Helicobacter pylori: Molecular insights into regulation of adhesion properties

Pär Gideonsson
Till morfar Torsten Forsberg 1921-2014
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table of Contents</td>
<td>i</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td>LIST OF PAPERS</td>
<td>iii</td>
</tr>
<tr>
<td>ABBREVIATIONS</td>
<td>iv</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>BACKGROUND</td>
<td>3</td>
</tr>
<tr>
<td><em>Helicobacter pylori</em></td>
<td>3</td>
</tr>
<tr>
<td>The gastric niche and host immunity</td>
<td>4</td>
</tr>
<tr>
<td>Epidemiology, gastric disease and treatment strategies</td>
<td>5</td>
</tr>
<tr>
<td>Strategies for persistent infection</td>
<td>8</td>
</tr>
<tr>
<td><em>H. pylori</em> gene regulation and two-component systems</td>
<td>10</td>
</tr>
<tr>
<td><em>H. pylori</em> adherence in the gastric environment</td>
<td>13</td>
</tr>
<tr>
<td>AIMS OF THE THESIS</td>
<td>17</td>
</tr>
<tr>
<td>RESULTS AND DISCUSSION</td>
<td>18</td>
</tr>
<tr>
<td>Paper I</td>
<td>18</td>
</tr>
<tr>
<td>Paper II</td>
<td>22</td>
</tr>
<tr>
<td>Paper III</td>
<td>26</td>
</tr>
<tr>
<td>Paper IV</td>
<td>29</td>
</tr>
<tr>
<td>CONCLUDING REMARKS AND OUTLOOK</td>
<td>33</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>37</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>39</td>
</tr>
</tbody>
</table>
**ABSTRACT**

*Helicobacter pylori* infects the human stomach and triggers an inflammatory response that damages the gastric tissue. This host-pathogen interplay has dire consequences as up to 20% of infected individuals develop peptic ulcer disease or gastric cancer. Given that half of the world’s population is infected, the number of afflicted humans is staggering and also tells that *H. pylori* is extremely efficient in spreading and maintaining infection. To enable persistent infection many factors play a role, but one important feature of *H. pylori* is its impressive ability to adhere to the slimy gastric mucus layer and the underlying epithelial cells. This occurs mainly via the BabA and SabA proteins that bind ABO/Le<sup>b</sup>- and sLe<sup>y</sup>/sLe<sup>a</sup>-antigens. I have in my thesis studied how these two proteins are utilized and regulated.

*H. pylori* transcription is in part controlled by two-component systems (TCSs) that use a sensor protein and a DNA-binding response regulator. We have studied how these systems control sabA and to some extent babA and indeed found a better map of how sabA and babA is regulated at the transcriptional level. We also found that variations in a polynucleotide T-tract located in the sabA promotor could fine-tune SabA expression/sLe<sup>a</sup>-binding. Thus we have exposed how strict regulation by TCSs combined with stochastic processes together shapes attachment in the bacterial population.

As the buffering mucus layer is constantly exfoliated, placing *H. pylori* in bactericidal acid, we hypothesized that low pH should abrogate adhesion. SabA expression was indeed repressed in low pH, however BabA expression remained unaffected. The BabA/Le<sup>b</sup>-binding was instead directly reversibly hampered by low pH and the degree of pH sensitivity was strain dependent and encoded in the BabA sequence. We believe that the pH dependent loss of binding is one key factor *H. pylori* utilizes to maintain persistent infection.

BabA is divided in generalists that bind ABO antigens and specialists that only bind blood group (bg) O. We co-crystalized BabA bound to these receptors and established the structural basis for generalist vs. specialist discrimination. We furthermore found a disulfide-clasped loop (CL2) in the center of the binding domain crucial for binding. Breaking CL2 with N-Acetylcysteine (NAC) disrupted binding and *H. pylori* infection mice experiments revealed inflammatory reduction upon NAC-treatment.

In sum, I have in my thesis dissected how *H. pylori* controls its adhesive abilities and how intrinsic properties in binding can be exploited for therapeutic purposes.
LIST OF PAPERS

I. A Repetitive DNA Element Regulates Expression of the Helicobacter pylori Sialic Acid Binding Adhesin by a Rheostat-like Mechanism

* Co-first authors

II. Structural Insights into Polymorphic ABO Glycan Binding by Helicobacter pylori

* Co-first authors

III. Acid Responsive Helicobacter pylori Adherence: Implications for Chronic Infection and Disease

Submitted manuscript
* Co-first authors

IV. The Helicobacter pylori Sialic Acid Binding Adhesin SabA is Regulated via a Network of Two-Component Systems
Åberg A., Gideonsson P., Brännström K., Argnqvist A

Manuscript
### ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>αCTD</td>
<td>α subunit C-terminal domain</td>
</tr>
<tr>
<td>ABO</td>
<td>ABO blood group system</td>
</tr>
<tr>
<td>ALe⁰/BLe⁰</td>
<td>A/B Lewis b antigen</td>
</tr>
<tr>
<td>ArsR</td>
<td>ArsRS response regulator</td>
</tr>
<tr>
<td>ArsRS</td>
<td>Acid-responsive (two-component) signaling system</td>
</tr>
<tr>
<td>BabA</td>
<td>Blood group antigen binding adhesin</td>
</tr>
<tr>
<td>CagA</td>
<td>Cytotoxin associated gene A</td>
</tr>
<tr>
<td>cagPAI</td>
<td>cag pathogenicity island</td>
</tr>
<tr>
<td>CL1-4</td>
<td>Cysteine loops 1 to 4 (in BabA)</td>
</tr>
<tr>
<td>CLR</td>
<td>C-Type lectin receptor</td>
</tr>
<tr>
<td>DL1/DL2</td>
<td>Diversity Loop 1/Diversity Loop 2 (in BabA)</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol (reducing agent)</td>
</tr>
<tr>
<td>EMSA</td>
<td>Electrophoretic mobility shift assay</td>
</tr>
<tr>
<td>FLA</td>
<td>Fragment Length Analysis</td>
</tr>
<tr>
<td>Gal</td>
<td>Galactose</td>
</tr>
<tr>
<td>GalNAc</td>
<td>N-Acetylgalactosamine</td>
</tr>
<tr>
<td>GlcNAc</td>
<td>N-Acetylglucosamine</td>
</tr>
<tr>
<td>Hop</td>
<td>Helicobacter outer membrane protein</td>
</tr>
<tr>
<td>Hor</td>
<td>Hop-related</td>
</tr>
<tr>
<td>ID</td>
<td>Insertion domain</td>
</tr>
<tr>
<td>Le⁰/Le⁰</td>
<td>Lewis a/Lewis b</td>
</tr>
<tr>
<td>LipoLLA</td>
<td>Liposomal linolenic acid</td>
</tr>
<tr>
<td>LPS</td>
<td>Lipopolysaccharide</td>
</tr>
<tr>
<td>MALT</td>
<td>Mucosa associated lymphoid tissue</td>
</tr>
<tr>
<td>NAC</td>
<td>N-acetylcysteine (N-acetyl-L-cysteine)</td>
</tr>
<tr>
<td>NLR</td>
<td>NOD-like receptors</td>
</tr>
<tr>
<td>NPHS</td>
<td>non-pylori Helicobacter species</td>
</tr>
<tr>
<td>OMP</td>
<td>Outer membrane protein</td>
</tr>
<tr>
<td>ORF</td>
<td>Open reading frame</td>
</tr>
<tr>
<td>PAGE</td>
<td>Polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PAMP</td>
<td>Pathogen-associated molecular pattern</td>
</tr>
<tr>
<td>PPR</td>
<td>Pathogen recognition receptor</td>
</tr>
<tr>
<td>PPI</td>
<td>Proton pump inhibitor</td>
</tr>
<tr>
<td>PsabA</td>
<td>sabA promoter</td>
</tr>
<tr>
<td>RIA</td>
<td>RadioImmunoAssay</td>
</tr>
<tr>
<td>RLR</td>
<td>RIG-like receptors</td>
</tr>
<tr>
<td>RNAP</td>
<td>RNA polymerase</td>
</tr>
<tr>
<td>RR</td>
<td>Response regulator</td>
</tr>
<tr>
<td>SabA</td>
<td>Sialic acid binding adhesin</td>
</tr>
<tr>
<td>SK</td>
<td>Sensor histidine kinase</td>
</tr>
<tr>
<td>sLe⁰/sLe⁰</td>
<td>Sialyl Lewis x/a</td>
</tr>
<tr>
<td>SAA</td>
<td>South American Amerindian</td>
</tr>
<tr>
<td>SPR</td>
<td>Surface Plasmon Resonance</td>
</tr>
<tr>
<td>SSM</td>
<td>Slipped strand mispairing</td>
</tr>
<tr>
<td>SSR</td>
<td>Simple sequence repeat</td>
</tr>
<tr>
<td>TCS</td>
<td>Two-component system</td>
</tr>
<tr>
<td>T4SS</td>
<td>Type 4 secretion system</td>
</tr>
<tr>
<td>TLR</td>
<td>Toll-like receptors</td>
</tr>
<tr>
<td>VacA</td>
<td>Vacuolating cytotoxin A</td>
</tr>
</tbody>
</table>
INTRODUCTION

Did anyone ever tell you that they are so stressed out that he or she is about to get an ulcer? This old misconception hangs on even though we know that *Helicobacter pylori* (*H. pylori*) causes the lion's share of stomach disease. However, the reasoning is not far-fetched as more benign syndromes e.g. functional dyspepsia can be triggered by life event stress, mimicking *H. pylori* induced disease, (reviewed in Mimidis and Tack, 2008). Also, for long the highly acidic stomach was considered sterile, despite several findings of spirochetes in stomach samples from animals and humans starting in the late 19th century (Bizzozero, 1893; Doenges, 1938; Freedberg and Barron, 1940; Steer and Colin-Jones, 1975). It would take until the 1980th before *H. pylori* and its connection to gastric disease was recognized by the Nobel prize awarded findings of Warren and Marshall (Marshall and Warren, 1983; 1984), where Koch’s postulates where demonstrated by self-experimentation.

Treatments of ulcerative diseases had so forth mainly consisted of surgical disconnection of the vagus nerve to the stomach, an approach with numerous side effects. Thus, the finding of *H. pylori* truly was a paradigm shift, in large replacing current surgical treatments with antibiotics. This naturally spurred a huge research interest, which has been ongoing, with currently >38,000 publications. The scientific attention is warranted as *H. pylori* infects roughly 50% of all humans, causing stomach disease in up to 20%, amounting to millions of afflicted individuals. By causing life-long infection that usually starts in childhood, chronic inflammation develops and can ignite a malignant process, i.e. stomach cancer, which is the third most common cause of cancer death in the world. Aside from sepsis, caused by a number of bacterial species, no single bacterial infection is even close to measure up in terms of caused fatalities.

Recently the 5300-year-old Iceman *H. pylori* genome was released (Maixner et al., 2016). In fact it seems *H. pylori*-infections date back much further than that to when the anatomically modern human left Africa about 60,000 years ago (Linz et al., 2007; Moodley et al., 2012). *H. pylori* has also been used to trace historic human migrations by comparing *H. pylori* DNA sequences (Moodley et al., 2009). This ancient *H. pylori*-human association has moreover shed light on co-adaptation between host and pathogen (Atherton and Blaser, 2009). Of special interest in this context, the now rapid declining *H. pylori* prevalence in the Western world (Sonnenberg, 2013) means humans are loosing a faithful companion that for time immemorial has stimulated a Th1/Th17 immunological response, reviewed
by (D'Elios and Czinn, 2014). Losing this bacterial hitchhiker could have implications for other disorders, such as increased risk of asthma as was shown in several publications (Arnold et al., 2011a; Engler et al., 2014) or loss of protection against TBC (Perry et al., 2010). Thus, *H. pylori* could have a Dr. Jekyll role complicating the well established Mr. Hyde role where *H. pylori* induces severe gastric diseases in concert with environmental factors and the hosts genetic predisposition. With more and more data surfacing, a debate on the *H. pylori* frenemy status is ongoing, where not all are convinced about the potential benefits of *H. pylori* (skeptical review by Graham, 2015).
BACKGROUND

*Helicobacter pylori*

The helix shaped, microaerophilic, 2-4 μm long Gram-negative bacterium *H. pylori* infects the human stomach and causes numerous diseases. The gastric environment is navigated by flagella-mediated chemotaxis in response to physicochemical gradients, positioning the bacteria close to the epithelial lining under a pH-neutral protecting mucus roof (Schreiber et al., 2004). Chronic *H. pylori* presence is further facilitated by specific acid acclimation, successful nutrient acquisition by the use of numerous virulence factors, immune modulation and firm attachment that is swiftly regulated, (reviewed by Salama et al., 2013).

Being part of the epsilonproteobacteria *H. pylori* is related to *Campylobacter jejuni* that causes acute gastroenteritis. The hallmark of *H. pylori* infection is in contrast to its relative, an often life-long infection, usually acquired in early childhood presumably by gastro-oral or possibly fecal-oral transmission, reviewed by (Delport and van der Merwe, 2007). *H. pylori* strains are broadly divided in genotypes that corresponds to geographic prevalence: hpEurope, hpEastAsia, hpAfrica1, hpAfrica2, hpSahul, hpAsia2, hpNEAfrica and hpSouthIndia (Falush et al., 2003; Kumar et al., 2015; Linz et al., 2007; Moodley et al., 2009). Beyond this classification, the small genome of about 1.5 million base pairs (Alm et al., 1999; Tomb et al., 1997) displays an extraordinary genetic diversity making strains unique. This diversity seems to be a consequence of *H. pylori*’s natural competence for DNA uptake (Hofreuter et al., 2001), fast recombination rate (Falush et al., 2001) coupled with phase variable restriction-modification systems (Lin et al., 2001; Gauntlett et al., 2014). Individuals are often infected with several strains and continuous genetic rearrangements additionally cause so-called quasi-panmictic populations (Suerbaum and Josenhans, 2007).

The *Helicobacter* genus is not restricted to *H. pylori* but also contains non-*pylori* *Helicobacter* species (NPHS) isolated both from humans (Ménard et al., 2014) and from a diverse range of animals, e.g. *Helicobacter valdiviensis* from birds (Collado et al., 2014), *Helicobacter cetorum* from dolphins (Davison et al., 2014) and *Helicobacter acinonychis* from cheetahs (Eaton et al., 1993). Many NPHS species seem to induce diseases both in humans and animals, however *H. pylori* is by far the most characterized.
The gastric niche and host immunity

The human stomach breaks down masticated food and forms an acid basin, which kills ingested microbes. From proximal to distal part it is divided in the topographic areas: fundus, corpus and antrum, where the last bridge to the duodenum via the pylorus segment. In the gastro-esophageal junction the commonly called cardia region is found, however it has been argued that the cardia only forms during disease conditions (Chandrasoma, 2005). The lining of the stomach consists of a single layer of columnar cells with continuous invaginations to form either of two gland unit types. Oxyntic glands – found in the fundus/corpus area (about 80% of the stomach), which produce hydrochloric acid via parietal cells, and simultaneously bicarbonate, released on the basolateral side for epithelial diffusion (McColl, 2012). Pyloric glands – found in the antrum, which release gastrin from G-cells to stimulate acid secretion. To prevent epithelial damage by the acidic gastric juice, cells are covered with a two-layered slimy mucus, where the inner layer is more firmly adherent compared to the outer looser layer. These layers are composed of large glycoproteins, held together in long chains by disulfide bonds forming a gel-like structure by binding water. Predominant glycoproteins are the cell surface associated MUC1, the surface epithelium released MUC5AC and MUC6, secreted from glands together with hydrochloric acid (HCl), which seems to make protective canals in the mucus for safe HCl deposit in the gastric lumen, (reviewed by Johansson et al., 2013). By mucin decoration of oligosaccharides, e.g. the Lewis blood group antigens, bottlebrush-looking structures are formed. This mucus meshwork provides a pH-gradient protecting the epithelial cells, and consequently a pH-neutral breeding ground for H. pylori, which adhere to mucin associated glycoreceptors (expanded on below). Of note, terminal α1,4-GlcNAc glycostructures, which can be found on gland-associated MUC6, was implicated as toxic to H. pylori (Kawakubo et al., 2004). The mucus layer further acts to concentrate inherent molecules, such as antimicrobial peptides/antibodies and to remove entrapped debris by the constant mucus turnover for epithelial replenishment. This clearance mechanism and bactericidal low pH forms an efficient pathogen barrier.

Innate immune cells scattered across the lamina propria form a second pathogen barrier. Both immune and epithelial cells display pathogen recognition receptors (PPRs) that upon target recognition trigger inflammatory responses and recruitment of immune cells, as seen during gastric inflammation. PPRs are divided in four classes for detection of different parts of infecting microbes (generally called pathogen-associated molecular patterns, PAMPs): Toll-like receptors (TLRs), C-Type lectin receptors (CLRs), RIG-like receptors (RLRs) and NOD-like receptors (NLR). Importantly, H. pylori is not efficiently targeted by these, e.g. TLR4
detection is essentially avoided as LPS evades recognition by removing phosphate groups from the 1- and 4’-positions of the Lipid A moiety making it less charged and also resistant to antimicrobial peptides (Cullen et al., 2011). *H. pylori* DNA is detected by TLR9 but induces an anti-inflammatory effect, which is also true for CLR (DC-SIGN) recognition of *H. pylori* associated fucosylated Lewis y (Peek et al., 2010; Salama et al., 2013). However, Nod1 recognition of peptidoglycan contributes to *H. pylori* killing via beta-defensin 2 (Grubman et al., 2010), strongly released in response to experimental *H. pylori* infection (Hornsby et al., 2008). Human beta-defensin 3 also targets *H. pylori* but its synthesis is down regulated in a CagA dependent manner (Bauer et al., 2012). Of further importance, *H. pylori* triggers release of antibodies in humans (Crabtree et al., 1991a) and vaccination of humans recently showed promise (Zeng et al., 2015), indicating that adaptive immunity is important, although usually not sufficient for pathogen clearance. The immune response typically displays Th1 and Th17 type T effector cells and an *H. pylori* induced tolerance mode, i.e. inflammation dampening and hampering of immune protection, via effects on dendritic cells and regulatory T cells (Treg) (Arnold et al., 2011b; Bamford et al., 1998; Oertli et al., 2012). Similarly, effects on Treg cells are often seen among other chronic infections, e.g. *Mycobacterium tuberculosis* (Sia et al., 2015), malaria (Adalid-Peralta et al., 2011) and HIV (Chevalier and Weiss, 2013). Furthermore, the *H. pylori* induced immune skewing was also suggested to suppress the risk of active *Mycobacterium tuberculosis* infection in monkeys (Perry et al., 2010) and showed protection against asthma in murine models (Koch et al., 2015). Oppositely could a Th2 dominant concurrent helminth infection attenuate *H. felis* induced gastric atrophy in mice (Fox et al., 2000).

**Epidemiology, gastric disease and treatment strategies**

The 50% global *H. pylori* prevalence is inversely correlated to socioeconomic status (Malaty, 2007) and furthermore varies with age and geographic region. Infection is usually contracted in childhood from another person, and although the fecal-oral route remains an option transmission likely occurs via the oral-oral route, reviewed in (Payão and Rasmussen, 2016). Furthermore, transcriptionally active *H. pylori* was found in vomitus, arguing for spread simultaneous with other common GI-infections (Janzon et al., 2009). Multiple risk factors have been found, e.g. poor hygiene, food/water contamination and household crowding (Khalifa et al., 2010). Since these factors have improved only recently in developed countries and infection is life-long, a birth cohort effect is seen with high prevalence in the oldest vs. low prevalence in the youngest strata of the population. Thus *H. pylori* incidence is dramatically decreasing in the developed world.
(Sonnenberg, 2013) due to lower acquisition, but is expected to stay high in the developing world, especially so considering that climate change here is estimated to worsen water/food availability and change demographics.

*H. pylori* infection is seemingly harmless in most infected individuals, but can cause ulcerative disease and gastric cancer in a minority of infected. Infection is usually asymptomatic, which leads to late detection contributing to the bad prognosis of gastric cancer. When scrutinizing gastric biopsies, histologic gastritis is seen in all that carry the bacteria, i.e. leukocyte infiltration of the gastric mucosa, commonly dominated by neutrophils and lymphoid tissue aggregates (Correa and Piazuelo, 2012). Chronic infection is typically established in the antrum region (Uemura et al., 2001), inducing gastrin release and hyperchlorhydria (more acid), which predisposes for duodenal ulceration. During the persistent infection gastric producing glands are prone for destruction, leading to hypochlorhydria (less acid). In these conditions, *H. pylori* easier spreads throughout the stomach and moves to the corpus region causing pangastritis, gastric ulcers or worse, malignant disease. This follows the so-called “gastric precancerous cascade” where the main steps, in order are: chronic active gastritis, chronic atrophic gastritis, intestinal metaplasia, dysplasia and invasive carcinoma (Correa, 2013). Although gastric adenocarcinoma is the dominant cancer form, occasionally the mucosa-associated lymphoid tissue is affected and gastric MALT-lymphoma develops, of which a majority can be cured by *H. pylori* eradication (Fischbach, 2014). The intermediate steps of atrophic disease can be estimated by comparing blood levels of pepsinogen I with pepsinogen II as these are secreted by the oxyntic and foveolar glands, respectively (Miki, 2006).

Whether an individual is afflicted by disease or not has been suggested to depend on the concerted action of bacterial strain virulence traits, host genetic predisposition, environmental factors, and importantly co-evolved or non-coevolved *H. pylori* strains vs. the host (Kodaman et al., 2014). Regarding the first, presence of the two cytotoxins cagA and vacA are important for disease development, (reviewed by Cover and Blanke, 2005; Hatakeyama, 2009), which has also been suggested for the adherence factors babA and sabA (Yamaoka, 2006) as well as other less characterized virulence traits, e.g. oipA (Yamaoka, 2006), dupA (Lu et al., 2005) and iceA (van Doorn et al., 1998). Simultaneous expression of CagA, VacA and BabA, commonly called “triple positive strains”, were associated to all major gastric disease types (Gerhard et al., 1999). Host factors contributing to disease are genetic polymorphisms in IL1β (also an acid suppressor) (El-Omar et al., 2000), TNF-α, IL-10 (El-Omar et al., 2003; Zambon et al., 2004), IL-8 (Taguchi et al., 2005) and TL4 (Zhou et al., 2014). HLA polymorphisms have
also been suggested to play a role, but disease associations seems to vary between populations and are hard to interpret. Environmental factors such as smoking and diet are also important for disease development, underscored in a rhesus macaque challenge study, were both *H. pylori* and a dietary carcinogen (ethyl-nitro-nitrosoguanidine) was needed to induce gastric neoplasia (Liu et al., 2009).

*H. pylori* infections are usually treated efficiently with a combination therapy consisting of a proton pump inhibitor (PPI) plus two or three of the following antibiotics: amoxicillin, clarithromycin and metronidazole, given for at least 14 days (Yuan et al., 2013). Bismuth salts are occasionally given as part of second-line therapies. Although PPI-treatment has proven its valuable role in eradication therapy it is extensively used for other purposes as well and has one major drawback, during PPI-induced hypochlorhydria the bacterial population can move from the stomach’s antrum to the corpus (Kuipers et al., 1995; 1996). Thus, the widespread use of PPIs could feasibly result in the more dangerous corpus infection predisposing gastric ulceration or cancer. Indeed *H. pylori* infected Mongolian gerbils receiving PPI treatment had significantly more adenocarcinomas than *H. pylori* only controls (Hagiwara et al., 2011).

Like the scenario for other bacterial pathogens, antibiotic resistant *H. pylori* strains are becoming increasingly more common and therapies must be varied locally since studies indicate population differences in resistance patterns – Amoxicillin: 0-15%, Clarithromycin: 2-24% and Metronidazole: 26-95% (O'Connor et al., 2014). An alternative approach to circumvent antibiotic resistance is to target the adhesive properties of the infecting microbe, which is also a strategy the body can adopt by e.g. releasable MUC1 mucin decoys (Lindén et al., 2009). One example has been to interfere with adhesion structure biogenesis, such as making pilicides or curlicides active against uropathogenic *E. coli* (Cegelski et al., 2009; Pinkner et al., 2006). Another example has been to compete with host receptors, such as using synthetic peptides mimicking a *Streptococcus mutans* adhesin (Kelly et al., 1999). A third possibility is to instead target the microbial adhesins, which avoids the possibility of causing unwanted downstream effects of host receptor binding. This can be done by e.g. using microbial adhesin proteins for immunization or, as was tested for *H. pylori*, using an adhesin competitor such as sialic acid oligosaccharides or cranberry juice components (Burger et al., 2002; Simon et al., 1997). Furthermore, the redox-sensitive binding of BabA, presented in Paper II could harbingers novel approaches in reducing *H. pylori* adherence to facilitate eradication. In fact the reducing agent N-acetylcysteine that was used in Paper II has already

7
shown therapeutic potential in clinical studies as an additive treatment against *H. pylori* infection (reviewed by Makipour and Friedenberg, 2011).

Although treatments of established *H. pylori* infections will in the foreseeable future play an important role, obliterating the need for such treatments by finding a prophylactic vaccine has for long been a holy grail in the field. After massive input in terms of publications and funding finally a large scale phase 3 study provided new hope for an *H. pylori* vaccine with 71.8% efficiency by immunizing children (6-15 years) with a fusion protein of the *E. coli* heat-labile enterotoxin subunit B and the *H. pylori* urease subunit B (Zeng et al., 2015). While not perfect, such protection rate could indeed provide a capable means of diminishing spread of the *H. pylori* menace into the next generation. Nevertheless, generalization of these results should initially be done with caution until the vaccine has been proven successful in other populations.

**Strategies for persistent infection**

The gastric environment poses many challenges that *H. pylori* deals with using a highly specialized toolbox of virulence mediators geared for the specific needs of the target niche. The first challenge is proper positioning and coping with stomach acidity. Human gastric biopsies and experimental infection of gerbils revealed that a substantial fraction of *H. pylori* cells are firmly adherent to the epithelium and the inner part of the mucus layer (Hessey et al., 1990; Schreiber et al., 2004). *H. pylori* bore themselves down by using polar flagella and the movement is likely facilitated by their helical shape, (Sycuro et al., 2010) and actively changing the rheological properties of mucus as they go (Celli et al., 2009). Since the mucus layer displays a low to high pH-gradient towards the epithelial layer, the deep *H. pylori* preferred niche reflects a nifty acid acclimation survival strategy. Furthermore, local pH gradients are likely formed in all possible directions in consequence of acid release from the deeper glands, seeping towards the lumen. To evade acid, *H. pylori* utilizes pH-guided chemotaxis, as was revealed in animal experiments controlling the mucus pH-gradient (Schreiber et al., 2004). In addition, the urea sensing TlpB chemoreceptor efficiently targets the bacteria to the epithelium, nicely shown in a gastric organoid system (Huang et al., 2015). Since urea is released in the extracellular space and diffuses towards the epithelium from below this could help explain why *H. pylori* targets gastric injury sites as was previously shown (Aihara et al., 2014). The continuously imported environmental urea is next converted into ammonia and bicarbonate, which buffers the cytoplasm and the periplasm (Sachs et al., 2011). This is achieved by selectively opening the inner membrane UreI-channel at low pH (Weeks et al., 2000), and urea processing by the Nickel-dependent urease enzyme and the α- and β-carbonic anhydrase enzymes.
Additionally, *H. pylori* responds to pH by regulating a large subset of genes by the acid-responsive signaling system, ArsRS (expanded on below).

The deep mucus penetration and establishment of the *H. pylori* infection close to the epithelium also facilitates nutrient acquisition (Tan et al., 2011), such as extracting cholesterol from epithelial membranes (Wunder et al., 2006) and gathering of nickel and other essential metals (De Reuse et al., 2013). Infection is however not only restricted to the surface epithelium but colonies can form further down in gastric glands (Sigal et al., 2015), close to capillaries (Aspholm et al., 2006) and possibly intracellularly (Amieva et al., 2002; Oh et al., 2005; Semino-Mora et al., 2003). In terms of positioning, *H. pylori* attachment plays a key role, and is elaborated on in a separate section.

Host cells are exploited using an arsenal of virulence mediators where the two most heavily studied are the released vacuolating cytotoxin A (VacA) and the cytotoxin associated gene A (CagA). CagA was discovered 20 years ago (Covacci et al., 1993; Crabtree et al., 1991b), has oncogenic potential when overexpressed in Mongolian gerbils (Ohnishi et al., 2008) and is translocated into host cells by a type 4 secretion system (T4SS)(Odenbreit et al., 2000), encoded on the cytotoxin-associated gene pathogenicity island (cagPAI) (Censini et al., 1996). Furthermore, the pathogenic effect of this T4SS is increased by simultaneous expression of BabA (Ishijima et al., 2011). The CagL subunit of the pilus apparatus sits at the tip of the protein assembly and makes contact with host cell integrins α5β1 and αvβ5, which is a prerequisite for CagA delivery (Kwok et al., 2007; Wiedemann et al., 2012). In addition, interactions with α5β1 were also found for the CagA, CagI and CagY proteins (Jiménez-Soto et al., 2009). As integrins are scattered on the basolateral side of the epithelial cells, the precise action of how CagA is delivered is somewhat obscured. However, access to the basolateral side was suggested to be mediated by disruption of cell-cell junctions by E-cadherin cleavage induced by the *H. pylori* HtrA protease (Hoy et al., 2010). Among the vital parts of the pilus structure, CagY stands out as it can turn on or off the T4SS function by changes in an inherent direct repeat structure during infection (Barrozo et al., 2013). Inside the cell CagA highjacks host cell kinases to phosphorylate tyrosines in repeats of EPIYA motifs, which obtains full protein activation (Odenbreit et al., 2000), although CagA also possesses proinflammatory potential when not phosphorylated (Suzuki et al., 2009). Studies on polarized epithelium showed that *H. pylori* depends strongly on CagA for replication close to the epithelial cells on the apical side, likely by affecting cell polarity (Tan et al., 2009).

VacA is present in a few isoforms labeled: s1-s2 (signal region), i1-i3 (intermediate-region) and m1-m2 (mid-region), where the s1 in combination
with m1 shows the strongest link to severe gastric disease (Atherton et al., 1995). The VacA protein has been shown to target the mitochondria (Galmiche et al., 2000) and ultimately causes host cell apoptosis. It is however seemingly Janus-faced and can also act immunomodulatory by entering T lymphocytes (Sewald et al., 2008), inhibiting them (Gebert et al., 2003) and stimulating Treg cells, a property also displayed by the secreted Gamma-glutamyl-transpeptidase (Oertli et al., 2013) and the neutrophil-activating protein (HP-NAP) (Sehrawat et al., 2015). Thus, *H. pylori* acts somewhat like a circus ringmaster and guides the performance of the immune system. In summary the concerted effects of CagA and VacA have profound impacts on epithelial host cells and their actions seems to co-depend on each other (reviewed by Hatakeyama, 2014; Palframan et al., 2012). Of note, *H. pylori* strains have classically been divided in Type I bacteria, which express either or both CagA and VacA and Type II bacteria that do not express CagA or VacA (Xiang et al., 1995).

An additional *H. pylori* defense mechanism is so called host mimicry where *H. pylori* exposes LPS bound Lewis type glycans, predominantly of Type 2 (Le^a^ and Le^β^), similar to the stomach counterpart, which could be important for host adaptation (Wirth et al., 1997). A more aggressive approach employed by *H. pylori* is the release of antimicrobial peptides (Pütsep et al., 1999) or the release of outer membrane vesicles (Olofsson et al., 2010) with multiple effects, e.g. induction of apoptosis or immune modulation (reviewed in Parker and Keenan, 2012).

**H. pylori gene regulation and two-component systems**

All *H. pylori* strains share a common core genome that is further diversified by strain specific additional genes, inversions, duplications, plasticity zones and frequent homopolymer simple sequence repeat (SSR) tracts of varying length (Alm et al., 1999; Tomb et al., 1997). Thus *H. pylori* displays a highly divergent genetic structure that is additionally complicated by individual gene variation, where e.g. no identical babA sequences have been found (Aspholm-Hurtig et al., 2004; Pride et al., 2001). This is likely a consequence of *H. pylori*’s fast mutation rate, which is additionally up to 10 times faster during the acute phase of infection than during chronic infection, as was suggested in studies with human volunteers (Linz et al., 2014). The source for *H. pylori*’s high mutation rates seems to be the bacterium’s DNA polymerase I, which lacks proofreading ability and makes mutagenic translesion synthesis (García-Ortíz et al., 2011) plus the absence of a MutSLH-like DNA mismatch repair system, which when present act to recognize and facilitate removal mismatched nucleotides (Björkholm et al., 2001; Pinto et al., 2005; Wang et al., 2005).
Another great contributor to *H. pylori*'s genetic diversity is the bacteria’s capacity to take up and incorporate foreign DNA into its genome. Thus *H. pylori*-genes from other strains can likely be imported during *Helicobacter* co-infections by natural transformation. Possibly non-*Helicobacter* genes can also be imported, as could have been the cause of cagPAI acquisition by horizontal gene transfer (Censini et al., 1996). However, *H. pylori* has a solid arsenal of restriction-modification systems that can degrade foreign DNA, so uptake of non-*Helicobacter* genes is probably unusual (Lin et al., 2001). Finally, the presence of SSRs, commonly present in bacterial pathogens with fairly small genome sizes, e.g. *N. meningitidis* and *H. influenzae*, represent genetic regions, which are subject to fast-paced changes. This occurs via slipped-strand mispairing (SSM), i.e. misalignment of complementary bases during DNA replication that become permanent mutations and further contributes to *H. pylori*'s genetic variability (reviewed by Dorer et al., 2009). Length changes in such repetitive DNA-tracts can have varying effects depending on the location of the SSR; such as phase variable On/Off shifts as the codons become changed when the SSR is present in an ORF. When SSRs instead are positioned in intergenic regions they can act to change the promoter appearances, commonly by changing the spacing of the -10 to -35 region, which alters the sigma factor/RNAP-binding.

The RNA polymerase (RNAP) is a protein complex where the $\alpha_2\beta\beta'\omega$ subunits are complemented with a $\sigma$-factor that recognizes what promotor to bind. *H. pylori* is divergent from many other RNAP’s as it has a $\beta\beta'$ fusion and shows large sequence variations in both $\sigma^{80}$ and the $\alpha$-subunits, which binds A/T rich sequences called UP-elements (Borin et al., 2014). In contrast to the genetic divergence, *H. pylori* seems more limited in its transcriptional control, with only two additional sigma factors: $\sigma^{54}(rpoN)$ and $\sigma^{28}(fliA)$ in addition to the housekeeping sigma factor $\sigma^{80}(rpoD)$ (Tomb et al., 1997). Adding to that is only a limited amount of transcription factors, which includes the stress response associated HspR/HrcA (Spohn and Scarlato, 1999; Spohn et al., 2004), the acid response implicated metal regulators Fur/NikR (Bury-Moné et al., 2004) and four two-component systems (TCS) plus two orphan RRs. The various factors often regulate each other and maintains transcriptional networks where e.g. Fur is a global regulator (reviewed by Danielli and Scarlato, 2010).

TCSs can achieve fast control of gene expression and are present in most bacteria, usually with increasing number of TCSs as the genome size increases. They operate by sensing an environmental signal with the N-terminal input domain of a sensor histidine kinase (SK) and upon activation autophosphorylates a conserved histidine positioned in the SK transmitter domain. As almost all SKs form homodimers, phosphorylation can occur in
cis or trans i.e. self- vs. cross-phosphorylation of the other monomer. The phosphoryl group is then transferred to the N-terminal receiver domain of a cognate response regulator (RR), activating its C-terminal output domain. Most response regulators act as dimers and use the output domain to bind DNA for gene regulation, but they can also interact with other proteins or act enzymatically. Some SKs can initiate phosphatase activity on their cognate RRs, which also aids in removing un-wanted phosphorylation as the TCSs occasionally act promiscuously and targets non-cognate RRs. In general, the half-life of phosphorylation on a RR is extremely varied with examples in different bacterial systems from seconds to days. TCSs are reviewed in Capra and Laub, 2012; Krell et al., 2010 and Salazar and Laub, 2015.

*H. pylori*'s repertoire of TCSs is limited with only 4 complete SK-RR pairs, reflecting its small genome. Among these the CheA/CheY-system, important for chemotaxis, stands out as here the phosphorylated CheY impacts on bacterial motility by establishing a protein interaction with the flagellar motor switch complex (Foynes et al., 2000). In contrast the remaining three TCSs act like transcriptional regulators. Of these the FlgRS (HP0703-HP0244) system is also involved in control of motility as whole-genome microarrays of the corresponding SK/RR-mutants showed impact on class 2 flagellar genes (Niehus et al., 2004). However later studies showed that it also impacts on a wider selection of genes, which was dependent on pH (Wen et al., 2009). Of note, FlgS is cytoplasmic, unlike the other *H. pylori* SK's, and FlgR activates σ54-mediated transcription without binding to enhancer DNA. The CrdRS (HP1365-HP1364), was initially thought to be important for acid acclimation, but these initial findings were not true for other *H. pylori* strains (Pflock et al., 2007). The major functions of the CrdRS system instead seems to be triggering of resistance against copper- and nitrosative stress (Hung et al., 2015; Waidner et al., 2005).

The last and most studied *H. pylori* TCS is the ArsRS (HP0166-HP0165). This system responds to acid changes in the periplasm by ArsS mediated pH-sensing and likely regulates >100 genes by ArsR-DNA interaction, including self-regulation (Bury-Moné et al., 2004; Loh and Cover, 2006; Pflock et al., 2004; 2006). Many of the genes predicted for ArsRS-regulation did not overlap in the aforementioned studies and only a few genes have been firmly verified to be regulated by ArsRS beyond these array-based studies, including the sabA gene (Goodwin et al., 2008; Harvey et al., 2014), confirming the SabA repression in response to pH 5.0 previously seen (Yamaoka, 2006). Besides functioning at low pH, the non-phosphorylated ArsR, presumed to be active at neutral pH, performs vital functions as ArsR deletion is lethal in contrast to a D52N mutant disrupting its phosphorylation acceptor site that is non-lethal (Schär et al., 2005). The
ArsS protein is on the contrary non-lethal and could potentially also have different functions as length changes in a repetitive poly-nucleotide C-tract located in the C-terminal ending alters the reading frame to form varying protein endings (Hallinger et al., 2012). All endings could nevertheless mediate kinase activity, thus the function is so far enigmatic (Beier and Frank, 2000). Underscoring the importance of in vivo regulation by the H. pylori TCSs, mice experiments revealed that mutants in ArsS, CrdS and FlgS could not establish infection (Panthel et al., 2003). Besides from the three described TCSs, two additional orphan RRs are encoded: HP1021 that binds to oriC and potentially impacts on chromosome replication (Donczew et al., 2015) and the homeostatic stress regulator, HsrA (HP1043) (Olekhnovich et al., 2014). Similar to ArsR both these RRs are essential for H. pylori survival (Beier and Frank, 2000).

To diversify transcriptional control H. pylori uses growth phase dependent supercoiling, i.e. alterations in DNA-winding, to affect transcription as was described for the flagellin gene flaA (Ye et al., 2007). Beyond classic transcriptional control, interestingly the H. pylori transcriptome revealed a more complex RNA-based posttranscriptional regulatory potential with antisense transcription, transcriptional start sites within ORFs and about 60 small RNAs (Sharma et al., 2010). One recently shown example of posttranscriptional regulation revealed how a small RNA impacts on regulation of TlpB that has a G-repeat tract in the leader sequence where the Regulator of polymeric G-repeats (RepG) binds (Pernitzsch et al., 2014). Depending on the G-tract length tlpB expression can be smoothly controlled similar to what the T-tract does for sabA in Paper I. Furthermore has the ability for RNA degradation become more acknowledged. H. pylori holds a small set of RNA degradation molecules where RNase J and RNA helicase RhpA assembles directly onto translating ribosomes, which is similar to eukaryotes (Redko et al., 2013). Degradation primarily focuses on mRNA and antisense RNAs, but was limited towards non-coding RNAs (Redko et al., 2016). I summary, the arena of posttranscriptional regulation in H. pylori is really beginning to show colors and will likely be a key field for understanding genetic control in H. pylori.

**H. pylori adherence in the gastric environment**

Bacteria and virus often utilize attachment to initiate and maintain persistent colonization (reviewed in Kline et al., 2009). Microbial attachment is usually highly specific towards its cognate receptors, which aids in tissue tropism to discriminate both species infected and the preferred niche. However, close adhesion can also mean that the bacteria come face to face with immune cells. This could help explain why many bacterial species display adhesins at the tip of fimbriae, which distances them from the host
tissue. Classic examples of such binding are the FimH adhesin that sits at the end of Type 1 fimbriae of various members of the *Enterobacteriaceae* family and binds mannosylated glycoproteins (Jones et al., 1995) and PapG that sits on P pili and binds globosides (Lindberg et al., 1987). In addition to adhesive structures that protrude from the bacteria there are also adhesins that sit closer to the bacterial cell surface, e.g. YadA from enteropathogenic *Yersinia* or UspA1/2 of *Moraxella catarrhalis*, as shown in electron micrographs (Hoiczyk et al., 2000).

*H. pylori* also displays adhesive properties, which seems logical considering that the infection certainly balances on a knife’s edge when being contrasted with strong immune responses at the epithelial lining and acidic pH in the gastric lumen. Interestingly it employs both adhesion that enables farther distance from the receptor to the bacterial cell using T4SS tip-associated proteins that attaches to host cell integrins α5β1 and αvβ5 (described above) and several membrane-embedded adhesin proteins that are part of a larger outer membrane protein (OMP) family. The OMP proteins share N- and C-terminus similarities, where the C-terminus has alternating hydrophobic and hydrophilic residues that ultimately constitutes a β-barrel structure inserted in the outer membrane, described in detail by (Alm et al., 2000). Among the OMP’s the following proteins have been implicated in adhesion but have no cognate receptor and/or are scarcely studied: AlpA/B (Odenbreit et al., 1999), HopQ (Loh et al., 2008), HopZ (Peck et al., 1999) and HorB (Snelling et al., 2007). The LabA(HopD) adhesin was recently found and thus belong to the more uncharacterized adhesins, however for this adhesin a cognate lacdiNAc receptor has been found (Rossez et al., 2014). Interestingly the protein sequence of LabA shows a YTE (TyrThrGlu) sequence also found in BabA where the Thr binds the Leb-receptor secretor fucose (Moonens et al., 2016).

Similar to e.g. norovirus and bacteria such as *Campylobacter jejuni* and *Salmonella typhimurium* the two extensively studied membrane-bound *H. pylori* adhesins BabA and SabA binds variants of histo-blood group antigens. The BabA protein was the first discovered *H. pylori* adhesin and binds type 1 chain ABO/Leb antigens (Aspholm-Hurtig et al., 2004; Borén et al., 1993; Ilver et al., 1998) that are present on MUC1 (Lindén et al., 2009) and MUC5AC (Lindén et al., 2002) mucins in the gastro-intestinal mucosal lining of so called “secretors”. I.e. individuals who express the secretor (fucosyl)transferase, which attaches the terminal “secretor” fucose typical for the H-antigen (that is further modified into A, B or O(unmodified)) present in saliva and the GI mucosa. For erythrocytes this terminal fucose is instead added by the H transferase, which is expressed by almost all people besides individuals of the rare Bombay phenotype. See table S1 in Paper II and
for a comprehensive view of the glycan receptors bound and the enzymes responsible for the various forms of glycans. BabA was additionally suggested to bind blood group antigens on type 4 chains, so called Globo H and Globo A (Benktander et al., 2012).

Certain BabA proteins cannot bind Lewis receptors that contain either the A- or the B-antigen and consequently only binds blood group O (bgO). Such forms of BabA, e.g. 60 % of South American Amerindian strains, are called specialists whereas BabA proteins that are not hindered by the A- or B-antigen are called generalists (Aspholm-Hurtig et al., 2004). The basis for binding discrimination between specialists and generalists is encoded in a few key amino acids present in the Diversity Loop 1 (DL1) domain, where specialist strains typically contain a bulky amino acid, which blocks the bgA or bgB determinants GalNAc/Gal (Moonens et al., 2016). Diversity Loops DL1 and DL2 are parts of the BabA carbohydrate binding domain (CBD) and represent BabA regions of increased sequence diversity (Aspholm-Hurtig et al., 2004). Binding of BabA proteins to their cognate receptors typically varies in the affinity range of $10^{-8}$-$10^{-11}$ M$^{-1}$ and depends on DL1, DL2 that binds the reducing end of the glycan plus the disulfide clasped loop 2 (CL2), which binds to the terminal “secretor” fucose. Together these three CBD regions make up BabA's three-pronged binding site where DL2 confers specificity towards lacto series type 1 chain ABO/Le$^a$-receptors (Moonens et al., 2016) that show a different core chain orientation than type 2 chains as the latter have a $\beta$1-4 Gal-GlcNAc linkage instead of a $\beta$1-3 linkage. Type 2 chains are also typically distributed deeper in the gastric glandular tissue (Mollicone et al., 1985).

The BabA mediated binding is varied in multiple ways: 1. Universal sequence variation in the central domain of BabA (Aspholm-Hurtig et al., 2004; Nell et al., 2014; Thorell et al., 2016). This variability alters the structural features of the protein, further discussed in Papers II and III, which also builds on the BabA structure report by (Hage et al., 2015). These structural modifications translate into distinct binding properties with regards to affinity, tolerance to pH and redox. 2. Variation of BabA expression levels (Bäckström et al., 2004; Hennig et al., 2004), likely caused by changes in promotor appearance, which should cause more/less “Velcro effect” in binding. 3. Translational On/Off shifts by phase variable changes in a CT-repeat tract present in some babA’s or homologous recombination of babA with babB, babC or another duplicated inactive babA, either present in any of the A/B/C loci that typically harbor these genes (Bäckström et al., 2004; Hennig et al., 2004; Solnick et al., 2004; Styer et al., 2010). Such recombinations are possible as the 5’ and 3’ end of the genes are highly similar among H. pylori OMP’s, in particular among the bab-variants, and give rise to chimeric genes
that either are turned off or changes attributes of the encoded protein in terms of binding strength etc. 4. Simultaneous expression of multiple BabA proteins, which can display different binding modes, (Moonens et al., 2016).

During inflammation sialylated structures are strongly up regulated, which is important for leucocyte homing. This is taken advantage of by many viruses such as influenza- and rotavirus and also the second *H. pylori* adhesin SabA that binds such sialylated sLe^x/sLe^a-antigens present in the stomach environment or on neutrophils (Mahdavi et al., 2002; Unemo et al., 2005). The SabA binding is typically of strong affinity (10^8-10^9 M^-1), but somewhat lower than BabA's affinity. However, similar to BabA the polypeptide sequence is divergent among different strains, leading to variations in binding specificity (Aspholm et al., 2006). Yet, not much is known about how the sequence changes translates into structural appearance, but with the extracellular part of the adhesin structure recently solved (Pang et al., 2013), the near future should uncover more structural insights.

SabA is extensively regulated similar to BabA. The gene is also turned On/Off via phase variation of a CT-repeat tract within the ORF and can be subject to homologous recombination with sabB and hopQ that are highly similar in the 5' and 3' ends of the gene (Lehours et al., 2004; Mahdavi et al., 2002; Talarico et al., 2012; Yamaoka, 2006). Also as we could show in Paper I, variations in a T-tract upstream of the -35-region gradually changes the sabA gene expression, with direct implications on sLe^x-binding (Åberg et al., 2014; Harvey et al., 2014). Furthermore, sabA transcription is delicately turned up or down by the TCSs, see above and Paper IV.
AIMS OF THE THESIS

The aim of the thesis was to study different aspects of the *H. pylori* BabA and SabA adhesins as is stated more specifically below:

I. To characterize how a T-tract in the *sabA* promoter region affects transcription and what implications this will have on sLe*-binding in the bacterial population.

II. To study the BabA binding site characteristics, the BabA sensitivity to reducing conditions and to determine if redox sensitivity can be therapeutically targeted.

III. To study how the BabA adhesin is rapidly regulated by changes in pH, aiming to understand how *H. pylori* avoids clearance during mucus layer turnover when using high-affinity binding.

IV. With focus on ArsRS, describe if and how two-component systems can affect expression of *sabA*.
RESULTS AND DISCUSSION

Paper I

A Repetitive DNA Element Regulates Expression of the Helicobacter pylori Sialic Acid Binding Adhesin by a Rheostat-like Mechanism

Phase variation is a mechanism common to many bacterial species to generate genetic diversity (reviewed in van der Woude, 2011). In this study we set out to identify how a T-tract adjacent to the -35 region upstream of the sabA gene impacts sabA expression and to find out whether this potentially regulatory sequence could be generalized to other similar genes. We also wanted to explore if the T-tract length varied in clinical H. pylori isolates in H. pylori isolated from mice experiments.

The study was initiated by comparing SabA expression and sLeα-binding in five H. pylori strains of different origin. Here we could see significant strain variation where mRNA expression, protein expression and sLeα-binding were comparable, indicating transcriptional regulation. These strains also had differences in the T-tract length and other parts of the promotor sequence. We next compared the sabA promoter (PsabA) from all publically available H. pylori genome sequences and found T-tract length variation from 5 to 22 T’s, indicating that either the T-tract length is completely randomly varying or that there is individual adaptation towards certain T-tract lengths. This finding was in line with the findings of (Kao et al., 2012) that found the T-tract in a set of Taiwanese isolates to vary from 10 to 28. Our subset of Peruvian strains also showed T-tract variation suggesting that geographic strain background did not account for the variations seen.

We next generated transcriptional PsabA::lacZ reporter fusions of varying T-tract lengths and revealed that promoter activity was multiphasic with e.g. high expression in T5 and low in T9. To expand on this finding we aimed to make otherwise isogenic H. pylori clones that had between 1 and 21 T’s using a method of contraselection as described by Dailidiene et al., 2006. In this method the DNA sequence of choice (PsabA) is replaced with a two-gene cassette that encodes chloramphenicol resistance and streptomycin sensitivity. The modified DNA sequence (PsabA with Tn) is next introduced to replace the two-gene cassette, selecting for streptomycin resistance. This however proved easier said than done, as this method was inherently leaky causing large numbers of false positive clones and was highly strain dependent. I solved this issue by making large dilutions of plated H. pylori
single colonies and calculating the number of false positives. I figured dilutions to get about 10 false clones should be good enough to use for transformation and could at that dilution obtain about 10x as many correct clones. This finding was critical to success of making otherwise isogenic \textit{H. pylori} mutants, which was important not only for this paper but for the entire PhD project.

Assaying the finally obtained clones exhibited strong multiphasic transcriptional output that produced corresponding amounts of SabA protein and similar levels of \textit{sLe}x-binding. The output levels were in the lower range for T variants 1, 9 and 18 and higher for 3 and 13, thus there was a striking sinus-wave like appearance that seemed to correspond to 10 nucleotides (nt), i.e. one turn of the DNA helix. We also fluorescently labeled these strains and applied them to human gastric tissue sections, which showed similar low/high binding. To study T-tract length changes over time we infected five mice with \textit{H. pylori} for two months. Fragment Length Analysis (FLA) of PCR amplifications from chromosomal DNA isolated from the input and the output bacteria obtained estimations of nucleotide polymorphisms in the entire bacterial populations. Here, we could indeed find T-tract length changes occurring in vivo over time, which was also confirmed by regular DNA-sequencing and similarly seen in bacteria isolated from three Swedish patients, which also showed differences between the antrum and the corpus region of the stomach arguing for local adaptation. Also for the FLA analysis the otherwise isogenic T-tract \textit{H. pylori} clones were critical as they were used to obtain the background “stutter” of the FLA method. I.e. when the fluorolabeled PCR-products are run on a gel they produce gradually decreasing peaks, which are inherently flawed. By comparing our samples to what was found in the T-tract clones, a better approximation of the “real” population could be obtained.

Having established that the T-tract length impacts on \textit{sabA} mRNA levels we wanted to determine the underlying mechanism and hypothesized that the T-tract somehow affected the RNA polymerase (RNAP) interaction with the \textit{PsabA}. Initially we used the \(\sigma^{80}\) \textit{H. pylori} sigma factor with the \textit{E. coli} RNAP in surface plasmon resonance experiments (SPR) to probe for binding of the \textit{PsabA}, including DNA fragments with varying T-tract length. However we could not produce solid binding using this combination and instead resorted to the \textit{E. coli} \(\sigma^{70}\)-RNAP, meaning that our obtained data presented below just presents an approximation of the real scenario. We could here indeed find differences in binding strength that compared to promoter activity for the given T-tract lengths. Furthermore did the T-tract not only function as a spacer as e.g. \(A_{13}\) differed significantly from \(T_{13}\) in both SPR experiments as well as in otherwise isogenic \textit{H. pylori} mutants. When comparing \textit{sLe}x-
binding of these *H. pylori* clones with *E. coli* promoter activity of *PsabA:*lacZ fusions of the corresponding T-tract length we found some discrepancies. One explanation to this could be the rather large divergence between the *E. coli* and the *H. pylori* RNAP’s (Zakharova et al., 1998) or other differences between the two bacteria, e.g. genome presentation or influence of additional transcription factors.

Next we set out to determine whether the T-tract affected binding patterns of the C-terminal domain of the RNAP α subunits (αCTDs) to A/T-rich sequences upstream of the core promoter, so called UP elements. DNase I footprint assays of *PsabA* with varying T-tract content showed core promoter binding, no interaction with the T-tract region but a protected region around -95 to -50 in *PsabA* variants with 13 nucleotides. This upstream binding indicated that the αCTDs indeed target DNA in this region and scrutinizing the DNA sequence showed a range of short A-tracts, which was highly interesting as αCTDs often bind to A/T sequences. In lack of homology to known UP-elements we chose to call these regions UP-like elements. To further assess interaction with the UP-like elements we removed the most proximal region and could then also see interaction to the distal part with A-tracts in SPR experiments. The overall binding strength between *PsabA* fragments of a set T-tract length and fragments lacking the proximal region was always altered, indicating an important function of αCTDs in the overall RNAP binding. In addition we made transcriptional *PsabA:*lacZ fusions and studied transcriptional activation in *E. coli* using deletion fragments, which showed a comparable trend as seen in the SPR. Furthermore, scrambling of the proximal, distal or both UP-like elements revealed drastic changes in *PsabA* activity. Our data thus point to the presence of several UP-like elements targeted by the RNAP αCTDs and that binding to these sites greatly impacts promoter activity.

A/T-rich regions are associated with DNA curvature, which can affect transcription in multiple ways, e.g. affecting binding of the RNAP or trans-acting factors and changing the melting temperature, reviewed by (Pérez-Martín et al., 1994). Therefore we hypothesized that the UP-like elements in *PsabA* could change the DNA structure and thereby impact on RNAP binding. To study this we made in silico predictions of the *PsabA* that showed substantial shift in structural appearance as the T-tract length was compressed or extended. This was also substantiated by different migration patterns of these DNA fragments upon PAGE-separation. We conclude that the T-tract acts to shift the UP-like elements into shorter/longer upstream distance, which thereby also changes the DNA structure, ultimately causing variations in transcription, discussed in detail in Paper I (Åberg et al., 2014).
In addition to A/T-rich sequences, so called nucleoid-associated proteins (NAPs), also cause bending of DNA and are frequently targeting A/T rich DNA where typical examples are H-NS, IHF, HU and Dps (reviewed in Dillon and Dorman, 2010). Not only are these proteins important for gene regulation they also seem to play a role in defense against foreign DNA as e.g. H-NS targets AT-rich regions in Salmonella (Navarre et al., 2006). We looked for NAP-homologous proteins in H. pylori and found the HU homolog Hup and the Dps homolog NapA. To determine whether these proteins are influencing the sabA expression we constructed the corresponding mutants in H. pylori strain SMI109 and assayed for mRNA expression, protein expression as well as sLeα-receptor binding. However, we found no evidence for sabA regulatory involvement of these proteins.

As the -35 adjacent T-tract strongly influenced transcriptional output we next wanted to find out whether this is a regulatory mechanism common also for other H. pylori genes. Thus we searched the H. pylori 26695 genome for polynucleotide tracts of >9 nucleotides similarly positioned as the T-tract in PsabA. Among the 25 candidate genes found, 15 encoded outer membrane proteins and 5 where positioned highly similar to the PsabA T-tract. When comparing all publically available genome sequences for these sites we also found variations in nucleotide tract length. Interestingly one such A/T-tract was placed between hp_0350 and pyrG (CTP synthase); here the A-tract was positioned just 3 nucleotides from the -35 region of hp_0350 whereas the T-tract was located about 30 nucleotides from pyrG. We then constructed transcriptional lacZ fusions with these two gene promoters and the corresponding variants with 5 nt shorter A/T-tract. Our β-galactosidase assays revealed distinct changes of transcriptional activity for hp_0350 but not for pyrG, adhering to our model of regulation caused by length changes in -35 adjacent polynucleotide tracts. Importantly we also found short A-tracts above the -35 region for hp_0350 again arguing for our model where the RNAP αCTD binding is influenced by UP-like elements.

In conclusion, we have revealed that variations in polynucleotide tracts upstream of -35 regions in H. pylori can have profound effects on gene expression by changing the DNA appearance/bending. Thus, H. pylori here employ a method of stochastic shifts to produce subsets of clones with variable expression. In the case of SabA, the Rheostat-like mechanism caused by the T-tract will generate a diverse population with low, medium and high SabA expressing clones. Thus, during gastric inflammation there will always be a clone that binds “just right”, obtaining the Goldilocks effect with the biologic implication being persistent infection.
Structural Insights into Polymorphic ABO Glycan Binding by Helicobacter pylori

The *H. pylori* adhesin BabA binds ABO blood group (bg) glycans, abbreviated Le\(^b\), in the gastric mucosa. However, how the interaction occurs and what specific features of BabA that forms the basis for binding was unknown. We here aimed to identify the structural prerequisites for BabA/Le\(^b\) binding. In doing so we hoped to find structural/functional determinants that could be therapeutically targeted.

A truncated form of BabA devoid of the C-terminal transmembrane region was recombinantly expressed and showed that binding affinity was dependent on oligomerization. Crystals of this BabA protein were obtained in the presence of BabA antibodies raised in alpacas (so called nanobodies) after immunization with BabA purified from *H. pylori* strain 17875 (Subedi et al., 2014). Two variant antibodies were chosen and these crystals yielded similar X-ray structures. Our structures were furthermore also comparable to the BabA structure displayed in another recent publication, although that BabA structure had a different truncation and was based on another strain (Hage et al., 2015). In more detail we found that the BabA ectodomain has a core domain with seven α-helices organized in two bundles (3+4 α-helices). As compared with other Hops BabA holds a unique insertion domain (ID) comprised of 80 amino acids with a four-stranded β-plate. Within the ID lie two closely spaced cysteine residues that with the exception of BabC are not present in the other Hops. This contrasts the other three cysteine pairs that are present in most Hops. Our crystals confirmed that all four cysteine pairs, in recombinant BabA, formed sequential disulfide bonds. To firmly establish that this was also the case in *H. pylori* we used non-reducing trypsination and MS-analysis of BabA purified from *H. pylori*, which produced an identical disulfide bond pattern compared with the crystal structures. Interestingly these disulfide constrained loops predominantly lie in regions of increased sequence diversity (Aspholm-Hurtig et al., 2004). Therefore we speculate that they display protruding features of the protein that are important for adaptation. A plausible role could be antigenic variation for immune evasion, i.e. an inherent invisibility cloak that changes the outer appearance of the protein to remain unseen.

To further determine the BabA carbohydrate-binding domain the protein was co-crystallized with a Le\(^b\) hexasaccharide. Multiple connections to the glycan receptor were found, with two general anchor points. In following X-ray structures with ALe\(^b\)/BLe\(^b\), described below, this was complemented with
a third binding position, which ultimately led us to denote the Le\(^{b}\) binding site “three-pronged”. In the BabA/Le\(^{b}\) X-ray structure, Gln207 and Thr246 together with the disulfide-clasped loop CL2-associated residues Cys189, Gly191 and Asn194 bound the terminal (secretor) fucose, which formed the first anchor point. The Asp233, Ser234 and Ser244, a triplet denoted the “triad”, bound the glycan core, forming the second anchor point. The triad binding helps explain how BabA binds to type 1 chains and not type 2 chains that have a \(\beta_{1-4}\) configuration of Gal-GlcNAc instead of \(\beta_{1-3}\) linkage thus displaying a completely different orientation of the core chain. Of the interacting residues, all amino acids within the CL2 are strictly conserved when comparing a large set of BabA sequences. On the other hand the Asp-Ser-Ser-triad are more divergent between strains. Interestingly when we grafted the CBD of two additional BabA sequences with less affinity in binding (\(H. pylori\) strains A730 and P436, also more described below), their structure diverged substantially from the high-affinity 17875 in the Asp-Ser-Ser-triad region DL2 and also in the DL1 region that shows great sequence variation between strains. Thus, CL2 binding forms the conserved structural basis for binding that is tweaked by sequence variation in Diversity Loops, which causes glycan re-orientation in these two ends. We conclude that the variations in DL1 and DL2 are functional and contribute to affinity variations possibly important for host adaptation and also specialist/generalist binding.

As we found that CL2 is strictly conserved, we hypothesized that it should be critical to a functional BabA protein. Thus, we first tested and showed that Le\(^{b}\)-binding in \(H. pylori\) is sensitive to reducing conditions that act to break-up disulfide bonds, and less so when the bacteria was allowed to bind the receptor before addition of DTT, arguing that the high-affinity binding interaction shields the CL2. We could also see that binding was recovered when placing \(H. pylori\) treated with DTT back into buffer free from reducing agent, i.e. the structure of the protein was not irreversibly damaged and the disulfide bond had the potential to reform. Although we did not substantiate this here, one could envision that during chronic inflammation as hypoxia spreads in the tissue, Baba mediated Le\(^{b}\)-binding is inactivated when the CL2 is reduced, and hence \(H. pylori\) can detach from the hostile mucosa. As we also found that \(H. pylori\) strains differ markedly in the redox-sensitivity for Le\(^{b}\)-binding one could further speculate that \(H. pylori\) strains have adapted the level of redox sensitivity to fit their respective host by mutating key amino acids in DL1 or DL2. To firmly establish that CL2 was instrumental for binding we used the contraselection approach mentioned above to make chromosomal Cys to Ala mutations in strain J166. This proved highly problematic as J166 almost always produce point mutations in the introduced gene, which meant extensive lab work to obtain a clone with intact sequence in the CBD. In hindsight, I would have made the same effort
in a multitude of strains to facilitate faster progress of the project. However, we wanted to be able to use this monkey-adapted strain in future animal experiments. The finally obtained chromosomal CL2 mutant clone indeed depleted the BabA function. We also studied the CL2 function during overexpression in *H. pylori*. For this I went to the lab of prof. Rainer Haas were we introduced the mutated *babA* gene into a newly developed shuttle vector system that enables manipulations in *E. coli* and subsequent conjugative gene transfer into *H. pylori*. However, this approach also rendered some serious trouble-shooting due to a repetitive CT tract in the ORF present in a number of *babA* genes, including *babA* from J166. Every attempt in expressing these genes led to rapid selection against BabA. Finally *babA* from strain 17875 without the CT repeats was over-expressed from the shuttle vector plasmid in *H. pylori* strain P1ΔbabA. Importantly, when we used this strain and measured binding in an SPR-system we could reveal that the sensitive part of binding was the association, as the small amount of bound protein remained bound at the dissociation phase of the experiment.

Knowing that BabA function could be disrupted with reducing agents we looked for similar drugs already approved for use in humans and found the N-acetylcysteine (NAC) reductant, a well-established treatment regime for several lung disorders as well as paracetamol overdose. We also found that NAC had shown some promising results as treatment against *H. pylori* infection, reviewed by (Makipour and Friedenberg, 2011). Our initial RIA-experiments revealed that Leb-binding was drastically reduced in the presence of NAC, which was not seen for the SabA mediated sLeα-binding (unpublished data). Similar inhibitory effect was also seen when applying fluorescently labeled *H. pylori* onto histio tissue sections of gastric mucosa. Quantitative MS analysis showed that NAC reduced the CL2 disulfide bond. Next we infected Leb transgenic mice (Falk et al., 1995) with *H. pylori* and treated them with NAC for 2 weeks. NAC treatment provided 2-fold reduction of epithelial adherence and 13-fold reduction in neutrophil recruitment to the gastric mucosa, likely reflecting the bacterial shift further out from the epithelial cells. Yet as disulfide bonds are widespread in proteins and mucins we cannot exclude that inflammatory reduction only relates to BabA CL2 domain inactivation.

BabA falls into two different binding categories, specialist BabA’s that only bind bg O and generalist BabA’s that can bind all ABO-variants (Aspholm-Hurtig et al., 2004). We next set out to determine the structural basis for generalist vs. specialist BabA binding. For this we obtained co-crystal structures of BabA with ALeβ5 (addition of GalNAc) or BLEβ7 (addition of Gal). This revealed similar bonding networks compared to what was obtained for the BabA/Leβ6 co-crystal, again showing that the terminal
secretor fucose is vital for the binding interaction. Interestingly, we found that the generalist BabA had a variable DL1 region producing a pocket where the bg A/B determinants could fit. When we instead grafted the DL1 of the specialist strain S831 into the same generalist BabA, this binding pocket was absent as a consequence of the 198/199 LeuPro to SerLys amino acid change immediately after CL2, i.e. the bulky Leu was pointing into the site for bg A/B determinants blocking their entrance. BabA sequence alignments revealed that specialists dominantly have a bulky amino acid (Asp, Asn, Leu) adjacent to the Pro residue. We also found four strains that displayed generalist binding although displaying specialist genotype in the 198/199 loci. However, we found that all these strains had mutations in the DL2 region responsible for binding to the core chain. We propose that this enforces structural changes, driving the bg A/B determinants out from clashing with the bulky 198 amino acid as shown with the X-ray structure of the BabA fusion protein grafted with the ID from strain P436. Additionally we also found one odd strain, A730, where the DL1 region contains a hairpin loop increasing binding to bg A/B determinants but also giving it properties of a poor Leb binder, this was consequently termed an A/B specialist.

We next wanted to find experimental evidence for receptor specificity shifts going from specialist to generalist by spontaneous mutations. We thus grew the specialist strain S831 to obtain single colonies and probed for clonal binding to Alexa 555-ALe. Here we obtained one clone that simultaneously expressed two BabA proteins, where the first BabA remained specialist whereas the other contained a Leu198Ser mutation that turned it into a generalist BabA. We thus uncovered a new expression mode where several BabA's can be simultaneously expressed. One can surmise that this is not exclusive for the 198/199 positions, but likely occurs elsewhere, making the *H. pylori* Lewis glycan binding mode considerably complex.

In summary we have in this paper determined the X-ray structures of BabA's interaction with cognate receptors, also revealing the basis of generalist vs. specialist binding. We have established that the receptor interaction is redox sensitive, which could have implications for redox-regulation during inflammation. Finally, we showed a way to exploit the redox sensitivity by therapeutically targeting *H. pylori*’s binding properties with N-Acetylcysteine. As antibiotic resistance in some parts of the world is becoming rather the rule than the exception, our findings open up the door for novel treatments, especially against hard-to-treat *H. pylori* infections, possibly saving lives. One could also envision optimizations of NAC treatment with e.g. combined treatments anchoring antibodies in the NAC opened CBD.
Paper III

Acid Responsive Helicobacter pylori Adherence: Implications for Chronic Infection and Disease

Stomach acid is released from the parietal cells, seeping towards the gastric lumen, forming small acid rivers through the mucus layer, which then forms an acid basin in the more collapsed mid section of the stomach. To adapt and survive such harsh conditions both rapid and slower regulation of attachment with regards to pH-changes is vital for persistent infection. As BabA is known for its strong affinity binding, it would be detrimental if rapid detachment could not take place, either when the bacteria are stuck to cells that are being shed off or in response to sudden exposure of acid from lateral directions. Furthermore, too strong adhesion can limit spread to close-by less hostile tissue. Given that BabA expression is not affected by pH-reductions, see Paper IV, we set out to define other modes of BabA regulation.

We found that bacterial binding to human gastric mucosal epithelium was increasingly sensitive to acid and sensitive RIA-assays showed rapid receptor dissociation and loss of affinity. Importantly the loss of binding was reversible, e.g. when incubating fluorescent H. pylori with Le\textsuperscript{b}-expressing cells and varying the pH. As dead bacteria were used we could exclude de novo synthesis of new functional adhesins. We hypothesize that this rapid detachment and following re-attachment enables a novel approach to evade acid and to “recycle” the infection.

Knowing that BabA is drastically divergent between H. pylori strains we next tested acid-sensitivity in Le\textsuperscript{b}-binding in a range of clinical isolates treating them with increasing amounts of acid. This showed tremendous variation of pH tolerance in Le\textsuperscript{b}-binding with “pH\textsubscript{50}” ranging from 2.35 to 4.87. To establish that inherent features of BabA could explain the variations in pH tolerance we introduced the babA gene from an acid tolerant strain, SW7, into the acid sensitive strain SW38 and found that the “Trans38-7” transformant had gained 75% of the Le\textsuperscript{b} tolerance compared to SW7. Together with experiments testing the pH tolerance of purified BabA proteins the transformation experiment strongly implied that BabA sequence determined a majority of the acid sensitivity in Le\textsuperscript{b}-binding. However, chromosomal DNA was used in the transformation so we could not exclude transfer of other genetic content influencing the results.

Knowing that pH tolerance patterns varied between strains we next set out to determine pH tolerance differences in clones isolated from the antrum (less
acidic) vs. the corpus (more acidic) regions of the stomach. In strains isolated from patient SO-2 we found a median difference in pH tolerance of 0.74 pH units. Gene sequence comparison revealed multiple sequence variations but dominantly these clones only differed at two positions: Leu199 and Glu428 in antrum vs. Pro199 and Gly428 in corpus. Numerous attempts to produce otherwise isogenic contraselection or shuttle-vector mutants were unsuccessfully made, highlighting the complexity of experimental _H. pylori_ genetics that seems highly strain dependent. We instead expressed the corresponding BabA proteins recombinantly and ran SPR analyses with acid titration, which revealed distinct differences in pH tolerance. We found that Pro199 was vital for binding affinity and denoted it “Key-position”, which also reflects its position in the CBD (Moonens et al., 2016). The second mutation in position 428 seemed to confer less difference in acid sensitivity, however its positioning was outside of the CBD, instead located in the “Velcro domain” region possibly important for dimerization. We also probed for antrum/corpus variations in a range of other strains, which revealed a subset of strains with oppositely less acid sensitive antrum clones compared to the corpus clones, which likely stems from BabA adaptation to local variations in the gastric environment with e.g. hypochlorhydria from corpus atrophy.

Globally a majority of _H. pylori_ strains display generalist BabAs, which binds to all type 1 ABO Lewis glycans. However BabA in e.g. South American Amerindian (SAA) strains are mostly specialists, restricted to the bg O antigen. Additionally it seems that gastric disease patterns vary where in India commonly antrum-predominant gastritis is seen with lower risk of gastric cancer, whereas the opposite pattern is found in Peru. We thus compared Le^b^-binding pH_{50}s of _H. pylori_ strains from these two regions and found 0.65 pH units mean difference, with Peruvian strains being more pH tolerant. We found the 199Pro in most of the Peruvian strain but only in one Indian strain arguing for its role in pH tolerance in Le^b^-binding. Furthermore, the 199-200 residues (Key-coil turn) were completely missing in most of the Indian strains. However some exceptional strains with acid resistant binding was found and these also included the Key-coil turn but with Lys or Arg in Key-position 199. We next expressed BabA from an Indian strain I9 and from the more acid resistant strain I18 on a shuttle vector in an otherwise isogenic _H. pylori_ strain and ran pH titrations assaying for Le^b^-binding. This experiment showed identical pH_{50}s compared to their corresponding “parent” strain and thus was instrumental in showing that acid sensitivity is encoded in the babA gene. We then grafted fragments of the region including the Key-coil turn with increasing downstream size and could shift the I9 strain into becoming increasingly more pH resistant, which was dependent on position 198 together with position 207 that forms a salt
bridge. Thus we propose that this salt bridge is broken during lower pH, abrogating binding, forming a pH-sensing action mechanism.

To study BabA evolution over time we looked for Le⁶-binding clones in strain USU101 after a long-term Rhesus macaque infection (Liu et al., 2009; 2015). A small number of clones where obtained and these bacteria were different in terms of acid sensitivity in Le⁶-binding when comparing antrum to corpus clones. They furthermore had amino acid replacements in the “Velcro domain” instead of substitutions in the CBD. We thus propose that the differences in acid sensitivity could be dependent on BabA oligomerization, which is expanded on below. As a second example of BabA evolution over time we obtained H. pylori from a Colombian patient after 0, 3 or 16 years of infection. Also here could we find variations in acid sensitivity between clones and accompanying amino acid sequence changes, albeit with the full series of substitutions located in the CBD.

As Le⁶-binding occurs with exceedingly high affinity we reasoned that this could be due to oligomeric binding and also found evidence for BabA multimerization when separating the BabA protein in 2D-fluorescence difference gel electrophoresis. We could furthermore see that acidification converted oligomeric forms of BabA into the monomeric protein suggesting that this transition could play a role in acid sensitivity and detachment. Also taking the Velcro domain mutations seen in the Rhesus macaque strain into account it is feasible that such amino acid replacements affect monomer/oligomer transitions as the Velcro domain acts like a hinge between the Head and the Stalk domains of the protein.

Given the indications of selection for certain BabA binding modes that related to host gastric acidity, we next probed for selection in real-time by mixing the pH sensitive SW38 strain with the pH tolerant SW7 strain and exposed them to cyclic variations of pH in a LigandTracer experiment. After first lowering the pH and then increasing it we found a 2-fold selection for reattachment of SW7.

In summary, our experiments reveal a novel protein-glycan binding mechanism where H. pylori rapidly detaches from Le⁶-receptors at low pH and then have the ability to re-activate binding and “recycle” the infection. This answers a long-standing question of how H. pylori can maintain persistent infection when simultaneously binding gastric receptors with antibody-like affinity. This capability completely depended on the inherently acid-tolerant BabA protein and we could furthermore expose regions on the BabA structure that are key for variations in acid sensitivity in Le⁶-binding. We believe that BabA sequence polymorphisms are a consequence of
individual selection for each and every \textit{H. pylori} strain. The medical and biological implications of BabA sequence variations are that there will be strains better or worse adapted to the antrum or corpus region of the stomach. This translates into predisposition to ulcerative disease (antrum infection) or gastric cancer (corpus infection).

**Paper IV**

\textbf{The Helicobacter pylori Sialic Acid Binding Adhesin SabA is Regulated via a Network of Two-Component Systems}

In Paper I we found that the T-tract greatly contributed to the control of \textit{sabA} expression/sLe\textsuperscript{-}-binding and we could also exclude that Hup and NapA influenced \textit{sabA} regulation. Knowing that SabA is repressed by low pH, in this project we were interested to determine how \textit{sabA} is controlled by other transcription factors. As described above in the section “\textit{H. pylori} gene regulation and two-component systems (TCSs),” \textit{H. pylori} only harbors few transcription factors and just three complete SK-RR pairs that have been implicated in DNA-binding.

We initially set out to study effects of the ArsRS system that is known to suppress SabA during acidic conditions. Here we could also confirm this suppressing effect for SabA in \textit{H. pylori} strain SM1109 upon bacterial incubation in pH 4 medium, which was not seen for BabA. An alternative interpretation is that the ArsRS system activates SabA at neutral conditions and that this activation is lost at lower pH. To further study the ArsRS effect we next mutated the \textit{arsS} gene (\textit{arsR} mutants are lethal) and found drastic SabA derepression at pH 4. This indicated that non-phosphorylated ArsR at neutral pH increases \textit{sabA} expression or possibly that ArsS acts on something else, e.g. regulating another TCS, that in turn controls \textit{sabA}. We also established that the repression was taking place during transcriptional initiation using \(\beta\)-galactosidase assays of \textit{PsabA::lacZ} promoter fusions introduced into the SM1109 and SM1109\textDelta{arsS} strains complemented with semi-quantitative RT-PCR experiments of the same strains. In future experiments it would also be interesting to study what happens to the \textit{sabA} mRNA at various pHs by e.g. determining the half-life adding rifampicin and sampling successively.

As \textit{H. pylori} are genetically diverse, as control we introduced the \textDelta{arsS} mutation into two additional \textit{H. pylori} strains and obtained similar results with strong derepression of SabA. However, low SabA expressing strains were more derepressed than higher SabA expressing strains, which was also true when we introduced the \textDelta{arsS} into otherwise isogenic strains only
varying in the T-tract, thereby having different sabA expression levels (Åberg et al., 2014). These results show that the repressing (or activating) potential in the ArsRS system is dependent on the PsabA appearance. Although the precise reason for this is clouded a combinatorial action of the T-tract appearance possibly facilitates ArsRS dependent repression.

Of further interest we found SabA derepression at neutral pH, indicating that the ArsRS system also represses sabA during neutral conditions or possibly that the non-phosphorylated ArsR can activate transcription at neutral pH. To discriminate the two possibilities future ArsR-antibody guided Western blot experiments from Phos-tag gel separation (or similar) will be instrumental in delineating at what pH and to what extent ArsR is phosphorylated. We have an ArsR antibody that recognizes native ArsR and experiments are in the pipeline. Another interesting experiment that did not work out was over-expression of ArsR, likely since alterations of a global regulator disrupts vital bacterial pathways. Intriguingly, at pH 4 the ΔarsS strains show more sabA repression than at pH 7, which in turn indicates that some other response regulator cross phosphorylates ArsR to mediate its repression or conceivably that some other trans-acting factor is regulating sabA.

We then turned our interest to the TCSs CrdRS and FlgRS and made additional H. pylori SK-mutants in ΔcrdS, ΔflgS and double mutants in ΔarsSΔcrdS, ΔarsSΔflgS and ΔflgSΔcrdS. Although the single mutants were fairly easy to obtain, the double mutants proved more strenuous, likely as multiple genetic control systems were being manipulated. Intra-strain comparison of SabA expression at pH 7 and pH 4 revealed that all of these strains were repressed to fairly similar levels at acidic pH, although the ΔflgS strain was slightly more repressed. Inter-strain comparison of SabA expression at pH 7 showed derepression in strain ΔarsS in contrast to the ΔflgS and ΔcrdS mutants that instead were repressed. This indicated that FlgR and CrdR in phosphorylated form either act as activators at neutral pH or that the non-phosphorylated FlgR and CrdR are repressors and possibly the opposite for ArsR. Looking at the SabA levels of the double mutants no additional effect was seen for the ΔflgSΔcrdS strain whereas the other two strains including the ΔarsS mutation was intermediate, but somewhat higher in SabA expression than the wt. This indicated that the ArsRS-system acts in concert with FlgRS and CrdRS to regulate sabA, although ArsRS seems to be the strongest component. Nevertheless, these data are hard to interpret without the knowledge of how the TCSs phosphorylate and regulate each other. Again, Western blot experiments revealing the degree of phosphorylation performed on the variations of SK-mutants will be highly informative to determine what phosphorylates what, during what pH this
occurs, the phosphorylation half-life, and the level of phosphorylation. Furthermore, one could also attempt to delete the *crdR* and *flgR* genes.

To establish if the RR sizes could interact directly with the PsabA we expressed and purified His-tagged ArsR, FlgR and CrdR and used them in electrophoretic mobility shift assays (EMSAs). As expected, FlgR that interacts with the σ54-RNAP does not bind PsabA DNA, however both ArsR and CrdR showed strong binding in this assay. These results indicate that ArsR and CrdR likely influence *sabA* expression via direct binding to the promoter DNA in contrast to FlgR that seems to act via some other non-direct mechanism. To obtain more information on in vivo binding e.g. bacterial chromatin immunoprecipitation experiments would provide a direct linkage to the region bound at various pH-conditions. Next we wanted to find out if the ArsR bound DNA both in its phosphorylated and non-phosphorylated form. Thus we purified a D52N mutant protein defective in phosphorylation and compared binding to the wt protein in EMSAs. Here we could see similar shifts, showing that the protein binds DNA in both forms.

A previous study looking into the structure of ArsR suggested that the protein is dimeric and furthermore suggested some amino acid residues that were predicted as important for DNA-binding (Gupta et al., 2009). To study the details in ArsR binding characteristics we purified the mutated K190R or R217A ArsR proteins, which both completely lost binding both in EMSA and SPR-experiments. We could thus experimentally confirm the DNA-interacting amino acids of ArsR. ArsR-dimerization was also suggested by another group that suggested that the ArsR-dimer binds PsabA DNA with dyad symmetry (Harvey et al., 2014). To verify this site and to find other possible interaction domains we ran EMSAs with a range of PsabA DNA-fragments of varying length, which indicated that the ArsR binds to several sites along the PsabA. To obtain more quantifiable data we shuffled the PsabA DNA sequence in 5 different regions and probed for ArsR binding to these shuffle-fragments in SPR. This revealed strongest binding to the site indicated by Harvey et al. However, in the sensitive SPR assay we also found less binding to DNA-fragments that were shuffled at the regions upstream from the -35 region. This suggested that ArsR not only binds to a highly specific site immediately downstream from the transcriptional start site but also that it can interact with the region where the RNAP αCTDs bind (Åberg et al., 2014). To complement these findings we made an otherwise isogenic *H. pylori* mutant in the PsabA site where ArsR bound strongest. This mutant strain revealed strong derepression confirming the key role of ArsR binding to “Site 4”.

31
Knowing that the ArsRS system acts more repressing in some strains than others we looked for sequence variations in the “Site 4” by multi-sequence alignments of numerous PsabA sequences. Indeed we found two different classes of sites that differed by the presence of a C residue interrupting either of two A/T-rich tracts and a third subtype, e.g. found in H. pylori strain 17875 that had no C present at these loci. In SPR experiments we found that the 17875 PsabA was more strongly bound by the ArsR, indicating that intact A/T-tracts are key for efficient ArsR binding, which was also confirmed when we obtained much lower binding upon introduction of C’s in the 17875 sequence.

Last we set out to identify if ArsR could dimerize. Here we ran non-reducing protein gels and Western blots of the purified protein, with and without pre-treatment with phosphorylation agent. We found ArsR monomers and dimers plus additionally larger multimers, which were also seen when analyzing H. pylori extracts. The tendency to dimerize was interestingly higher in the DNA-binding ArsR mutants contrasted with the seemingly decreased multimerization upon phosphorylation. Western blots from native Phos-tag gels ran at various pH conditions should also here help determine the impact pH and phosphorylation has on multimerization in vivo.

In conclusion, our results show that the TCSs ArsRS, CrdRS and FlgRS all act to control sabA expression via different mechanisms that to a large extent still are shrouded. To our knowledge such co-operative control by TCSs is not exceedingly common and likely reflects H. pylori’s slim repertoire of TCSs. Contrasted with the slower stochastic control of SabA expression in the bacterial population, the faster regulation mediated by the TCSs is likely needed as SabA is employed to bind sialic acids that surface during inflammation when too tight adherence possibly means ingestion by the immune system. As to the general question of why SabA is repressed during acidic conditions, I believe that the answer lies partially in the conclusion of Paper III; it is a survival mechanism to maintain infection. However, to complete the transcriptional network surrounding sabA, additional experiments needs to be performed, including how the remaining regulatory proteins: HrcA, HspR, Fur and NikR impacts sabA expression. Then of course it would be highly interesting to study if sabA is subject to post-transcriptional regulation, possibly similar to ureAB that seems to be controlled by both ArsRS and antisense sRNA in conjunction (Wen et al., 2013; 2011).
CONCLUDING REMARKS AND OUTLOOK

I have in my thesis focused on how *H. pylori* regulate the adhesin proteins BabA and SabA. Although this is just one small piece in a bigger puzzle portraying the host-pathogen interplay, bacterial attachment is likely key for persistent infection. Yet, intriguingly one study examining OMP expression in 200 *H. pylori* isolates found that only around 2/3 of the strains expressed BabA, 1/3 expressed SabA and furthermore was BabA rapidly turned off in animal infection experiments (Liu et al., 2015; Odenbreit et al., 2009; Ohno et al., 2011; Solnick et al., 2004; Styer et al., 2010). However, another study focusing on human infections showed quite stable Leb-binding over time, in contrast to what is seen in animal models (Nell et al., 2014). Taken together all these studies certainly merits some critical thinking about how and why *H. pylori* expresses its attachment factors. It also argues that animal models are sub-optimal for studies of *H. pylori*. Hopefully the future will reveal functional alternative methods, e.g. by using novel thinking in cell line production, such as human gastric organoids (McCracken et al., 2014), recently used to probe CD44s role in gastric disease (Bertaux-Skeirik et al., 2015).

The purpose of *H. pylori* attachment seems multifaceted, with obvious motives such as nutrient acquisition or avoidance of clearance but also facilitation of virulence mediator delivery as shown for BabA (Ishijima et al., 2011). In support of the latter a comparative study between *H. pylori* and *Helicobacter felis* could not identify babA, sabA, vacA or the cagPAI in *H. felis* although large parts of the genome was similar, indicating that all those virulence factors influence each other (Arnold et al., 2011c). Possibly the bacterium needs to be close to the target tissue upon delivery of immunomodulatory or toxic virulence mediators to avoid their neutralization by antibodies. All of the reasons stated above imply the need for stringent control of adhesin expression.

As described in the introductory section, BabA and SabA can turn on or off expression in a slower fashion via phase variation mechanisms. Although studies as the one by Odenbreit et al., 2009 have paved the way towards better understanding of how the OMPs are expressed, what is found by the lab-bench is not necessarily what is found in vivo. Accordingly, with some few exceptions, as long as the babA or sabA genes are present one can surmise that there will be a time or place when these genes are turned on (or off), such as for SabA during *H. pylori*-induced inflammation with up regulation of sLe[x] receptors (Mahdavi et al., 2002). We reasoned that *H.
H. pylori should also possess a more delicate way to control the amount of sabA expression to respond to both slower and faster environmental changes. In Paper I, we could show that T-tract length variations in the sabA promoter allows for dramatic variations in the degree of SabA expression, promoting an H. pylori population that carries a sum of many expression phenotypes. Although SSM is a well-known phenomenon, our study sheds light on a new mechanism where the appearance of a polynucleotide tract changes the DNA curvature, which limits or facilitates αCTD RNAP binding to the PsabA, largely independent of other transcription factors. We could also see that this SSM mechanism was at play for other H. pylori genes, indicating a general more slow control mechanism. In Paper IV we further complete the picture of sabA regulation and showed a complex network of transcriptional faster control by TCSs. Thus the combined action of direct control by trans-acting transcription factors with a heterogeneous population of different basal transcription of sabA make up a tremendous diversity in SabA output, thus there will always be a clone best fit for the present conditions.

In terms of expressional control by TCSs we could only find that the CrdRS system influenced BabA expression, which could argue that BabA somehow is important for metal homeostasis. This is a largely unexplored field, which would be interesting to follow up with e.g. binding studies of nkrR and fur mutants. As BabA expression is not affected by low pH, Paper III aimed to determine how H. pylori avoids clearance upon epithelial shedding into the acidic lumen by some other means of fast BabA control. Here we found that the BabA protein itself is reversibly sensitive to low pH, which argues that the infection can be “re-cycled” when cognate receptors are released close to the acidic gastric lumen. As electron micrographs reveal a multitude of cagT4SS structures protruding from H. pylori cells, it seems that this protein complex also is highly important for adherence, especially in regions where the tissue is distorted exposing integrin receptors. Interestingly, the cagT4SS CagL binding was also abolished by low pH (Bonsor et al., 2015). Taken together it seems plausible that H. pylori adhesins are sensitive to acid, supporting the “re-cycling” hypothesis. Extrapolating from this, one role of proton pump inhibitor (PPI) treatment could be that higher pH leads to better H. pylori binding and thus loss of infection due to mucus turnover with adherent bacteria.

In Paper II we determined the BabA structure and obtained key insights into a three-pronged binding site; for discussion on receptor specificity and comparison to other blood group binding adhesins see (Moonens et al., 2016). As more and more structures emerges, in the future it should be possible make in silico models of wanted binding interactions and next synthetically produce cheap fusion-proteins that display high affinity binding
towards a given receptor with an additional payload of choice. The possible implications of this are mindboggling, spanning from targeting infectious disease to cancer treatments. Similar to our finding that either redox or pH controls the function of the BabA adhesin, designer proteins can be made to only activate themselves in certain environments. Thus, in a bigger context, studies like ours will likely impact the future in ways that are hard to foresee.

Zooming in on our structural findings on BabA, there were three protruding disulfide clasped loops (CL1, CL3, CL4), which display significant sequence variation and does not seem to play a major role in binding. Therefore one could suspect that these regions instead function as a decoy for antibodies that can be rapidly varied. We also found that CL2 was instrumental for BabA binding. Strikingly, this cysteine pair was also found in the babC gene, but not in the other OMP’s, which makes it highly likely that BabC is an undiscovered H. pylori adhesin. As the CL2 here is more widely spaced with >13 amino acids between the Cys-residues, not similar to BabA, one can envision that this protein targets another gastric receptor. The other suggested adhesins lacks the CL2; one alternative could be that these adhesins target some non-human environmental receptor. Although this seems somewhat far-fetched, biology has a tendency for surprise.

We could in Paper II exploit the CL2 and disrupt binding by using the redox active NAC. It could therefore be beneficial to look for disulfide bond dependence in other virulence mediators. One example is the multimeric Gal/GalNAc Lectin of Entamoeba histolytica that has a heavy (H) subunit, an intermediate (I) subunit and a light (L) subunit, where H and L are connected via a disulfide bond, reviewed in (Aguirre García et al., 2015). The heavy subunit, which seems to be active in binding is cysteine rich (Pillai et al., 1999). Upon antisense RNA inhibition of the light subunit, cytotoxic activity in hamsters was inhibited indicating that the subunits need to be simultaneously active for optimal function. Therefore, breaking the L-H subunits disulfide bond could possibly, similar to BabA, be targeted by NAC.

NAC treatment not only reduced Leb-binding but also reduced inflammation in the mouse stomach. This is likely caused by reduced H. pylori adherence, but could possibly be an isolated effect of NAC itself. Keeping in mind that BabA expression is only turned on in about half of infected it could be argued that NAC is non-functional against non-Leb-binding strains. However, it is likely that NAC also acts against other OMPs containing disulfide bonds, and also thins the mucus layer, which makes it universally active, although maybe less so against non-BabA expressing strains.
A dual effect against *H. pylori* and inflammation was also recently shown for liposomal linolenic acid (LipoLLA) that was tested with success in mice (Thamphiwatana et al., 2014). In times of increasing antibiotic resistance, adding additional drugs to the *H. pylori* eradication arsenal is needed, especially ones that are anti-inflammatory. Therefore it would be highly interesting to see clinical trials combining the effects of NAC, LipoLLA, PPIs and variations of antibiotics. To this, hopefully new drugs will be developed against novel targets such as the Hpn and Hpn-2 proteins involved in nickel accumulation that was recently to be essential for mouse colonization (Vinella et al., 2015). With a broader setup of possible treatment combinations it will be easier to tailor the therapy towards a specific *H. pylori* strain. Possibly there will also be situations where multi-resistant strains are found that prove impossible to eradicate. In these situations anti-inflammatory treatments and suppression of infection by drugs such as NAC and LipoLLA could prove useful, however more studies are needed to determine their anti-inflammatory role. Yet another benefit of not using classic antibiotics is that it to larger extent avoids killing commensals whose role in terms of keeping the immune system in check and acting as a bio-barrier is becoming more and more appreciated. Furthermore it limits the spread of antibiotic resistance in other microbial inhabitants. Taken together the possibilities of using NAC in larger clinical trials with other novel treatments could certainly be a future clinical research subject for me.

Finally, although the PhD-work has been a struggle with hundreds of failed experiments, I will always remember when I picked up NAC at the pharmacy, went into the lab and found a drastic reduction in Leb-binding in a RIA assay. That was some of those days when the life of a PhD-student was extraordinary!
ACKNOWLEDGEMENTS

This thesis would not have happened if it weren’t for the help and support of many fantastic people that work or have worked at medchem during my years as a PhD-student, thank you all!

First and foremost I would like to thank my main supervisor Anna Arnqvist for having me as a PhD student. I highly appreciate all of our scientific as well as casual conversations. Your patience has been endless and I could not have wished for a better and more caring supervisor! You always strive for the better and really pushed all of us in the group to excel, but never at the cost of family. I really felt that you wanted the best for me all the time.

Second I would like to thank my co-supervisor Thomas Borén. Your scientific determination is a source of inspiration. We had some great months in Davis, which started a fruitful collaboration all through my PhD period. Thank you for letting me take part on your projects that indeed turned out great, culminating in the structure-paper.

Past and present members of the Arnqvist and Borén groups: Thank you Anna Vallström for introducing me to the H. pylori world and for teaching me how to grow vegetables =). You and me have gone through a similar journey combining MD/PhD studies and children, thanks for all help and discussions! Also thank you Carina Öhman, Jenny Nordén and Rolf Sjöström for introduction to the lab-work. The king, Annelie Olofsson, thank you for everything in the lab and outside. You are inspirational in so many ways; I have never met such a self-less thoughtful person, and to top it up you are a fantastic scientist with a great sense of humor. Thanks for being such a good friend and making me laugh! Speaking of which, thank you Christer Larsson for fun times and good luck with your group! Also big thank you to Göran Bylund for all help in the lab and for hundreds of lunch breaks and the trips to your island! Olena Rakhimova, I will always remember our bicycle rides in the underground, thanks for all chats in the Hood! Also thanks Alexej Schmidt for helping with constructive ideas! Melissa Mendez, you really are a sunshine, thanks for brightening up my days and best of luck with your future endeavours. Sara Henriksson, thanks for all help and fun talks, remember that you are great scientist and a gifted writer.
This thesis would certainly not have had the same shape without the help from two highly talented scientists: Anna Shevtsova and Anna Åberg. Both of you are so incredibly persistent and thorough, I’m sure that either of you would be a great asset to any supervisor in almost any field! Special thanks to Anna Shevtsova for helping me with variations of RIA-assays, statistics and for all lovely cakes! Special thanks to Anna Åberg for all help with both practical issues in the lab and fruitful rounds of passing manuscripts around. I have learned so much from you! Also, Jeanna Bugaytsova, thanks for fun discussions and all the cat-pictures =). Ikenna Obi, keep up the good work with OMVs and Olivia Ohlsson, welcome to the group and thanks for some good lunch talks. Marie Westermark, it was really nice to have you in the group for a while. Kristoffer Brännström thanks a lot for helping me with Biacore experiments and for nice talks about fishing, economics and other stuff!

The guys in Davis: Jay Solnick, Mary Kable and Lori Hansen. The stay in your lab was highly educational and I’m very grateful to the hospitality you all showed me! I miss California! Also thanks to the people of Rainer Haas and his group for two nice stays in Munich. Special thanks to Evelyn Weiss for making my stay wonderful e.g. with our trip to the Neuschwanstein castle. I would also like to thank Guiu Xuclà Serés, you were brave venturing into the wild with us!

Huge thanks to the administrative and service staff: Clas, Ingrid, Anna S, Elisabeth and Jenny.

I have been blessed with the best family there is, thank you for all support through the years: Staffan, Berith and Karin with family. Although I may not have shown it all the time, I appreciate all of you enormously! My wife Ida, I could not have done this without you! Thank you for your patience during my thesis work and for your support since our little star Moa arrived! I love you! Also, thank you Pippi and Doris for all hours of purring in my lap while working. There are of course many other friends important to me, thank you all for being who you are!

Finally to the microbes that knocked me out with gastroenteritis on the last week of writing, a very ironic thanks and golf clap, nice job.
REFERENCES


