Journal of Proteomes, Interactomes and Biological Networks

PIBN

Review

CagA as an indicator for *Helicobacter Pylori* infections.

Gaia Corona [1,*], Samuele Storari [1] and Laura Claudia Verdesca [1]

[1] Department of Pharmacy, Biotechnology and Sport and Exercise Science; segfarbiomot@unibo.it
* Correspondence: gaia.corona@studio.unibo.it

Abstract: *Helicobacter pylori* infection highly correlates with gastric cancer. Interaction between bacterial factors such as CagA and host signal transduction pathways seems to be critical for mediating cell transformation, cell proliferation, invasion, apoptosis/anti-apoptosis, and angiogenesis. In particular CagA, which is secreted inside epithelial gastric cells by a Type IV Secretion System (TFSS), undergoes tyrosine phosphorylation, enabling many interactions inside the cells including the one with SHP2, a pro-oncogenic protein. The SHP2 dysregulation will cause morphological changes (hummingbird phenotype) in the host cell, correlating with gastric cancer. SHP2-CagA interactions occur at different sites, called EPIYA sequences, sequence & copy number variation play an important role in determining gastric cancer's presence. Through bioinformatic tools and databases, the goal of the review is to shed light on the link between the interactions of CagA and gastric infections, paying attention to the type of proteins that CagA interacts with and the diversity contained in these interactions in various populations.

Keywords: *Helicobacter pylori,* gastric cancer, CagA, SHP2, *Helicobacter pylori* infections

1. Introduction

Helicobacter pylori infection is considered to be the main cause of gastric cancer [1], the fifth most common malignancy, with approximately 950,000 new cases registered each year [2]. It has been demonstrated that the bacterium is able to adapt to the extreme acidic conditions of the gastric environment, to establish persistent infection and to deregulate host functions, leading to gastric pathogenesis and cancer [3]. In this review the main focus is the ***CagA***, an oncoprotein that seems to be the first indicator for *Helicobacter Pylori* infections, indeed involved in the carcinogenic process and in the generated inflammation [1].

2. Materials and Methods

In this review we have used bioinformatic tools such as ChimeraX, which is a protein visualization tool in order to visualize the CagA protein and its interaction with SHP2; Python, in which we have written some codes that analyze the *Helicobacter pylori*'s proteome; Linux, to pre-analyze the interactions of CagA; Mega, in order to construct the phylogenetic tree; Inkscape, which is an editor tool in order to beautify our tree.

We retracted information by many databases such as Kegg, which contains metabolic pathways; PDB, which is a database for 3D protein structures; UniProt, from which we retracted protein sequences; IntAct, from which we downloaded molecular interaction data; String, which is a database of protein-protein interactions; EBI search, that also contains biological data on proteins.

Citation: To be added by editorial staff during production.

Academic Editor: Gaia Corona, Samuele Storari, Laura Claudia Verdesca

Received: 2023
Accepted: 2023
Published: 2023

3. Results

3.1. Subsection

3.1.1. **Introduction on *Helicobacter pylori***

Helicobacter pylori (*H. pylori*), part of the ***Helicobacteraceae*** family, is a micro-aerophilic, spiral-shaped Gram-negative bacterium discovered in the epithelial lining of the stomach by Marshall and Warren.

It is defined as a ***gastric Helicobacter***, a human-stomach-dwelling organism that causes many gastric illnesses, including gastritis, ulcer, and gastric cancer, affecting approximately half of the entire human population.

H. pylori is a very heterogeneous species but characteristics important for colonization and pathogenesis are found in most of the strains [1].

Considering the *Helicobacter pylori strain ATCC 700392 / 26695* from *UniProt*, the analysis (see *Supplementary Materials*) of its proteome led to the following insights:

- a number of 1554 sequences are present, among which 612 coming from *SwissProt* and 942 from *Tremble;*
- the proteome is made of 703729 amino acids, whose relative abundance is shown in *Fig. 1.*

In 1994 the *International Agency for Research on Cancer, World Health Organization* (IARC/WHO) classified the *H. pylori* as a **group I carcinogen**. Indeed, the stomach pathogen is estimated to be the cause of almost 75% of all gastric cancers, a percentage that may increase when considering East Asian countries.

Although all *H. pylori* strains are involved in gastric infections (pathway shown in *Fig. 2*), *cag+ H. pylori*, expressing the *cag pathogenicity island*, are linked to an increase in risk for severe gastritis and gastric cancers compared to *cag-deficient mutants* [1].

The *cag pathogenicity island* (***cag* PAI)** contains 27-31 putative genes, among which at least 18 are involved in the Type IV Secretion System (TFSS), by encoding the proteins that serve as building block for the definition of this syringe-like structure that is capable of delivering macromolecules across two bacterial membranes. (*Fig. 2*) [1].

- An indicator for *cag+ H. pylori* is the presence of the **cagA**, a gene localized at one end of the cag PAI encoding for the CagA protein. Based on the presence of this gene, *H. pylori* can be divided into *cagA-positive* and *cagA-negative* strains.
- Approximately 60-to-70% of Western *H. pylori* strains and almost 100% of East Asian strains express CagA [1].

The infection by *H. pylori* also causes a change in the morphology of the cells, defined as *hummingbird phenotype* (shown in *Fig. 3*(a)).

The ***H. pylori*-mediated hummingbird phenotype** involves cell elongation and migration.

- While non-infected gastric epithelial AGS cells show a round morphology, infection with *H. pylori* wild type induced loss of cell-to-cell contacts, cell elongation and migration, as shown in *Fig. 3*(b).
- The elongated cell morphology in response to *H. pylori* is dependent on the injection of CagA, since AGS cells infected with a cagA-deficient *H. pylori* mutant do not elongate [4].

3.1.2. **CagA introduction**

Due to the genetic heterogeneity present within *H. pylori* genomes, bacterial virulence factors likely play an important role in determining the outcome of *H. pylori* infection.

The *cag pathogenicity island* (*cag* PAI) is a 40-kb DNA insertion element which contains 27 to 31 genes flanked by 31-bp direct repeats, including our protein of interest CagA.

CagA and its correlation with gastric cancer were first identified in the early 90s, now *cag* PAI is a well-studied virulence factor, and the presence of CagA is frequently used as an indicator of the presence of the entire *cag* PAI.

Among the other genes, at least 18 genes encode proteins serving as building blocks of a **type IV secretion system** (TFSS).

The *H. pylori* CagA protein is a 120- to 140-kDa protein. CagA is delivered inside the cell by the TFSS, then it's able to interact with many different molecular players. In particular CagA is tyrosine phosphorylated at the EPIYA motifs, which induces morphological changes inside the cell, called the hummingbird phenotype, in which cells elongate abnormally.

3.1.3. **Structure of CagA**

CagA has a unique tertiary structure, consisting of a solid N-terminal region (70% of the entire CagA) and an intrinsically disordered C-terminal tail (30% of the entire CagA).

The crystal structure of CagA has overall dimensions 110 × 80 × 55 Å. We can distinguish three different domains (as to see in *Fig. 4*):

- **Domain I** (highlighted in cyan) is the extreme N-terminal domain of CagA, it has a small interacting surface with Domain II, but no contact occurs with Domain III. Domain I is highly mobile and flexible.
  The intrinsically disordered proline-rich region of ASPP2 (this is a protein that stimulates apoptosis for the protein p53) binds to the pocket formed by the three-helix bundle present in Domain I.
  RUNX3 (recurrent domain in transcription regulator, it's a tumor repressor protein) interacts with CagA, supposedly through the WW domain, but there is no structural evidence of a WW domain, which is a modular protein domain that mediates specific interactions with protein ligands.

Domain II and Domain III form a protease-resistant structural CagA core, meaning that a protease won't be able to break the intramolecular bonds of CagA.

- **Domain II** also contains a large anti-parallel β-sheet (highlighted in light pink), with which CagA binds to β1-integrin (cell surface receptor) for its delivery into the host cell. Domain II contains a basic patch constituting the PS-binding K-Xn-R-X-R (a small basic patch is also shown in pink) motif. The basic patch plays an important role in the interaction of CagA with PS.
- In **Domain III** we have the N-terminal binding sequence, that forms a four-helix bundle, and the C-terminal binding sequence (CBS), which is placed in the disordered C-terminal tail. The CBS forms a sort of lariat loop that strengthens the interaction of CagA with PAR1 and SHP2.

CagA contains a particular motif called **EPIYA** (glutamate-proline-isoleucine-tyrosine-alanine) located in within the C-terminal region of CagA, the tyrosine is phosphorylated, which induces the *hummingbird* phenotype. To date, four distinct EPIYA motifs (EPIYA-A, -B, -C, and -D) have been identified, and they are distinguished by different amino acid sequences surrounding the EPIYA motif.

The EPIYA copy number and disposition of the different motifs are correlated with the strain's virulence potential.

The highly diverged and disordered structure of the C-terminal region of CagA might be explained by the disordered nature of EPIYA, which can be aligned with different structural constraint and variable copy numbers without affecting CagA function.

In CagA there is also a CM motif (16 aa), located downstream of the EPIYA-D segment. The K-Xn-R-X-R motif in the central region is required for the binding of CagA with the membrane phospholipid, phosphatidylserine (PS).

The CM motif is highly conserved, there are 5 amino-acid alterations between East Asian and Western CagA species [1].

3.1.4. **Structural diversity of CagA**

CagA structure can vary depending on the population of origin, also the copy number can change, all of these factors contribute in determining gastric cancer.

The C-terminal EPIYA-repeat region of Western CagA includes the EPIYA-A, EPIYA-B and a variable number (mostly 1–3) of EPIYA-C segments.

The C-terminal EPIYA-repeat region of East Asian CagA includes the EPIYA-A, EPIYA-B and EPIYA-D segments.

East Asian CagA has a single CM (CME) motif downstream of the EPIYA-D segment. Western CagA possesses at least two CM motifs, one in the EPIYA-C segment, which is unique to Western CagA (CMW), and the other located distal to the last EPIYA-C segment (either CMW or CME).

3.1.5. **CagA tethering**

CagA is delivered into gastric epithelial cells through the TFSS mechanism, then it's tethered to the inner leaflet of the membrane via two distinct mechanisms depending on the cell polarity.

1. In polarized epithelial cells, the tethering occurs through the interaction of the membrane and the central region of CagA, which contains multiple basic amino acids.
2. In non-polarized cells, however, the C-terminal region is primarily responsible for the membrane tethering of CagA

Membrane-localized CagA then undergoes tyrosine phosphorylation [1].

3.1.6. **CagA Network of interactions**

CagA is able to interact with many different proteins inside the cell.

By selecting CagA as protein and *Helicobacter pylori* as reference organism, the network in *Fig. 5* is produced.

In particular C694_02700 is involved in the Type IV secretion system, as well as C694_027020; they correspond respectively to Cag5 and Cag7. Also cagE is involved in the TFSS as well as in the DNA transfer.

By selecting instead *Homo sapiens* as reference organism, important functional partners are shown in *Fig. 6*, as:

- SRC, a proto-oncogene tyrosine-protein kinase;
- PTPN11, tyrosine-protein phosphatase non-receptor type 11;
- GRB2, growth factor receptor-bound protein 2 containing an SH domain.

From a biological process point of view, the proteins MET, SRC GRB2 and CDH1 are all involved with the entry of bacterium into host cell, while MET, SRC, PTPN11, TJP1 are involved in Epithelial cell signaling in *Helicobacter pylori* infection.

This review intends to focus especially on the interaction between CagA and PTPN11, the gene encoding SHP2, that contains a SH2 domain, a recognition domain in signaling pathways that acts through the phosphorylation of tyrosines [1].

- The cagA-SHP2 interaction (shown in *Fig. 7.*), mediated by tyrosine-phosphorylated EPIYA-C or EPIYA-D, is one of the key interactions through which CagA exerts its pro-oncogenic action. In fact, CagA, through deregulation of SHP2, is involved in activating mutation of PTPN11 and indeed playing an important role in carcinogenesis.
    - Since SHP-2 is involved in both cell growth and cell motility, deregulation of SHP-2 by CagA may have a role in the induction of abnormal proliferation and movement of gastric epithelial cells, a cellular condition eventually leading to gastritis and gastric carcinoma [5].
- CagA is delivered into gastric epithelial cells and, on tyrosine phosphorylation, specifically binds and activates the SHP2 oncoprotein, thereby inducing the formation of an elongated cell shape known as the 'hummingbird' phenotype [6].

It is possible to see this interaction through IntAct, selecting cagA in *Helicobacter pylori* with the ID P80200 as shown in *Fig. 8*. The interaction between cagA and PTPN11 can also be seen by dealing with the IntAct database that contains the interactome, through bash commands (see *Supplementary Materials*).

If we select the ID P55980 for the CagA of *Helicobacter pylori*, the network in *Figure 9.* is produced.

The interaction with MARK2 is of particular interest, since it could be the cause of the development of another gastric disease.

- MARK2 kinase helps maintain gastric cell polarity.
- CagA inhibits MARK2 by blocking the substrate-binding site, leading to loss of polarity in cells of the gastric epithelium, which is thought to lead to development of ulcers [7].

3.1.7. **Phylogenetic tree**

We have also created a phylogenetic tree, shown in *Figure 10.* using Mega based on the SHP2 protein in order to see its evolutionary pattern across different animals.

3.2. Figures, Tables and Schemes

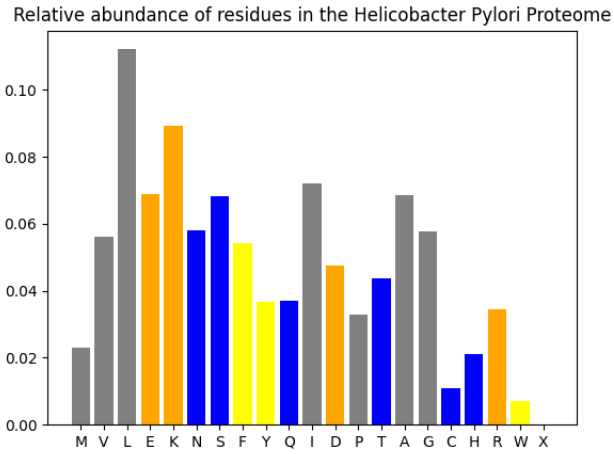


Figure 1. *Characterization of the H. pylori proteome through amino acid composition*. The gray bars represent non polar amino acids, the orange is for charged amino acids, the blue bars are for the polar amino acids and the yellow is for aromatic amino acids.

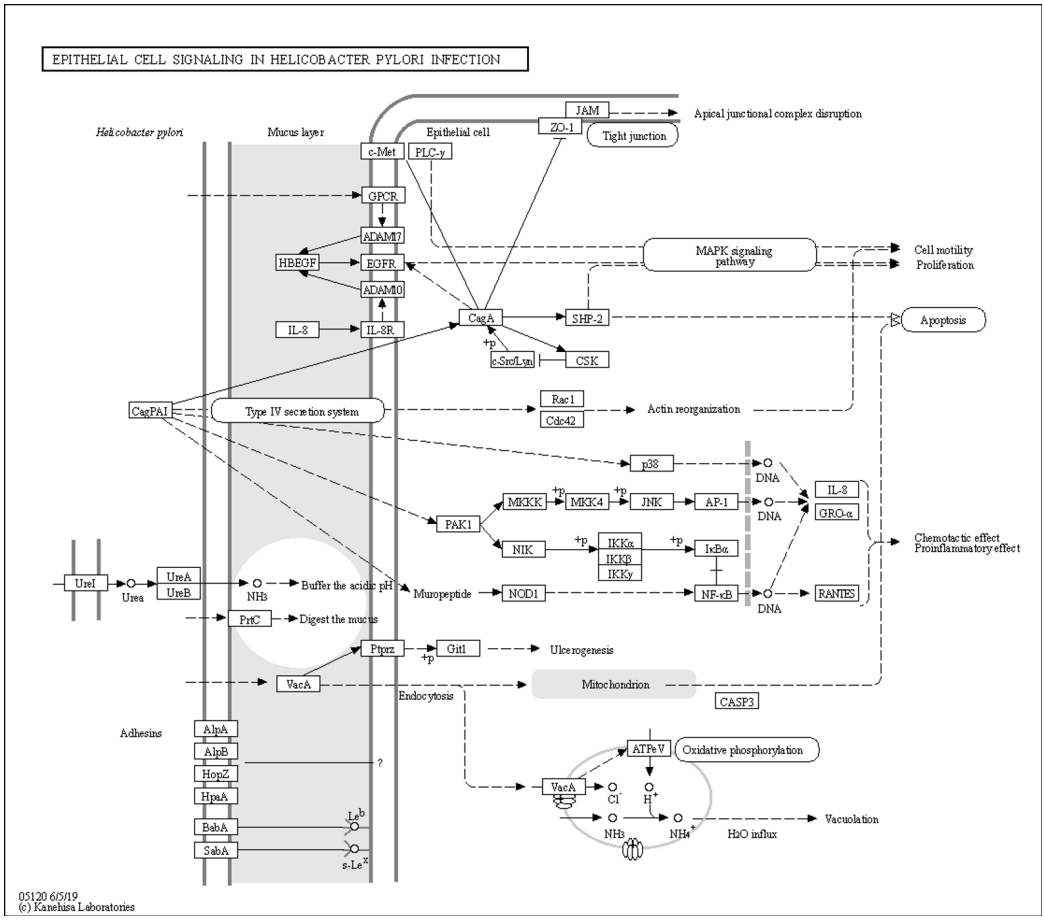
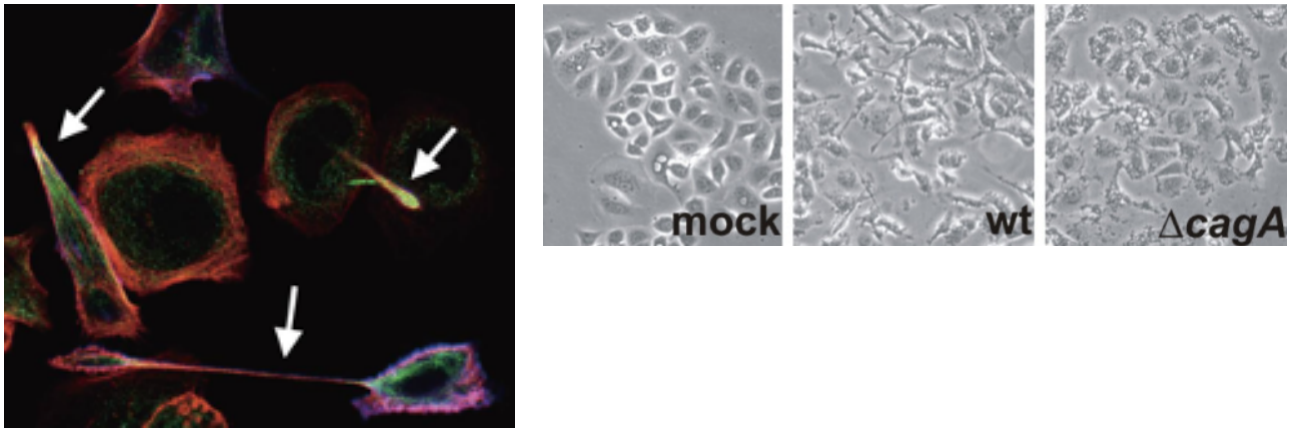
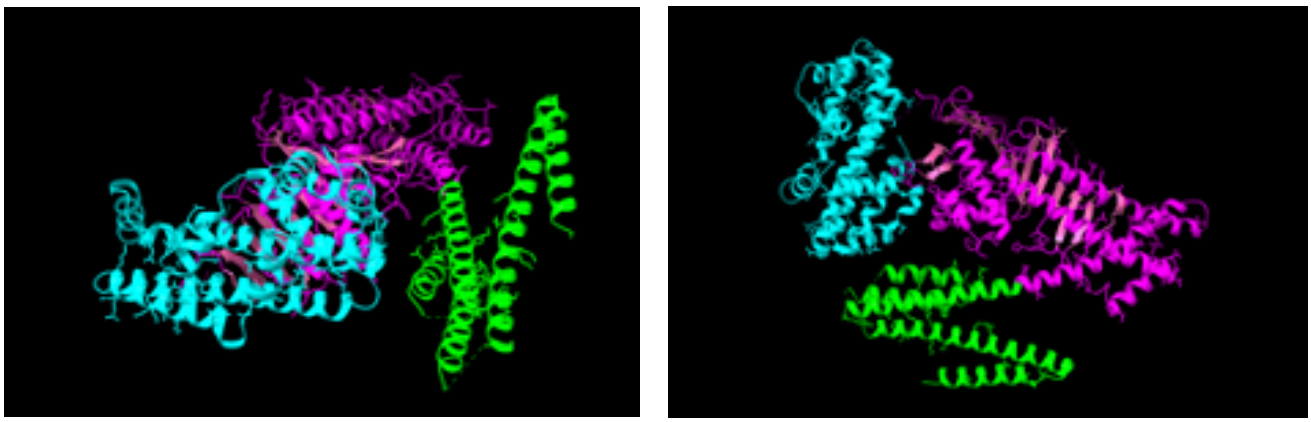


Figure 2. *Epithelial cell signaling in Helicobacter pylori infection*. It was detected through the database resource KEGG, in which it is noticeable the role of cagPAI and the type IV secretion system (T4SS) to deliver CagA into the host cells.



(**a**) (**b**)

Figure 3. *The hummingbird phenotype induced by H. pylori CagA*. (a): the white arrows indicate the extremely elongated cells in response to the H. pylori infection. (b): Differences between the morphology of the non-infected and infected gastric epithelial AGS cells. mock stands for non-infected gastric epithelial AGS cells, wt represents infection with H. pylori and ΔcagA is for cagA-deficient H. pylori mutant. Notice the differences in morphology: from a round shape in mock to a more elongated in wt.



(**a**) (**b**)

Figure 4. *CagA structure*. Domain I (cyan) constitutes the N-terminus, while Domain II (Magenta) tethers CagA to the inner plasma membrane through electrostatic interaction between the basic patch and the acidic phosphatidylserine (PS). While Domain III (Lime green) forms many intramolecular interactions.

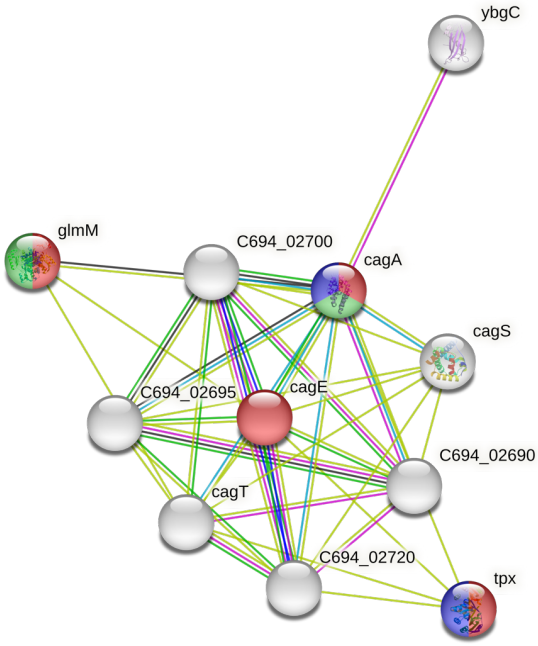


Figure 5. *Network of interactions between CagA and the predicted functional partners*. The nodes in red are the proteins involved in gastric cancers following the paper "*Helicobacter pylori* Pathogenicity Factors Related to Gastric Cancer" of 2017 (DOI:10.1155/2017/7942489). The nodes in green are related to the paper "*Risk assessment of gastric cancer in the presence of Helicobacter pylori cagA and hopQII genes*" of 2022 (DOI: 10.14715/cmb/2021.67.4.33). The proteins in blue are the ones analysed in the paper "*Helicobacter pylori colonization of the human gastric epithelium: a bug's first step is a novel target for us*" of 2010 (DOI: 10.1111/j.1440-1746.2009.06141.x). The network has been obtained through String.

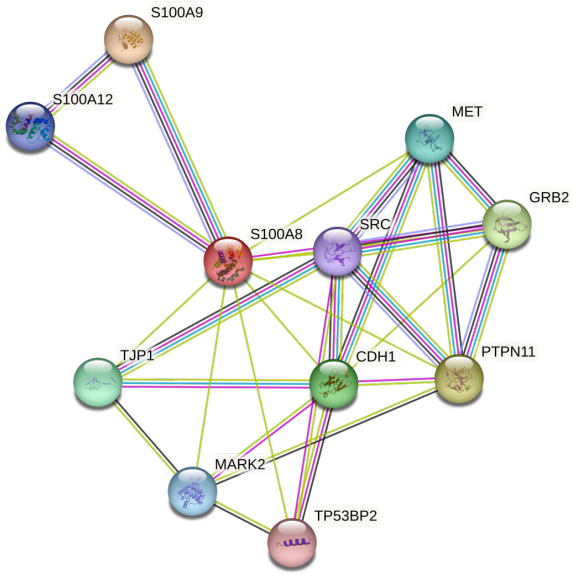


Figure 6. *Network of interactions between CagA and the predicted functional partners*. The network has been obtained through String.

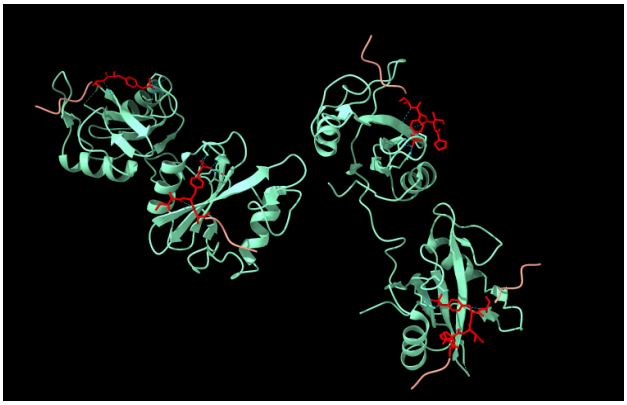


Figure 7. *Interaction between SHP2 and CagA*. In this image, obtained through ChimeraX, it is visible the interaction of SHP2 (green) with just the EPIYA sequences (red) of CagA, in particular four EPIYA segments binding in different points of the SHP2 protein.

Figure 8. *Interaction between PTPN11 and cagA.* The number of interactions present are 8, as explained by the IntAct database and found through a Linux analysis (see *Supplementary Materials*).

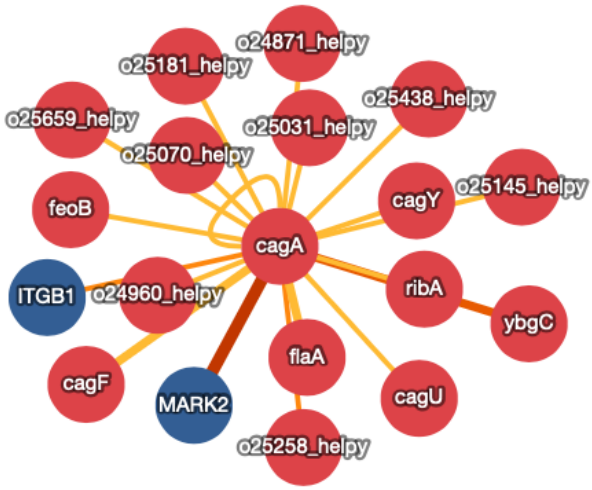


Figure 9. *Network of interaction of cagA with proteins of other bacteria and Homo Sapiens*. ITGB1 and MARK2, the nodes in blue, are the two proteins from *H. sapiens*. In particular the number of

interactions between cagA and MARK2 are 4, as described in Intact and through the Linux Analysis (see *Supplementary Materials*).

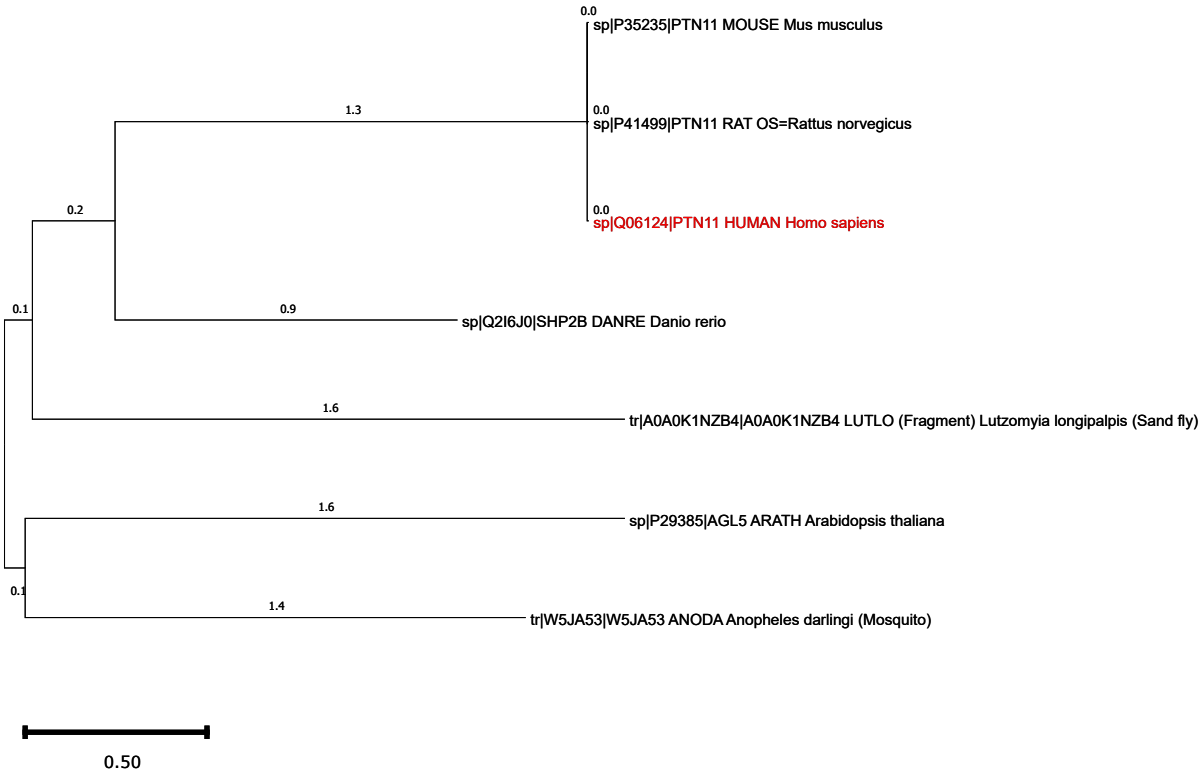


Figure 10. *Phylogenetic tree of SHP2 protein.* The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree is shown. The tree is drawn to scale, with branch lengths (next to the branches) in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. This analysis involved 7 amino acid sequences. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 1226 positions in the final dataset. Evolutionary analyses were conducted in MEGA11.

4. Discussion and Conclusions

The main focus of this review was to study the CagA protein and its interaction to assess its pro-oncogenic action.

The results obtained suggest that CagA induces a signaling pathway capable of conferring abnormal physical characteristics to the cell, that subsequently alter cell normal functioning. This is linked to the idea that CagA is involved in the development of many infections, being indeed a virulence factor.

Therefore, we can affirm that H. *pylori*'s CagA does in fact correlate with gastric infections, agreeing also with statistics about gastric cancer and *H. pylori*'s presence. In fact, thanks to the analysis performed on the structure of interaction between CagA and SHP2 and to the study of CagA's network of interactions, we can define *H. pylori CagA* as an **exogenous cancer-promoting protei**n that is injected intermittently into gastric epithelial cells by *H. pylori*.

Further investigations are needed in order to uncover structural and molecular biology details of CagA and the pathway that it induces. Further studies must be conducted also on the impending resistance to antibiotics of *Helicobacter pylori* [8].

This is a starting point for new therapies and drugs against cancer, for example in eradication therapy [9].

New insights have been discovered on the implication of CagA copy number and its correlation with gastric cancer, further studies are needed [10].

Supplementary Materials: Any additional material including the ChimeraX files, Python codes, fasta file of the protein sequences and additional images can be found at the GitHub at this link.

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: G. Corona, L.C. Verdesca, S. Storari; data collection: G. Corona, L.C. Verdesca, S. Storari; analysis and interpretation of results: G. Corona, L.C. Verdesca, S. Storari; draft manuscript preparation: G. Corona, L.C. Verdesca, S. Storari. All authors reviewed the results and approved the final version of the manuscript.

Acknowledgments: This work was supported by Alma Mater Studiorum - University of Bologna.

1. Masanori Hatakeyama. Structure and function of *Helicobacter Pylori CagA*, the first identified bacterial protein involved in human cancer. *Review* **2017** *Volume 93, Issue 4,* P 196-219. DOI: 10.2183/pjab.93.013.
2. F. Mégraud; E. Bessède; C. Varon. *Helicobacter pylori* infection and gastric carcinoma. *Review* **2015** *Volume 21, Issue 11*, P 984-990. DOI: 10.1016/j.cmi.2015.06.004.
3. V. Camilo; T. Sugiyama; E. Touati. Pathogenesis of *Helicobacter pylori* infection. *Review* **2017** *Volume 22, Issue S1*. DOI: 10.1111/hel.12405.
4. S. Schneider; C. Weydig & S. Wessler. Targeting focal adhesions: *Helicobacter pylori*-host communication in cell migration. *Review* **2008** *Cell Communication and Signaling 6, Article number: 2*. DOI: 10.1186/1478-811X-6-2
5. Higashi, H.; Tsutsumi, R.; Fujita, A.; Yamazaki, S.; Asaka, M.; Azuma, T.; Hatakeyama, M. Biological activity of the Helicobacter pylori virulence factor CagA is determined by variation in the tyrosine phosphorylation sites. Proc. Natl. Acad. Sci. USA **2002**, 99, 14428–14433. [Google Scholar] [CrossRef][Green Version]. DOI: 10.1073/pnas.222375399.
6. I. Saadat; H. Higashi; C. Obuse; M. Umeda; N. Murata-Kamiya; Y. Saito; H. Lu; N. Ohnishi; T. Azuma; A. Suzuki; S. Ohno & M. Hatakeyama. *Helicobacter pylori* CagA targets PAR1/MARK kinase to disrupt epithelial cell polarity. ***Nature* 447**, 330-333(**2007**). DOI: 10.1038/nature05765.

7. G. Farley; J. Issroff; R. Kaplan; D. Alatras; C. Cherston; J. Drucker; M. Durkin; S. Edelstein; Liang-Liang Feng; C. Jacobson; M. Klabin; C. Lovett; T. Shapiro; L. Susman; J. Hackett; C. Erec Stebbins. Modeling the inhibition of human MARK2 kinase *by H. pylori* CagA virulence factor. ***2011*** *Volume 25, Issue S1.* https://doi.org/10.1096/fasebj.25.1_supplement.lb162**.**
8. Rizzato, C., Torres, J., Obazee, O. et al. Variations in cag pathogenicity island genes of Helicobacter pylori from Latin American groups may influence neoplastic progression to gastric cancer. Sci Rep 10, 6570 (**2020**). DOI: https://doi.org/10.1038/s41598-020-63463-0.
9. Suzuki, H., Matsuzaki, J. Gastric cancer: evidence boosts Helicobacter pylori eradication. Nat Rev Gastroenterol Hepatol 15, 458–460 (**2018**). DOI: 10.1038/s41575-018-0023-8.
10. Su, H., Tissera, K., Jang, S. et al. Evolutionary mechanism leading to the multi-cagA genotype in Helicobacter pylori. Sci Rep 9, 11203 (**2019**). DOI: https://doi.org/10.1038/s41598-019-47240-2.