The Cytotoxin-Associated Gene A (CagA) of *Helicobacter pylori*: the Paradigm of an Oncogenic Virulence Factor

Alfiia I. Iunusovaa, Irina S. Litvinovaa, Polina A. Karpenoka and Abolghasem Tohidpoura,b,c,*

Siberian Federal University
Institute of Fundamental Biology and Biotechnology
Department of Biophysics
79 Svobodny, Krasnoyarsk, 660041, Russia

Krasnoyarsk State Medical University
named after Prof. V.F. Voino-Yasenetsky
Research Institute of Molecular Medicine and Pathobiocchemistry
1 Partizan Zheleznyak Str., Krasnoyarsk, 660022, Russia

Monash University
Monash Biomedicine Discovery Institute
Department of Microbiology, Infection and Immunity Program
Clayton, Victoria 3800, Australia

Received 10.03.2017, received in revised form 01.04.2017, accepted 05.04.2017

* Corresponding author E-mail address: a.tohidpour@gmail.com

*Corresponding author E-mail address: a.tohidpour@gmail.com

Helicobacter pylori is a microaerophilic, spiral-shaped and gram-negative microorganism that produces various virulence factors such as CagA, VacA, urease, and host cells adhesins, which in a synchronous concert, allow H. pylori to colonize and infect the host gastric epithelium. *H. pylori* infection is associated with some severe side effects in human, such as gastritis, peptic ulcer, non-Hodgkin’s lymphoma and adenocarcinoma. CagA is the most notorious virulence factor of *H. pylori*. It is known as the first bacterial oncoprotein. The gene encoding CagA is localized on the cag pathogenicity island (cagPAI), a 40kbp DNA segment which also carries genes for the type four secretion system (T4SS) of *H. pylori*. The interaction of CagA with intracellular partner proteins leads to some irreversible alteration of host cells by increasing cell size, elevating motility, phenomena known as the “hummingbird phenotype”. CagA also disrupts the epithelium apical junctions and thereby destroys the normal epithelial architecture. A tyrosine phosphorylation site, named EPIYA motif, helps CagA to bind to cytosolic proteins in a phosphorylation-dependent manner. CagA is also interacts with host proteins in a phosphorylation-independent fashion, which altogether will assist to develop adenocarcinoma in infected cells. This review summarizes the core data on the structure and function of CagA and its role in conferring the main pathophysiologic
effects of *H. pylori* infection as well as suggesting a therapeutic option for treatment of *H. pylori* infection based on CagA virulence.

**Keywords:** adenocarcinoma, cagPAI, EPIYA, T4SS, hummingbird phenotype.

**DOI:** 10.17516/1997-1389-0015.

---

Цитотоксин-ассоциированный ген А (CagA)

*Helicobacter pylori*: парадигма онкогенного фактора вирулентности

А.И. Юнусоваа, И.С. Литвиноваа,
[П.А. Карпенок]а, А. Тохидпура,б,в

а Сибирский федеральный университет

Институт фундаментальной биологии и биотехнологии

Факультет биофизики

Россия, 660041, Красноярск, пр. Свободный, 79

б Красноярский государственный медицинский университет

им. проф. В.Ф. Войно-Ясенецкого

Научно-исследовательский институт молекулярной медицины и патобиохимии

Россия, 660022, Красноярск, ул. Партизана Железняка, 1

в Университет Монаш

Монаш институт биомедицинских открытый

Факультет микробиологии

Программа по инфекции и иммунитету

Клейтон, Виктория 3800, Австралия

*Helicobacter pylori* — это микроаэрофильная, спиралевидная, грамотрицательная бактерия, которая производит различные факторы вирулентности, такие как CagA, VacA, уреаза, а также адгезины, которые обеспечивают адгезию к клетке-хозяину. Синхронизированное взаимодействие факторов вирулентности позволяет *H. pylori* колонизировать и инфицировать эпителий желудка хозяина. Инфицирование организма человека *H. pylori* вызывает ряд побочных эффектов, таких как гастрит, язвенная болезнь желудка и двенадцатиперстной кишки, неходжкинская лимфома и аденоакарцинома. CagA является наиболее печально известным фактором вирулентности *H. pylori* и признан первым бактериальным онкогеном. Он расположен на островке патогенности cag (cagPAI) — сегменте ДНК размером 40 т.н., который также содержит гены системы секреции четвертого типа (T4SS) *H. pylori*. Взаимодействие CagA с внутриклеточными белками-партнерами приводит к некоторым необратимым изменениям в клетках хозяина (увеличение их размера, повышение подвижности...
Introduction

Since the discovery of a spiral bacterium named *Helicobacter pylori* in 1982 by Marshall and Warren (Marshall and Warren, 1984), a significant amount of attention is given to this pathogen of various human gastric diseases. *H. pylori* is a microaerophilic and gram-negative microorganism which colonizes the mucosal layer of gastric tissue (Tohidpour, 2016). Two original morphological shapes, bacillary and coccoid, have been described for *H. pylori*. Although the bacillary form is clearly the predominant form, the coccoid morphology is the only form of *H. pylori* observed in vivo. It is thought that the bacillary form is virulent and the coccoid form mainly serves to protect the microorganism. The bacillary form of *H. pylori* is highly motile with multiple unipolar flagella (Chan et al., 1994; Covacci et al., 1999).

*H. pylori* has successfully colonized the stomach of around 50% of the world population. However, the prevalence of its infection varies within and between countries. The rate of infection depends on the socioeconomic levels as well as the genetic background and lifestyle of the hosts that are affected. In such regions as the East Asia and some parts of Latin America, the prevalence of infection tends to be high. In these areas, the initial encounter with the bacteria usually occurs in childhood, so that about 80% of the adult population will develop the infection by the age of about 20. In contrast, developed countries such as Europe, USA, or Australia, show less prevalence of *H. pylori* infection in children (aged below 10) and adults show a maximum 40% rate of infection (aged 30 to 40) (Konturek et al., 2005; Eusebi et al., 2014; Crew and Neugut, 2006; Graham et al., 1991; Yamamoto, 2001). In 1955, the overall prevalence of *H. pylori* infection in children in St. Petersburg, Russia, was 44%. Ten years later this rate decreased to 13%, probably due to the significant improvement of household hygienic practices and use of anti- *H. pylori* eradication therapies. In fact, several studies showed that the rate of *H. pylori* infection is significantly higher in rural areas compared to the urban sites, which further approves the effect of lifestyle on the disease incidence by *H. pylori* (Malaty, 2007; Tkachenko et al., 2007; Dore et al., 2002).

*H. pylori* produces many enzymes that facilitate colonization; such as catalase, phospholipase, thioredoxin reductase, and urease. Urease is a hydrolysis enzyme, which breaks the urea into ammonium and carbon dioxide and therefore provides an ammonia foam which protects the bacteria from the harsh acidic environment of the stomach (Chan et al., 1994; Covacci et al., 1999; Cover and Blaser, 1992;...
The bacterium usually enters the stomach through the fecal-oral route, but it can also contaminate food or water. However, the person-to-person contact, especially within the families is known as the primary source of contamination (Nurgalieva et al., 2002; Klein et al., 1991; Hopkins et al., 1993). *H. pylori* has a colonization site specificity to the gastric epithelium and uses its polar flagella to actively attach to the epithelium and resist the gastric flow which normally removes the hostile pathogens from the gastric lumen. *H. pylori* then migrates to the surface of epithelium, the pyloric antrum, which has a higher pH (6-7) and is optimal for the growth and colonization of *H. pylori* strains (Eaton et al., 1989; Scott et al., 1998).

### Major virulence factors of *H. pylori*

Adaptation of *H. pylori* to survive in the acidic niche of gastric epithelium has enabled it to induce severe pathological outcomes such as gastritis, peptic ulcer, and gastric cancer. *H. pylori* can produce various virulence factors, which help to colonize the stomach tissue and damage the epithelial mucosa such as cytotoxin-associated gene A antigen (CagA), vacuolating toxin (VacA), Urease, blood group antigen-binding adhesin (BabA), outer inflammatory protein (OipA), and induced by contact with epithelium protein (IceA).

### Cytotoxin associated gene A antigen (CagA)

CagA is the most important virulence factor of *H. pylori* with a well-established role in the induction of mucosal inflammation. It is a 120-145 kDa protein secreted by the virulent strains of *H. pylori*. CagA function is associated with some of the most notorious pathophysiologic outcomes of *H. pylori* infection. The gene encoding CagA is located at one end of a large 40 kbp DNA segment called cag Pathogenicity Island (cagPAI).

It is assumed that cagPAI has been inherited by *H. pylori* from unknown ancestors through the horizontal gene transfer mechanism. The cagPAI comprises about 27-31 genes, which mainly encode CagA and subunits (18 genes) of the type IV secretion system (TFSS) of *H. pylori* (which upon assembly, form a tunneling apparatus for delivery of CagA into the host cells) (Covacci et al., 1993). Based on the ability to produce CagA, the strains of *H. pylori* are divided into two subpopulations: cagA-positive and cagA-negative.

The cagA-positive *H. pylori* strains are associated with a higher degree of gastric inflammation and are more virulent than the cagA-negative strains (Kuipers et al., 1995; Parsonnet et al., 1997).

### Interaction of CagA with host proteins and development of adenocarcinoma

Upon translocation into the host cytoplasm using the TFSS, CagA localizes itself to the inner leaflet of the plasma membrane and undergoes tyrosine phosphorylation by several intracellular kinases such as c-Src, Fyn, and Lyn (Selbach et al., 2002; Stein et al., 2002). Tyrosine phosphorylation of cytosolic proteins plays a crucial role in transmitting intracellular signals for growth, movement or differentiation in the mammalian cell. Studies showed that tyrosine phosphorylation of some bacterial proteins enables them to intervene the intracellular signal transduction and induce cellular dysfunction which can lead to cell transformation and malignancy (Hatakeyama and Higashi, 2005). The tyrosine phosphorylation site of CagA consists of a unique conserved array of amino acids: Gly-Pro-Ile-Tyr-Ala, so-called EPIYA motif. Based on the flanking amino acid sequences surrounding the EPIYA, four different types of EPIYA segments (EPIYA: -A,
-B, -C, and -D), are found. EPIYA-A (32 amino acids (a.a)) and EPIYA B (40 a.a) are commonly found in all types of CagA. EPIYA-C (34 a.a) is mainly found in strains isolated from the western regions (Europe, North America, and Australia) and few Asian countries such as India and Malaysia (Western \textit{H. pylori} strains). The number of EPIYA-C motifs on the C-terminus of Western-\textit{H. pylori} CagA can vary up to maximum three. EPIYA-D, on the other hand, is solely found as a single repeat on the C-terminus of CagA in \textit{H. pylori} strains isolated from the eastern countries such as China, South Korea, and Japan (East Asian \textit{H. pylori} CagA) (Higashi et al., 2002a, 2002b; Hatakeyama, 2004; Higashi et al., 2005). CagA can interact with intracellular partner proteins in either phosphorylation-dependent or independent fashion. In the phosphorylation-dependent manner, upon tyrosine phosphorylation, CagA interacts with various host cytoplasmic proteins such as c-terminal Src kinase (Csk), SHP-2, Grb2, CrkII, PI3k, and SHP-1 (Hatakeyama and Higashi, 2005; Higashi et al., 2002a; Hatakeyama, 2006). SHP-2 is one the most important targets that undergo interaction with CagA and plays an essential role in signal transduction pathways of growth factor/cytokine receptors and regulates cellular responses such as proliferation, morphogenesis, and cell motility. Interaction of CagA with SHP-2 disrupts its physiologic functions and leads to a morphologic transformation of cells, named the \textit{hummingbird} phenotype. The CagA-SHP-2 complex is found in the atrophic gastric mucosa so that the complex can play a crucial role in the development of atrophic gastritis and the transition from atrophy to intestinal metaplasia (Ohnishi et al., 2008; Pattis et al., 2007; Hatakeyama, 2006; Neel et al., 2003; Tegtmeyer et al., 2011).

SHP-2 has two repeated Src homology (SH2) domains (N-SH2 and C-SH2) on the N-terminal and a protein tyrosine phosphatase (PTP) domain on the C-terminal region. The N-SH2 domain of SHP-2 includes the catalytic cleft of the PTP domain, which blocks the substrate access. Interaction of CagA with the SH2 domains induces a conformational change in SHP-2, so the inhibitory effect of the PTP is suppressed, resulting in the activation of SHP-2 phosphatase activity. The SHP-2 specifically binds to the tyrosine-phosphorylated EPIYA-C, and EPIYA-D sites of Western and East Asian \textit{H. pylori} strains CagA. However, the EPIYA-D motif shows stronger binding affinity to SHP-2 than EPIYA-C (Higashi et al., 2002a, 2002b; Hof et al., 1998; Naito et al., 2006). The CagA-SHP-2 complex dephosphorylates the activating tyrosine phosphorylation sites of focal adhesion kinase (FAK) (Tyr396, Tyr574, and Tyr575) and down-regulates the FAK kinase activity (Higashi et al., 2002a, 2002b; Tegtmeyer et al., 2011; Hatakeyama, 2006; Higashi et al., 2004; Higuchi et al., 2004). SHP-2 activates the extracellular signal-regulated kinase (Erk) mitogen activated protein (MAP) kinase pathway by both Ras-dependent and Ras-independent mechanisms (Ras are GTPses which role as molecular switches to regulate some intracellular signaling pathways). CagA-SHP-2 mediated deregulation of Erk kinase deregulates the normal cell cycle and triggers the development of hummingbird phenotype in infected cells (Tegtmeyer et al., 2011; Neel et al., 2003; Tsutsumi et al., 2003).

CagA can also interact with host proteins in a phosphorylation-independent fashion. One of the most well-described proteins which interacts with CagA in this manner is an adaptor protein named Grb2 (growth factor binding receptor protein 2). CagA binding to Grb2 activates Ras and elevates the cell motility, and induces a scattering phenotype (Hatakeyama, 2003; Mimuro et al., 2002). Development of gastric cancer (adenocarcinoma) is a multistep process,
which includes deregulation of intracellular pathways, changing the expression rate of the oncogenic genes and production of inflammatory responses. Successful colonization with CagA-positive strains of \textit{H. pylori} and persistence of CagA within the host cells can eventually lead to irreversible changes in the host cells, causing anomalous signals for growth, cell motility, and development of abnormalities in the infected cells (Hatakeyama, 2006). Moreover, chronic infection with cagA-positive strains of \textit{H. pylori} triggers some histopathological changes in the gastric mucosa which show exemplary steps of so called the intestinal-type gastric adenocarcinoma: starting from the superficial gastritis to atrophic gastritis, intestinal metaplasia, dysplasia and finally ending with the adenocarcinoma (Hatakeyama, 2006; Yamazaki et al., 2003; Correa, 1992).

\textbf{Analysis of CagA crystal structure}

Determining the three-dimensional structure of CagA has been the main burden to identify its effect on the host cells. Due to the large size, lack of significant homology to other prokaryotic proteins and a high degree of flexibility, it has been difficult to establish a stable structural analysis of CagA, a hydrophilic protein with no transmembrane sequence. The most hydrophilic region of the CagA includes some region of amino acid repeats such as EFKNGKNKDFS, EPIYA and a stretch of six asparagine amino acids (NNNNNN), which are on the C-terminus of CagA (Covacci et al., 1993; Tsutsumi et al., 2006). Currently, there are few successful reports of the three-dimensional (3D) structure of CagA. Kaplan-Turkoz et al. (2012) presented a crystal structure of CagA consisting of four domains (D1-D4). Domain 1 (D1) (amino acid residues 1-270) was not in the model because the electron density map of the latter region was in low quality. Domain 2 (D2) (305-642) is the central domain of CagA and consists of three subdomains. The central part of D2 forms a so-called SLB (single-layer β-sheet) region, which is known as CagA binding site to β1 integrin on the surface of the host cell (plays a key role in CagA translocation). SLB consists 11 antiparallel strands (β1–β5 and β8–β13) with a left-hand twist at β8. Next subdomain, D2’ is inserted between strands β5 and β8 and formed by β-hairpin (β6, β7) and helices from α8 to α10. Third subdomain, D2’’, comprises three helices, α11, α12 and α13. D2’’ stabilizes the upper part of the β-sheet and also looks like a hairpin structure. Domain 3 (D3) (648-705) is a single helix (α14) that links D2 and D4. Domain 4 (722-822) contains four antiparallel helices, α15-α18, which are located at the carboxyl terminal of CagA (Fig. 1).

Another study by Hayashi et al. (2012) studied the 3D structure of the N-terminus of CagA from \textit{H. pylori} ATCC 26695. They reported that about 70% of CagA has a stable structure which mainly includes the N-terminus of the protein. The rest 30% is mainly located on the C-terminus and is significantly disordered. Their crystal structure of the N-terminal CagA consisted three domains (I-III). Domain I (residues 24-221) connects with domain II (residues 303-644) by a disordered region which includes about 80 amino acids. The interaction between the surfaces of domains II and I is small which implies that domain I could readily dissociate from domain II and therefore could be highly mobile in solution. Domains II (residues 303-644) and III (residues 645-824) are connected by a long α-helix (α-19) and form an “N-shaped” structure in the center of CagA. This region includes 13 α-helices and a large β-sheet structure (Hayashi et al., 2012).

\textbf{Optimum strategies to manage the infection of \textit{H. pylori}}

Currently, there are several types of therapies to treat the \textit{H. pylori} infection. Some
of the recommendations for the eradication of \textit{H. pylori} were adopted in 2010 by the scientific society of Gastroenterologists of Russia (Standards for the diagnosis..., 2010) and the Maastricht IV (Malferttheiner et al., 2012). Accordingly, selection of the eradication scheme depends on the availability of particular patients whose bodies respond to the medication, as well as the sensitivity of \textit{H. pylori} strains to these drugs. Currently, three main antibiotics are widely used to treat the \textit{H. pylori} infections. These include clarithromycin, metronidazole, and amoxicillin (Kim et al., 2015). The term “A single triple therapy” refers to an elimination method, which consists of administration of two antibiotics (usually a mix of clarithromycin and amoxicillin) and a proton pump inhibitor (Molina-Infante and Gisbert, 2014; Gisbert et al., 2000). Although the latter method proved significantly effective, due to various reasons such as genetic polymorphism, smoking habits, and bacterial resistance to antibiotics (Graham and Fischbach, 2010; Gasparetto et al., 2012; Xie and Lu, 2015) its efficacy has noticeably decreased. Another more efficient regimen of therapy is called bismuth quadruple therapy, which includes a combination of bismuth subcitrate potassium, tetracycline, and a proton pump inhibitor (Kim et al., 2015). The quadruple therapy method has been remarkably successful in eradicating clarithromycin-resistance cases of \textit{H. pylori} infection (Papastergiou et al., 2014; 

---

**Fig. 1.** The general schematic structure of the N-terminal domain of CagA. A – Illustration of the location of the domains (D1 is shown in red, the SLB is colored pink, subdomain D2’ is in cyan, D2’’ is shown in blue, D3 is green, and D4 is yellow. A, B, and C represent the location of EPIYA motifs. The PAR1-MARK kinase binding regions (Nesic et al., 2010) are shown as MKI1 and MKI2). B – 3D ribbon structure of the N-terminal domain of CagA. C – A cartoon representation of the domains of CagA (Figure source: Kaplan-Turkoz et al., 2012, Fig. 1).
Kim et al., 2015). The use of clarithromycin in the eradication schemes is only possible in areas where resistance is up to 15-20%. In regions with clarithromycin resistance higher than 20%, clarithromycin is suitable for determination of the sensitivity of *H. pylori* to antibiotics (Malfertheiner et al., 2012).

Vaccination seems a reliable option to control the reduced efficacy of antimicrobial-based therapy and prevent the development of *H. pylori*-related malignancies such as adenocarcinoma. The currently available anti-*H. pylori* vaccine contains recombinant CagA, VacA, and *H. pylori*-NAP (neutrophil activating protein). This vaccine has shown a good degree of immunogenicity and safety during the clinical trials, however it still requires further analysis before receiving the approval for clinical use (Malfertheiner et al., 2008). Recent studies have found some antigens of *H. pylori* for a new vaccine so-called pan-vaccine, which comprises a set of antigens to protect against *H. pylori* strains from different geographic regions (Walduck et al., 2015).

The development of a successful anti-*H. pylori* vaccine is restricted due to several factors (Sutton and Chionh, 2013). A major issue is the lack of knowledge about the interaction of immune system against *H. pylori* infection. Most of the studies in animal models could only decrease the colonization rate of *H. pylori* but hardly achieved a complete eradication or protection against re-emergence of the infection. Due to the high rate of *H. pylori* infection amongst human population, an ideal anti-*H. pylori* vaccine should be able to provide prophylaxis and treatment simultaneously. What would further limit the development of a *H. pylori* vaccine is the possible benefits from the colonization of *H. pylori* to some hosts (Atherton and Blaser, 2009; Arnold et al., 2012). However, it is suggested that such advantages of *H. pylori* would be exerted in the early stages of the host life, while malicious effects of *H. pylori* infection start to appear over the adulthood stage (Atherton and Blaser, 2009). Therefore despite the solid knowledge that *H. pylori* can develop very adverse effects on its host, the potential of having some benefits in the early stages of host colonization shows that it is critical to carefully design vaccines which are based on the most virulent factors of *H. pylori*. Perhaps the best candidate for such vaccine strategy is CagA, which is only found in the pathogenic strains of *H. pylori*. Therefore, by eliminating the CagA-positive *H. pylori* strains and allowing the CagA-negative strains of *H. pylori* to colonize, the vaccination therapy would allow to reduce the adverse pathogenic effects but also to receive the advantage of commensalism between the non-pathogenic *H. pylori* strains and human host. In other words, a valuable *H. pylori* vaccine would be able to prevent gastric cancer, even without providing the sterilizing immunity.

**Acknowledgements**

This work was funded and supported by the Siberian Federal University SibFU International Competitiveness Enhancement Program (Project “5-100” [Grant № M 2.2.3]).

With thanks to Professor Valentina Kratasyuk (Siberian Federal University, Krasnoyarsk, Russia), and Professor Alla Salmina (Research Institute of Molecular Medicine and Pathobiochemistry, V.F. Voino-Yasenetsky Krasnoyarsk State Medical University, Krasnoyarsk, Russia) for their support.
Alfiia I. Iunusova, Irina S. Litvinova… The Cytotoxin-Associated Gene A (CagA) of Helicobacter pylori: the Paradigm…

References


