Investigation of the CagY Protein in *Helicobacter Pylori*

Chris LeBailly

November 20, 2013

1 Introduction

The objective of this paper is to study a specific protein using both bioinformatic tools in addition to a literature search. I was given a DNA sequence of a PCR product that contains a gene for the protein I wanted to study. In order to identify the gene I used the NCBI's Translated BLAST: blastx web search [1]. This identified the protein as the CagY protein found in Helicobacter pylori. Before researching the CagY protein in *H. pylori*, lets see what else we can determine using existing bioinformatics tools.

2 Investigation

After identifying the given sequence, I used different tools to try and learn more about the CagY protein. To do this I wrote a program to find the long ORF in the gene we were given. The protein which is coded by the long ORF was found in the third reading frame on the reverse strand with a length of 2215 amino acids. I took this protein and first looked for homologs of the protein. After that I looked to see what families were contained in the protein. Finally I looked for transmembrane helices.

2.1 Homologs

To identify homologs of the protein, I ran the protein through HHpred. This identified two homologous proteins. The first was 3jqo_A, which is a TraF Transport protein used in type IV secretion systems. This had an E-value of 4.2×10^{-31}. The second was 2bhv_A, which is a bacterial protein that is part of a type IV secretion system. This had an E-value of 1.1×10^{-27}. No other results were significant (the next smallest E-value was 13) [18].

2.2 Protein Families

To identify different protein families contained with in the CagY protein, I used Pfam. Pfam found three different types of protein domain families. The first is CagY_I, which is a type 1 repeat.

There are two CagY_I repeats. The first occurrs between residues 9 and 73. The second occurrs between residues 137 and 204. Next there are 40 CagY_M's, which are DC-EC repeats. These occur between residues 310 and 1799. Finally there is TrbI, which is a Bacterial conjugation TrbI-like protein family, located between residues 1978 and 2175. These results are summarized in Figure 1.

Figure 1: CagY_I Repeat (Green), CagY_M Repeat (Red), and TrbI Family (Blue) [17].

2.3 Transmembrane Helices

To detect transmembrane helices I first used TMHMM2.0 [12]. This program found one transmembrane helix, located between residues 214 and 236 of the protein. It stated residues 1 to 213 were located inside the membrane and 237 to 2215 were located outside the membrane. Its results are summarized in Figure 2. We can clearly see the transmembrane helix from residues 214 to 236. We can see there is a small chance that another transmembrane helix occurs just beyond the 2000^{th} residue, accord this hidden Markov model. This warrants further analysis.

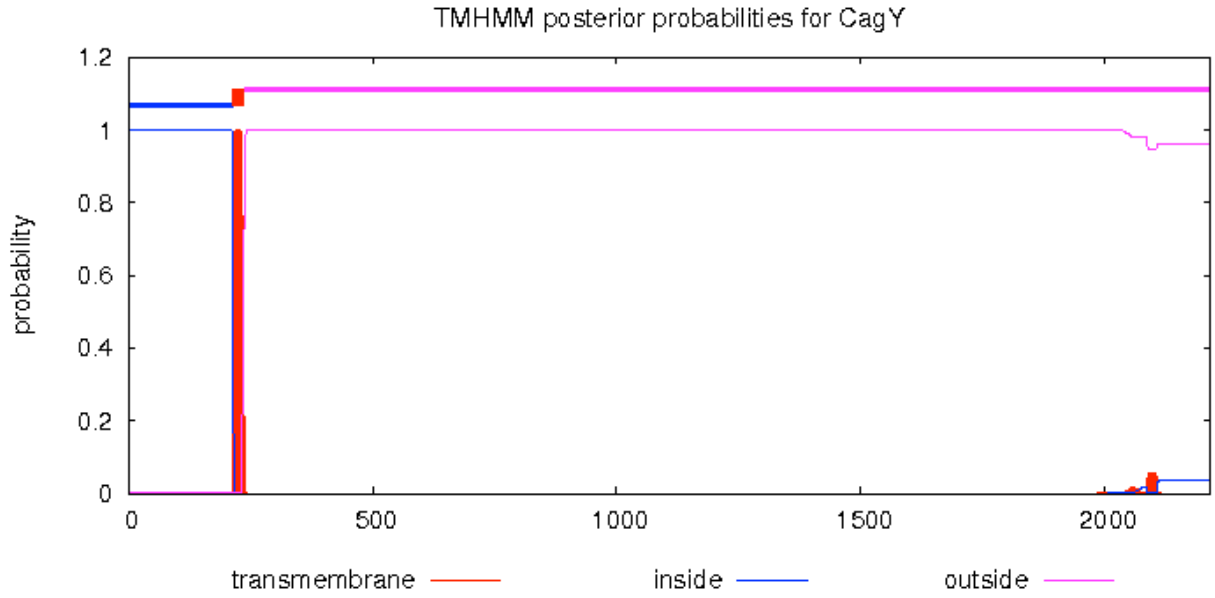
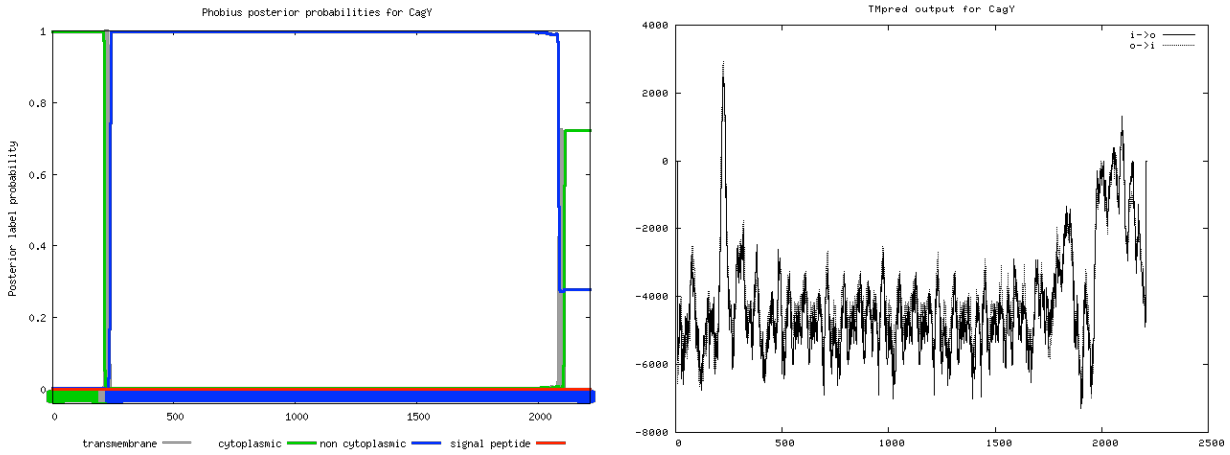


Figure 2: TMHMM Analysis for Transmembrane Helices [12].

Next I ran the protein through Phobius [11]. It predicted a transmembrane helix between residues 214 and 239 with probability 1. It also detected a 75% chance of having another transmembrane helix just over 2000, as shown in Figure 3a. These odds are much higher than those predicted by the TMHMM program. To help clear this up I used a third program, TMpred [10]. Figure 3b shows scores of this model versus the residue. There is a clear spike at the transmembrane helix we have established and a less pronounced spike just beyond the 2000^{th} residue. The TMpred program states there are "2 strong transmembrane helices, total score : 4282" [10]. The first goes

from residues 214 to 235 and the second goes from residues 2086 to 2108. Given the results from these three programs it seems likely there are two transmembrane helices.



(a) Phobius Analysis [11].

(b) TMpred Analysis [10].

Figure 3: Phobius and TMpred Looking for Second Transmembrane Helix

3 Relevant Literature

Helicobacter Pylori is a gram-negative bacterium which is found in the stomach of over half of the humans in the world [9, 15]. In most cases is cause chronic, asymptomatic gastritis. In more sever cases it can cause peptic ulcer disease, gastric cancer (which is the second most common cause of cancer worldwide), duodenal ulceration, and a MALT lymphoma [2,3,7,15]. The virulence is most strong associated with the presence of the cytotoxin associated gene pathogenicity island (*cag*PAI). The *cag*PAI contains roughly 27 genes, many which encode a type IV secretion system (T4SS) [3]. The T4SS translocates cytotoxin-associated gene A (CagA) into the gastric epithelial cells, which according to [3] causes "epithelial cell elongation [8], disruptions of tight junctions [14], and alteration of cell polarity [4, 19]."

The *cagY* gene is part of the *cag*PAI. The CagY protein is believed to manage contact between the inner and outer bacterial membrane [6]. Compared to other genes with similar functions, *cagY* is surprising large. This is because of the highly repetitive structure we saw in Figure 1. This could allow selection or duplication within the *cagY* gene and might change how the *cag*PAI functions [3]. Furthermore, in [3], they showed that "infection with *H. pylori* leads to host immunity-dependent recombinations in *cagY* that is sufficient to eliminate the functionality of the T4SS." This means that changing the *cagY* gene could affect the functionality of the T4SS which injects CagA into the host cell. In [3] they were able to produce alleles of the *cagY* gene which translocated CagA and others which could not translocate CagA.

If the T4SS pilus in *H. pylori* is no longer created or if its structure has change when the *cagY* gene is altered, this could explain why it loses its ability to translocate CagA. To test this theory, [3] used a field emission scanning electron microscope to image *H. pylori*. They found that

with the *cag*PAI removed, *H. pylori* did not form a T4SS pilus. This is consistent with previous studies [13, 16, 20, 21]. They then examined strains of *H. pylori* where the *cagY* gene was replaced with alleles that did or did not translocate CagA. They also examined strains of *H. pylori* which had the *cagY* gene deleted. In all of these cases pilus structures were observed (see Figure 4). Hence the T4SS pills forms regardless of the *cagY* gene. They conclude that "the loss of function that occurs with changes in CagY results from a functional change in the T4SS without any detectable structural defect in the T4SS pilus."

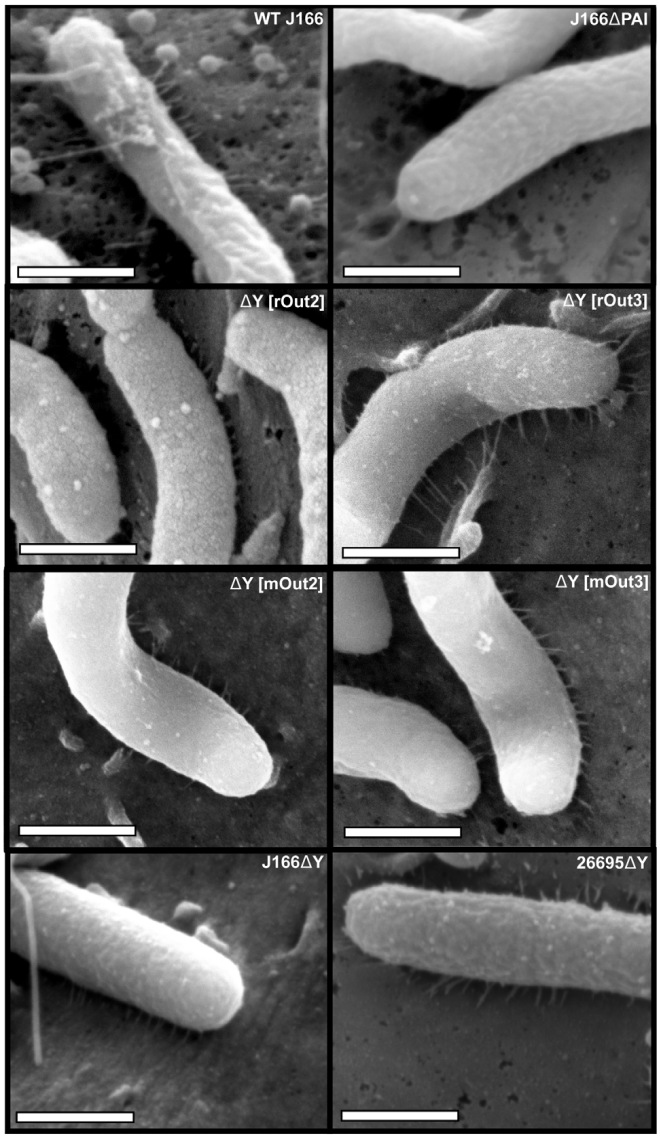


Figure 4: The pilus structure is visible on WT J166, but not visible when the *cag*PAI is removed (J166ΔPAI). Pili were visible when the *cagY* gene was replaced functioning (rOut3, mOut3) or non-functioning (rOut2, mOut2) alleles. Pili were also seen when the *cagY* gene was removed in J116ΔY and 26695Δ*Y*. Bars show 500 nm. Image from [3].

It was suggested in earlier studies (such as [5]) that the highly repetitive nature of *cagY* was to evade immune responses. However, in the studies preformed by [3], the theory is not sufficient to explain their results. They propose that "CagY variants serve not to evade the host immune response, but rather to tune it so as to establish the optimal homeostatic conditions of inflammation under which H. pylori is most fit." CagY is varied by recombinations. These alter the functionality of T4SS in *H. pylori* to maximize persistent infection.

4 Conclusions

In my investigation I found a number of structural components of the CagY protein which turned out to be important. Looking for homologous showed us that CagY was related to a protein used in type 4 secretion systems. Looking for transmembrane helices showed us where the CagY protein would likely intersect the bacterial membrane. We also saw that the CagY protein structure had a very repetitive nature, which in reading the literature turned out to play an important role, as recombinations of the CagY protein affected the translocation of CagA. Further studies are still needed to fully understand how how the CagY protein and it's different recombinations alter the T4SS and it's ability to translocate CagA.

References

[1] Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, 25:3389–3402, 1997.

[2] Steffen Backert, Marguerite Clyne, and Nicole Tegtmeyer. Molecular mechanisms of gastric epithelial cell adhesion and injection of CagA by *Helicobacter pylori*. *CellCommunSignal*, 9:28, 2011.

[3] Roberto M. Barrozo, Cara L. Cooke, Lori M. Hansen, Anna M. Lam, Jennifer A. Gaddy, Elizabeth M. Johnson, Taryn A. Cariaga, Giovanni Suarez, Richard M. Peek, Jr., Timothy L. Cover, and Jay V. Solnick. Functional plasticity in the type IV secretion system of *Helicobacter pylori*. *PLoSPathog*, 9(2):e1003189, February 2013.

[4] B. Biswas, R. Vemulapalli, and S. K. Dutta. Molecular basis for antigenic variation of a protective strain-specific antigen of *Ehrlichia risticii*. *InfectImmun*, 66(8):3682–3688, August 1998.

[5] M. J. Blaser, G. I. Perez-Perez, H. Kleanthous, T. L. Cover, R. M. Peek, P. H. Chyou, G. N. Stemmermann, and A. Nomura. Infection with *Helicobacter pylori* strains possessing cagA is associated with an increased risk of developing adenocarcinoma of the stomach. *CancerRes*, 55(10):2111–2115, May 15 1995.

[6] J. E. Crabtree. Role of cytokines in pathogenesis of *Helicobacter* pylori-induced mucosal damage. *DigDisSci*, 43(9 Suppl):46S–55S, September 1998.

[7] Robin M. Delahay, Graham D. Balkwill, Karen A. Bunting, Wayne Edwards, John C. Atherton, and Mark S. Searle. The highly repetitive region of the *Helicobacter pylori* CagY protein comprises tandem arrays of an alpha-helical repeat module. *JMolBiol*, 377(3):956–971, March 28 2008.

[8] J. Dworkin and M. J. Blaser. Nested DNA inversion as a paradigm of programmed gene rearrangement. *ProcNatlAcadSciUSA*, 94(3):985–990, February 4 1997.

[9] Masanori Hatakeyama and Hideaki Higashi. *Helicobacter pylori* CagA: A new paradigm for bacterial carcinogenesis. *CancerSci*, 96(12):835–843, December 2005.

[10] K. Hofmann and W. Stoffel. TMbase - A database of membrane spanning proteins segments. *Biol. Chem. Hoppe-Seyler*, 374:166, 1993.

[11] Lukas Käll, Anders Krogh, and Erik L. L. Sonnhammer. Advantages of combined transmembrane topology and signal peptide prediction–the Phobius web server. *NucleicAcidsRes*, 35(Web Server Issue):W429–W432, July 2007.

[12] A. Krogh, B. Larsson, G. von Heijne, and E. L. Sonnhammer. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *JMolBiol*, 305(3):567–580, January 19 2001.

[13] E. J. Kuipers, D. A. Israel, J. G. Kusters, M. M. Gerrits, J. Weel, A. van Der Ende, R. W. van Der Hulst, H. P. Wirth, J. Höök-Nikanne, S. A. Thompson, and M. J. Blaser. Quasispecies development of *Helicobacter pylori* observed in paired isolates obtained years apart from the same host. *JInfectDis*, 181(1):273–282, January 2000.

[14] R. McCulloch, G. Rudenko, and P. Borst. Gene conversions mediating antigenic variation in *Trypanosoma brucei* can occur in variant surface glycoprotein expression sites lacking 70-base-pair repeat sequences. *MolCellBiol*, 17(2):833–843, February 1997.

[15] Patrick Olbermann, Christine Josenhans, Yoshan Moodley, Markus Uhr, Christiana Stamer, Marc Vauterin, Sebastian Suerbaum, Mark Achtman, and Bodo Linz. A global overview of the genetic and functional diversity in the *Helicobacter pylori* cag pathogenicity island. *PLoSGenet*, 6(8):e1001069, August 2010.

[16] R. M. Peek, Jr., G. G. Miller, K. T. Tham, G. I. Perez-Perez, X. Zhao, J. C. Atherton, and M. J. Blaser. Heightened inflammatory response and cytokine expression *in vivo* to cagA+ *Helicobacter pylori* strains. *LabInvest*, 73(6):760–770, December 1995.

[17] M. Punta, P.C. Coggill, R.Y. Eberhardt, J. Mistry, J. Tate, C. Boursnell, N. Pang, K. Forslund, G. Ceric, J. Clements, A. Heger, L. Holm, E.L.L. Sonnhammer, S.R. Eddy, A. Bateman, and R.D. Finn. The Pfam protein families database. *Nucleic Acids Research*, Database Issue 40:D290–D301, 2012.

[18] Johannes Soding, Andreas Biegert, and Andrei N. Lupas. The hhpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Research*, 33 (suppl 2):W244–W248, March 2005.

[19] B. Stevenson, S. Casjens, and P. Rosa. Evidence of past recombination events among the genes encoding the Erp antigens of *Borrelia burgdorferi*. *MicrobiologyReadingEngl*, 144 (Pt 7):1869–1879, July 1998.

[20] A. van der Ende, E. A. Rauws, M. Feller, C. J. Mulder, G. N. Tytgat, and J. Dankert. Heterogeneous *Helicobacter pylori* isolates from members of a family with a history of peptic ulcer disease. *Gastroenterology*, 111(3):638–647, September 1996.

[21] M. E. Woolhouse, L. H. Taylor, and D. T. Haydon. Population biology of multihost pathogens. *Science*, 292(5519):1109–1112, May 11 2001.