. J., Baaske, P., Rothbauer, U., Braun, D. & Duhr, S. Protein-binding assays in biological liquids using microscale thermophoresis. *Nat. Commun.* **1,** 100 (2010).

126. Marchler-Bauer, A. *et al.* CDD: NCBI's conserved domain database. *Nucleic Acids Res.* **43,** D222–6 (2014).

127. Patti, G. J., Tautenhahn, R. & Siuzdak, G. Results from multiple profiling experiments. *Nat. Protoc.* **7,** 508–516 (2013).

128. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 10037–41 (2001).

129. 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).